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

Justin Doty and Suyong Song

July 24, 2020

Introduction

- Identification issues: optimal input choices are functions of unobserved productivity leads to transmission bias
- Popular control function approaches
  - Olley and Pakes (1996): Simultaneity and selection, investment policy proxy, application to telecommunications industry
  - Levinsohn and Petrin (2003): Intermediate input proxy, application to Chilean manufacturing firms
  - Ackerberg, Caves, and Frazer (2015): Structural value-Added production with intermediate input Proxy, identification under different DGPs
- These approaches have focused on estimates of output elasticities on location of conditional firm-size distribution
- There could be considerable heterogeneity not captured by average estimates

Levinsohn and Petrin (2003)

- Brief review of LP approach
- LP consider the following gross output production function

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_\iota \iota_{it} + \omega_{it} + \eta_{it} \tag{1}$$

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

1. Information Set: $\mathcal{I}_{it}$ includes current and past productivity shocks, but not future productivity. $\mathbb{E}[\eta_{it}|\mathcal{I}_{it}] = 0$
2. First Order Markov: Productivity shocks evolve according to the distribution $p(\omega_{it}|\omega_{it-1})$
3. Capital Accumulation: $k_{it} = \kappa(k_{it-1}, l_{it-1})$
4. Scalar Unobservability: $\iota_{it} = \iota_t(k_{it}, \omega_{it})$
5. Strict Monotonicity: $\iota_{it} = \iota_t(k_{it}, \omega_{it})$ is strictly increasing in ω_{it}

Levinsohn and Petrin (2003)

- Given these assumptions intermediate input demand can be inverted $\omega_{it} = \omega_t(k_{it}, \iota_{it})$ and substituted into the production function

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_t(k_{it}, \iota_t) + \eta_{it} = \beta_l l_{it} + \Phi(k_{it}, \iota_{it}) + \eta_{it} \quad (2)$$

  where $\Phi(k_{it}, \iota_{it}) = \beta_0 + \beta_k k_{it} + \beta_\iota \iota_{it} + \omega_t(k_{it}, \iota_{it})$
- $\Phi(k_{it}, \iota_{it})$ can be estimated nonparametrically
- Labor coefficient and other variable coefficients are identified in the first stage
- First stage estimates are $\hat{\beta}_l$ and $\hat{\Phi}(k_{it}, \iota_{it})$
- Let $y^* = y - \hat{\beta}_l l_{it}$ denote the output net of labor contribution

Levinsohn and Petrin (2003)

- Two moment conditions identify β_k and β_ι
- Capital does not respond to the innovation in productivity
- Last period's input choice should not be correlated with the innovation in productivity this period

$$\mathbb{E}[(\eta_{it} + \xi_{it})Z_{it-1}] = \mathbb{E}[(y^*_{it} - \beta_k k_{it} - \beta_\iota \iota_{it} - \hat{\mathbb{E}}[\omega_{it}|\omega_{it-1}])Z_{it-1}] \quad (3)$$

- $\hat{\mathbb{E}}[\omega_{it}|\omega_{it-1}]$ can be estimated using the estimates of ω_{it} from first stage estimates evaluated at (β_k, β_ι)
- Z_{it-1} includes k_{it} and ι_{it-1} as well as additional instruments
- LP minimize GMM criterion function and bootstrap standard errors

Heterogeneous Coefficient Models

- Kasahara, Schrimpf, and Suzuki (2017) propose two methods to estimate a random coefficient production function and find evidence of unobserved heterogeneity beyond a Hicks-neutral technology term
- Balat, Brambilla, and Sasaki (2018) extend the control function approach to multi-dimensional coefficients with an application to international trade
- Li and Sasaki (2017) develop an identification strategy for heterogeneous elasticities using first order conditions for intermediate inputs
- Gandhi, Navarro, and Rivers (2020) consider a nonparametric production function and a new identification argument in critique of OP/LP and explore heterogeneity in resulting TFP estimates
- Dermirer (2020) uses a similar control variable approach to estimate output elasticities and markups. He finds that markup estimates using standard approaches are biased when both heterogeneity and factor-augmenting productivity are ignored

Production Functions and Quantile Regression

- Allowing firm technology to vary over the conditional output distribution is also valid for estimating a production function with unobserved heterogeneity
- Applications of quantile regression are however limited due to simultaneity bias
- Typical IV approaches such as Chernozhukov and Hansen (2005) are also limited due to the normal arguments of using input prices as IVs
- Possible to exploit the panel data structure and use fixed effects or correlated random effects such as Koenker (2004), Lamarche (2010), or Canay (2011)
- The fixed effect shifts the location of the conditional distribution (not the quantiles)
- Incidental Parameters
- Independence of unobserved heterogeneity and productivity

Production Functions and Quantile Regression

- Quantile regression and production function estimation is not entirely new and has been used in the frontier literature
- Production frontier models assume firms are not always efficient, but there is a frontier of maximal production in which they deviate from
- More specifically, stochastic frontier (SFA) models assume this frontier is stochastic (inefficiency + iid noise). These models are often characterized by strong distributional assumptions on these variables
- Firms can vary in production technology for different levels of inefficiency making this model a strong candidate for quantile regression analysis (Bernini, Freo, & Gardini, 2004), (Liu, Laporte, & Ferguson, 2008)
- Some difficulties of this approach is determining which quantiles correspond to the efficient frontier
- We believe we are the first to explore quantile regression in the standard production function literature, while also considering simultaneity bias

A Nonseparable Production Function

We seek to extend the control function approach to a random coefficient value-production function defined as

$$y_{it} = \beta_k(\eta_{it})k_{it} + \beta_l(\eta_{it})l_{it} + \omega_{it} \tag{4}$$

