

# A Dynamic Panel Data Framework for Identification and Estimation of Nonlinear Production Functions

Justin Doty\*

## Abstract

This paper studies identification and estimation of a non-linear model for production functions with unobserved heterogeneity. Non-parametric identification results are established for the production function under stationarity assumptions. This paper then estimates the conditional quantiles of firm production using non-linear quantile regression. This paper finds that objects of interest, such as output elasticities with respect to inputs vary considerably with respect to the rank of unobserved technology shocks. In the application to US firm-level manufacturing data, this paper considers a Translog production function with non-Hicksian neutral productivity shocks and a productivity process that features non-linear persistence.

## 1 Introduction

Recent advancements in production function estimation have addressed the simultaneity problem of unobserved productivity under various timing assumptions on firm's input decisions. While some of these advancements have come at the cost of restrictive parametric assumptions on the firm's productivity-production relationship as well as the presence of additional unobservables (other than productivity), there is undoubtedly been extensive use of these methods for empirical applications. These methods, often called control functions or proxy variables model an input decision for the firm as an increasing function of productivity which is then inverted and substituted into the production function which is estimated in a two-step approach. For example, Olley and Pakes (1996) (hereafter OP) address simultaneity and selection by introducing an investment demand function that is solved from a firm's dynamic optimization problem. More recently, Levinsohn and Petrin (2003) (hereafter LP) and Akerberg *et al.* (2015) (hereafter ACF) show that instead of using investment, an intermediate input demand function can be used under various timing assumptions. This alleviates the data issues with the OP approach, namely that investment is often censored at zero so the strict monotonicity assumption is violated.

Consistent estimates of the production function are used to obtain measures of output elasticities which are used to construct estimates of total factor productivity (TFP), returns to scale, capital intensity, mark-ups, and other features of firm production. It is well known, that even after controlling for productivity, there is still large unexplained variation in firm productivity. Part of this variation can be accounted for by model specification. For example, the control function approaches mentioned earlier typically use a log-linear Cobb-Douglas production function, but other specifications, such as Translog, could be used instead. These parametric assumptions can also be augmented with firm-specific production functions in a random-coefficient framework<sup>1</sup>. Nonparametric estimation, such as the procedure proposed by Gandhi *et al.* (2020) (hereafter GNR) shows that choice of the production function matters. Estimates from a value-added production function will be different from those from a gross-output production function since the latter conditions on intermediate inputs. Estimates of TFP and dispersion ratios of TFP will appear more disperse under a value-added approach.

---

\*Department of Economics, University of Iowa, S321 Pappajohn Business Building, 21 E Market St, Iowa City, IA 52242.  
Email: justin-doty@uiowa.edu

<sup>1</sup>See for example Kasahara *et al.* (2017), Balat *et al.* (2018) and Li and Sasaki (2017)

Another challenge of the control function approach that of GNR is the restriction on unobservables generating input demand functions. In the case of OP, investment demand is a function of observed state variables such as capital and age of the firm. It is also a function of the scalar unobserved productivity. Intuitively, if the investment equation depended on other unobservables we would not be able to infer values of productivity from different levels of investment. These unobservables can be unobserved investment prices among firms or unobserved demand shocks. It is possible to extend the control function approach to this situation, however it requires demand side data which is often not available. The same logic for inverting intermediate input demand to control for unobserved productivity although ACF extends the LP approach to show their model is identified when there are unobservable factors affecting labor demand. Other unobservables, such as measurement error in capital also violates the monotonicity assumption of the control function. il Kim *et al.* (2016) allow for measurement error in capital and other inputs using identification arguments from Hu and Schennach (2008) in the framework of the control function approaches. Hu *et al.* (2019) take a similar identification approach, but propose a new generalized method of moment (GMM) estimator.

The nonparametric identification results of Hu and Schennach (2008) are flexible and can be used to identify any feature of the conditional densities of interest with the trade-off of high-level assumptions that can sometimes be difficult to verify in practice. Their identification arguments are constructive, they propose using a sieve MLE estimator for the conditional densities. I propose adopting the identification arguments of their paper by extending the assumptions of Hu *et al.* (2019) and estimating the *conditional quantiles* instead of the conditional means in the presence of non-separable unobservables in the production function, productivity process, and the input demand functions.

## 2 Literature Review

The paper most closely related to mine in terms of identification arguments for the production function is Hu *et al.* (2019). They utilize two proxies for unobserved productivity such as investment and intermediate inputs (materials or energy) or investment and a lead value of output. One of these proxies can be used to instrument for the other and the results from Hu and Schennach (2008) for non-classical measurement models is used to identify the structural parameters of the production function. Even though their results can accommodate non-linear or non-separable models for the production function and input demand functions, an additively separable model facilitates familiar sufficient conditions required for identification. For example, the injectivity assumption reviewed in a later section requires invertibility of operators corresponding to the conditional densities. Under conditional independence assumptions, additive separability of productivity, and non-vanishing characteristic functions, deconvolution arguments can be used to satisfy high-level injectivity assumptions. The feature of additive separability also satisfies normalization assumptions that are commonly used in the production function literature. In my paper, I assume a more flexible specification for production and inputs which allows for non-linearities in productivity and state variables such as capital. This comes at the cost of more general assumptions, which I provide interpretation for in the next section.

This paper proposes to estimate the production function using quantile regression. The literature relating production and its conditional quantiles is small relative to the literature using conditional means. Quantile regression and production functions has been used in the context of frontier models. These assume that firms deviate from some optimal level of efficiency known as the production frontier. One way to characterize this frontier is at the highest percentiles of the conditional output distribution. Then any firm below this frontier is one that corresponds to a lower rank along this distribution in the unit interval<sup>2</sup>. To my knowledge, the quantile frontier literature has not addressed the issue of the simultaneity of productivity. In the standard production function model, a parametric random coefficient

---

<sup>2</sup>See for example Aragon *et al.* (2005) for a formal characterization of the production frontier in this context

model can be used while also controlling for the unobserved productivity. Doty and Song (2021) consider the Skorohod representation of a value-added production function:

$$y_{it} = \beta_k(\eta_{it})k_{it} + \beta_l(\eta_{it})l_{it} + \omega_{it}, \quad \text{where } \eta_{it} \sim U[0, 1], \quad (1)$$

where  $y_{it}$  denotes value-added output for firm  $i$  at time  $t$ ,  $l_{it}$  denotes labor input,  $k_{it}$  denotes capital input,  $\omega_{it}$  is unobserved productivity and  $\eta_{it}$  denotes an iid shock to production. They show that a conditional mean representation of Equation (1) can be used to estimate productivity in the control function approach of ACF which is then plugged in to generate a new dependent variable  $\hat{y}_{it} = y_{it} - \omega_{it}$  which is used in a quantile regression on  $k_{it}$  and  $l_{it}$ .

