Productivity and Quality Estimation

Zi Wang
HKBU

Spring 2026

Productivity and Quality

- Productivity: input $\Rightarrow$ output
- Quality: given price, variation in demanded quantity
- Descriptive analysis: estimated productivity/quality
  - Variations
  - Correlation with other variables
  - Controls
- Policies/shocks $\Rightarrow$ Estimated productivity/quality $\Rightarrow$ Outcomes
  - Full structural model in need
  - *e.g.* Trade liberalization $\Rightarrow$ Productivity $\uparrow$ $\Rightarrow$ Reallocation across firms

Productivity Estimation

- Production function:

$$Y_i = e^{\beta_0} K_i^{\beta_1} L_i^{\beta_2} e^{\varepsilon_i} \tag{1}$$

- Taking logs:

$$y_i = \beta_0 + \beta_1 k_i + \beta_2 l_i + \varepsilon_i \tag{2}$$

- Extensions:
  - Additional inputs, including intermediates, different types of labor/capital, and intangible capital
  - More flexible models:

$$\begin{aligned} y_i &= f\left(k_i, l_i; \beta\right) + \varepsilon_i \\ y_i &= f\left(k_i, l_i, \varepsilon_i; \beta\right) \\ y_i &= \beta_0 + \beta_{1i} k_i + \beta_{2i} l_i + \varepsilon_i \end{aligned} \tag{3}$$

Endogeneity

- Firm chooses k_i and l_i to maximize profits, depending on ε_i–firm's information structure
- Example:

$$\max_{K_i, L_i} p_i e^{\beta_0} K_i^{\beta_1} L_i^{\beta_2} e^{\varepsilon_i} - r_i K_i - w_i L_i \tag{4}$$

- Decompose ε_i:

$$y_i = \beta_0 + \beta_1 k_i + \beta_2 l_i + \omega_i + \epsilon_i, \tag{5}$$

  where ω_i is an unobservable that is predictable to the firm when it decides k_i and l_i, and ϵ_i is an unobservable that the firm has no information about when making input decisions
- Productivity ω_i is causing the endogeneity problem, not ϵ_i

Instrumental Variables

- Input and output prices, w_i, r_i, and p_i, affect firms' optimal choices of k_i and l_i
- These prices are excluded from the production function as they do not directly determine y_i conditional on the inputs:
  - Are w_i, r_i, and p_i uncorrelated with ω_i?
  - Require perfect competition: more believable for input markets than for output markets
- Practical restrictions:
  - Difficult to get data on w_i and r_i
  - Very little variation in w_i and r_i across firms
  - Variation in w_i: different input prices vs. unobserved labor quality
- If one can find a market where there is convincing exogenous input price variation, IV approach is convincing without many auxiliary assumptions: *e.g.* randomized experiments

Fixed Effects

- $\omega_{it} = \omega_i$. So it can be controlled by the firm fixed effect. Or

$$\Delta y_{it} = \beta_1 \Delta k_{it} + \beta_2 \Delta l_{it} + \Delta \epsilon_{it} \tag{6}$$

- Extensions: dynamic panel models with richer error structures
  - $\omega_{it} = \rho \omega_{i,t-1} + \xi_{it}$
  - $\omega_{it} = \alpha_i + \lambda_{it}$ where $\lambda_{it} = \rho \lambda_{i,t-1} + \xi_{it}$

Olley and Pakes (OP, 1996)

- Setup:
$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{7}$$

- Idea: recover ω_{it} using the observed firms' investment decisions i_{it}

- Assumption:
  - ω_{it} is first-order Markov: $\omega_{i,t+1} = g(\omega_{it}) + \xi_{i,t+1}$, where $E_t(\xi_{i,t+1}) = 0$
  - Labor is a perfectly variable input, *i.e.* l_{it} is chosen by the firm at t after observing ω_{it}
  - Labor has no dynamic implications
  - Capital is accumulated as $K_{it} = i_{i,t-1} + \delta K_{i,t-1}$, where i_{it} is the investment chosen by the firms in period t after observing ω_{it}