- The constant $\beta_0(\eta_{it})$ is subsumed into ω_{it}
- The interpretations of the variables in the above equation are the same as the LP model introduced earlier
- In this specification, the unobservable ranking variable, η_{it} is responsible for heterogeneity in output conditional on firms using the same input combinations and latent productivity ω_{it}
- If we had valid instruments, say Z_{it} and the conditions of Chernozhukov and Hansen (2005) are met, we could use quantile IV approaches using the identification condition

$$P(y_{it} \leq \beta_0(\tau) + \beta_k(\tau)k_{it} + \beta_l(\tau)l_{it} | k_{it}, l_{it}, Z_{it}) = \tau \tag{5}$$

Quantile Production Functions

- A special case of (4) is the location-scale model

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + (\mu_k k_{it} + \mu_l l_{it} + \mu_\omega \omega_{it})\eta_{it} \tag{6}$$

- So the conditional quantiles of y_{it} are given by

$$Q(y_{it}|k_{it}, l_{it}, \omega_{it}) = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + (\mu_k k_{it} + \mu_l l_{it} + \mu_\omega \omega_{it})F^{-1}(\tau) \tag{7}$$

- Equation (6) is not new to the production function literature and is often used as a model for production risk (Just & Pope, 1978)
- They provide a set of postulates for which the above specification is empirically valid
- The quantile specification could be seen as an extension of the higher-order moment estimation of risk initiated by Antle (1983)

Production Risk and Quantiles

- We do not seek to develop new theories of firm behavior under risk
- If one wanted to use production risk as an explanation of firm heterogeneity then it would be possible to use quantile utility maximization as the framework (Manski, 1988), (Bhattacharya, 2009), (Rostek, 2009), (Chambers, 2007), (Castro & Galvao, 2017)
- It would then be possible to use the first order condition approach to estimate production function parameters using quantile regression techniques
- This approach would be more difficult to synthesize in the proxy variable approach
- Which is why we consider production variation in η_{it} to be an econometric misspecification issue rather than one that has any economic meaning
- Since most econometric tests would not be able to differentiate production risk from misspecification, we believe our interpretation to be equally as valid

Quantile Production Function: Assumptions

- We proceed with this interpretation and consider the production function specified in equation (4)
- We maintain several timing assumptions that are standard in the control function literature

Assumption 1

1. Information Set: $\mathcal{I}_{it}$ includes current and past productivity shocks, but not future productivity. η_{it} is independent of $\mathcal{I}_{it}$
2. Firm's productivity follows an AR(1) process

$$\omega_{it} = g(\omega_{it-1}, \xi_{it}) \tag{8}$$

   where ξ_{it} are iid productivity innovations that is independent of $\mathcal{I}_{it-1}$
3. Capital Accumulation: $k_{it} = \kappa(k_{it-1}, i_{it-1})$
4. Scalar Unobservability: $\iota_{it} = \iota_t(k_{it}, \omega_{it})$
5. Strict Monotonicity: $\iota_{it} = \iota_t(k_{it}, \omega_{it})$ is strictly increasing in ω_{it}

Quantile Production Function: Estimation

- Given these assumptions we invert intermediate input demand $\omega_{it} = \omega_t(k_{it}, \iota_{it})$ and substitute into the production function

$$y_{it} = \beta_l(\eta_{it})l_{it} + \beta_k(\eta_{it})k_{it} + \omega_t(k_{it}, \iota_{it}) = \beta_l(\eta_{it})l_{it} + \Phi(k_{it}, \iota_{it}, \eta_{it}) \quad (9)$$

- Using Assumption 1 we have:

$$P\left(y_{it} \leq \beta_l(\tau)l_{it} + \Phi(k_{it}, \iota_{it}; \tau)\middle|\mathcal{I}_{it}\right) = \tau \quad (10)$$

- If one uses a polynomial approximation to $\Phi_\tau(k_t, \iota_t; \tau)$, $\hat{\beta}_l(\tau)$ and $\hat{\Phi}(k_t, \iota_t; \tau)$ can be estimated using a polynomial quantile regression
- Substituting into the production

$$y_{it} = \beta_k(\eta_{it})k_{it} + \hat{\beta}_l(\tau)l_{it} + g(\hat{\Phi}(k_{it-1}, \iota_{it-1}; \tau) - \beta_k(\eta_{it})k_{it-1}, \xi_{it}) \quad (11)$$

Quantile Production Function: Estimation

We can write a second stage restriction as

$$P\left(y_{it} \leq \beta_k(\tau)k_{it} + \hat{\beta}_l(\tau)l_{it} + g(\hat{\Phi}(k_{it-1}, \iota_{it-1}; \tau) - \beta_k(\tau)k_{it-1}; \tau)|\mathcal{I}_{it-1}\right) = \tau \tag{12}$$

This can be represented by conditional moment restrictions

$$\begin{aligned}\mathbb{E}[\mathbb{1}\{y_{it} - \beta_k(\tau)k_{it} + \hat{\beta}_l(\tau)l_{it} \\ + g(\hat{\Phi}(k_{it-1}, \iota_{it-1}; \tau) - \beta_k(\tau)k_{it-1}; \tau) \leq 0\} - \tau|\mathcal{I}_{it-1}] = 0\end{aligned} \tag{13}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function To estimate the production function parameters we use the unconditional moments

$$\begin{aligned}\mathbb{E}[Z_{it-1}(\mathbb{1}\{y_{it} - \beta_k(\tau)k_{it} + \hat{\beta}_l(\tau)l_{it} \\ + g(\hat{\Phi}(k_{it-1}, \iota_{it-1}; \tau) - \beta_k(\tau)k_{it-1}; \tau) \leq 0\} - \tau)] = 0\end{aligned} \tag{14}$$

where Z_{it-1} includes the instruments used in the firm's information set at time $t-1$

