Heterogeneity in Firms: A Proxy Variable Approach to Quantile Production Functions

Justin Doty and Suyong Song

October 5, 2019

Introduction

- Identification issues: optimal input choices are functions of unobserved productivity leads to transmission bias
- Popular control function approaches and issues
  - OP (1994): Simultaneity and selection, investment policy proxy, application to telecommunications industry
  - LP (2003): Intermediate input proxy, gross-output production, application to Chilean manufacturing firms
  - **ACF (2015)**: Value-Added production with material Input Proxy, identification under different DGPs
- Previous approaches have focused on estimates of output elasticities on location of conditional output distribution
- Similarly, productivity heterogeneity are estimated average TFP measurements, there could be considerable heterogeneity not captured by quantile estimates

Ackerberg, Caves, Frazer (2015)

- ACF consider the following value-added production function (subscript i omitted)

$$y_t = \beta_0 + \beta_k k_t + \beta_l l_t + \omega_t + \varepsilon_t \tag{1}$$

Where ω_t is productivity observed by the firm, but unobserved by the researcher (e.g, management quality, expected defect rates, etc.) and ε_t represents iid shocks to production after making input choices at time t With the following assumptions

1. Information Set: $\mathcal{I}_t$ includes current and past productivity shocks, but not future productivity. $\mathbb{E}[\varepsilon_t|\mathcal{I}_t] = 0$
2. First Order Markov: Productivity shocks evolve according to the distribution $p(\omega_t|\omega_{t-1})$
3. Capital Accumulation: $k_t = \kappa(k_{t-1}, l_{t-1})$
4. Scalar Unobservability: $m_t = \tilde{f}_t(k_t, l_t, \omega_t)$
5. Strict Monotonicity: $\tilde{f}_t(k_t, l_t, \omega_t)$ is strictly increasing in ω_t

- Given these assumptions intermediate input demand can be inverted $\omega_t = \tilde{f}_t^{-1}(k_t, l_t, m_t)$ and substituted into the production function

$$y_t = \beta_0 + \beta_k k_t + \beta_l l_t + \tilde{f}_t^{-1}(k_t, l_t, m_t) + \varepsilon_t = \tilde{\Phi}(k_t, l_t, m_t) + \varepsilon_t \quad (2)$$

The first stage moment conditions

$$\mathbb{E}[\varepsilon_t | \mathcal{I}_t] = \mathbb{E}[y_t - \tilde{\Phi}(k_t, l_t, m_t) | \mathcal{I}_t] = 0 \quad (3)$$

For the second stage moment condition we need the following from assumption 2

$$\omega_t = \mathbb{E}[\omega_t | \mathcal{I}_{t-1}] + \xi_t = g(\omega_{t-1}) + \xi_t \quad (4)$$

Where ξ_t represents innovation in productivity and by construction $\mathbb{E}[\xi_t | \mathcal{I}_{t-1}] = 0$

Ackerberg, Caves, Frazer (2015)

- So we arrive at the second stage moment condition

$$
\begin{aligned}
&\mathbb{E}[\xi_t + \varepsilon_t] \\
&= \mathbb{E}[y_t - \beta_0 - \beta_k k_t - \beta_l l_t \\
&\quad - g(\tilde{\Phi}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}) - \beta_0 - \beta_k k_{t-1} - \beta_l l_{t-1}) | \mathcal{I}_{t-1}] = 0
\end{aligned} \tag{5}
$$

- Estimate by approximating $\tilde{\Phi}(k_t, l_t, m_t)$ with a high-order polynomial
- A simple model, $\omega_t = \rho\omega_{t-1} + \xi_t$
- Translate into a set of unconditional moment restrictions

$$
\begin{aligned}
&= \mathbb{E}\Bigg[y_t - \beta_0 - \beta_k k_t - \beta_l l_t \\
&\quad - \rho(\tilde{\Phi}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}) - \beta_0 - \beta_k k_{t-1} - \beta_l l_{t-1}) \otimes \begin{pmatrix} 1 \\ k_t \\ l_{t-1} \\ \tilde{\Phi}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}) \end{pmatrix} \Bigg] = 0
\end{aligned} \tag{6}
$$

