IOWA

A Dynamic Framework for Identification and Estimation of Nonseparable Production Functions

Job Market Presentation

Justin Doty

Introduction

- This paper identifies and estimates a **nonseparable** production function with **unobserved heterogeneity**
- Two important contributions to the literature
  1. New nonparametric estimates of firm-specific production functions
     - Nonseparable model allows interactions between unobserved production shocks and inputs
     - Captures sources of unobserved heterogeneity arising from factor-specific productivity changes
     - Important implications for models that use production function estimates
  2. New framework to capture heterogeneous productivity dynamics
     - Incorporates **asymmetric persistence** in productivity history
     - Driven by size and sign of productivity shocks
     - Can explain variation in productivity rankings across time

Production Functions

- Production functions are important in many economic models
- They link outputs to inputs (e.g. capital, labor) and represents firm technology
- Estimates can be used in the following applications:
  1. Measuring returns to scale
  2. Gains from trade
  3. Capital misallocation
  4. Estimates of market power (e.g. markups)
- A correctly specified production function is crucial for correct inference in these areas
- For example, biased flexible input elasticities are transmitted to markup estimates

Simultaneity Bias

- Significant progress in solving one type of bias in production function estimation: **simultaneity bias**
- Researchers don't observe productivity
- Firm chooses inputs depending on their productivity
- A more productive firm may hire more/less workers
- In this case, labor estimates will be biased
- **Proxy variable** approaches of OP, LP, ACF remain a popular tool
- Basic idea: A policy function (e.g. material demand) is inverted as a function of productivity
- Substitute inverted function into production function and estimate in two-step approach ACF Estimator

Limitations of Proxy Variables

- This approach relies on many crucial assumptions
- No unobserved errors in the policy function
  - No measurement error
  - No unobserved demand shocks
- Productivity and an unobserved production shock are additive (e.g. log-Cobb-Douglas)
  - Implies productivity is "factor-neutral"
  - Technology is fixed across firms
  - Estimates only capture average firm production
- Productivity process is subject to additive shocks
  - Empirical evidence is in favor of more flexible productivity dynamics
  - Since productivity is transmitted to inputs it is important to understand its dynamics to test model validity

Why Nonseparable Models

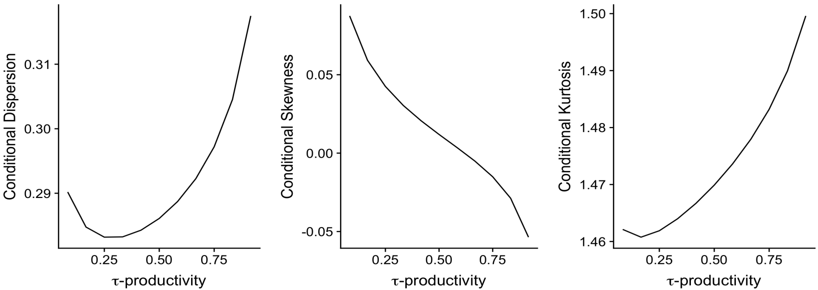
- Empirical research points to missing heterogeneity from technological change, which favors inputs such as labor
- Labor augmenting productivity explains much of the variation in markups over time (Dermirer, 2020)
- It is also the primary driver of the fall in labor shares across many advanced economies (Doraszelski and Jaumandreu, 2018)
- Recent advancements have used a structural approach to estimating factor-specific productivity
- These rely on either a parametric or nonparametric inversion of policy functions
- These techniques are still invalid in the presence of unobservables in the policy functions

Why Nonseparable Models

- Productivity is significantly heterogeneous across firms even in narrowly defined industries
- The cross-sectional distribution of productivity varies over time
- These dynamic effects can change the production function at different points in time
- A more flexible productivity process may provide insight to how firms asymmetrically adjust inputs and output in accordance to shocks
- Large degree of persistence in productivity rankings in the U.S. (Bartelsman and Dhrymes, 1998)
- High and low productivity firms may change productivity rankings depending on the size of sign of shocks

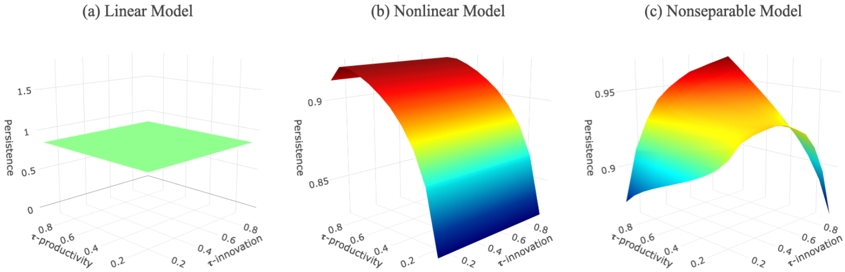
Preview of Results

Figure 1: Summary Statistics of Productivity Dynamics



Preview of Results

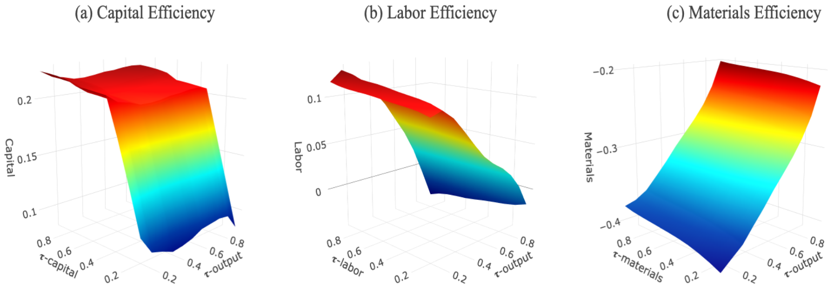
Figure 2: Productivity Persistence



*Panel (a): Productivity persistence from a linear model. Panel (b): Productivity persistence for a nonlinear model that is separable in unobserved shocks. Panel (c): Productivity persistence estimated in the nonseparable model.

Preview of Results

Figure 3: Non-Hicks Neutral Elasticities



*Panel (a): Capital efficiency evaluated at τ_η and percentiles of capital τ_k averaged over values of (l_{it}, m_{it}) that correspond to τ_k. Panel (b): Labor efficiency evaluated at τ_η and percentiles of labor τ_l averaged over values of (k_{it}, m_{it}). Panel (c): Materials efficiency evaluated at τ_η and percentiles of materials τ_m averaged over values of (k_{it}, l_{it}).

Summary of Findings

Data: U.S. Compustat public manufacturing firms

- Asymmetric persistence in productivity
  1. Positive shocks for low productivity firms
  2. Negative shocks for high productivity firms

  have lower persistence of productivity
- Asymmetric adjustments of inputs with respect to productivity and shocks
- Length of time until recovery from bad productivity shocks varies by firm
- Nonlinearities in production function estimates
  1. Positive capital productivity effects
  2. Positive and negative labor productivity effects
  3. Negative material productivity effects

Outline for the Rest of Talk

1. ~~Introduction~~
2. Economic Model
3. Econometric Identification
4. Econometric Procedure and Quantile Modelling
5. Results
6. Conclusions

The Production Function

- Consider a nonseparable model for a firm's gross-output production function (in logs)

$$y_{it} = Q_t^y(k_{it}, l_{it}, m_{it}, \omega_{it}, \eta_{it}), \quad \eta_{it} \sim \mathit{Uniform}(0,1) \qquad (1)$$

Skorohod Representation

- Allows for non-linear interactions between inputs and unobserved productivity
- Assume the following

Assumption 1 (Production Function)

1. *The unanticipated production shocks η_{it} are i.i.d. over firms and time.*
2. *The unanticipated production shock η_{it} follows a standard uniform distribution independent of $(k_{it}, l_{it}, m_{it}, \omega_{it},)$.*
3. $\tau \to Q_t^y(k_{it}, l_{it}, m_{it}, \omega_{it}, \tau)$ *is strictly increasing on* $(0,1)$.