Smoothed GMM for Quantile Models

- The indicator function makes estimation of the production function parameters intractable
- We smooth the indicator function and rely on estimation procedure of Kaplan and Sun (2016)
- Castro, Galvao, Kaplan, and Liu (2018) develop theory for feasible estimators of parameters in general conditional quantile restrictions that include non-linear IVQR
- This approach is computationally attractive compared to approaches such as the MCMC approach proposed by Chernozhukov and Hong (2003)
- To fix notation let
    - Z_{it-1} the set of instruments discussed earlier
    - x_{it} the set of exogenous and endogenous covariates
    - $\Lambda(\cdot)$ denote the residual function that defines the conditional quantile restriction that is known up to $\beta_k(\tau)$

Smoothed GMM for Quantile Models

- The sample analog of (12) can be written as:

$$\hat{M}_n(\beta,\tau) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}\left[\tilde{I}\left(\frac{\Lambda(y_{it}, x_{it}, \beta(\tau))}{h_n}\right) - \tau\right], \tag{15}$$

where h_n is a bandwidth (sequence) and $\tilde{I}(\cdot)$ is a smoothed version of the indicator function $\mathbb{1}\{\cdot \leq 0\}$ used by Horowitz (1998), Whang (2006), and Kaplan and Sun (2016):

$$\tilde{I}(u) = \mathbb{1}\{-1 \leq u \leq 1\}\left[0.5 + \frac{105}{64}\left(u - \frac{5}{3}u^3 + \frac{7}{5}u^5 - \frac{3}{7}u^7\right)\right] + \mathbb{1}\{u > 1\}. \tag{16}$$

The smoothed GMM estimator minimizes a weighted quadratic norm of the smoothed sample moment vector

$$\hat{\beta}_{GMM} = \underset{\beta}{\operatorname{argmin}} \hat{M}_n(\beta,\tau)^\top \hat{W} \hat{M}_n(\beta,\tau). \tag{17}$$

Smoothed GMM for Quantile Models

- The optimal weighting matrix is the estimator of the inverse long-run variance of the sample moment $\hat{W}^* = \bar{\Omega}^{-1} \xrightarrow{p} \Omega^{-1}$
- In simulation and in application we use an estimate of the long-run variance using a truncated kernel with bandwidth choice of Andrews (1991). We find this helps with scaling.
- The asymptotic results for this estimator cannot be applied here due to the semi-parametric nature of the two-step procedure
- The asymptotic results of Ai and Chen (2007) could be applied here.
- This gives us extra conditions to show what rate gives an asymptotically negligible bandwidth. The smoothed sample moments are differentiable so we can take a Taylor expansion around $\beta(\tau)$. We leave this for future research agenda
- We use bootstrap for our empirical application to be discussed later
- Optimal choice of bandwidth is an ongoing research topic for these types of estimators

Quantile Production Function: Simulation

- Simulations follow a location-scale version of LP and ACF using ACF's original set of DGPs
- Parameters are chosen to match a couple of key moments in the Chilean data used by LP
- Productivity follows a first order AR(1) process with persistence $\rho = 0.7$
- 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_{it} is chosen either at t or $t - b$
- Possible optimization error in labor input decision

Monte Carlo Simulation

- Production function is assumed Leontief in materials, the location-scale specification is

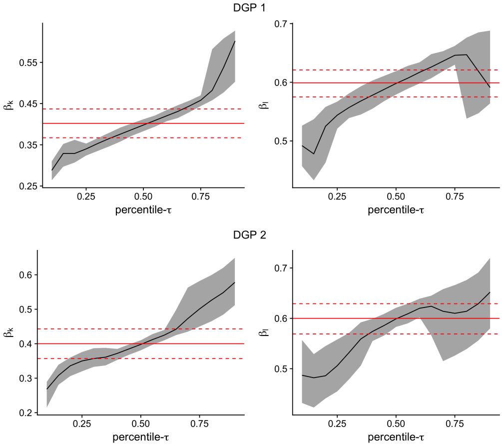
$$y_t = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + (0.7k_{it} + 0.6l_{it} + 0.1\omega_{it})\eta_{it} \qquad (18)$$

  with $\beta_k = 0.4$, $\beta_l = 0.6$
  Simulation 1: ACF Estimated Quantile Production Function
  - Firms face different wages, where the wage process following an AR(1) process
  - Assume labor is chosen at time $t - b$ where $b = 0.5$

  Simulation 2: LP Estimation Quantile Production Function
  - No wage variation across firms and labor chosen at time t
  - Added optimization error in labor
- Both simulations include 2 DGP's for the error term η_t. In DGP 1 $\eta_{it} \sim N(0, 0.1)$ and in DGP 2 $\eta_{it} \sim Laplace(0, 0.1)$

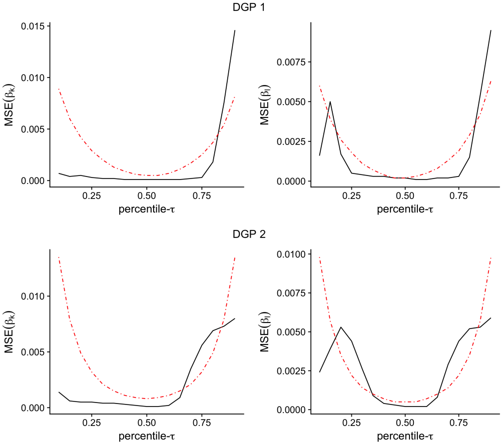
Monte Carlo Simulation 1

Figure: Simulated estimators of $\beta_k(\tau)$ and $\beta_l(\tau)$. Dotted line is ACF estimator



