Production Functions and Productivity

Paul T. Scott
NYU Stern
ptscott@gmail.com

PhD Empirical IO
Fall 2021

Overview

- Brief background on industry dynamics
- Estimating production functions:
  - Olley Pakes (1995)
  - Ackerberg, Caves, and Frazer (2015)
- Foster, Haltiwanger, and Syverson (2008): Revenue vs. Physical TFP
- Production, markups and market power
  - De Loecker and Warzynski (2012)
  - More recent papers applying this method to study market power

Some Questions

- Role of entry and exit in driving growth?
- Impact of events like trade liberalization and deregulation on productivity?
- What are the factors driving plant/firm-level changes in productivity and growth?
- What can we say about changes in markups/market power based on production data? Can we learn about recent trends in market power by analyzing production data?

The firm size distribution

- A very robust finding: the firm size distribution has a long upper tail.
- ... this holds within the vast majority of industries, countries, and after conditioning on observable characteristics.
- Typically, the size distribution is approximated with a lognormal or Pareto distribution.
- Broader theme: almost any variable we look at exhibits tremendous heterogeneity across firms

Gibrat's Law, bare-bones model of growth and heterogeneity

- Gibrat's law states that if the growth rate of a variable is independent of its size and over time, it will have a log-normal distribution in the long run.

- Let Y_{it} denote firm i's size (employment or output) in year t . Suppose it evolves according to the following process:

$$(Y_{i,t+1} - Y_{it})/Y_{it} = \varepsilon_{it}$$

  where ε_{it} is i.i.d. across i and t

- Then, after allowing a large group of firms to evolve for a while, the cross-sectional distribution of Y_{it} will have a log-normal distribution.

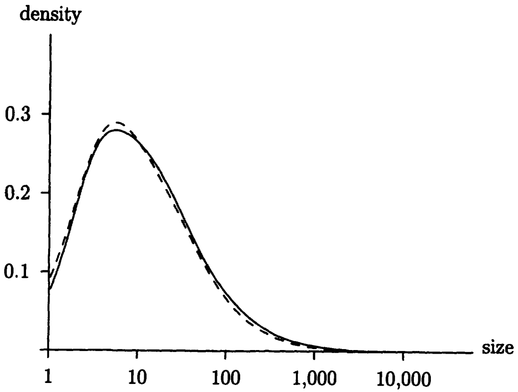

FIGURE 2. FIRM SIZE DISTRIBUTION IN 1983 (SOLID LINE) AND 1991 (DASHED LINE), BASED ON EMPLOYMENT DATA FROM THE *QUADROS DO PESSOAL* DATA SET

Source: Cabral and Mata (2003)

Theoretical models of industry dynamics

- While the "Gibrat Model" of industry dynamics is way too simple to explain all the data, it's still in the background. Most modern models of heterogeneous firms are based on the assumption that each firm experiences a series of unpredictable and persistent shocks, generating lots of heterogeneity.
- In modern models, rather than just having random growth/size, there is a heterogenous *productivity* variable that determines firm size
- Jovanovic (1982), entry and exit model that explains some facts:
  - Small firms have higher and more variable growth rates
  - Smaller firms are more likely to exit
  - See Dunne, Roberts, and Samuelson (1989) for empirical evidence.
- Hopenhayn (1992) - equilibrium model with stochastic productivity
- Melitz (2003) - Equilibrium trade model in which high-productivity firms select into exporting and low-productivity firms exit.

What is productivity, and how do we measure it?

- The two most popular measures of productivity are labor productivity and total factor productivity (TFP).
- Labor productivity is defined as the ratio of output to labor inputs (Y_t/L_t).
- TFP is defined as the residual of a production function. For example, with the Cobb-Douglas Production function

$$Y_t = e^{\omega_t} L_t^{\alpha} K_t^{\beta},$$

which we can rewrite in logs,

$$y_t = \alpha l_t + \beta k_t + \omega_t.$$

TFP is ω_t (lower case variables represent logs of uppercase variables).

Concerns with productivity definitions

- Labor productivity can change due to changes in the capital-labor ratio without any changes in technology (e.g., due to wage changes). Consequently, TFP is typically the object of choice for studies on technological change or firm performance.
- That said, TFP is not without its own conceptual and practical limitations.
  - Unlike labor productivity, TFP is defined in terms of a specific functional form and does not have units.
  - TFP relies on measurement of capital stocks, which is typically difficult.