Quantile Function

- For a given τ, the conditional quantile function for the random variable $Y|X$ is defined as

$$Q_\tau(Y|X) = \inf\{y \in \mathbb{R} : \tau \leq F_{Y|X}(y|x)\}, \quad \tau \in (0,1)$$

  where $F_{Y|X}$ is continuous and strictly increasing
- This quantile assumption will be important later on as it allows me to recover estimates of firm specific production functions, input demands, and productivity
- The standard production function and quantiles is an unexplored area
- Doty and Song, 2021 consider the case for a simple additive-in-productivity model and discuss some of its implications

Productivity

- Productivity evolves according to an exogenous first-order Markov process given by

$$\omega_{it} = Q_t^{\omega}(\omega_{it-1}, \xi_{it}), \quad \xi_{it} \sim \mathit{Uniform}(0,1), \tag{2}$$

  where $\xi_{i1}, \ldots, \xi_{iT}$ are independent uniform random variables which represent innovation shocks to productivity
- The function Q^{ω} is a function that allows the persistence in productivity in firms to be nonlinear across different quantiles

Assumption 2 (Productivity)

1. *The productivity innovation shocks ξ_{it} are i.i.d. across firms and time.*
2. *ξ_{it} follows a standard uniform distribution independent of previous period productivity ω_{it-1}.*
3. *$\tau \to Q_t^{\omega}(\omega_{it-1}, \tau)$ is strictly increasing on $(0,1)$.*

Flexible Inputs

- Labor and material inputs are chosen to maximize current period profits
- Therefore they are a function of current period state variables

$$l_{it} = Q_t^{\ell}(k_{it}, \omega_{it}, \epsilon_{\ell,it}), \quad \epsilon_{\ell,it} \sim Uniform(0,1), \tag{3}$$

$$m_{it} = Q_t^{m}(k_{it}, l_{it}, \omega_{it}, \epsilon_{m,it}), \quad \epsilon_{m,it} \sim Uniform(0,1), \tag{4}$$

- $\epsilon_{\ell,it}$ and $\epsilon_{m,it}$ are i.i.d. unobservable input demand shocks that are assumed to be independent of current period state variables
- In the control function approach, with material inputs as a proxy, this function could not be inverted as an expression of productivity only
- This can also be extended to the case where labor has adjustment frictions Labor Adjustments

Flexible Inputs

Assumption 3 (Flexible Inputs)

1. *The unobserved input demand shocks* $\epsilon_{\ell,it}$ *and* $\epsilon_{m,it}$ *are i.i.d. across firms and time.*
2. $\epsilon_{\ell,it}$ *and* $\epsilon_{m,it}$ *follow a standard uniform distribution independent of* (k_{it}, ω_{it}) *and* $(k_{it}, l_{it}, \omega_{it})$*, respectively.*
3. $\tau \to Q_t^{\ell}(k_{it}, \omega_{it}, \tau)$ *and* $\tau \to Q_t^{m}(k_{it}, l_{it}, \omega_{it}, \tau)$ *are strictly increasing on* $(0, 1)$*.*

Capital and Investment

- Capital accumulates to the following generalized law of motion

$$K_{it} = \kappa_t(K_{it-1}, I_{it-1}, \upsilon_{it-1}) \tag{5}$$

  where I_{it-1} denotes firm investment in the prior period
- Eliminates the deterministic relationship of capital with respect to previous period state and choice variables
- Assume this error term is independent of the arguments in the capital accumulation law
- In each period, a firm chooses investment to maximize its discounted future profits:

$$I_{it} = \iota_t(K_{it}, \omega_{it}) = \underset{I_t \geq 0}{\operatorname{argmax}} \left[\Pi_t(K_{it}, \omega_{it}) - c(I_{it}, \omega_{it}) + \beta \mathbb{E}\left[V_{t+1}(K_{it+1}, \omega_{it+1}) | \mathcal{I}_t \right] \right], \tag{6}$$

Capital and Investment

- $\pi_t(\cdot)$ is current period profits as a function of the state variables $c(\cdot)$ is the cost function, $\mathcal{I}_t$ is information set
- Empirical investment rule is

$$i_{it} = Q_t^i(k_{it}, \omega_{it}, \zeta_{it}), \quad \zeta_{it} \sim Uniform(0,1). \tag{7}$$

Assumption 4 (Capital Accumulation and Investment)

1. *The unobserved investment demand shocks ζ_{it} is i.i.d. across firms and time.*
2. *ζ_{it} follows a standard uniform distribution independent of (k_{it}, ω_{it}).*
3. *The production shock η_{it} and ζ_{it} are independent conditional on $(k_{it}, l_{it}, m_{it}, \omega_{it})$. In addition, υ_{it} is independent of η_{it} conditional on $(k_{it}, l_{it}, m_{it}, \omega_{it})$*
4. *$\tau \to Q_t^i(k_{it}, \omega_{it}, \tau)$ is strictly increasing on $(0,1)$*

Nonparametric Identification

- I show that the conditional densities corresponding to production, inputs, and productivity are nonparametrically identified
- Let $Z_t = (l_t, k_t, m_t, k_{t+1})$ denote conditioning variables

Assumption 5 (Conditional Independence)

1. $f(y_t | y_{t+1}, I_t, \omega_t, Z_t) = f(y_t | \omega_t, Z_t)$
2. $f(y_{t+1} | I_t, \omega_t, Z_t) = f(y_{t+1} | \omega_t, Z_t)$

- First equality states that conditional on ω_t and Z_t, y_{t+1} and I_t do not provide any additional information about y_t
- Second equality states that conditional on ω_t and Z_t, I_t does not provide any additional information about y_{t+1}
- Satisfied by mutual independence assumptions on η_t and ζ_t conditional on $(\omega_t, k_t, l_t, m_t)$

Nonparametric Identification

- Begin by relating a conditional density as a function of observable to densities containing unobserved productivity
- Using the conditional independence assumption, I can write

$$f_{y_t, I_t | y_{t+1}, Z_t} = \int f_{y_t | Z_t, \omega_t} f_{I_t | Z_t, \omega_t} f_{\omega_t | y_{t+1}, Z_t} d\omega_t \tag{8}$$

- The identification strategy follows HS by using a eigenvalue-eigenfunction decomposition of integral operators of (8)

Definition 1

(Integral Operator) *Let a and b denote random variables with supports $\mathcal{A}$ and $\mathcal{B}$. Given two corresponding spaces $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(\mathcal{B})$ of functions with domains $\mathcal{A}$ and $\mathcal{B}$, let $L_{b|a}$ denote the operator mapping $g \in \mathcal{G}(\mathcal{A})$ to $L_{b|a} g \in L_{b|a} \mathcal{G}(\mathcal{B})$ defined by*

$$[L_{b|a} g](b) \equiv \int_{\mathcal{A}} f_{b|a}(b|a) g(a) da,$$

where $f_{b|a}$ denotes the conditional density of b given a.