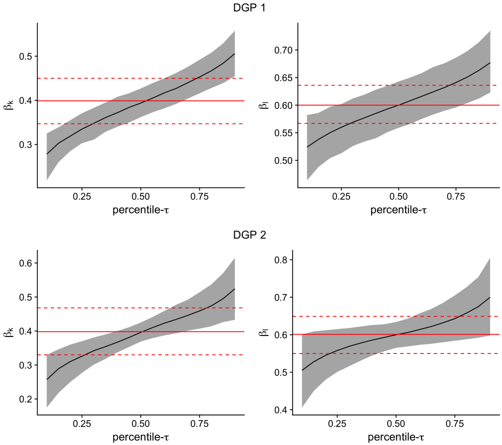
Monte Carlo Simulation 1

Figure: Simulated precision estimators of $\beta_k(\tau)$ and $\beta_l(\tau)$. Dotted line is ACF estimator



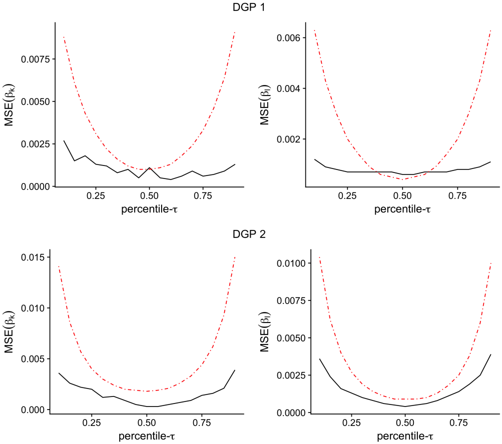
Monte Carlo: Simulation 2

Figure: Simulated estimators of $\beta_k(\tau)$ and $\beta_l(\tau)$. Dotted line is LP estimator



Monte Carlo: Simulation 2

Figure: Simulated precision of estimators of $\beta_k(\tau)$ and $\beta_l(\tau)$. Dotted line is LP estimator



Application

- We apply this estimator to three country manufacturing datasets: US, Chile, and Colombia
- The US data comes from Compustat which contains data on sales, capital expenditure, number of employees, and other items from firm's financial statements
- Compustat has been used in prior production function estimation although the data is of lower quality compared to the Census of Manufactures which is restricted data
- Our Compustat sample contains publicly traded manufacturing firms between 1961 and 2011
- Manufacturing data from Chile is collected by the Instituto Nacional de Estadística (INE) which contains plant level data from 1979 to 1996
- Manufacturing data from Colombia come from the manufacturing census which contains plant level data from 1981 to 1991

US Manufacturing: Compustat

Table: Summary Statistics for the US Compustat Manufacturing Industries

Industry (NAICS code)		1st Qu.	Median	3rd Qu.	Mean	sd
31 (N=3271)	Output	19.05	20.24	21.57	20.3	1.77
	Capital	18.66	20.37	21.76	20.19	2.12
	Labor	17.42	19.08	20.61	19.02	2.21
	Materials	17.96	19.59	21.15	19.54	2.21
32 (N=7207)	Output	15.67	17.04	18.51	17.01	2.05
	Capital	15.65	17.51	19.13	17.31	2.41
	Labor	14.44	16.01	17.57	16.01	2.29
	Materials	14.89	16.53	18.25	16.52	2.37
33 (N=13978)	Output	7.38	8.58	9.8	8.5	1.67
	Capital	6.67	8.29	9.74	8.15	1.95
	Labor	6.01	7.42	8.91	7.48	1.93
	Materials	6.33	7.82	9.29	7.82	1.95
All (N=24456)	Output	18.58	19.78	21.23	19.85	1.79
	Capital	18.14	19.86	21.26	19.67	2.16
	Labor	16.98	18.59	20.13	18.56	2.17
	Materials	17.49	19.12	20.66	19.06	2.2

US Manufacturing: Compustat

Table: Coefficient Estimates and Standard Errors for US Manufacturing Firms

		Capital		Labor		Returns to Scale	
Industry (NAICS code)	τ	Coef.	s.e.	Coef.	s.e.	Coef.	s.e
31	0.10	0.059	0.0308	0.612	0.0328	0.671	0.0369
	0.25	0.099	0.0213	0.558	0.0358	0.657	0.0374
	0.50	0.119	0.0275	0.474	0.0352	0.594	0.0404
	0.75	0.082	0.0247	0.441	0.0356	0.523	0.0401
32	0.10	0.039	0.0274	0.687	0.0422	0.726	0.0466
	0.25	0.079	0.0244	0.627	0.0311	0.706	0.0359
	0.50	0.154	0.0326	0.589	0.0252	0.742	0.0399
	0.75	0.146	0.0307	0.548	0.0230	0.695	0.0356
33	0.10	0.136	0.0366	0.210	0.0579	0.346	0.0624
	0.25	-0.031	0.0343	0.357	0.0401	0.327	0.0539
	0.50	0.174	0.0416	0.446	0.0289	0.620	0.0485
	0.75	0.100	0.0531	0.470	0.0203	0.571	0.0570
All	0.10	-0.091	0.0959	0.365	0.0350	0.274	0.0997
	0.25	0.072	0.0370	0.412	0.0226	0.485	0.0422
	0.50	0.192	0.0498	0.475	0.0181	0.667	0.0547
	0.75	0.420	0.0457	0.485	0.0149	0.905	0.0474