There are several limitations to this approach. In order to estimate productivity from the conditional mean representation,  $\omega_{it}$  must follow a location-shift assumption. That is, productivity only shifts the conditional mean of output. This paper allows for the possibility that  $\omega_{it} = \omega_{it}(\tau)$ . Also, since that paper also follows ACF in estimating unobserved productivity, it is restricted to scalar unobservables in the input demand functions. In the control function approach of ACF, LP, or OP and the GNR approach the productivity process can be nonparametrically estimated. Separability of the innovation shocks to productivity are required for moment conditions used in the second stage of their models. In Doty and Song (2021) productivity is assumed to follow a linear AR(1) process in order for their identification arguments to hold. In this paper, I show that not only is a nonparametric productivity process identified, the persistence of productivity can be allowed to depend on the magnitude and sign of current and future innovation shocks.

This paper uses the estimation procedure of Arellano and Bonhomme (2016) to simulate draws of productivity from its posterior density which depends on a continuum of parameters specified in the next section. These draws are used in a sequence of quantile regressions which updates the parameters in a stochastic EM algorithm. This paper is similar in concept to Arellano *et al.* (2017) which uses the same framework to estimate non-linear earnings and consumption functions. I introduce the econometric model in Section 3. In Section 4, I discuss nonparametric identification. In Section 5 and 6, I discuss estimation based on the econometric restrictions and practical implementation using a stochastic EM algorithm. In Section 7, I apply this estimator to US publicly-listed manufacturing firms. Section 8 concludes with extensions and directions for future research.

### 3 The Model of Firm Production

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

$$y_{it} = f_t(a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it}, \eta_{it}) \quad (2)$$

where  $y_{it}$  is firm  $i$ 's output at time  $t$ ,  $a_{it}$  denotes age of the firm and  $l_{it}, m_{it}, k_{it}$  denotes optimal input choices for labor, materials, and capital respectively. The unobserved productivity is denoted by  $\omega_{it}$  which is correlated to input choices of the firm at time  $t$ . The unobserved production shocks are denoted by  $\eta_{it}$  which are assumed to be independent of input choices and productivity at time  $t$ .

Without loss of generality I normalize  $\eta_{it} \sim U[0, 1]$ . This model corresponds to a nonlinear random coefficient model where the outcome  $y_{it}$  is assumed to be strictly monotonic in  $\eta_{it}$ . In practice I can allow for nonlinear interactions between inputs and unobserved productivity at different quantiles so that marginal effects can be modeled as non-Hick's neutral. The function  $f_t$  is an unknown nonlinear function which can depend on time. I summarize the restrictions on the production function with the following assumption

**Assumption 3.1.** (*Production Function*)

- (a) *The unanticipated production shocks  $\eta_{it}$  are iid over firms and time*
- (b) *The unanticipated production shock  $\eta_{it}$  follows a standard uniform distribution independent of  $(\omega_{it}, a_{it}, k_{it}, l_{it}, m_{it})$*
- (c)  *$\tau \rightarrow Q_t(y_{it}|a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it}, \tau)$  is strictly increasing on  $(0, 1)$*

In this model, heterogeneity in production technology across firms is driven by the rank of the unobserved production shocks  $\eta_{it}$ . Productivity evolves according to the exogenous first order Markov process:

$$\omega_{it} = h_t(\omega_{it-1}, \xi_{it}) \quad (3)$$

where  $\xi_{i1}, \dots, \xi_{iT}$  are independent uniform random variables which represent innovation shocks to productivity. I assume  $\omega_{it}$  is monotonic in  $\xi_{it}$ . I let  $h$  be another unknown nonlinear function that allows the persistence in productivity in firms to be nonlinear across different quantiles. The exogeneity of the productivity process can be relaxed when I consider productivity enhancing activities such as R&D similar to Doraszelski and Jaumandreu (2013). The flexibility in the specification in (3) allows for the interaction of the rank of innovation shock and lagged productivity. Non-linear persistence for each firm can be calculated from

$$\rho_t(\omega_{it-1}, \tau) = \frac{\partial Q_t(\omega_{it}|\omega_{it-1}, \tau)}{\partial \omega_{it-1}} \quad (4)$$

where  $Q_t(\omega_{it}|\omega_{it-1}, \tau)$  is the  $\tau$ -th conditional quantile of  $\omega_{it}$ . As I will later show, heterogeneity in productivity persistence could imply that different firms adjust inputs at different rates following a shock to productivity at a given time. I summarize the restrictions on productivity with the following assumption

**Assumption 3.2.** *(Productivity)*

- (a) *The productivity innovation shocks  $\xi_{it}$  are iid across firms and time*
- (b)  *$\xi_{it}$  follows a standard uniform distribution independent of previous period productivity  $\omega_{it-1}$*
- (c)  *$\tau \rightarrow Q_t(\omega_{it}|\omega_{it-1}, \tau)$  is strictly increasing on  $(0, 1)$*

Labor inputs are chosen to maximize current period profits and therefore are a function of current period state variables

$$l_{it} = \ell_t(a_{it}, k_{it}, \omega_{it}, \epsilon_{\ell, it}) \quad (5)$$

where  $\epsilon_{\ell, it}$  is iid and independent of current period state variables. I assume the labor demand function is strictly increasing in  $\epsilon_{\ell, it}$  which is normalized to be uniformly distributed on the interval  $[0, 1]$ . Material inputs are chosen to maximize current period profits and therefore are a function of current period state variables

$$m_{it} = \mu_t(a_{it}, k_{it}, \omega_{it}, \epsilon_{m, it}) \quad (6)$$

where  $\epsilon_{m, it}$  is iid and independent of current period state variables. I assume the materials demand function is strictly increasing in  $\epsilon_{m, it}$  which is normalized to be uniformly distributed on the interval  $[0, 1]$ . I can extend this to the case where labor is chosen prior to choosing material inputs in which case I would include  $l_{it}$  as a state variable in equation (6). I interpret the flexible input demand shocks as a Skorohod disturbance so that they describe the ranking of the firm's across the conditional input distributions.

**Assumption 3.3.** *(Flexible Inputs)*

- (a) *The unobserved input demand shocks  $\epsilon_{l, it}$  and  $\epsilon_{m, it}$  are iid across firms and time*

- (b)  $\epsilon_{l,it}$  and  $\epsilon_{m,it}$  follow a standard uniform distribution independent of  $(a_{it}, k_{it}, \omega_{it})$   
(c)  $\tau \rightarrow Q_t(l_{it}|a_{it}, k_{it}, \omega_{it}, \tau)$  and  $Q_t(m_{it}|a_{it}, k_{it}, \omega_{it}, \tau)$  are strictly increasing on  $(0, 1)$