Nonparametric Identification

- The observed density in (8) can be written in operator notation

$$L_{y_t,I_t|y_{t+1},Z_t} = L_{y_t|Z_t,\omega_t}\Delta_{I_t|Z_t,\omega_t}L_{\omega_t|y_{t+1},Z_t} \tag{9}$$

- Will show that under a set of assumptions, the conditional density is identified from an eigenvalue-eigenfunction decomposition of (9)

Assumption 6 (Injectivity)

The operators $L_{y_t|Z_t,\omega_t}$ and $L_{y_{t+1}|Z_t,\omega_t}$ are injective

- The above assumption allows us to take inverses of the operators.
- Injectivity of $L_{y_t|Z_t,\omega_t}$ can be interpreted as its corresponding density $f_{y_t|Z_t,\omega_t}(I_t|K_t,\omega_t)$ having sufficient variation in ω_t given Z_t.
- Type of nonparametric IV rank condition

Nonparametric Identification

Assumption 7 (Uniqueness)

For any $\bar{\omega}_t, \tilde{\omega}_t \in \Omega$, *the set* $\{f_{I_t|\omega_t,Z_t}(I_t|\bar{\omega}_t, Z_t) \neq f_{I_t|\omega_t,Z_t}(I_t|\tilde{\omega}_t, Z_t)\}$ *has positive probability whenever* $\bar{\omega}_t \neq \tilde{\omega}_t$.

- This assumption is relatively weak
- Satisfied if there is conditional heteroskedasticity in $f_{I|\omega,Z}$
- Satisfied if any functional of its distribution is strictly increasing in ω_t
- I assume $E[I_t|\omega_t, Z_t]$ is strictly increasing in ω_t
- Similar to the invertibility condition in Olley and Pakes, 1996

Nonparametric Identification

Assumption 8 (Normalization)

There exists a functional Γ *such that* $\Gamma[f_{y_t|\omega_t,Z_t}(y_t|\omega_t, Z_t)] = \omega_t$.

- This functional does not need to be known
- Sufficient to consider a known function of the data distribution as shown by Arellano and Bonhomme, 2016
- In my empirical application, I consider a nonseparable translog production function
- The assumption can be satisfied by the normalization $E[y_t|\omega_t, 0] = \omega_t$
- For more generalized production functions, if $E[y_t|\omega_t, Z_t]$ is strictly increasing in ω_t, then one could normalize $\omega_t = E[y_t|\omega_t, Z_t]$
- These restrictions are easily adaptable in estimation as it amounts to centering the coefficients in the model

Nonparametric Identification

Theorem 1 (Identification)

Under Assumptions 5, 6, 7, and 8, given the observed density $f_{y_t, I_t | y_{t+1}, Z_t}$*, the equation*

$$f_{y_t, I_t | y_{t+1}, Z_t} = \int f_{y_t | \omega_t, Z_t} f_{I_t | \omega_t, Z_t} f_{\omega_t | y_{t+1}, Z_t} d\omega_t \tag{10}$$

admits a unique solution for $f_{y_t | \omega_t, Z_t}$*,* $f_{I_t | \omega_t, Z_t}$*, and* $f_{\omega_t | y_{t+1}, Z_t}$

- The proof follows using Hu and Schennach, 2008
- However it does not directly identify the Markov transition function for productivity $f_{\omega_{it+1} | \omega_{it}}(\omega_{it+1} | \omega_{it})$

Nonparametric Identification

Corollary 1 (Identification of Markov Process: Stationarity Case)

Suppose that the production function is stationary, $f_{y_t|\omega_t, Z_t} = f_{y_1|\omega_1, Z_1}, \forall t \in \{1, \cdots, T\}$*. Then, under Assumptions 5, 6, 7, and 8, the observed density* $f_{y_t, I_t|y_{t+1}, Z_t}$*, uniquely determines the density* $f_{\omega_{t+1}|\omega_t}, \forall t \in \{1, \ldots, T-1\}$

Corollary 2 (Identification of Markov Process: Non-Stationarity Case)

Under Assumptions 5, 6, 7, and 8, the observed density $f_{y_{t+1}, I_{t+1}|y_{t+2}, Z_{t+1}}$*, uniquely determines the density* $f_{\omega_{t+1}|\omega_t}, \forall t \in \{1, \ldots, T-2\}$

Econometric Procedure: Production

- The production function is specified as Translog with non-Hicks neutral effects

$$
\begin{aligned}
&Q_t^y(k_{it}, l_{it}, m_{it}, \omega_{it}, \tau) = \\
&\quad \gamma_0(\tau) + (\gamma_k(\tau) + \sigma_k(\tau)\omega_{it})k_{it} + (\gamma_l(\tau) + \sigma_l(\tau)\omega_{it})l_{it} + (\gamma_m(\tau) + \sigma_m(\tau)\omega_{it})m_{it} \\
&\quad + (\gamma_{kl}(\tau) + \sigma_{kl}(\tau)\omega_{it})k_{it}l_{it} + (\gamma_{lm}(\tau) + \sigma_{lm}(\tau)\omega_{it})l_{it}m_{it} + (\gamma_{km}(\tau) \\
&\quad + \sigma_{km}(\tau)\omega_{it})k_{it}m_{it} + (\gamma_{kk}(\tau) + \sigma_{kk}(\tau)\omega_{it})k_{it}^2 + (\gamma_{ll}(\tau) + \sigma_{ll}(\tau)\omega_{it})l_{it}^2 \\
&\quad + (\gamma_{mm}(\tau) + \sigma_{mm}(\tau)\omega_{it})m_{it}^2 + \sigma_\omega(\tau)\omega_{it}.
\end{aligned}
\tag{11}
$$

- Similar model was estimated by Ackerberg and Chen (2015)
- In my approach I can simulate productivity from estimated intital conditions and Markov process to compute average derivative effects
- Provides a better picture of heterogeneity instead of reporting individual coefficients

Econometric Procedure: Productivity

- I specify productivity using 3rd order polynomial

$$Q^{\omega}(\omega_{it-1}, \tau) = \rho_0(\tau) + \rho_1(\tau)\omega_{it-1} + \rho_2(\tau)\omega_{it-1}^2 + \rho_3(\tau)\omega_{it-1}^3. \tag{12}$$

Selection Bias

- Initial productivity

$$Q^{\omega_1}(k_{i1}, \tau) = \sum_{j=1}^{J} \rho_{\omega_1,j}(\tau)\phi_{\omega_1,j}(k_{i1}), \tag{13}$$

- I can also consider the case where productivity may evolve endogenously as Doraszelski and Jaumandreu, 2013

$$Q^{\omega}(\omega_{it-1}, r_{it-1}, \tau) = \mathbb{1}\{R_{it-1} = 0\}Q^{\omega}(\omega_{it-1}, \tau) + \mathbb{1}\{R_{it-1} > 0\}Q^{\omega,r}(\omega_{it-1}, r_{it-1}, \tau). \tag{14}$$

R&D Firms

Econometric Procedure: Flexible Inputs

- I specify the labor input demand equation as follows:

$$Q_t^{\ell}(k_{it}, \omega_{it}, \tau) = \sum_{j=1}^{J} \alpha_{\ell,j}(\tau)\phi_{\ell,j}(k_{it}, \omega_{it}), \qquad (15)$$

  where $\phi_{\ell,j}$ can be another non-linear function

- Material inputs are specified as

$$Q_t^{m}(k_{it}, l_{it}, \omega_{it}, \tau) = \sum_{j=1}^{J} \alpha_{m,j}(\tau)\phi_{m,j}(k_{it}, l_{it}, \omega_{it}), \qquad (16)$$

- Again, $\phi_{m,j}$ can be a non-linear function

Econometric Procedure: Investment

- The investment demand function is specified as

$$i_{it} = Q_t^i(k_{it}, \omega_{it}, \tau) = \sum_{j=1}^{J} \delta_j(\tau)\phi_{i,j}(k_{it}, \omega_{it}), \qquad (17)$$

  where $\phi_{\iota,j}$ is specified similarly as the labor and material input decision rule.