Quantile Production Function

Consider the following production function

$$Q_\tau(y_t|k_t, l_t) = \beta_k(\tau)k_t + \beta_l(\tau)l_t + \omega_t \tag{7}$$

- l_t is a freely variable input like labor and k_t denotes the state variable as capital input
- ω_t is an unobserved state variable to the econometrician that impacts firm's optimal input decisions
- We allow for inputs to affect quantiles of the output distribution
- We stick with the value-added production function for two main reasons:
  1. The OP dynamic programming problem is too complicated for a proxy in a quantile production function
  2. We avoid identification issues of using intermediate input demand in a gross-output production function (Gandhi, Navarro, Rivers, 2016)

Quantile Production Function: Assumptions

We maintain assumptions that are similar to ACF (2015)

1. Information Set: $\mathcal{I}_t$ includes current and past productivity shocks, but not future productivity. $Q_\tau(\varepsilon_t|\mathcal{I}_t) = 0$
2. Firm's productivity follows a a Quantile Autoregressive(1) Process

$$Q_\alpha(\omega_t|\mathcal{I}_{t-1}) = Q_\alpha(\omega_t|\omega_{t-1}) = \rho(\alpha)\omega_{t-1} \tag{8}$$

   where $Q_\alpha(\xi_t|\mathcal{I}_{t-1}) = 0$
3. Capital Accumulation: $k_t = \kappa(k_{t-1}, l_{t-1})$
4. Scalar Unobservability: $m_t = \tilde{f}_t(k_t, l_t, \omega_t; \tau)$
5. Strict Monotonicity: $\tilde{f}_t(k_t, l_t, \omega_t; \tau)$ is strictly increasing in $\omega_t, \forall \tau \in (0, 1]$

Quantile Production Function: Estimation

We maintain assumptions that are similar to ACF (2015)

- Given these assumptions we invert intermediate input demand $\omega_t = f^{-1}(k_t, l_t, m_t; \tau)$ and substitute into the production function

$$Q_\tau(y_t|k_t, l_t) = \beta_l(\tau)l_t + \beta_k(\tau)k_t + f_t^{-1}(k_t, l_t, m_t; \tau) = \tilde{\Phi}_t(k_t, l_t, m_t; \tau) \tag{9}$$

- The first stage moment restrictions comes from our assumption $Q_\tau(\varepsilon_t|\mathcal{I}_t) = 0$

$$Q_\tau(\varepsilon_t|\mathcal{I}_t) = \mathbb{E}[\mathbb{1}\{y_t - \tilde{\Phi}_t(k_t, l_t, m_t; \tau) \leq 0\} - \tau|\mathcal{I}_t] = 0 \tag{10}$$

- We estimate the parameters of the unconditional quantile restrictions implied by (10) using the feasible estimators of the smoothed moments (de Castro, Galvao, Kaplan, and Liu, 2018)

Smoothed GMM for Quantile Models

- Develops theory for feasible estimators of parameters in general conditional quantile restrictions that include non-linear IVQR
- Borrowing their notation, they consider a non-linear conditional quantile model

  $$Q_\tau(\Lambda(Y_i, X_i, \beta_0\tau)|Z_t) = 0 \tag{11}$$

  Which is represented by the conditional moment restriction

  $$\mathbb{E}[\mathbb{1}\{\Lambda(Y_i, X_i, \beta_{0\tau}) \leq 0\} - \tau|Z_t] = 0 \tag{12}$$

- They estimate the parameters of the unconditional moment restrictions implied by (12) and a smoothed version of the discontinuous indicator function
- Establish local identification of $\beta_{0\tau}$ as well as large-sample properties, consistency and asymptotic normality
- This will is a useful application to the possibly non-linear approximation in the first step, and the (non-linear) second step (next slide)