Capital accumulates to the following generalized law of motion

$$K_{it} = \kappa(K_{it-1}, I_{it-1}, v_{it-1}) \quad (7)$$

where  $I_{it-1}$  denotes firm investment in the prior period. Under this specification, capital is determined in period  $t - 1$ . I introduce a random error term  $v_{it-1}$  which eliminates the deterministic relationship of capital with respect to previous period state and decision variables. I will show later that this step is crucial for my nonparametric identification result which is used by Hu *et al.* (2019). Similar to Hu *et al.* (2019) I specify a data-generating process for investment given by  $I_{it}^* = \iota_t(\omega_{it}, k_{it}, \zeta_{it})$  where  $I_{it} = \max\{0, I_{it}^*\}$ . This reflects that investment demand is often censored because firms do not always make new capital investments in each period. Here,  $\zeta_{it}$  is an unobservable that affects firm's investment demand. I assume that investment demand is strictly increasing in this component. A possible interpretation for this comes from a dynamic model of firm investment which is a slight modification of Ericson and Pakes (1995). In each period, a firm chooses investment to maximize its discounted future profits:

$$I_{it}^* = \iota_t(a_{it}, K_{it}, \omega_{it}, \zeta_{it}) = \operatorname{argmax}_{I_{it} \geq 0} \left[ \Pi_t(a_{it}, K_{it}, \omega_{it}, \zeta_{it}) - c(I_{it}) + \beta \mathbb{E}[V_{t+1}(a_{it+1}, K_{it+1}, \omega_{it+1}, \zeta_{it+1}) | \mathcal{I}_t] \right], \quad (8)$$

where  $\pi_t(\cdot)$  is current period profits as a function of the state variables and an unobservable demand shock  $\zeta_{it}$ . These are shocks to a firm's product demand which are privately observed by each firm and i.i.d across  $i$  and  $t$ . I assume these shocks are independent from the firm's state variables. Current costs to investment are given by  $c(I_t)$  and  $\beta$  is the firm's discount factor. Pakes (1994) provides specific conditions for which the investment policy function is strictly increasing in its unobservable components. Without loss of generality, I normalize  $\zeta_{it} \sim U[0, 1]$ . This dynamic model can be augmented with a selection rule similar to Olley and Pakes (1996). Since I use an unbalanced panel in my application, selection bias is possible and I show how to extend my estimation procedure in the presence of an optimal exit decision rule. To summarize the restrictions on the capital process and investment, I assume the following:

**Assumption 3.4.** (*Capital Accumulation and Investment*)

- (a) The unobserved investment demand shocks  $\zeta_{it}$  is iid across firms and time  
(b)  $\zeta_{it}$  follows a standard uniform distribution independent of  $(a_{it}, k_{it}, \omega_{it})$   
(c) The production shock  $\eta_{it}$  and  $\zeta_{it}$  are independent conditional on  $(a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it})$ . In addition,  $v_{it}$  is independent of  $\eta_{it}$  conditional on  $(a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it})$   
(d)  $\tau \rightarrow Q_t(I_{it}|a_{it}, K_{it}, \omega_{it}, \tau)$  is strictly increasing on  $(0, 1)$

## 4 Identification

In this section I show that the conditional densities corresponding to the production function, productivity, input decisions and investment are nonparametrically identified using Hu and Schennach (2008). To show this, I introduce some new notation. Let  $Z_t = (l_t, k_t, m_t, k_{t+1})$  denote conditioning variables where I have dropped the  $i$  subscript for convenience. Assume the following:

**Assumption 4.1.** (*Conditional Independence*):

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

The first equality of Assumption 4.1 states that conditional on  $\omega_t$  and  $Z_t$ ,  $y_{t+1}$  and  $I_t$  do not provide any additional information about  $y_t$ . The second equality states that conditional on  $\omega_t$  and  $Z_t$ ,  $y_{t+1}$  does not provide any additional information about  $I_t$ . These are satisfied by mutual independence assumptions on  $\eta_t$  and  $\zeta_t$  conditional on  $(\omega_t, k_t, l_t, m_t)$  and the fact that  $\eta_{it}$  is assumed to be conditionally independent over time. The next assumption is more technical.

**Assumption 4.2.** (*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. Consider the operator  $L_{y_t|Z_t, \omega_t}$ , following Hu and Schennach (2008), injectivity of this operator can be interpreted as its corresponding density  $f_{y_t|Z_t, \omega_t}(y_t|Z_t, \omega_t)$  having sufficient variation in  $\omega_t$  given  $Z_t$ . This assumption is often phrased as completeness condition in the nonparametric IV literature on the density  $f_{y_t|k_t, \omega_t}(y_t|Z_t, \omega_t)$ . More formally, for a given  $Z_t \in \text{Supp}(Z_t)$

$$\int f_{y_t|Z_t, \omega_t}(y_t|Z_t, \omega_t)g(\omega_t)d\omega_t = 0 \quad (9)$$

for all  $y_t$  implies  $g(\omega_t) = 0$  for all  $\omega_t$

For injectivity of the second operator  $L_{y_{t+1}|Z_t, \omega_t}$  it is intuitive to consider  $y_{t+1}$  having sufficient variation for different values of  $\omega_t$  given  $Z_t$ . Since productivity is specified as an AR(1) process and is highly persistent over time, this assumption is intuitive.

This assumption is more restrictive than that of Hu *et al.* (2019). Since their model is separable in  $\omega_t$  they are able to utilize convolution type arguments which require mild conditional independence assumptions as well as assumptions regarding characteristic functions. Injectivity is the price we pay for assuming a more general production function specification. I also require two additional assumptions for uniqueness of the identification strategy.

**Assumption 4.3.** (*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 and is satisfied if there is conditional heteroskedasticity in  $f_{I_t|\omega_t, Z_t}$  or if any functional of its distribution is strictly increasing in  $\omega_t$ . For example, this assumption is satisfied if  $E[I_t|\omega_t, Z_t]$  is strictly increasing in  $\omega_t$  which is similar to the invertibility conditions required in Olley and Pakes (1996). This assumption is problematic if investment is censored, in which case  $E[I_t|\omega_t, Z_t]$  can only be inverted for positive values of investment. In this case we can consider for some. Therefore, I can assume  $E[\mathbb{1}\{I_{it} > 0\}|\omega_t, Z_t]$  as the functional that satisfies the monotonicity assumption. Since I will estimate input decision rules in my model using quantile regression, correcting for censoring is straightforward as I will show later. However, in my application censoring is not an issue since I use publicly listed firms, most of which who report positive capital expenditure in their income statements.

**Assumption 4.4.** (*Normalization*): There exists a functional  $\Gamma$  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, it is sufficient to consider a known function of the data distribution as shown by Arellano and Bonhomme (2016). For a general nonseparable panel model this assumption is satisfied if  $E[y_t|\omega_t, Z_t]$  is strictly increasing in  $\omega_t$ , then one could normalize  $\omega_t = E[y_t|\omega_t, Z_t]$ . In my empirical application, we use a non-separable translog production function. In this case our normalization can be achieved by  $E[y_t|\omega_t, 0] = \omega_t$ . With these assumptions I can now state the first part of our identification results.

**Theorem 4.1.** Under Assumptions 4.1, 4.2, 4.3, and 4.4, given the observed density  $f_{y_t, I_t|y_{t+1}, Z_t}$ , the equation

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

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

*Proof:* See Appendix.

This result identifies the conditional density of output and investment. It also identifies the marginal distribution for productivity since  $f_{\omega_t} = \int f_{\omega_t|y_{t+1}, Z_t} f_{y_{t+1}, Z_t} d(y_{t+1}, Z_t)$  where  $f_{y_{t+1}, Z_t}$  is known from the observed data. It also identifies the input decision rules for the labor and material inputs. We need additional results to identify the productivity process  $f_{\omega_t|\omega_{t-1}}(\omega_t|\omega_{t-1})$  and the initial condition  $f_{\omega_{t-1}}(\omega_{t-1})$ . The requirements for identifying these distributions are different if the production function is stationary or non-stationary. Therefore I claim two separate results.