Bartelsman and Doms (2003): overview

Bartelsman and Doms (2003) review some empirical work on productivity. Stylized facts:

- Large productivity dispersion across firms.
- Within firm, productivity is highly but imperfectly persistent.
- There is considerable reallocation within industries;
  "the aggregate data belie the tremendous turmoil underneath."

De Loecker and Syverson (2021) report that 90-10 percentile TFP ratios of 2:1 are typical.

What to make of these residuals?

- “I found the spectacle of economic models yielding large residuals rather uncomfortable, even when the issue was fudged by renaming them technical change and claiming credit for their 'measurement.' ”
  – Zvi Griliches
- Bad data could be one reason could be one source of TFP dispersion, but we observe large dispersion everywhere we have data,
  and measured productivities are connected to real outcomes:
  - more productive firms are less likely to exit
  - more productive firms are more likely to be exporters
  - productivities of entrants tend to be lower than average incumbents

Simultaneity

- $y_t = \alpha l_t + \beta k_t + \omega_t$
- Generally, we should expect input use to respond to ω_t. For example, if capital is set at $t-1$ and labor can be adjusted at t, we should expect labor to respond to the current realization of productivity.
- Input prices as instruments are a potential solution, but we often don't observe any variation in them, and if they do vary, you might question whether the variation is exogenous.

Selection

- The firms that exit are those that have low productivity draws.
- Selection will be an issue if we want to estimate how the productivity process evolves or how endogenous variables like exporter status impact productivity (e.g., because of Melitz's selection story).

Measuring Y

$$y_t = \alpha l_t + \beta k_t + \omega_t$$

- Typically, we would like y to be a measure of physical output, but often revenue is all that's available.
  - "TFPQ" and "TFPR"
- Even if we have quantity data, multi-produce establishments make it difficult to use
  - In principle, we should think in terms of transformation functions, but they are difficult to analyze econometrically.
  - The vast majority of establishments are multi-product

"The Dynamics of Productivity in the Telecommunications Equipment Industry"
Olley and Pakes (1996)

Overview

- Analyzes effects of deregulation in telecommunications equipment industry.
- Deregulation increases productivity, primarily through reallocation toward more productive establishments.
- Estimation approach deals with simultaneity and selection issues.

Background I

- AT&T had a monopoly on telecommunications services in the US throughout most of the 20th century (note: a telecommunications network is a classic example of a natural monopoly).
- Before the regulatory change, AT&T required that equipment attached to their network must be supplied by the AT&T, and virtually all of their equipment was supplied by their subsidiary, Western Electric. Thus, they leveraged their network monopoly to a monopoly on phones.
- Entering the telecom equipment market would have required also entering as a network operator, which was generally prohibited by regulatory barriers.

Background II

- A change in technology opened up new markets for telecommunications equipment (e.g., fax machines)
- Meanwhile, the FCC (regulatory agency) decided to begin allowing the connection of privately-provided devices to AT&T's network.
- A surge of entry into telecommunications equipment manufacturing followed in the late 1960's and 1970's.

TABLE I

CHARACTERISTICS OF THE DATA

Year	Plants	Firms	Shipments (billions 1982 $)	Employment
1963	133	104	5.865	136899
1967	164	131	8.179	162402
1972	302	240	11.173	192248
1977	405	333	13.468	192259
1982	473	375	20.319	222058
1987	584	481	22.413	184178

Background III

- AT&T continued purchasing primarily from Western Electric into the 1980's (although consumers were free to purchase devices from other companies).
- The divestiture (breakup) of AT&T created seven regional Bell companies that were no longer tied to Western Electric, and they were prohibited from manufacturing their own equipment.
- The divestiture was implemented in January 1984. Western Electric's share dropped dramatically.

TABLE II

BELL COMPANY EQUIPMENT PROCUREMENT
(PERCENT PURCHASED FROM WESTERN ELECTRIC)

1982	1983	1984	1985	1986[E]
92.0	80.0	71.8	64.2	57.6

[E] Estimated for 1986.
Source: NTIA (1988, p. 336, and discussion pp. 335–337).