OP (1996)

- The firm solves the dynamic programming problem to get $i_{it} = f_t\left(k_{it}, \omega_{it}\right)$
- The investment function can be inverted, *i.e.* $\omega_{it} = f_t^{-1}\left(k_{it}, i_{it}\right)$ under the following conditions:
  - Strict monotonicity: f_t is strictly monotonic in ω_{it}
  - Scalar unobservable: ω_{it} is the only econometric unobservable in the investment equation
- Under the conditions above, we have

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + f_t^{-1}\left(k_{it}, i_{it}\right) + \epsilon_{it} \tag{8}$$

- Non-parametric approximation for f_t^{-1}: polynomial

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \gamma_{0t} + \gamma_{1t} k_{it} + \gamma_{2t} i_{it} + \gamma_{3t} k_{it}^2 + \gamma_{4t} i_{it}^2 + \gamma_{5t} k_{it} i_{it} + \epsilon_{it} \tag{9}$$

OP (1996)

- β_1 is not separable with γ_{1t}

- A two-stage estimate of (β_1, β_2)
  - OLS $y_{it} = \beta_2 l_{it} + \underbrace{\tilde{\gamma}_{0t} + \tilde{\gamma}_{1t} k_{it} + \gamma_{2t} i_{it} + \gamma_{3t} k_{it}^2 + \gamma_{4t} i_{it}^2 + \gamma_{5t} k_{it} i_{it}}_{\hat{\Phi}_{it} \equiv \beta_0 + \widehat{\beta_1 k_{it}} + \omega_{it}} + \epsilon_{it}$ to get β_2 and $\hat{\Phi}_{it}$

  - Recover β_1 from $\hat{\Phi}_{it}$ and $\omega_{it} = g(\omega_{i,t-1}) + \xi_{it}$
    - $\hat{\omega}_{it}(\beta_1) = \hat{\Phi}_{it} - \beta_1 k_{it}$

    - Polynomial approximation on $g(.) \Rightarrow \hat{\xi}_{it}(\beta_1)$

    - Moment condition: $E[\xi_{it}(\beta_1) k_{it}] = 0$

- Identification:
  - First stage: Compare output of firms with same i_{it} and k_{it} (which implies the same ω_{it}), but different l_{it}

  - Second stage: Compare output of firms with same $\omega_{i,t-1}$ but different k_{it}

Levinsohn and Petrin (LP, 2003)

- LP's worries about OP (1996)
  - OP assumes that investment is strictly monotonic in ω_{it}
  - In many datasets, a large fraction of firms have zero i_{it} in many years
- LP (2003) use intermediate inputs to learn about ω_{it} because these inputs rarely take the value 0:

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 \underbrace{m_{it}}_{\text{intermediate input}} + \omega_{it} + \epsilon_{it} \tag{10}$$

  - $m_{it} = f_t\left(k_{it}, \omega_{it}\right)$
  - Strict monotonicity $\Rightarrow y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + f^{-1}\left(k_{it}, m_{it}\right) + \epsilon_{it}$
  - Two-stage estimator: $E\left[\xi_{it} m_{i,t-1}\right] = 0$ in the second stage

Ackerberg, Caves, and Frazer (ACF, 2015)

- The first stage in LP (2003):
$$y_{it} = \beta_2 l_{it} + np\left(k_{it}, m_{it}\right) + \epsilon_{it} \tag{11}$$

- Critique: Can two firms with the same k_{it} and m_{it} have different l_{it}?
  - Intermediate input: $m_{it} = f_t\left(k_{it}, \omega_{it}\right)$
  - Labor input: $l_{it} = h_t\left(k_{it}, \omega_{it}\right)$
  - So l_{it} is a deterministic function of k_{it} and m_{it} $\Rightarrow$ Collinearity in the first stage of LP