**Corollary 4.1.** (*Stationarity*): Suppose that the production function is stationary i.e.  $f_{y_t|\omega_t, Z_t} = f_{y_1|\omega_1, Z_1} \forall t \in \{1, \dots, T\}$ . Then, under Assumptions 4.1, 4.2, 4.3, and 4.4, the observed density  $f_{y_t, I_t|y_{t+1}, Z_t}$  uniquely determines the density  $f_{\omega_{t+1}|\omega_t}$  and the initial condition  $f_{\omega_t}$  for any  $t \in \{1, \dots, T-1\}$

*Proof:* See Appendix.

**Corollary 4.2.** (*Non-Stationary*): Under Assumptions 4.1, 4.2, 4.3, and 4.4, the observed density  $f_{y_{t+1}, I_{t+1}|y_{t+2}, y_t, Z_{t+1}}$  uniquely determines the density  $f_{\omega_{t+1}|\omega_t}$  and the initial condition  $f_{\omega_t}$  for any  $t \in \{1, \dots, T-2\}$

*Proof:* See Appendix.

The main conclusion of these two corollaries state that under the condition of stationarity, the productivity process can be identified with  $T = 2$  observations per firms whereas under non-stationarity, the productivity process is identified with  $T = 3$  observations per firm. The number of time periods required for identification increases with the length of the auto-regressive process. It is standard to estimate an AR(1) process for productivity. The data requirements are similar to the control function approach where the instrument set often includes secondary lags of inputs. The main identification assumptions are stronger than those in Hu *et al.* (2019), but necessary for the generality of the model I present in this paper.

## 5 Estimation Strategy

I specify the following functional forms for estimation of the model in Section 3.

*Output:*

$$\begin{aligned} Q_t(y_{it}|a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it}, \tau) \\ = \beta_0(\tau) + \beta_a(\tau)a_{it} + \sum_{0 < n_x \leq n} \beta_{n_k, n_l, n_m}(\tau) k_{it}^{n_k} l_{it}^{n_l} m_{it}^{n_m} \omega_{it} \end{aligned} \quad (11)$$

where  $t$  denotes the age of firm  $i$  at time  $t$ ,  $n_x = n_k + n_l + n_m$ . In my application I take  $n_k = n_l = n_m = 2$ . In this case the production function is specified as Translog with first-order interactions of productivity, in a sense, allowing for productivity to affect input usage beyond a Hicks-neutral effect. Therefore, there are a total of 21 parameters to estimate.

*Persistent Productivity:*

I specify productivity to transition according to:

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

Later I consider the case where productivity may evolve endogenously. In this example, firms engage in activities hoping to increase future productivity. Activities such as R&D, exporting activity, or even spill-over effects from other firms are omitted variables in  $\xi_{it}$ . Consider a productivity enhancing activity such as R&D, in logs as

$$r_{it} = R_t(k_{it}, \omega_{it}, \varrho_{it}), \quad (13)$$

where  $\varrho_{it}$  is a shock to R&D that is independent from the state variables and normalized to be  $U[0, 1]$ . It is straightforward to include this process in the identification argument. In this case we could rewrite 12 as

$$\omega_{it} = g_t(\omega_{it-1}, \tau) = \sum_{j=1}^J \rho_j(\tau) \phi_{\omega,j}(\omega_{it-1}, r_{it-1}) \quad (14)$$

where  $\phi_{\omega,j}$  can capture interaction effects between productivity and R&D expenditure. Another extension for the productivity process allows can also control for endogenous exit of firms from the sample. This insight is built from the Olley and Pakes (1996) model which provides an optimal exit rule for firms whose productivity falls below some threshold. In the Appendix, I show how a propensity score can be estimated in my framework and used in the productivity process as an additional control variable in  $\phi_{\omega,j}$ .

*Flexible Inputs:*

I specify the labor input demand equation as follows:

$$Q_t(l_{it}|a_{it}, k_{it}, \omega_{it}, \tau) = \sum_{j=1}^J \alpha_{l,j}(\tau) \phi_{l,j}(a_{it}, k_{it}, \omega_{it}) \quad (15)$$

where  $\phi_{l,j}$  is a tensor product linear sieve. One could use basis functions as Fourier, B-splines, or Hermite polynomials. In practice I approximate this function by a third-order polynomial with interactions in  $\omega_{it}$  and  $k_{it}$  and additive in firm age. I specify the material input demand equation as follows:

$$Q_t(m_{it}|a_{it}, k_{it}, \omega_{it}, \tau) = \sum_{j=1}^J \alpha_{m,j}(\tau) \phi_{m,j}(a_{it}, k_{it}, \omega_{it}) \quad (16)$$

where  $\phi_{m,j}$  is specified similarly as the labor decision rule. In the case where labor inputs are chosen prior to material inputs, one could include  $l_{it}$  as an additional state variable in the material input rule specified above.

*Investment Demand:*

I specify the investment demand equation (in logs) as:

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

where  $\phi_{i,j}$  is specified similarly as the labor and material input decision rule. In the case where investment is censored, I can write

$$I_{it} = \max\{0, I_{it}^*\} = \max\{0, I^*(a_{it}, K_{it}, \omega_{it}; \zeta_{it})\} \quad (18)$$

in which case the conditional quantiles (in logs) can be written

$$Q_\tau(i_{it}|a_{it}, k_{it}, \omega_{it}) = \max\{0, \sum_{j=1}^J \delta_j(\tau) \phi_{i,j}(a_{it}, k_{it}, \omega_{it})\}, \quad (19)$$

due to the equivariance property of quantiles. The censored quantile regression model avoids distributional assumptions in estimation at the cost of computational complexity.



## 6 Implementation

The following conditional moment restrictions hold as an implication of Assumptions 3.1-3.4.

$$\mathbb{E} \left[ \Psi_\tau \left( y_{it} - \beta_0(\tau) - \beta_a(\tau) a_{it} - \sum_{0 < n_x \leq n} \beta_{n_k, n_l, n_m}(\tau) k_{it}^{n_k} l_{it}^{n_l} m_{it}^{n_m} \omega_{it} \right) \middle| a_{it}, k_{it}, l_{it}, m_{it}, t \right] = 0 \quad (20)$$

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

$$\mathbb{E} \left[ \Psi_\tau \left( m_{it} - \sum_{j=1}^J \alpha_{m,j}(\tau) \phi_{m,j}(a_{it}, k_{it}, \omega_{it}) \right) \middle| a_{it}, k_{it}, \omega_{it} \right] = 0 \quad (22)$$

$$\mathbb{E} \left[ \Psi_\tau \left( i_{it} - \sum_{j=1}^J \delta_j(\tau) \phi_{l,j}(k_{it}, \omega_{it}, a_{it}) \right) \middle| a_{it}, k_{it}, \omega_{it} \right] = 0 \quad (23)$$