Quantile Production Function: Estimation

- The first step moment produces estimates of $\tilde{\Phi}_t(k_t, l_t, m_t; \tau)$
- To write the second stage moment restrictions we recall $Q_\tau(\varepsilon_t|\mathcal{I}_{t-1}) = 0$ and $Q_\alpha(\xi_t|\mathcal{I}_{t-1}) = 0$
- One caveat of forming the second stage restrictions is that only under certain assumptions:

$$Q_\tau(\xi_t + \varepsilon_t|\mathcal{I}_{t-1}) = Q_\tau(\xi_t|\mathcal{I}_{t-1}) + Q_\tau(\varepsilon_t|\mathcal{I}_{t-1}) = 0 \tag{13}$$

  Definition: *Comonotonicity*

  Two random variables $X, Y : \Omega \to \mathbb{R}$ if there exists a random variable $Z : \Omega \to \mathbb{R}$ and increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$
- An important consequence of comonotonicity is that the sum of the τ-quantiles of X and Y can be written as the τ-quantile of their sum

Quantile Production Function: Estimation

- Essentially, assuming comonotonicty requires a strong assumption about the random shocks ξ_t and ε_t being entirely dependent on some increasing transformation of some other random shock, say η_t
- Can firm-specific innovations in productivity and the ex-post output shocks be entirely determined by some other random shock? Perhaps
- This also restricts us to look at when the quantiles of the conditional output distribution and the individual-firm productivity process match
- Need to explore this assumption further for economic context
- Is it possible to re-parameterize the model so that the variables are comonotonic?

Quantile Production Function: Estimation

- With comonotonicity, the second stage moment restrictions are

$$
\begin{aligned}
&Q_\tau(\xi_t + \varepsilon_t | \mathcal{I}_{t-1}) \\
&= \mathbb{E}[\mathbb{1}\{y_t - \beta_k(\tau)k_t - \beta_l(\tau)l_t \\
&\quad - \rho(\tau)(\tilde{\Phi}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}; \tau) - \beta_k(\tau)k_{t-1} - \beta_l(\tau)l_{t-1}) \leq 0\} - \tau | \mathcal{I}_{t-1}] = 0
\end{aligned}
\tag{14}
$$

Which is translated to

$$
\begin{aligned}
&= \mathbb{E}\Bigg[\Bigg(\mathbb{1}\{y_t - \beta_k(\tau)k_t - \beta_l(\tau)l_t \\
&\quad - \rho(\tau)(\hat{\tilde{\Phi}}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}; \tau) - \beta_k(\tau)k_{t-1} - \beta_l(\tau)l_{t-1}) \leq 0\} - \tau\Bigg) \\
&\otimes \begin{pmatrix} k_t \\ l_{t-1} \\ \hat{\tilde{\Phi}}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}) \end{pmatrix}\Bigg] = 0
\end{aligned}
\tag{15}
$$

Quantile Production Function: Estimation

- Possible issues with sample selection bias of $\hat{\Phi}_{t-1}(k_{t-1}, l_{t-1}, m_{t-1}; \tau)$ entering the second stage moment restriction
- Searching for solutions similar to Lee (2007) and Buchinksy (1998)
- Estimation (like in ACF (2015)) may have the "spurious minimum" problem
- Can be addressed with solutions proposed by Kim and Luo (2018) using additional instruments or a sequential search algorithm

Quantile Production Function: Simulation

- Simulations follow a location-scale model of ACF (2015) original set of DGPs
- Parameters are chosen to match a couple of key moments in the Chilean data used by LP
- Productivity follows a ARCH process
- Firms make optimal choices of investment in the capital stock to maximize the expected discounted value of future profits
- Convex capital adjustment costs
- Labor input l_t is chosen either at t or $t-b$ (if this is the case then labor is chosen only knowledge of ω_{t-b})