US Manufacturing: Compustat

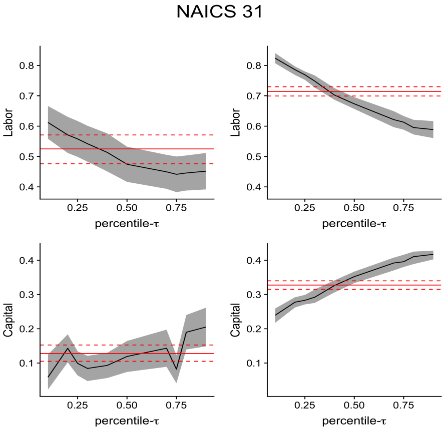


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

US Manufacturing: Compustat

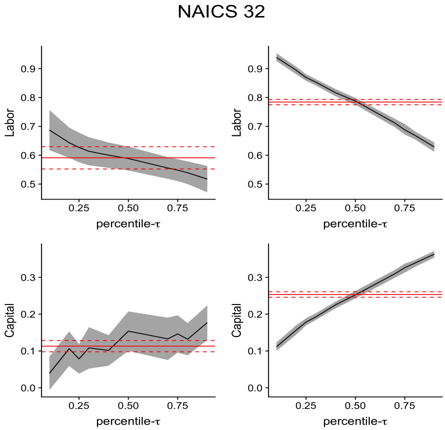


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

US Manufacturing: Compustat

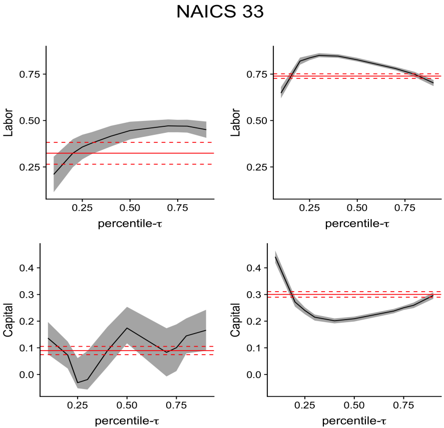


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

US Manufacturing: Compustat

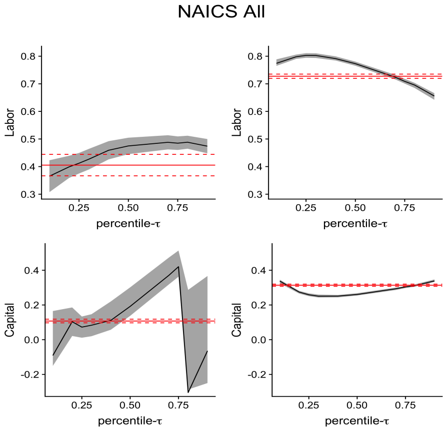


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

US Manufacturing: Compustat

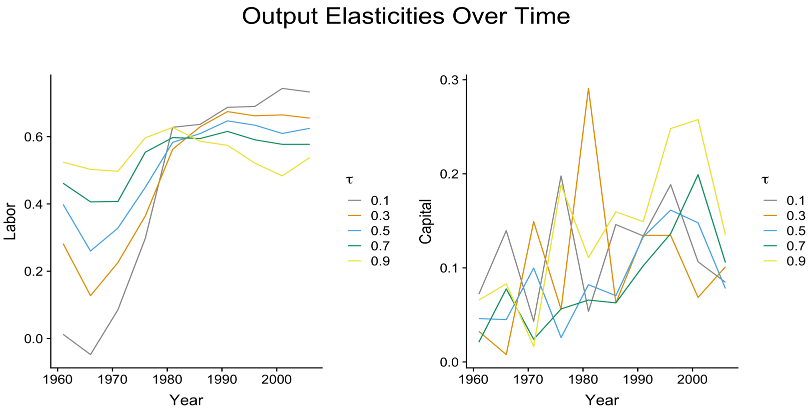


Figure: Estimated values of production function coefficients over time. Estimated at 5 year intervals