For  $t \geq 2$ ,

$$\mathbb{E} \left[ \Psi_\tau \left( \omega_{it} - \rho_0(\tau) - \rho_1(\tau) \omega_{it-1} - \rho_2(\tau) \omega_{it-1}^2 - \rho_3(\tau) \omega_{it-1}^3 \right) \middle| \omega_{it-1} \right] = 0, \quad (24)$$

where  $\Psi_\tau(u) = \tau - \mathbb{1}\{u < 0\}$ . Estimating the above conditional moment restrictions is infeasible due to the unobserved productivity component. Therefore, I use the following unconditional moment restrictions and posterior distributions for  $\omega_{it}$  to integrate out the unobserved productivity. To fix ideas, consider the following unconditional moment restriction corresponding to the production function:

$$\mathbb{E} \left[ \int_{\Omega} \Psi_\tau \left( y_{it} - \beta_0(\tau) - \beta_a(\tau) a_{it} - \sum_{0 < n_x \leq n} \beta_{n_k, n_l, n_m}(\tau) k_{it}^{n_k} l_{it}^{n_l} m_{it}^{n_m} \omega_{it} \right) \otimes \begin{pmatrix} 1 \\ a_{it} \\ k_{it} \\ l_{it} \\ m_{it} \\ \omega_{it} \end{pmatrix} g_i(\omega_{it}; \theta(\cdot)) d\omega_{it} \right] = 0 \quad (25)$$

The posterior density  $g_i(\omega_{it}; \theta(\cdot)) = f(\omega_{it} | y_{it}, k_{it}, l_{it}, m_{it}, I_t, \theta(\cdot))$  involves the entire set of model parameters. This is due to the equivalence between the density of a random variable and the inverse of the derivative of its quantile function. Therefore, it is impossible to estimate the model parameters in a  $\tau$ -by- $\tau$  procedure. To eliminate the intractability of this problem, the continuous model parameters are approximated by piece-wise linear splines. For example, the function  $\beta_j(\tau_q)$  is modeled as a piecewise-polynomial interpolating splines on a grid  $[\tau_1, \tau_2], [\tau_3, \tau_4], \dots, [\tau_{Q-1}, \tau_Q]$ , contained in the unit interval and is constant on  $[0, \tau_1]$  and  $[\tau_Q, 1]$ . The intercept coefficient  $\beta_0$  is specified as the quantile of an exponential distribution on  $(0, \tau_1]$  (indexed by  $\lambda^-$ ) and  $[\tau_{Q-1}, 1]$  (indexed by  $\lambda^+$ ). The remaining functional parameters are modeled similarly. I take  $Q = 11$  and  $\tau_q = \frac{q}{Q+1}$ .

The usefulness of the piece-wise linear spline is that the posterior density has a closed form expression without relying on strong distributional assumptions. 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 problem. 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. 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

$$g_i(\omega_{it}; \hat{\theta}^{(s)}) = f(\omega_{it}|y_{it}, a_{it}, k_{it}, l_{it}, m_{it}, I_t; \hat{\theta}^{(s)}) \propto \prod_{t=1}^T f(y_{it}|a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it}; \hat{\beta}^{(s)}) f(l_{it}|a_{it}, k_{it}, \omega_{it}; \hat{\alpha}_l^{(s)}) f(m_{it}|a_{it}, k_{it}, \omega_{it}; \hat{\alpha}_m^{(s)}) \\ \times f(i_{it}|a_{it}, k_{it}, \omega_{it}; \hat{\delta}^{(s)}) \prod_{t=2}^T f(\omega_{it}|\omega_{it-1}; \hat{\rho}^{(s)}) f(\omega_{i1}; \hat{\rho}_1^{(s)})$$

2. *Maximization Step*: For  $q = 1, \dots, Q$ , solve

$$\hat{\beta}(\tau_q)^{(s+1)} = \underset{\beta(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \psi_{\tau_q} \left( y_{it} - \beta_0(\tau_q) - \beta_a(\tau_q) a_{it} - \sum_{0 < n_x \leq n} \beta_{n_k, n_l, n_m}(\tau_q) k_{it}^{n_k} l_{it}^{n_l} m_{it}^{n_m} \omega_{it}^{(m)} \right) \\ \hat{\alpha}_l(\tau_q)^{(s+1)} = \underset{\alpha_l(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \psi_{\tau_q} \left( l_{it} - \sum_{j=1}^J \alpha_{l,j}(\tau_q) \phi_{l,j}(a_{it}, k_{it}, \omega_{it}^{(m)}) \right) \\ \hat{\alpha}_m(\tau_q)^{(s+1)} = \underset{\alpha_m(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \psi_{\tau_q} \left( m_{it} - \sum_{j=1}^J \alpha_{m,j}(\tau_q) \phi_{m,j}(a_{it}, k_{it}, \omega_{it}^{(m)}) \right) \\ \hat{\delta}(\tau_q)^{(s+1)} = \underset{\delta(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \psi_{\tau_q} \left( i_{it} - \sum_{j=1}^J \delta_j(\tau_q) \phi_{l,j}(a_{it}, k_{it}, \omega_{it}^{(m)}) \right) \\ \hat{\rho}(\tau_q)^{(s+1)} = \underset{\rho(\tau_q)}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \psi_{\tau_q} \left( \omega_{it}^{(m)} - \rho_0(\tau_q) - \rho_1(\tau_q) \omega_{it-1}^{(m)} - \rho_2(\tau_q) \omega_{it-1}^{(m)2} - \rho_3(\tau_q) \omega_{it-1}^{(m)3} \right),$$

where  $\psi_\tau(u) = (\tau - \mathbb{1}\{u < 0\})u$  is the “check” function from quantile regression. I specify initial productivity to be normally distributed so that  $\hat{\rho}_1 = (\hat{\mu}_{\omega_1}, \hat{\sigma}_{\omega_1}^2)$ . It is computationally efficient to 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. There are many components to this model that complicates asymptotic inference such as the dimension of the series approximations and the number of knots in the interpolating spline. Therefore, I do not discuss asymptotic properties of the estimator in this paper.

## 7 Application

I apply the estimator to US manufacturers from the Standard and Poors Compustat database. The sample covers publicly traded firms and contains data from their financial statements. I collect a sample between 1990 and 2016 on sales, capital expenditures, property, plant, and equipment, employees, and other expenses to construct output, investment, capital stock, labor, and material inputs. The financial data is deflated using 3-digit deflators from the NBER-CES Manufacturing Industry Database. For firm’s age I use the date the firm first appeared in the database. The sample is unbalanced to avoid artificially reducing the sample size by truncating observations for firms who exit.

I use the stochastic EM algorithm presented in this paper. I run the estimation procedure with 500 random walk Metropolis-Hastings steps and 200 EM steps with  $M = 1$ . I experimented with many different types of proposal distributions for the Metropolis-Hastings step. It is difficult to tune the rate of accepted draws for each firm, however I found that using the optimal variance rule  $(2.38^2)/n_t$ , where  $n_t$  is the number of time-period observations per firm, worked well. The final estimators are used to simulate productivity from its initial conditions and the decision rules for investment, labor and materials. In order to simulate output, I need to specify an accumulation process for capital. The process specified in Equation 7 is flexible and I use it to accumulate capital using the perpetual inventory method with industry-average time-varying depreciation rates. I have found that estimates are similar using different types of capital accumulation processes.