- In the case where investment is censored, I can write

$$Q_t^{i^*}(k_{it}, \omega_{it}, \tau) = \max\{0, \sum_{j=1}^{J} \delta_j(\tau)\phi_{i,j}(k_{it}, \omega_{it})\}, \qquad (18)$$

  due to the equivariance properties of quantiles

- The censored quantile regression model avoids distributional assumptions at the cost of computational complexity
- Censored investment levels are not an issue in Compustat

Econometric Restrictions

The following conditional moment restrictions hold:

$$\mathbb{E}\left[\Psi_\tau\left(y_{it} - Q_t(y_{it}|k_{it}, l_{it}, m_{it}, \omega_{it}; \beta(\tau))\right)\middle| k_{it}, l_{it}, m_{it}\right] = 0 \tag{19}$$

$$\mathbb{E}\left[\Psi_\tau\left(l_{it} - \sum_{j=1}^{J} \alpha_{l,j}(\tau)\phi_{l,j}(k_{it}, \omega_{it})\right)\middle| k_{it}, \omega_{it}\right] = 0 \tag{20}$$

$$\mathbb{E}\left[\Psi_\tau\left(m_{it} - \sum_{j=1}^{J} \alpha_{m,j}(\tau)\phi_{m,j}(k_{it}, \omega_{it})\right)\middle| k_{it}, \omega_{it}\right] = 0 \tag{21}$$

$$\mathbb{E}\left[\Psi_\tau\left(i_{it} - \sum_{j=1}^{J} \delta_j(\tau)\phi_{\iota,j}(k_{it}, \omega_{it})\right)\middle| k_{it}, \omega_{it}\right] = 0 \tag{22}$$

For $t \geq 2$,

$$\mathbb{E}\left[\Psi_\tau\left(\omega_{it} - \rho_0(\tau) - \rho_1(\tau)\omega_{it-1} - \rho_2(\tau)\omega_{it-1}^2 - \rho_3(\tau)\omega_{it-1}^3\right)\middle| \omega_{it-1}\right] = 0, \tag{23}$$

$$\mathbb{E}\left[\Psi_\tau\left(\omega_{i1} - \sum_{j=1}^{J} \rho_{\omega_1,j}(\tau)\phi_{\omega_1,j}(k_{i1})\right)\middle| k_{i1}\right] = 0, \tag{24}$$

Econometric Restrictions

- The function $\Psi_\tau(u) = \tau - \mathbb{1}\{u < 0\}$
- Estimating the above conditional moment restrictions is infeasible due to the unobserved productivity component
- Use the unconditional moment restriction and integrate out productivity
- Let the finite and functional parameters be indexed by a finite dimensional parameter vector $\theta(\cdot)$.
- To fix ideas, consider the unconditional moment restriction corresponding to the production function

$$\mathbb{E}\left[\int \Psi_\tau\left(y_{it} - Q_t^y(k_{it}, l_{it}, m_{it}, \omega_{it}; \beta(\tau))\right) \otimes \begin{pmatrix} k_{it} \\ l_{it} \\ m_{it} \\ \omega_{it} \end{pmatrix} g_i(\omega_i^T; \theta(\cdot)) d\omega_i^T\right] = 0, \tag{25}$$

- The posterior density $g_i(\omega_i^T; \theta(\cdot)) = f(\omega_i^T | y_i^T, k_i^T, l_i^T, m_i^T, i_i^T, \theta(\cdot))$ involves the entire set of model parameters

Implementation

- Therefore, it is impossible to estimate the model parameters in a τ-by-τ procedure
- To eliminate the intractability of this problem, the continuous model parameters are approximated by piece-wise linear splines
- θ is a piecewise-polynomial interpolating splines on a grid $[\tau_1, \tau_2], [\tau_3, \tau_4], \ldots, [\tau_{Q-1}, \tau_Q]$, contained in the unit interval and is constant on $[0, \tau_1]$ and $[\tau_Q, 1)$
- The intercept coefficient β_0 is specified as the quantile of an exponential distribution on $(0, \tau_1]$ (indexed by λ^-) and $[\tau_{Q-1}, 1)$ (indexed by λ^+).
- The remaining functional parameters are modeled similarly.
- With piece-wise linear splines, the posterior density has a closed form expression without relying on strong distributional assumptions for estimation

Implementation

- In order to estimate the model, the integral inside the expectation of Equation (25) needs to be approximated
- This can be done using quadrature methods or Monte Carlo integration by converting the problem into a weighted quantile regression
- Due to the high-dimensionality of my application I choose to use a random-walk Metropolis Hastings algorithm to compute the integral
- This is known as a Monte Carlo Expectation Maximization (MCEM) procedure where the maximization step is performed using quantile regression
- This type of estimator is used by Arellano and Bonhomme, 2016 and Arellano, Blundell, and Bonhomme, 2017

Implementation

Given an initial parameter value $\hat{\theta}^0$. Iterate on $s = 0, 1, 2, \dots$ in the following two-step procedure until converge to a stationary distribution:

1. *Stochastic E-Step*: Draw M values $\omega_i^{(m)} = (\omega_{i1}^{(m)}, \omega_{i2}^{(m)}, \dots, \omega_{iT}^{(m)})$ from

$$g_i(\omega_i^T; \hat{\theta}^{(s)}) = f(\omega_i^T | y_i^T, k_i^T, l_i^T, m_i^T, i_i^T, ; \hat{\theta}^{(s)}) \propto$$

$$\prod_{t=1}^{T} f(y_{it} | k_{it}, l_{it}, m_{it}, \omega_{it}; \hat{\beta}^{(s)}) f(l_{it} | k_{it}, \omega_{it}; \hat{\alpha}_l{}^{(s)}) f(m_{it} | k_{it}, l_{it}, \omega_{it}; \hat{\alpha}_m^{(s)})$$

$$\times f(i_{it} | k_{it}, \omega_{it}; \hat{\delta}^{(s)}) \prod_{t=2}^{T} f(\omega_{it} | \omega_{it-1}; \hat{\rho}^{(s)}) f(\omega_{i1} | k_{i1}; \hat{\rho}_{\omega_1}^{(s)})$$

2. *Maximization Step*: For $q = 1, \dots, Q$, solve (e.g for production function)

$$\hat{\boldsymbol{\beta}}(\tau_q)^{(s+1)} = \underset{\boldsymbol{\beta}(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \psi_{\tau_q} \left(y_{it} - Q_t^y(k_{it}, l_{it}, m_{it}, \omega_{it}^{(m)}; \beta(\tau_q)) \right)$$