Entry

TABLE III

ENTRANTS ACTIVE IN 1987

	Number	Share of Number Active in 1987 (%)	Share of 1987 Shipments (%)	Share of 1987 Employment (%)
Plants: New since 1972	463	79.0	32.8	36.0
Firms: New since 1972	419	87.0	30.0	41.4
Plants: New since 1982	306	52.0	12.0	13.5
Firms: New since 1982	299	60.1	19.4	27.5

Exit

TABLE IV

INCUMBENTS EXITING BY 1987

	Number	Share of Number Active in Base Year (%)	Share of Shipments in Base Year (%)	Share of Employment in Base Year (%)
Plants active in 1972 but not in 1987	181	60.0	40.2	39.0
Firms active in 1972 but not in 1987	169	70.0	13.8	12.1
Plants active in 1982 but not in 1987	195	41.2	26.0	24.1
Firms active in 1982 but not in 1987	184	49.1	17.3	16.1

Where we're going

- What was impact of liberalization on the telecom equipment market? Two major events:
  - Registration and certification program allowing entry in late 70's
  - Breakup of AT&T, decreed in January 1982 and implemented January 1984
- How was productivity affected?
- What were channels for productivity impacts (entry, exit)?

The model

- Incumbent firms (i) make three decisions:
  - Whether to exit or continue. If they exit, they receive a fixed scrap value Ψ and never return.
  - If they stay, they choose labor l_{it},
  - and investment i_{it}.
- Capital accumulation:

$$k_{t+1} = (1-\delta)\,k_t + i_t$$

- Another state variable is age: $a_{t+1} = a_t + 1$

Production

- They assume the following Cobb-Douglas production function:

$$y_{it} = \beta_0 + \beta_a a_{it} + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \eta_{it}$$

  where y_{it} is output, k_{it} is capital, l_{it} is labor, ω_{it} is a persistent component of productivity, and η_{it} is a transient shock to productivity.
- Productivity evolves according to a Markov process: $F(\cdot|\omega)$.
- η is either measurement error, or there is no information about it when labor decisions are made.

Equilibrium behavior

- They assume the existence of a Markov perfect equilibrium. Market structure and prices are state variables in the MPE, but they are common across firms, so they can be absorbed into time subscripts for the value function:

$$V_t(\omega_t, a_t, k_t) = \max\Big\{\Psi, \sup_{i_t \geq 0} \pi_t(\omega_t, a_t, k_t) - c(i_t) \\ + \beta E\left[V_{t+1}(\omega_{t+1}, a_{t+1}, k_{t+1}) \,|J_t\right]$$

  where J_t represents the information set at time t.

- Equilibrium strategies can be described by functions $\underline{\omega}_t(a_t, k_t)$ and $i_t(\omega_t, a_t, k_t)$.
  - A firm will continue if and only if $\omega \geq \underline{\omega}_t(a_t, k_t)$.
  - Continuing firms invest $i_t = i_t(\omega_t, a_t, k_t)$

Aside: Markov perfect equilibrium

- Loosely, it means a subgame perfect equilibrium in which strategies are functions of "real" (payoff relevant) state variables. Formally defined by Maskin and Tirole (1988)
- This rules out conditioning on variables that don't impact present or future payoffs. For example, in the repeated prisoner's dilemma, cooperation with grim trigger punishments is ruled out. The only MPE of the repeated prisoner's dilemma is repeated static Nash.
- Markov perfect equilibrium is to dynamic games what perpetual static Nash is to repeated games.

Thinking about bias

- How does the simultaneity of the input decision bias the labor coefficient?

Thinking about bias

- How does the simultaneity of the input decision bias the labor coefficient?
  - Up: when productivity is high, a firm will use more labor

Thinking about bias

- How does the simultaneity of the input decision bias the labor coefficient?
  - Up: when productivity is high, a firm will use more labor
- How does selection due to exit bias the capital coefficient estimate?

Thinking about bias

- How does the simultaneity of the input decision bias the labor coefficient?
  - Up: when productivity is high, a firm will use more labor
- How does selection due to exit bias the capital coefficient estimate?
  - Down: firms with high capital have lower cutoffs $\underline{\omega}_t$ for exit. Thus, conditional on survival, there is a negative correlation between k and ω

Thinking about bias

- How does the simultaneity of the input decision bias the labor coefficient?
  - Up: when productivity is high, a firm will use more labor
- How does selection due to exit bias the capital coefficient estimate?
  - Down: firms with high capital have lower cutoffs $\underline{\omega}_t$ for exit. Thus, conditional on survival, there is a negative correlation between k and ω
- Another potential source of bias: measurement error. See Collard-Wexler and De Loecker (2016)

Productivity inversion

- In a technical paper, Pakes (1994) shows that optimal investment $i_t(\omega_t, a_t, k_t)$ is monotonically increasing in ω_t, provided $i_t > 0$.
- Given monotonicity, optimal investment can be inverted for productivity:
$$\omega_{it} = h_t(i_{it}, a_{it}, k_{it}) .$$
- We're going to talk more about the $i_t > 0$ requirement with Levinsohn and Petrin (2003).