Using the simulated data, I construct a number of structural estimates. For example consider the *average quantile marginal effect* of capital on output:

$$\bar{\beta}_k(\tau) = \mathbb{E} \left[ \frac{\partial Q_t(y_{it} | a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it}, \tau)}{\partial k_{it}} \right]$$

where the expectation is taken with respect to the marginal distributions of  $(a_{it}, k_{it}, l_{it}, m_{it}, \omega_{it})$ . I compute this estimator for each input and the non-Hicks neutral effects. I use the same formulation to compute estimates of average quantile productivity persistence in addition to computing *individual quantile effects* of the production function with respect to inputs (capital) by

$$\beta(\tau_\eta, \tau_k) = \mathbb{E} \left[ \frac{\partial Q_{\tau_\eta}(y_{it} | Q_{\tau_k}(k_{it}), l_{it}, m_{it}, \omega_{it})}{\partial k_{it}} \right]$$

where the expectation is taken with respect to the distributions of  $(a_{it}, l_{it}, m_{it}, \omega_{it})$  conditional on  $k_{it}$  evaluated at different percentiles of capital. I also calculate the individual quantile effects of productivity given by

$$\hat{\rho}(\omega_{it-1}, \tau) = \hat{\rho}_1(\tau) + 2\hat{\rho}_2(\tau)\omega_{it-1} + 3\hat{\rho}_3(\tau)\omega_{it-1}^2$$

Also of interest are the dynamic effects of productivity innovation shocks on inputs. Similar to Hu *et al.* (2019) I estimate how quickly firms respond to shocks to current productivity. This analysis would show whether input decision rules for labor, materials and capital are subject to adjustment frictions. For example, if I find that labor responds positively to increases in productivity, then policies designed to increase productivity will have a faster effect depending on how quickly the firm is able to adjust how many workers it can hire. My model allows me to examine this effect on two different dimensions: the size of the shock to labor demand and the size of the innovation shock. The estimator is given by

$$\hat{l}(\tau_\epsilon, \tau_\xi) = \mathbb{E} \left[ \frac{\partial Q_{\tau_\epsilon}(l_{it} | k_{it}, Q_{\tau_\xi}(\omega_{it} | \omega_{it-1}))}{\omega_{it}} \times \left( \frac{\partial Q_{\tau_\xi}(\omega_{it} | \omega_{it-1})}{\xi_{it}} \right) \right]$$

where  $\partial Q_{\tau_\xi}(\omega_{it} | \omega_{it-1}) / \partial \xi_{it}$  can be approximated by finite differences. In practice I simulate impulse response functions under various innovation shocks to productivity and input demand under some initial conditions which I will discuss in the next section.