Empirical Implementation

- $\psi_\tau(u) = (\tau - \mathbb{1}\{u < 0\})u$ is the "check" function from quantile regression
- Repeat Step 2 for estimating the productivity process, input decision rules and investment
- Take $M = 1$ in the MCEM algorithm and the report estimates as the average of the last $\tilde{S} = S/2$ draws
- This is known as the stochastic EM algorithm (stEM) of Celeux and Diebolt, 1985
- The sequence of maximizers $\hat{\theta}^{(s)}$ is a time-homogeneous Markov chain which, if ergodic, will converge to its stationary distribution
- Nielsen, 2000 provides sufficient conditions for ergodicity and asymptotic properties of the estimator when the "M-step" is solved using maximum likelihood
- Arellano and Bonhomme, 2016 discuss the asymptotic properties of the estimator when the M-step is solved using quantile regression

Empirical Implementation

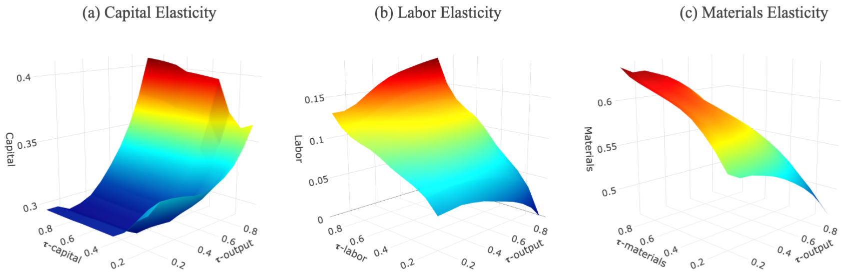
- Estimation procedure is ran with 500 random walk Metropolis-Hastings steps keeping the last draw for estimation
- 200 EM steps where the average is taken over half the draws
- $Q = 11$ for grid size for the interpolating spline
- Experimented with many different proposal distributions and initial values
- Normal distribution centered at the current draw of productivity with variance equal 0.01
- Acceptance rate $\approx 10\%$
- Initial values for productivity are simulated from TFP estimated from the LP model
- Replication code is available on author's Github

Application

- Standard and Poors Compustat database 1997 – 2016
- Productivity is simulated from its estimated parameters and used to construct investment, inputs, and output using their estimated parameters
- Capital is simulated from a linear accumulation process with constant depreciation rate 0.02
- Results are not too different from reasonable specifications for the capital accumulation process
- Interested in a variety of average and individual marginal quantile effects
- Using these estimates, can analyze how firms react to latest shocks to production, inputs, and productivity
- How long does it take for firms to recover from bad shocks to productivity?

Production Elasticities

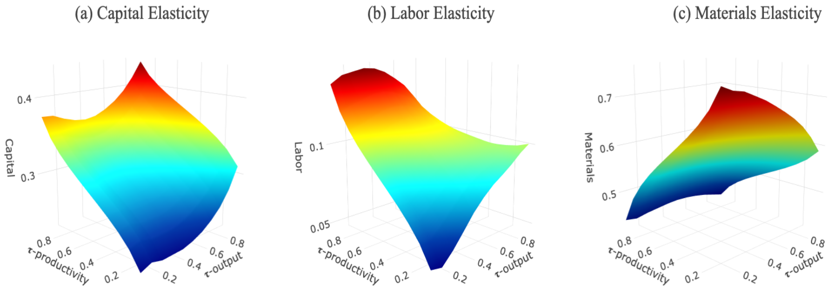
Figure 4: Output Elasticities



*Panel (a): Capital elasticity evaluated at τ_η and percentiles of capital τ_k averaged over values of (l_{it}, m_{it}) that correspond to τ_k. Panel (b): Labor elasticity evaluated at τ_η and percentiles of labor τ_l averaged over values of (k_{it}, m_{it}). Panel (c): Materials elasticity evaluated at τ_η and percentiles of materials τ_m averaged over values of (k_{it}, l_{it}).

Production Elasticities

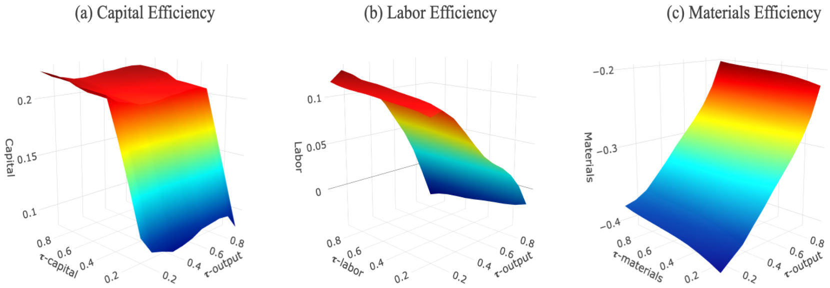
Figure 5: Output Elasticities



*Panel (a): Capital elasticity evaluated at τ_η and τ-productivity averaged over values of (k_{it}, l_{it}, m_{it}) that correspond to τ-productivity. Panel (b): Labor elasticity evaluated at evaluated at τ_η and τ-productivity averaged over values of (k_{it}, l_{it}, m_{it}) that correspond to τ-productivity. Panel (c): Materials elasticity evaluated at τ_η and τ-productivity averaged over values of (k_{it}, l_{it}, m_{it}) that correspond to τ-productivity.

Production Elasticities

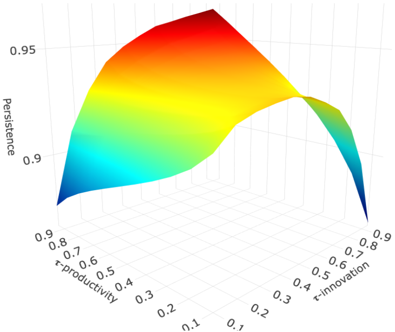
Figure 6: Non-Hicks Neutral Elasticities



*Panel (a): Capital efficiency evaluated at τ_η and percentiles of capital τ_k averaged over values of (l_{it}, m_{it}) that correspond to τ_k. Panel (b): Labor efficiency evaluated at τ_η and percentiles of labor τ_l averaged over values of (k_{it}, m_{it}). Panel (c): Materials efficiency evaluated at τ_η and percentiles of materials τ_m averaged over values of (k_{it}, l_{it}).

Productivity Persistence

Figure 7: Productivity Persistence

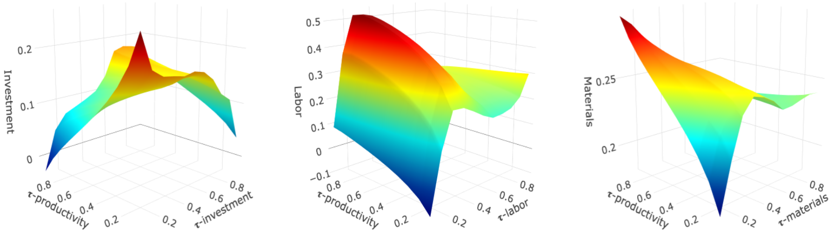


*Estimates of average productivity persistence evaluated at τ_ξ and percentiles of previous productivity.