- ACF: identify β_2 with the capital coefficient in the second stage
  - $m_{it} = f_t\left(k_{it}, \omega_{it}, l_{it}\right)$
  - $y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + f_t^{-1}\left(k_{it}, m_{it}, l_{it}\right) + \epsilon_{it}$
  - Identify the composite term $\hat{\Phi}_{it} = \beta_0 + \beta_1 \widehat{k_{it} + \beta}_2 l_{it} + \omega_{it}$
  - Moment conditions to identify (β_1, β_2): $E\left[\xi_{it}\left(\beta_1, \beta_2\right) k_{it}\right] = E\left[\xi_{it}\left(\beta_1, \beta_2\right) l_{i,t-1}\right] = 0$

Piveteau and Smagghue (2019): Estimating firm product quality using trade data

- CES utility of good g in destination d at period t: aggregate over firms $f \in \Omega_{gdt}$

$$C_{gdt} = \left[\sum_{f \in \Omega_{gdt}} \left(\underbrace{\lambda_{fgdt}}_{\text{quality}} \underbrace{q_{fgdt}}_{\text{quantity}} \right)^{\frac{\sigma_j - 1}{\sigma_j}} \right]^{\frac{\sigma_j}{\sigma_j - 1}} \tag{12}$$

- Demand function

$$q_{fgdt} = \underbrace{p_{fgdt}^{* - \sigma_j}}_{\text{individual price}} \lambda_{fgdt}^{\sigma_j - 1} P_{gdt}^{\sigma_j - 1} E_{gdt} \tag{13}$$

- Individual price

$$p^*_{fgdt} = \frac{\tau_{gdt}}{\underbrace{e_{dt}}_{\text{nominal exchange rate}}} p_{fgdt} \tag{14}$$

Piveteau and Smagghue (2019)

- Estimation:

$$\log q_{fgdt} = -\sigma_j \log p_{fgdt} + \tilde{\lambda}_{fgdt} + \mu_{gdt}, \tag{15}$$

  - $\tilde{\lambda}_{fgdt} \equiv (\sigma_j - 1)\left(\log \lambda_{fgdt} - \bar{\log}\,\lambda_{gdt}\right)$ is the de-mean quality
  - $\mu_{gdt} \equiv -\sigma_j \log\left(\frac{\tau_{gdt}}{e_{gdt}}\right) + (\sigma_j - 1)\log P_{gdt} + \log E_{gdt} + (\sigma_j - 1)\bar{\log}\,\lambda_{gdt}$
- Instrument for prices at the firm-level: real exchange rates fluctuations faced by importing firms
  - Real exchange rate shocks on a firm's imports are cost shocks; firms pass these cost shocks through to its export prices

Piveteau and Smagghue (2019)

- Import-weighted log real exchange rates:

$$\log \bar{rer}_{ft_0 t} = \sum_c \omega_{cft_0} \times \log\left(\widetilde{rer}_{ct}\right), \quad \omega_{cft_0} \equiv \frac{m_{cft_0}}{\sum_{c'=1}^{C} m_{c' ft_0}} \tag{16}$$

- Instrument: interact $\log \bar{rer}_{ft_0 t}$ with the share of these imports in the operating costs of the firm

$$R\bar{E}R_{ft_0 t} = \log \bar{rer}_{ft_0 t} \times \frac{\sum_t m_{ft}}{\sum_t OC_{ft}} \tag{17}$$

Piveteau and Smagghue (2019)