US Manufacturing: Compustat

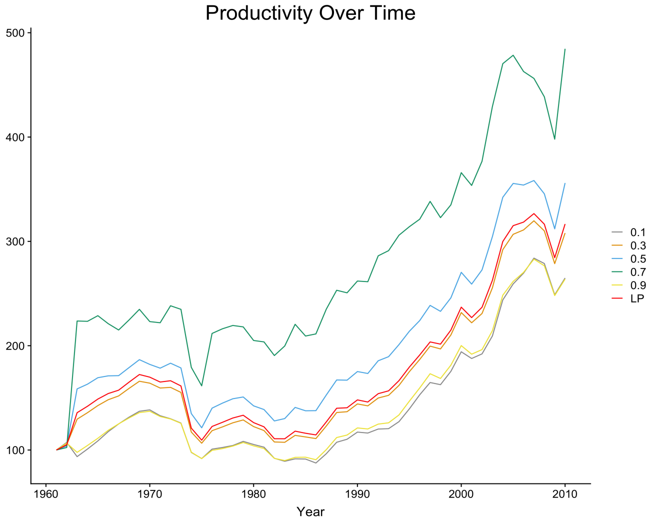


Figure: Estimated average TFP. Base productivity in 1961 is set to 100

Chilean Manufacturing

Table: Summary Statistics for Chile Manufacturing Data

Industry (ISIC code)		1st Qu.	Median	3rd Qu.	Mean	sd
311 (N=13838)	Output	10.21	10.84	12.22	11.36	1.58
	Capital	10.56	11.4	12.4	11.52	1.37
	Labor	10.49	11.4	12.54	11.53	1.43
	Materials	10.38	11.28	12.53	11.56	1.6
381 (N=4311)	Output	6.69	7.66	9.06	8.02	1.98
	Capital	7.52	8.51	9.7	8.65	1.68
	Labor	7.21	8.34	9.56	8.4	1.72
	Materials	7.22	8.35	9.72	8.54	1.92
321 (N=4302)	Output	2.77	3.22	3.91	3.49	0.99
	Capital	2.89	3.47	4.22	3.71	1.08
	Labor	2.94	3.48	4.37	3.69	0.95
	Materials	2.89	3.43	4.28	3.67	1.02
All (N=22451)	Output	9.84	10.46	11.81	10.94	1.56
	Capital	9.91	10.75	11.79	10.86	1.41
	Labor	9.68	10.62	11.75	10.73	1.48
	Materials	9.81	10.68	11.89	10.93	1.62

Chilean Manufacturing

Table: Coefficient Estimates and Standard Errors for Chile Manufacturing Firms

		Capital		Labor		Returns to Scale	
Industry (ISIC code)	τ	Coef.	s.e.	Coef.	s.e.	Coef.	s.e
311	0.10	0.367	0.1054	0.390	0.0438	0.757	0.1137
	0.25	0.214	0.0442	0.411	0.0325	0.565	0.0520
	0.50	0.243	0.0366	0.413	0.0249	1.066	0.0409
	0.75	0.167	0.0428	0.415	0.0220	0.721	0.0480
381	0.10	0.246	0.1005	0.733	0.0687	0.728	0.1215
	0.25	0.249	0.0593	0.624	0.0449	0.616	0.0715
	0.50	0.266	0.0404	0.564	0.0369	0.802	0.0495
	0.75	0.093	0.0599	0.520	0.0390	0.715	0.0709
321	0.10	0.356	0.1264	0.710	0.0761	0.625	0.1419
	0.25	0.320	0.0778	0.628	0.0523	0.979	0.0840
	0.50	0.201	0.0686	0.582	0.0457	0.948	0.0685
	0.75	0.193	0.0586	0.554	0.0418	0.827	0.0681
All	0.10	0.301	0.0469	0.526	0.0205	0.647	0.0482
	0.25	0.315	0.0360	0.502	0.0174	0.912	0.0387
	0.50	0.292	0.0370	0.466	0.0140	0.839	0.0387
	0.75	0.194	0.0404	0.436	0.0150	0.880	0.0434