Marginal Productivities

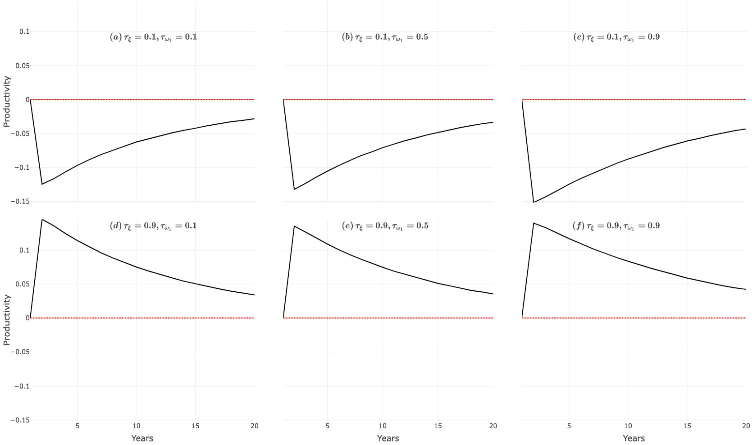
Figure 8: Input Demand Responses to Productivity

(a) Investment Response (b) Labor Response (c) Materials Response



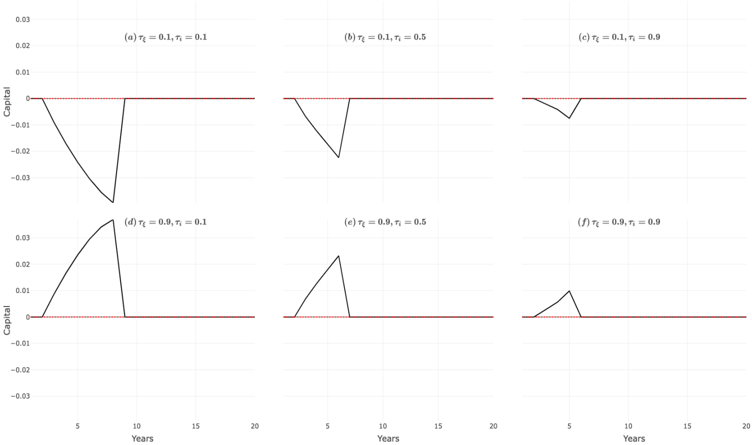
*Panel (a): Investment demand evaluated at τ_{ζ} and percentiles of productivity τ_{ω} averaged over values of k_{it}. Panel (b): Labor demand evaluated at $\tau_{\epsilon\ell}$ and percentiles of productivity τ_{ω} averaged over values of k_{it}. Panel (c): Material demand evaluated at τ_{ϵ_m} and percentiles of productivity τ_{ω} averaged over values of k_{it} and l_{it}

Productivity Innovation Shocks



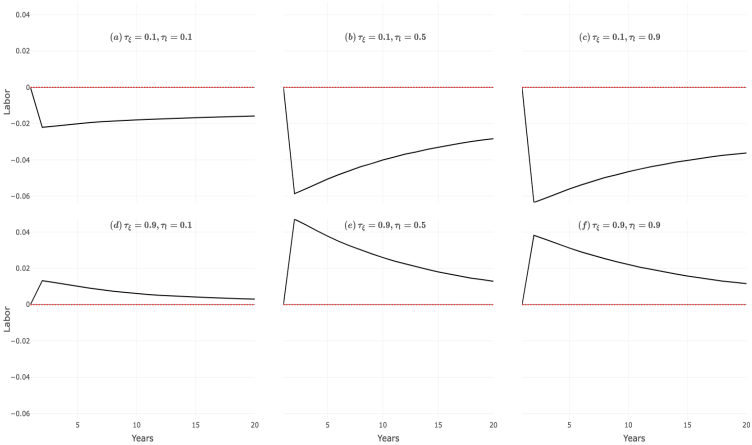
*Top row: Differences between firms hit with low productivity shock $\tau_\xi = 0.1$ and medium shock $\tau_\xi = 0.5$ at different levels of initial productivity. Bottom row: Differences between firms hit with high productivity shock $\tau_\xi = 0.9$ and medium shock $\tau_\xi = 0.5$ at different levels of initial productivity.

Productivity Innovation Shocks to Capital



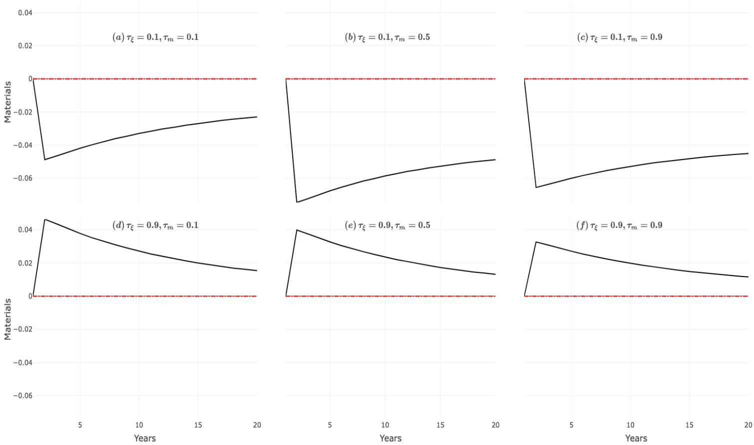
*Top row: Differences between firms hit with low productivity shock $\tau_\xi = 0.1$ and medium shock $\tau_\xi = 0.5$ at different levels of investment demand. Bottom row: Differences between firms hit with high productivity shock $\tau_\xi = 0.9$ and medium shock $\tau_\xi = 0.5$ at different levels of investment demand

Productivity Innovation Shocks to Labor



*Top row: Differences between firms hit with low productivity shock $\tau_\xi = 0.1$ and medium shock $\tau_\xi = 0.5$ at different levels of labor demand. Bottom row: Differences between firms hit with high productivity shock $\tau_\xi = 0.9$ and medium shock $\tau_\xi = 0.5$ at different levels of labor demand.

Productivity Innovation Shocks to Materials



*Top row: Differences between firms hit with low productivity shock $\tau_\xi = 0.1$ and medium shock $\tau_\xi = 0.5$ at different levels of materials demand. Bottom row: Differences between firms hit with high productivity shock $\tau_\xi = 0.9$ and medium shock $\tau_\xi = 0.5$ at different levels of materials demand.

Conclusion

- Nonparametric identification of production function, input demand, and productivity
- Proposed an estimation framework to document firm heterogeneity in production, input decisions, and productivity
- Asymmetric productivity persistence depends on size of innovation shock and levels of previous productivity
- Some firms respond to productivity shocks by using more inputs, but this affect is asymmetric for different productivity levels input demand size
- Asymmetric impact of innovation shocks to inputs after bad/good shocks
- Extension to multi-dimensional unobservables: fixed effects, labor-augmenting productivity
- Implications for TFP estimates and market power

ACF Estimator

- ACF procedure for estimating a *value-added* production function (in logs):

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it}, \tag{26}$$

- y_{it} is value-added output for firm i and time t
- l_{it} denotes labor input
- k_{it} denotes capital input
- ω_{it} is unobserved productivity
- ε_{it} denotes an independent and identically distributed (i.i.d) shock to production
- The constant β_0 is omitted since it is not separately identified from the mean of productivity.