Table 2
Results on pooled data.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS		2SLS			Reduced Form	
Panel A (1^{st} STAGE)			*log price*			*log qty*	
$\overline{RER}_{ft_0t}$		0.16 ***	0.25 ***	0.23 ***	0.25 ***	-1.06 ***	-1.16 ***
		(0.041)	(0.065)	(0.076)	(0.067)	(0.30)	(0.34)
$\overline{RER}_{ft_0t-1}$				0.029			0.16
				(0.068)			(0.29)
$\overline{RER}_{ft_0t} \times ms_{fpdt_0}$					-0.41		
					(0.83)		
$Entry_{fpdt}$		0.0023 ***	0.0015	0.0014	0.0015	-0.95 ***	-0.95 ***
		(0.00070)	(0.0020)	(0.0020)	(0.0020)	(0.012)	(0.012)
$\overline{gpdc}_{ft}^{imp}$		0.0037 ***	0.0060 **	0.0060 **	0.0061 **	0.020 *	0.020 *
		(0.0014)	(0.0025)	(0.0025)	(0.0025)	(0.012)	(0.012)
$\overline{gpdc}_{ft}^{exp}$		0.0033 *	0.0053	0.0053	0.0054	0.24 ***	0.24 ***
		(0.0017)	(0.0041)	(0.0042)	(0.0042)	(0.021)	(0.021)
Panel B (2^{nd} STAGE)			*log qty*				
Log price ($-\hat{\sigma}$)	**-0.78** ***	**-3.03** **	**-4.26** ***	**-4.18** **	**-4.39** ***		
	(0.0080)	(1.39)	(1.65)	(1.65)	(1.69)		
$Entry_{fpdt}$	-0.32 ***	-0.31 ***	-0.95 ***	-0.95 ***	-0.94 ***		
	(0.0041)	(0.0053)	(0.014)	(0.014)	(0.015)		
$\overline{gpdc}_{ft}^{imp}$	0.012 *	0.020 **	0.045 ***	0.045 ***	0.046 ***		
	(0.0066)	(0.0090)	(0.017)	(0.017)	(0.018)		
$\overline{gpdc}_{ft}^{exp}$	0.15 ***	0.16 ***	0.26 ***	0.26 ***	0.27 ***		
	(0.010)	(0.012)	(0.028)	(0.028)	(0.029)		
Sample	Full	Full	>6 yrs	>6 yrs	>6 yrs	>6 yrs	>6 yrs
Kleibergen-Paap F-stat		14.3	14.5	7.3	7.6		
Hansen *p*-value				0.4	0.00		

Notes: The full sample contains 10,762,689 observations while the restricted sample contains 3,481,154 observations. Firm×prod×dest×spell and prod×dest×year fixed effects are included in all regressions. Standard errors in parentheses are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Piveteau and Smagghue (2019)

Table 3
Price-elasticity estimates ($-\hat{\sigma}$) for different product categories.

Product categories	OLS		IV			IV (single FS)		
	Coef	SE	Coef	SE	F-stat	Coef	SE	N
Animal Products	−0.88 ***	(0.042)	1.82	(5.17)	1.21	**−10.1** *	[5.63]	190,887
Vegetable Products	−0.75 ***	(0.028)	10.4	(24.2)	0.29	**−3.28**	[5.01]	195,596
Foodstuffs	−0.94 ***	(0.018)	−1.03	(4.24)	0.81	**−1.70**	[5.34]	409,242
Mineral Products	−0.84 ***	(0.083)	−171.6	(5.8e3)	0.00	**−3.78**	[7.79]	24,125
Chemicals and Allied	−0.90 ***	(0.021)	−1.34	(1.61)	2.11	**−4.12**	[2.97]	374,169
Plastics, Rubbers	−0.92 ***	(0.025)	−1.26	(1.20)	7.78	**−2.43**	[2.95]	227,886
Skins, Leather	−0.74 ***	(0.042)	−18.8	(41.9)	0.20	**−5.90** **	[3.00]	56,251
Wood, Wood products	−0.86 ***	(0.023)	−3.06 *	(1.75)	5.52	**−1.47**	[2.83]	178,783
Textiles	−0.70 ***	(0.038)	−5.82	(4.45)	3.91	**−4.42**	[2.72]	663,856
Footwear, Headgear	−0.68 ***	(0.061)	−7.09	(5.21)	2.30	**−6.79** **	[3.37]	65,454
Stone, Glass	−0.84 ***	(0.034)	−1837.6	(5.5e5)	0.00	**−4.97**	[3.03]	80,316
Metals	−0.81 ***	(0.025)	−2.43	(3.03)	1.82	**−3.01**	[2.69]	260,784
Machinery, Electrical	−0.87 ***	(0.021)	−2.87 **	(1.32)	6.14	**−3.78**	[2.57]	392,429
Transportation	−0.79 ***	(0.031)	−6.02	(5.65)	1.26	**−8.95** **	[4.41]	113,832
Miscellaneous	−0.79 ***	(0.023)	−4.96	(3.52)	2.87	**−4.02** *	[2.41]	247,544