Chilean Manufacturing

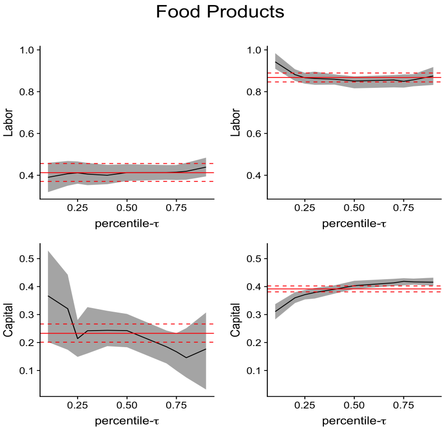


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Chilean Manufacturing

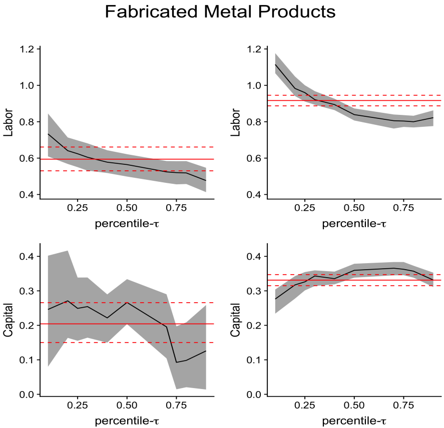


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Chilean Manufacturing

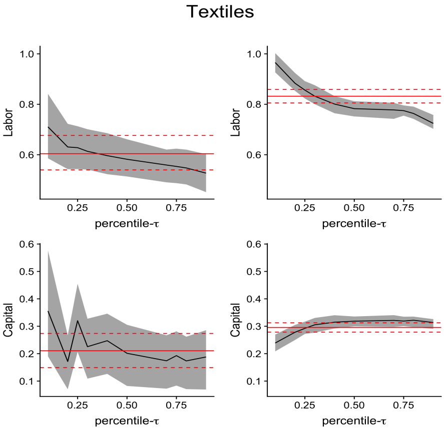


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Chilean Manufacturing

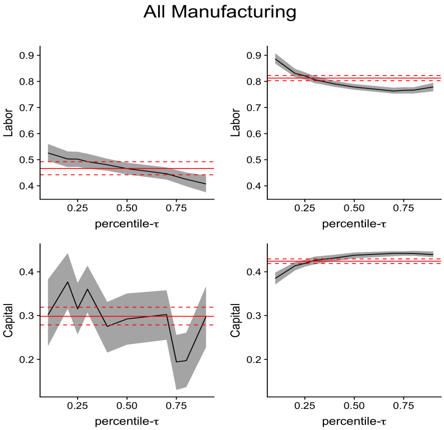


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Colombian Manufacturing

Table: Summary Statistics for Colombia Manufacturing Data

Industry (ISIC code)		1st Qu.	Median	3rd Qu.	Mean	sd
311 (N=13215)	Output	9.03	10.21	11.59	10.42	1.8
	Capital	8.69	9.37	10.22	9.49	1.18
	Labor	8.52	9.3	10.33	9.54	1.43
	Materials	8.7	9.62	10.88	9.92	1.67
322 (N=12182)	Output	6.02	7.07	8.35	7.24	1.78
	Capital	5.47	6.14	6.93	6.23	1.21
	Labor	5.89	6.75	7.81	6.93	1.55
	Materials	5.9	6.89	8.16	7.12	1.77
381 (N=7411)	Output	2.56	3.09	3.97	3.36	1.1
	Capital	2.77	3.3	3.95	3.42	0.92
	Labor	2.64	3.18	3.91	3.37	0.98
	Materials	2.71	3.3	4.11	3.5	1.09
All (N=32808)	Output	8.39	9.73	11.26	9.87	2
	Capital	7.62	8.53	9.46	8.48	1.51
	Labor	7.77	8.65	9.72	8.8	1.58
	Materials	7.89	8.93	10.26	9.15	1.88

Colombian Manufacturing

Table: Coefficient Estimates and Standard Errors for Colombia Manufacturing Firms

		Capital		Labor		Returns to Scale	
Industry (ISIC code)	τ	Coef.	s.e.	Coef.	s.e.	Coef.	s.e
311	0.10	0.155	0.0601	0.757	0.0284	0.912	0.0629
	0.25	0.347	0.0561	0.627	0.0181	0.709	0.0572
	0.50	0.314	0.0352	0.517	0.0185	1.555	0.0377
	0.75	0.297	0.0322	0.426	0.0204	1.028	0.0358
322	0.10	0.237	0.0632	0.829	0.0264	0.929	0.0659
	0.25	0.377	0.0418	0.754	0.0223	0.560	0.0423
	0.50	0.302	0.0304	0.643	0.0241	1.347	0.0338
	0.75	0.312	0.0370	0.522	0.0291	0.884	0.0417
381	0.10	0.374	0.1131	1.181	0.0563	0.974	0.1139
	0.25	0.294	0.0406	0.948	0.0304	1.067	0.0451
	0.50	0.269	0.0525	0.759	0.0315	1.243	0.0547
	0.75	0.464	0.0621	0.641	0.0292	1.027	0.0661
All	0.10	0.156	0.1969	0.872	0.0152	0.929	0.1983
	0.25	0.150	0.0252	0.744	0.0100	1.145	0.0264
	0.50	0.239	0.0236	0.642	0.0083	1.227	0.0243
	0.75	0.323	0.0244	0.567	0.0092	0.893	0.0258

Colombian Manufacturing

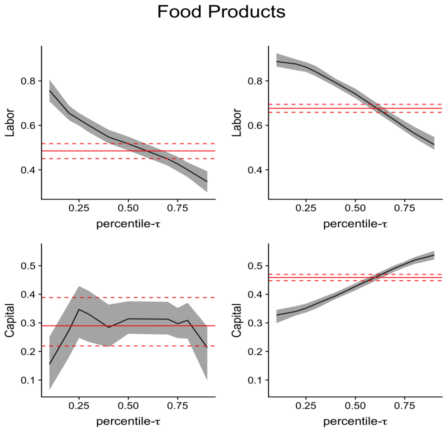


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Colombian Manufacturing

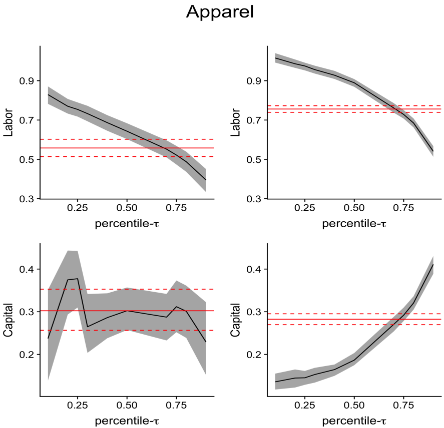


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Colombian Manufacturing

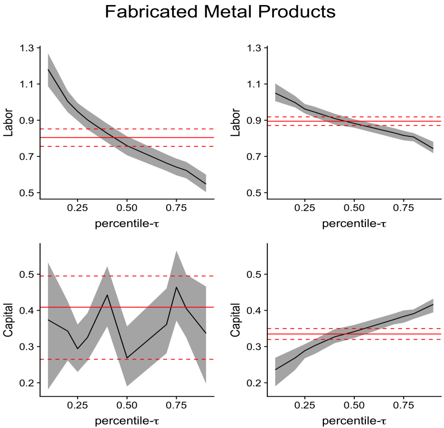


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Colombian Manufacturing

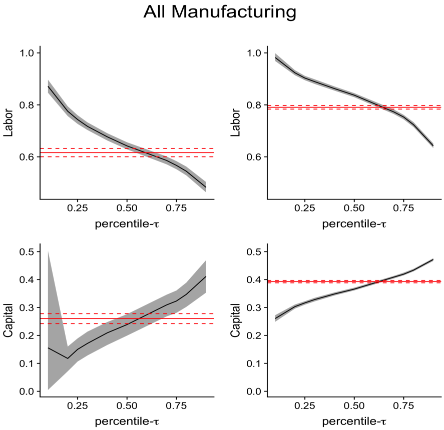


Figure: Estimated values of production function coefficients and their 90% confidence interval. The plots on the LHS are the QLP and LP estimates. The plots on the RHS are quantile regression and OLS estimates

Conclusion

- Most estimates are consistent with prior empirical results and economic theory
- Output elasticity of labor decreases with firm size
- Output elasticity of capital increases with firm size
  - Dermirer (2020)
  - Holmes and Mitchell (2008)
  - Rajan, Zingales, and Kumar (1999)
- Without controlling for productivity, estimates tend to be biased upwards
- Still exploring implications for TFP estimates
- Still working on better bandwidth selection mechanisms
- Extensions to a nonlinear model using Arellano and Bonhomme (2016)

References I

Ackerberg, Daniel, Kevin Caves, and Garth Frazer (2015). "Identification Properties of Recent Production Function Estimators". In: *Econometrica* 83.6, pp. 2411–2451. DOI: 10.3982/ecta13408.