ACF Estimator

- ACF introduces an intermediate input demand function defined as

$$m_{it} = m_t(k_{it}, l_{it}, \omega_{it}) \tag{27}$$

- The function m is assumed to be strictly increasing in ω_{it} for all k_{it} and l_{it}.
- Productivity can then be expressed as

$$\omega_{it} = m_t^{-1}(k_{it}, l_{it}, m_{it}) \tag{28}$$

- Substituting this equation into the production function

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + m_t^{-1}(k_{it}, l_{it}, m_{it}) + \varepsilon_{it} = \Phi_t(k_{it}, l_{it}, m_{it}) + \varepsilon_{it}. \tag{29}$$

- The function, $\Phi_t(k_{it}, l_{it}, m_{it})$, is identified by the following first stage moment restriction

$$\mathbb{E}[\varepsilon_{it} | \mathcal{I}_{it}] = 0 \tag{30}$$

- $\mathcal{I}_{it}$ denotes the firm's information at time t.
- The first stage estimate of Φ_t can be obtained by a local

ACF Estimator

- For the second stage, assume that productivity follows an AR(1) process given by

$$\omega_{it} = \mathbb{E}[\omega_{it}|\omega_{it-1}] + \xi_{it} = \rho\omega_{it-1} + \xi_{it}, \tag{31}$$

- ξ_{it} denotes an innovation to productivity which satisfies $\mathbb{E}[\xi_{it}|\mathcal{I}_{it-1}] = 0$.
- Plugging into the production function gives

$$\begin{aligned} y_{it} &= \beta_k k_{it} + \beta_l l_{it} + \rho\omega_{it-1} + \xi_{it} + \varepsilon_{it} \\ &= \beta_k k_{it} + \beta_l l_{it} + \rho(\Phi_{t-1}(k_{it-1}, l_{it-1}, m_{it-1}) - \beta_k k_{it-1} - \beta_l l_{it-1}) + \end{aligned}$$

- The production function parameters β_k, β_l and ρ are identified from the moment restrictions given by

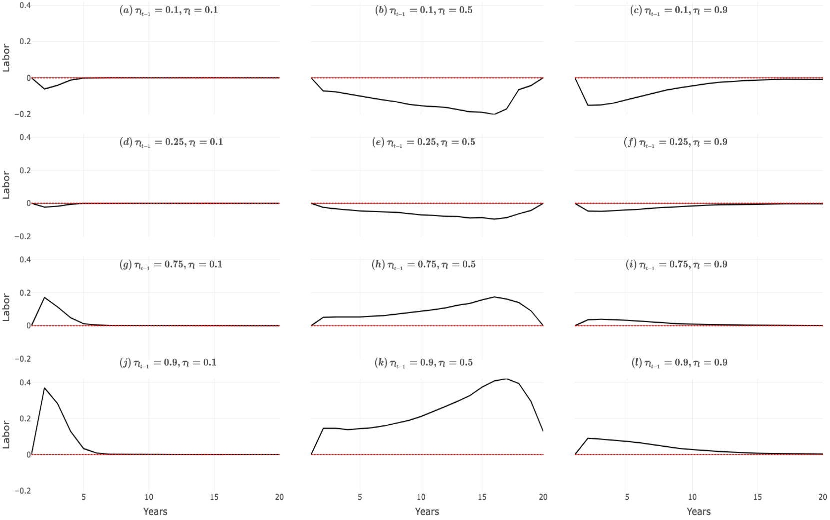
$$\mathbb{E}[\xi_{it} + \varepsilon_{it}|\mathcal{I}_{it-1}] = 0. \tag{32}$$

- Estimation of the second stage coefficients proceeds by plugging in first stage estimates $\hat{\Phi}_{t-1}$ and forming a Generalized Method of Moments (GMM) criterion function.

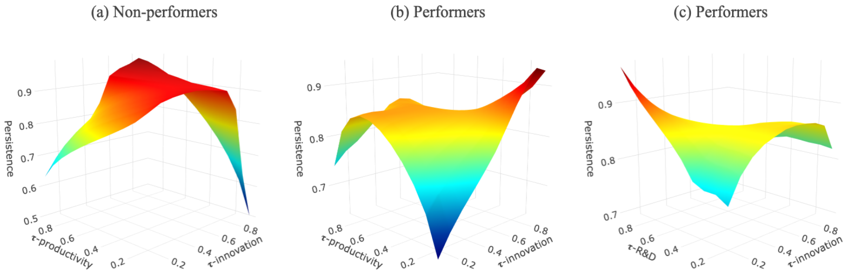
Skorohod Representation

- This representation comes from the fact that η_{it} can be defined as
$$\eta_{it} = F(y_{it}|k_{it}, l_{it}, m_{it}, \omega_{it}) \tag{33}$$
- η_{it} is then uniformly distributed independently of $(k_{it}, l_{it}, m_{it}, \omega_{it})$ on $(0, 1)$
$$y_{it} = F^{-1}(k_{it}, l_{it}, m_{it}, \omega_{it}, \eta_{it}), \quad \eta_{it} \sim Uniform(0, 1) \tag{34}$$
which is the quantile function

Labor Adjustments

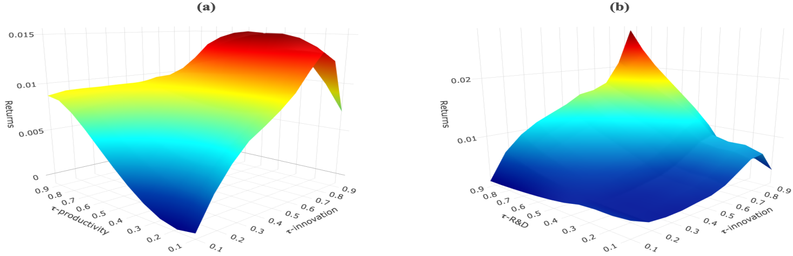


R&D Firms



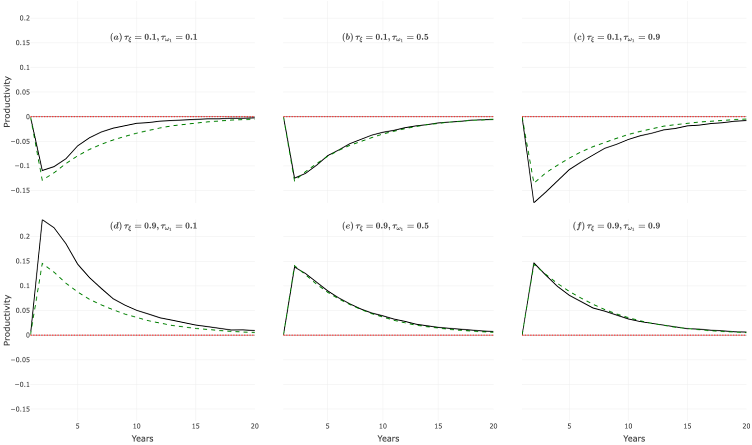
*Panel (a): Estimates of average productivity persistence for non R&D firms evaluated at τ_ξ and percentiles of previous productivity. Panel (b): Estimates of productivity persistence for R&D firms evaluated at τ_ξ and percentiles of previous productivity averaged over R&D. Panel (c): Estimates of productivity persistence for R&D firms evaluated at τ_ξ and percentiles of R&D averaged over productivity.

R&D Firms



*Panel (a): Returns to R&D for firms evaluated at τ_ξ and percentiles of previous productivity averaged over R&D. Panel (b): Returns to R&D for firms evaluated at τ_ξ and percentiles of R&D averaged over productivity.