Quantile Production Function: Simulation

- Production function is assumed Leontief in materials

$$y_t = \beta_k k_t + \beta_l l_t + \omega_t + (\gamma_0 + \gamma_1 k_t)\varepsilon_t \quad (16)$$

- $\beta_k = 0.4$, $\beta_l = 0.6$, $\gamma_0 = 1$, $\gamma_1 = \frac{1}{2}$
- Productivity follows an autoregressive conditional heteroskedastic process

$$\omega_t = \rho\omega_{t-1} + \xi_t \quad (17)$$

- $\text{Var}(\xi_t) = \lambda_0 + \lambda_1 \text{Var}(\xi_{t-1})$ where $\rho = 0.7$, $\lambda_0 = 0.2$, $\lambda_1 = 0.5$
- Also allow a wage process for each firm determined by an AR(1) process
- Labor is chosen at time $t - b$ where $b = 0.5$

Quantile Production Function: Simulation

Table: No Heteroskedasticity

τ	β_k		β_l		ρ	
0.1	0.399	(0.0293)	0.601	(0.0174)	0.699	(0.017)
0.2	0.4	(0.023)	0.599	(0.0141)	0.7	(0.0143)
0.3	0.399	(0.0224)	0.6	(0.0133)	0.7	(0.0135)
0.4	0.4	(0.022)	0.599	(0.0126)	0.7	(0.0131)
0.5	0.399	(0.0217)	0.6	(0.0126)	0.7	(0.0127)
0.6	0.4	(0.0217)	0.6	(0.0124)	0.699	(0.0128)
0.7	0.4	(0.0218)	0.6	(0.0125)	0.699	(0.0129)
0.8	0.399	(0.0245)	0.601	(0.0142)	0.698	(0.014)
0.9	0.401	(0.0291)	0.6	(0.0166)	0.699	(0.0175)

Table: 1000 replications. Standard deviations reported are of parameter estimates across the 1000 replications

Quantile Production Function: Simulation

Table: Location Scale Model $(1 + \frac{k_{it}}{2})$, No ARCH Process

τ	β_k		β_l		ρ	
0.1	0.028	(0.0885)	0.597	(0.0271)	0.746	(0.0337)
0.2	0.176	(0.052)	0.598	(0.0228)	0.729	(0.0255)
0.3	0.268	(0.0407)	0.598	(0.0211)	0.717	(0.0229)
0.4	0.339	(0.0361)	0.599	(0.0212)	0.708	(0.0216)
0.5	0.399	(0.0344)	0.6	(0.0207)	0.699	(0.0208)
0.6	0.455	(0.0347)	0.602	(0.0208)	0.691	(0.021)
0.7	0.511	(0.0379)	0.603	(0.0223)	0.683	(0.0221)
0.8	0.574	(0.0475)	0.605	(0.0275)	0.672	(0.0308)
0.9	0.656	(0.0578)	0.607	(0.0318)	0.661	(0.0301)

Table: 1000 replications. Standard deviations reported are of parameter estimates across the 1000 replications

Quantile Production Function: Simulation

- Productivity can be estimated in the standard way

$$p_t = \exp(\omega_t + \varepsilon_t) = \exp(y_t - \hat{\beta}_k(\tau)k_t - \hat{\beta}_l(\tau)l_t) \qquad (18)$$

- Aggregate productivity can be calculated as the share-weighted average of firm-level productivity, using firm-level output shares as weights at time t (OP, 1996)
- Can also construct measures of dispersion as in Syverson (2004), Hsieh and Klenow (2009), Gandhi, Navarro, Rivers (2014), and many others
- Example, 75/25, 90/10 ratio tells us with the same amount of inputs how much more productive a 75th percentile plant is over a 25th percentile plant
- How do we compare dispersion of productivity from our estimates? Interpretation?

Quantile Production Function: Simulation