Antle, John M. (1983). "Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach". In: *Journal of Business & Economic Statistics* 1.3, pp. 192–201. DOI: 10.1080/07350015.1983.10509339.

Arellano, Manuel and Stéphane Bonhomme (2016). "Nonlinear panel data estimation via quantile regressions". In: *The Econometrics Journal* 19.3, pp. C61–C94. DOI: 10.1111/ectj.12062.

Balat, Jorge, Irene Brambilla, and Yuya Sasaki (2018). "Heterogeneous firms: skilled-labor productivity and the destination of exports". Working paper.

Bernini, Cristina, Marzia Freo, and Attilio Gardini (2004). "Quantile estimation of frontier production function". In: *Empirical Economics* 29.2, pp. 373–381. DOI: 10.1007/s00181-003-0173-5.

Bhattacharya, Debopam (2009). "Inferring Optimal Peer Assignment From Experimental Data". In: *Journal of the American Statistical Association* 104.486, pp. 486–500. DOI: 10.1198/jasa.2009.0015.

References II

Canay, Ivan A. (2011). "A simple approach to quantile regression for panel data". In: *The Econometrics Journal* 14.3, pp. 368–386. DOI: 10.1111/j.1368-423x.2011.00349.x.

Castro, Luciano I. de and Antonio F. Galvao (2017). "Dynamic Quantile Models of Rational Behavior". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2932382.

Castro, Luciano de et al. (2018). "Smoothed GMM for Quantile Models". In: *Journal of Econometrics*. forthcoming.

Chambers, Christopher P. (2007). "Ordinal aggregation and quantiles". In: *Journal of Economic Theory* 137.1, pp. 416–431. DOI: 10.1016/j.jet.2006.12.002.

Chernozhukov, Victor and Christian Hansen (2005). "An IV Model of Quantile Treatment Effects". In: *Econometrica* 73.1, pp. 245–261. DOI: 10.1111/j.1468-0262.2005.00570.x.

Chernozhukov, Victor and Han Hong (2003). "An MCMC approach to classical estimation". In: *Journal of Econometrics* 115.2, pp. 293–346. DOI: 10.1016/s0304-4076(03)00100-3.

References III

Dermirer, Mert (2020). "Production Function Estimation with Factor Augmenting Technology: An Application to Markups". In:

Gandhi, Amit, Salvador Navarro, and David A. Rivers (2020). "On the Identification of Gross Output Production Functions". In: *Journal of Political Economy*, pp. 000–000. DOI: 10.1086/707736.

Holmes, Thomas J. and Matthew F. Mitchell (2008). "A theory of factor allocation and plant size". In: *The RAND Journal of Economics* 39.2, pp. 329–351. DOI: 10.1111/j.0741-6261.2008.00017.x.

Just, Richard E. and Rulon D. Pope (1978). "Stochastic specification of production functions and economic implications". In: *Journal of Econometrics* 7.1, pp. 67–86. DOI: 10.1016/0304-4076(78)90006-4.

Kaplan, David M. and Yixiao Sun (2016). "Smoothed Estimating Equations for Instrumental Variables Quantile Regression". In: *Econometric Theory* 33.1, pp. 105–157. DOI: 10.1017/s0266466615000407.

Kasahara, Hiroyuki, Paul Schrimpf, and Michio Suzuki (2017). "Identification and Estimation of Production Function with Unobserved Heterogeneity". Working Paper. Working paper.

Koenker, Roger (2004). "Quantile regression for longitudinal data". In: *Journal of Multivariate Analysis* 91.1, pp. 74–89. DOI: 10.1016/j.jmva.2004.05.006.

Lamarche, Carlos (2010). "Robust penalized quantile regression estimation for panel data". In: *Journal of Econometrics* 157.2, pp. 396–408. DOI: 10.1016/j.jeconom.2010.03.042.

Levinsohn, James and Amil Petrin (2003). "Estimating Production Functions Using Inputs to Control for Unobservables". In: *Review of Economic Studies* 70.2, pp. 317–341. DOI: 10.1111/1467-937x.00246.

Li, Tong and Yuya Sasaki (2017). "Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function". Working paper.

Liu, Chunping, Audrey Laporte, and Brian S. Ferguson (2008). "The quantile regression approach to efficiency measurement: insights from Monte Carlo simulations". In: *Health Economics* 17.9, pp. 1073–1087. DOI: 10.1002/hec.1398.

Manski, Charles F. (1988). "Ordinal utility models of decision making under uncertainty". In: *Theory and Decision* 25.1, pp. 79–104. DOI: 10.1007/bf00129169.

References V

Olley, G. Steven and Ariel Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". In: *Econometrica* 64.6, p. 1263. DOI: 10.2307/2171831.

Rajan, Raghuram G., Luigi Zingales, and Krishna B. Kumar (1999). "What Determines Firm Size?". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.170349.

Rostek, Marzena (2009). "Quantile Maximization in Decision Theory". In: *The Review of Economic Studies* 77.1, pp. 339–371. DOI: 10.1111/j.1467-937x.2009.00564.x.