R&D Firms



*Top row: Differences in productivity between firms hit with low productivity shock $\tau_\xi = 0.1$ and medium shock $\tau_\xi = 0.5$ at different levels of initial productivity. Bottom row: Differences in productivity between firms hit with high productivity shock $\tau_\xi = 0.9$ and medium shock $\tau_\xi = 0.5$ at different levels of initial productivity.

Correction for Selection Bias

- The exit rule can be written as

$$
\begin{aligned}
h_t(\omega_{it-1},\xi_{it}) \geq \underline{\omega}_t(k_{it}), \\
\xi_{it} &\geq h_t^{-1}(\omega_{it-1}, \underline{\omega}_t(k_{it})), \\
\xi_{it} &\geq h_t^{-1}(\omega_{it-1}, k_{it}), \\
\xi_{it} &\geq \underline{\omega}_t(\omega_{it-1}, k_{it})
\end{aligned}
\tag{35}
$$

- Then if ξ_{it} is independent of (k_{it}, ω_{it-1}), $\xi_{it}|(k_{it}, \omega_{it-1}) \sim U(0,1)$
- The cutoff for which firms stay in operation can be estimated from

$$\underline{\omega}_t(\omega_{it-1}, k_{it}) = \text{Prob}(\chi_{it} = 1|\omega_{it-1}, k_{it}) \equiv p(\omega_{it-1}, k_{it}) \tag{36}$$

- Firms that receive an innovation shock greater than $p(\omega_{it-1}, k_{it})$ continue to operate
- Then, the distribution of $\xi_{it}|(k_{it}, \omega_{it-1}, \chi_{it} = 1)$ is

$$\xi_{it}|(k_{it}, \omega_{it-1}, \chi_{it} = 1) \sim U(p(\omega_{it-1}, k_{it}), 1) \tag{37}$$

Correction for Selection Bias

- Consider a simple process for productivity $\omega_{it} = \rho(\xi_{it})\omega_{it-1}$

$$\begin{aligned} \text{Prob}(\omega_{it} &\leq \rho(\tau)\omega_{it-1}|\omega_{it-1}, k_{it}, \chi_{it} = 1) \\ &= \text{Prob}(\xi_{it} \leq \tau|\omega_{it-1}, k_{it}, \chi_{it} = 1) \\ &= \frac{\tau - p(\omega_{it-1}, k_{it})}{1 - p(\omega_{it-1}, k_{it})} \equiv G(\tau, p) \end{aligned} \tag{38}$$

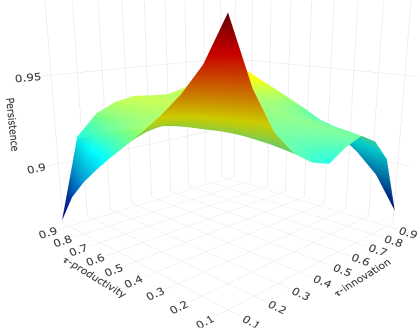
- If ω_{it} is known, then $\rho(\tau)$ is the solution to a rotated quantile regression problem

$$\hat{\rho}(\tau) = \underset{\rho}{\text{argmin}} \sum_{i=1}^{N} \sum_{t=2}^{T} \chi_{it} \Big[G(\tau, \hat{p})(\omega_{it} - \rho\omega_{it-1})^{+} + (1 - G(\tau, \hat{p}))(\omega_{it} - \rho\omega_{it-1})^{-} \Big] \tag{39}$$

  where $a^{+} = max(a, 0)$ and $a^{-} = max(-a, 0)$
- $p(\omega_{it-1}, k_{it})$ can be estimated using a probit model
- This approach is similar to that of Arellano and Bonhomme, 2017

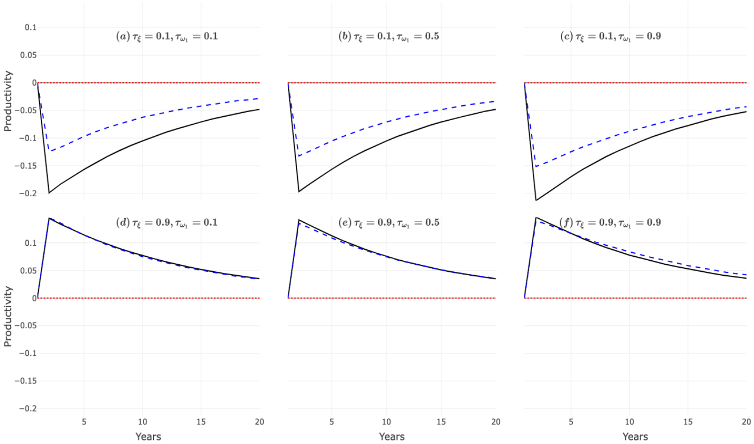
Correction for Selection Bias

Figure 9: Productivity Persistence (Selection Corrected)

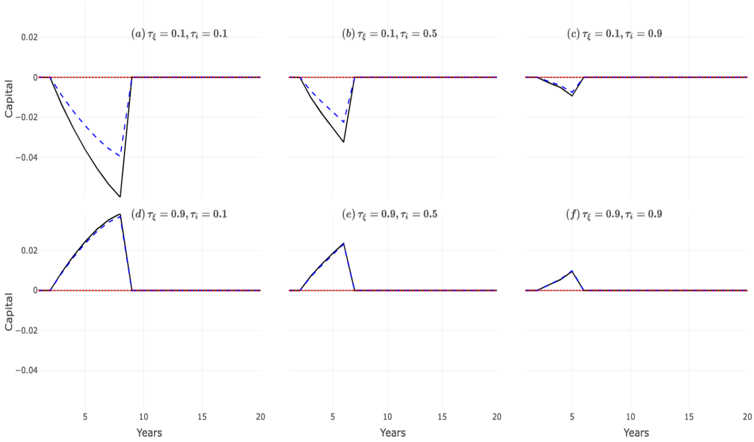


*Estimates of average productivity persistence evaluated at τ_ξ and percentiles of previous productivity.

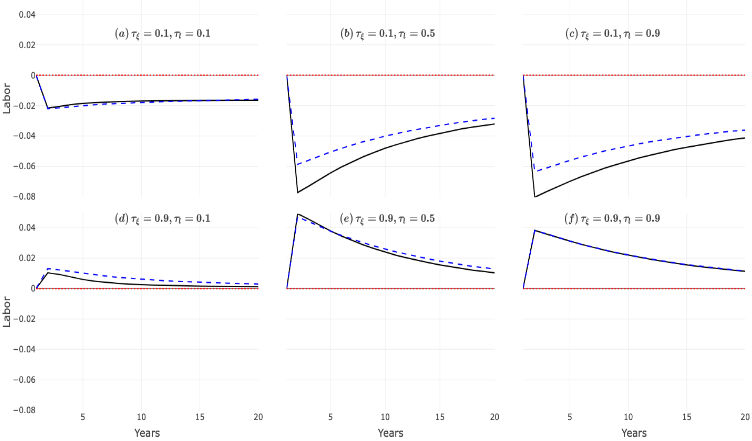
Productivity Innovation Shocks (Selection Corrected)



Productivity Innovation Shocks to Capital (Selection Corrected)



Productivity Innovation Shocks to Labor (Selection Corrected)



Productivity Innovation Shocks to Materials (Selection Corrected)