## 7.1 Empirical Results

Figure 1: Average Output Elasticities

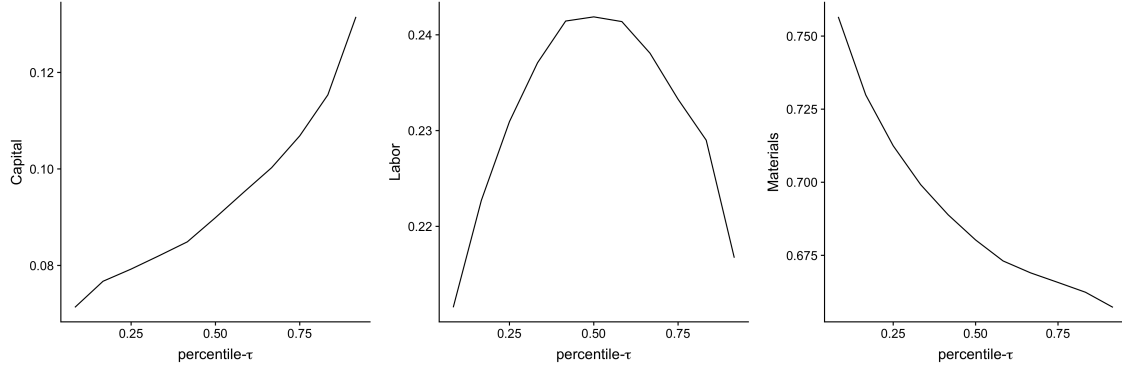


Figure 2: Individual Output Elasticities

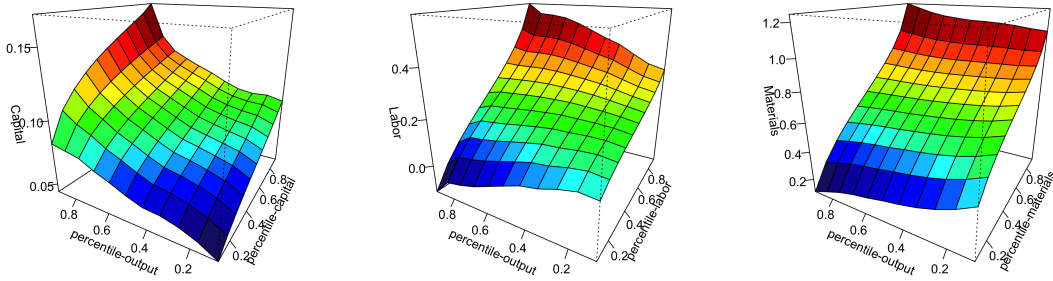


Figure 3: Average Non-Hicks Neutral Elasticities

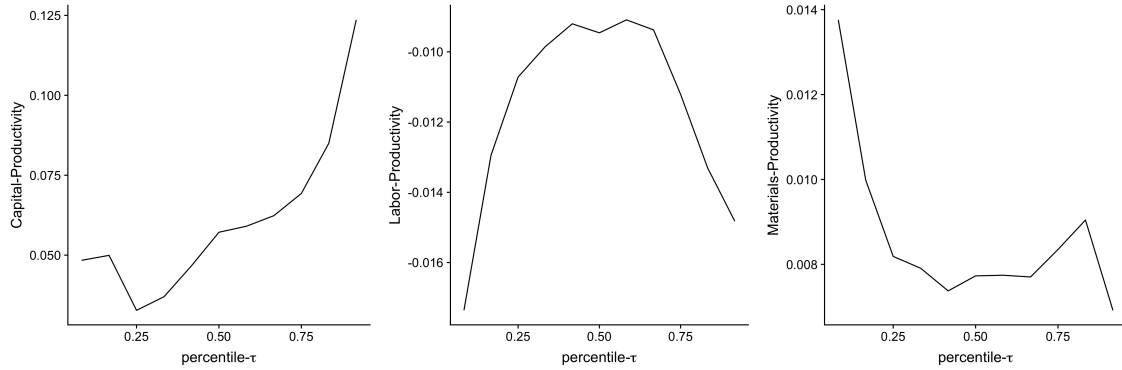


Figure 4: Individual Non-Hicks Neutral Elasticities

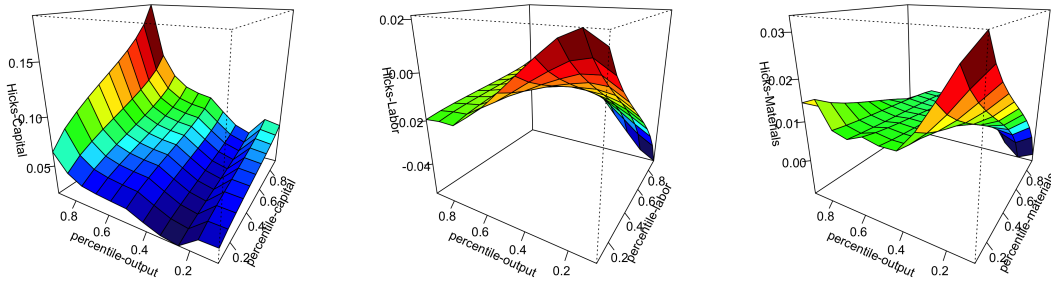


Figure 5: Productivity Persistence

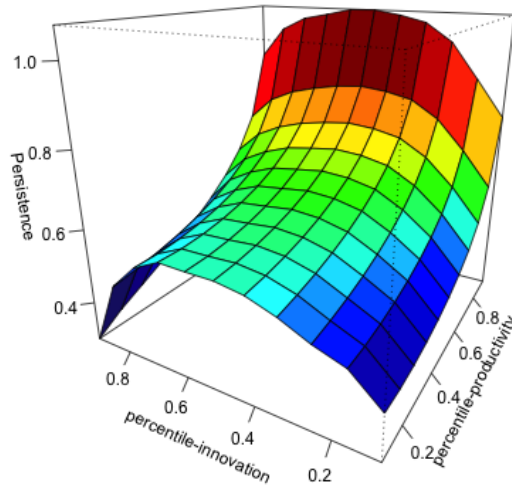


Figure 6: Marginal Productivity of Inputs

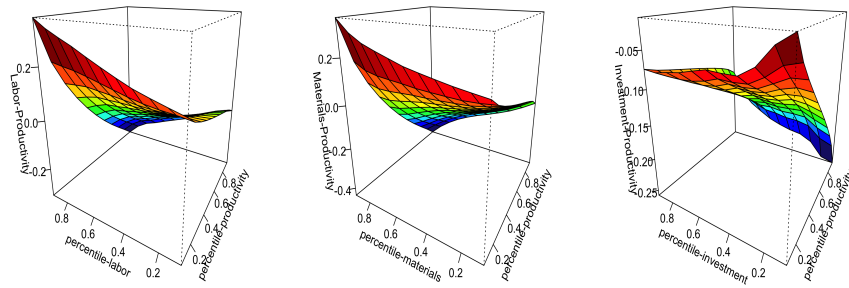


Figure 7: Impulse Response of an Innovation Shock to Labor

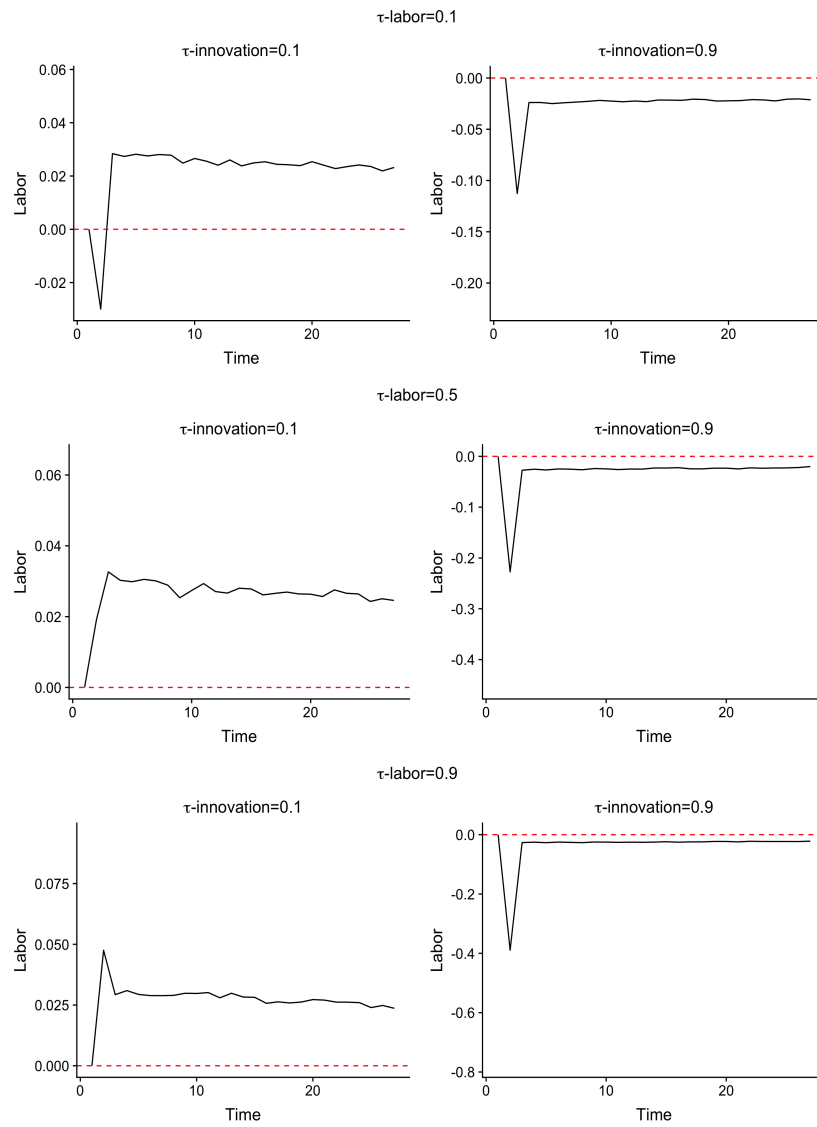


Figure 8: Impulse Response of an Innovation Shock to Materials

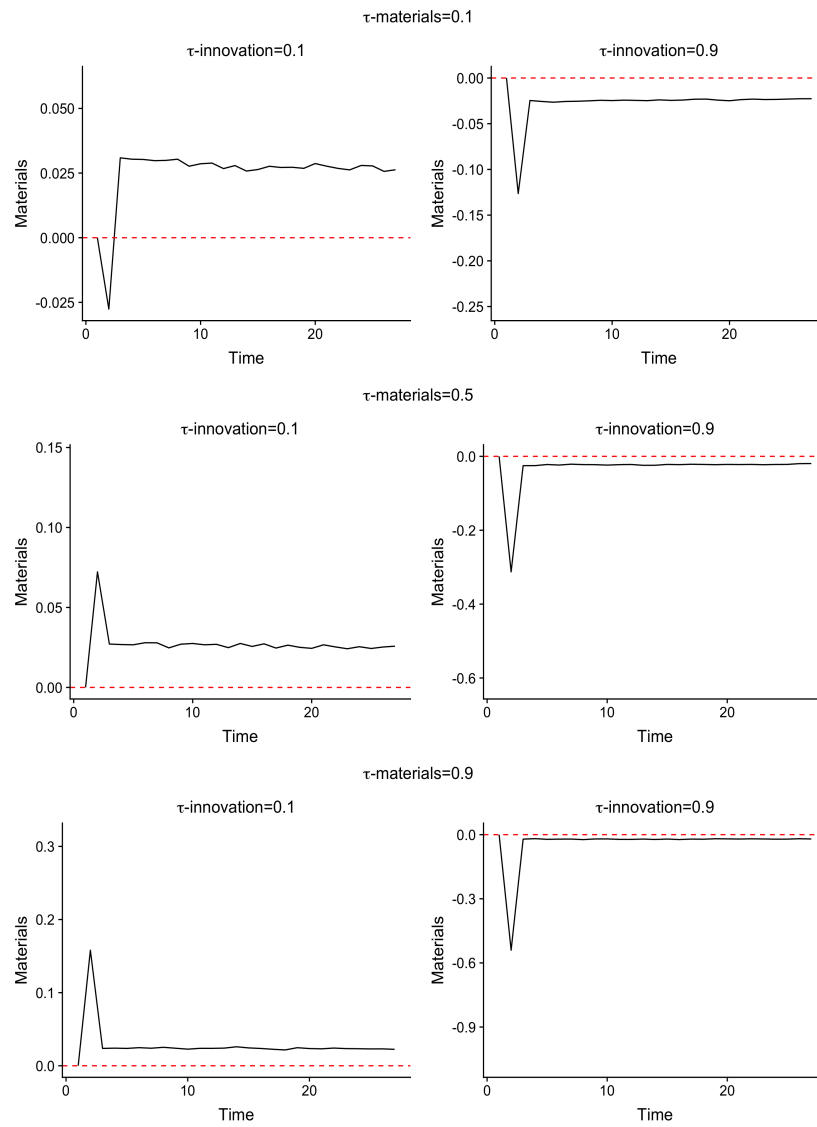
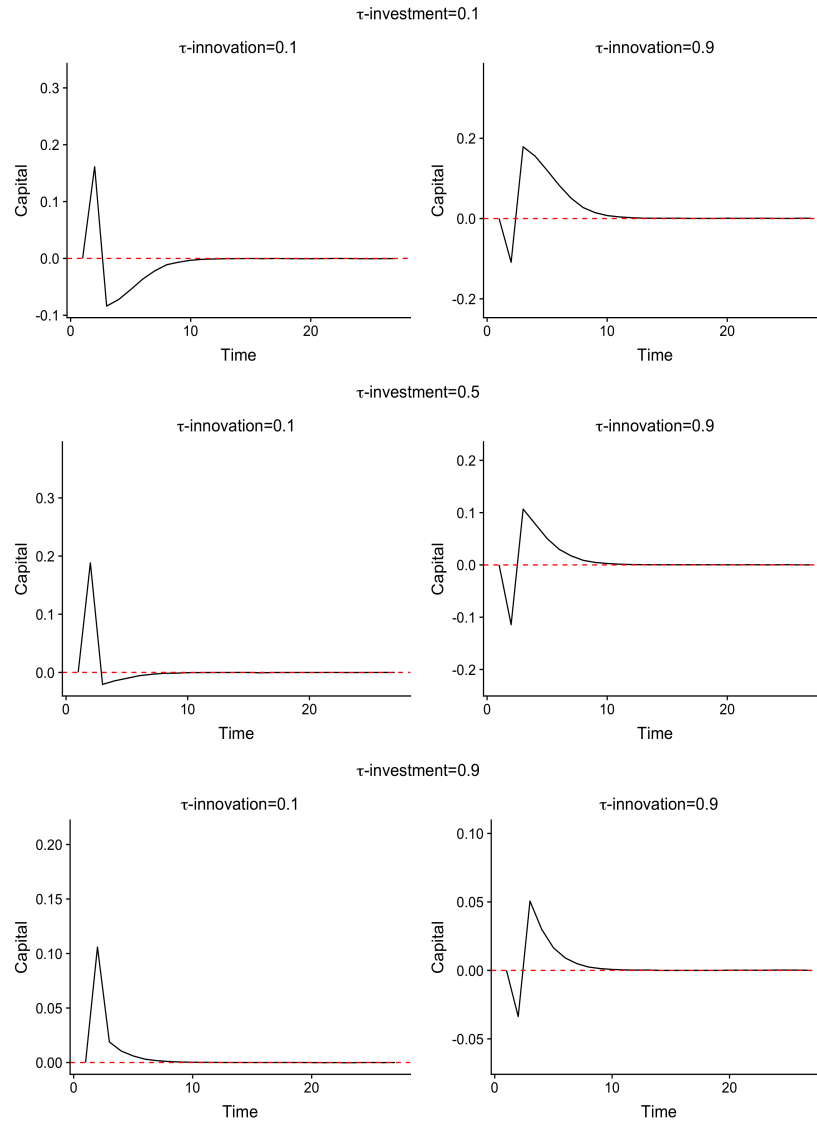




Figure 9: Impulse Response of an Innovation Shock to Capital



## 8 Conclusion

## References

- ACKERBERG, D. A., CAVES, K. and FRAZER, G. (2015). Identification properties of recent production function estimators. *Econometrica*, **83** (6), 2411–2451.
- ARAGON, Y., DAOUIA, A. and THOMAS-AGNAN, C. (2005). Nonparametric frontier estimation: A conditional quantile based approach. *Econometric Theory*, **21** (02).
- ARELLANO, M., BLUNDELL, R. and BONHOMME, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, **85** (3), 693–734.
- and BONHOMME, S. (2016). Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal*, **19** (3), C61–C94.
- BALAT, J., BRAMBILLA, I. and SASAKI, Y. (2018). Heterogeneous firms: skilled-labor productivity and the destination of exports, Working paper.
- CELEUX, G. and DIEBOLT, J. (1985). The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistics Quarterly*, (2), 73–82.