Notes: Estimates in columns "OLS" and "IV" are obtained by estimating eq. (4) separately for each industry, respectively by OLS and 2SLS. Estimates in column "IV (single FS)" is obtained by estimating a single first stage and a second stage where the price-elasticity is allowed to vary across industries. Controls for GDP per capita ($\overline{gpdc}_{ft}^{exp}$ and $\overline{gpdc}_{ft}^{imp}$) and for partial-year effect ($Entry_{fpdt}$) are included in all regressions. Firm×Prod×Dest×Spell and Prod×Dest×Year fixed effects are included in all regressions. IV specifications use $\overline{RER}_{ft_0t}$ as instrument. Standard errors are clustered at the firm level and standard errors for the "IV (Single FS)" specification are obtained through 1000 bootstrap replications using firm as the sampling unit. Column "F-stat" reports the value of the Kleibergen-Paap F-stat. * $p < .1$, ** $p < .05$, *** $p < .01$.

Piveteau and Smagghue (2019)

Table 6
Correlation with firms' characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: estimated quality ($\tilde{\lambda}_{fpdt}$)					
	No fixed effects		Dest, prod and year FE		Dest×prod×year FE	
log(wage)	0.83 ***	0.76 ***	0.97 ***	0.94 ***	1.09 ***	1.06 ***
	(0.020)	(0.018)	(0.021)	(0.020)	(0.023)	(0.022)
log(employment)		0.067 ***		0.10 ***		0.12 ***
		(0.0098)		(0.010)		(0.012)
log(capital)		0.053 ***		0.061 ***		0.070 ***
		(0.0069)		(0.0073)		(0.0081)
N	13,391,548	13,201,466	13,391,528	13,201,442	13,297,957	13,096,862
R^2	0.0068	0.012	0.0094	0.016	0.022	0.032

Notes: The variable log(*wage*) is obtained by taking the logarithm of the total wage bill divided by the number of employees. Specifications (1) and (2) have a non-reported constant. Standard errors in parentheses are clustered at the firm-year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Piveteau and Smagghue (2019)

Table 8
Correlation between prices and quality.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: log price					
	All markets		>5 firms		>20 firms	
Quality	0.67 ***	0.59 ***	0.66 ***	0.59 ***	0.65 ***	0.58 ***
	(0.00063)	(0.00071)	(0.00077)	(0.00083)	(0.0011)	(0.0012)
Quality × quality lad.	0.16 ***	0.16 ***	0.17 ***	0.15 ***	0.18 ***	0.18 ***
	(0.00088)	(0.00064)	(0.0011)	(0.00084)	(0.0020)	(0.0016)
Firm×Prod×year FE	No	Yes	No	Yes	No	Yes
N	20,048,513	13,830,187	16,147,196	11,683,523	10,225,420	7,608,688
R^2	0.49	0.88	0.50	0.88	0.50	0.89

Notes: Quality ladder is the difference between the 95th and 5th percentiles of the quality distribution within a market, normalized to have a mean of zero and a variance of one. Quality measures and prices are also normalized to have zero mean and a standard deviation of one within markets. Each regression includes product×dest×year fixed effects. Standard errors in parentheses are clustered at the product×dest×year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Summary

- Productivity estimation: residuals in estimating production function
  - Input $\Rightarrow$ Output
  - Identification issue: input is correlated with the unobserved productivity
- Quality estimation: residuals in estimation demand function
  - Price $\Rightarrow$ Demanded quantity
  - Identification issue: price is correlated with the unobserved quality
- Caution:
  - Productivity and quality are estimated using structural models
  - When applying estimated productivity/quality in empirical settings, think about whether the empirical specifications are consistent with the structural models used to estimate productivity/quality