- DORASZELSKI, U. and JAUMANDREU, J. (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies*, **80** (4), 1338–1383.
- DOTY, J. and SONG, S. (2021). Estimating quantile production functions: A control function approach.
- ERICSON, R. and PAKES, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, **62** (1), 53.
- GANDHI, A., NAVARRO, S. and RIVERS, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, pp. 000–000.
- HU, Y., HUANG, G. and SASAKI, Y. (2019). Estimating production functions with robustness against errors in the proxy variables. *Journal of Econometrics*.
- and SCHENNACH, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, **76** (1), 195–216.
- IL KIM, K., PETRIN, A. and SONG, S. (2016). Estimating production functions with control functions when capital is measured with error. *Journal of Econometrics*, **190** (2), 267–279.
- KASAHARA, H., SCHRIMPF, P. and SUZUKI, M. (2017). Identification and estimation of production function with unobserved heterogeneity, Working paper.
- LEVINSOHN, J. and PETRIN, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, **70** (2), 317–341.
- LI, T. and SASAKI, Y. (2017). Constructive identification of heterogeneous elasticities in the cobb-douglas production function, Working paper.
- NIELSEN, S. F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, **6** (3), 457.
- OLLEY, G. S. and PAKES, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, **64** (6), 1263.
- PAKES, A. (1994). Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interactions. In C. A. Sims (ed.), *Advances in Econometrics*, Cambridge University Press, pp. 171–274.