First stage model

- Substituting in the inversion function,

$$y_{it} = \beta_l l_{it} + \phi_t\left(i_{it}, a_{it}, k_{it}\right) + \eta_{it}$$

where

$$\phi_t\left(i_{it}, a_{it}, k_{it}\right) = \beta_0 + \beta_a a_{it} + \beta_k k_{it} + h_t\left(i_{it}, a_{it}, k_{it}\right)$$

- We can estimate this equation using a semiparametric regression. This may identify β_l, but not the other coefficients.
- With Ackerberg, Caves, and Frazer (2015), we will think more carefully about what's identifying β_l, but don't worry about it for now.

First stage output

- With $\hat{\beta}_l$, we can also estimate ϕ:

$$\hat{\phi}_{it} = y_{it} - \hat{\beta}_l l_{it}$$

- So far we have estimates of β_l and ϕ. $\beta_k k$ and ω are both in the control function ϕ, and we would like to separate them. We're going to use the Markov assumption on ω for identification.

Identifying β_k,β_a

- Let's first think about how to do this without worrying about exit. Define

  $$g\left(\omega_{i,t-1}\right) = E\left[\omega_{i,t}|\omega_{i,t-1}\right],$$

  so that

  $$\omega_{i,t} = g\left(\omega_{i,t-1}\right) + \xi_{i,t}$$

  where $\xi_{i,t+1}$ is the innovation (unexpected change) to productivity.

- We can write out a second stage regression equation:

  $$\phi_{i,t} = \beta_k k_{it} + \beta_a a_{it} + g\left(\omega_{i,t-1}\right) + \xi_{i,t}$$

  and note that $\omega_{i,t-1}$ can also be written as a function of (β_k, β_a):

  $$\phi_{i,t} = \beta_k k_{it} + \beta_a a_{it} + g\left(\phi_{i,t-1} - \beta_k k_{i,t-1} - \beta_a a_{i,t-1}\right) + \xi_{i,t}$$

Identifying β_k,β_a

- Second stage regression equation:

$$\phi_{i,t} = \beta_k k_{it} + \beta_a a_{it} + g\left(\phi_{i,t-1} - \beta_k k_{i,t-1} - \beta_a a_{i,t-1}\right) + \xi_{i,t}$$

- One way to think about this: once we specify a parametric function for g, this basically becomes OLS.
- NLLS: we can guess values of (β_k, β_a), (nonparametrically) estimate g conditional on those value of (β_k, β_a), and then back out $\xi_{i,t}(\beta_k, \beta_a)$. Search over (β_k, β_a) to minimize sum of squares of $\xi_{i,t}(\beta_k, \beta_a)$.

Selection

- Essentially a Heckman selection correction.
- Let $P_t = Pr\left(\chi_{t+1} = 1 | \underline{\omega}_{t+1}\left(k_{t+1}, a_{t+1}\right), J_t\right)$ be the propensity score for exit.
- As long as the conditional density of ω_{t+1} has full support, this can be inverted to express $\underline{\omega}_{t+1} = f\left(P_t, \omega_t\right)$

The second stage with selection

- Write the expectation of $y_{t+1} - \beta_l l_{t+1}$ conditional on survival:

$$E\left[y_{t+1} - \beta_l l_{t+1} | a_{t+1}, k_{t+1}, \chi_{t+1} = 1\right]$$

$$= \beta_a a_{t+1} + \beta_k k_{t+1} + g\left(\underline{\omega}_{t+1}, \omega_t\right)$$

where $g\left(\underline{\omega}_{t+1}, \omega_t\right) = E\left[\omega_{t+1} | \omega_t, \chi_{t+1} = 1\right]$

- Using the inversion of the selection probability, we can write

$$g\left(\underline{\omega}_{t+1}, \omega_t\right) = g\left(f\left(P_t, \omega_t\right), \omega_t\right)$$

which can be written more simply as $g\left(P_t, \omega_t\right)$.

Final step

- Conditional on values of (β_a, β_k), we can construct an estimate of $\omega_t = \phi_t - \beta_a a_t - \beta_k k_t$
- Finally, write

$$\begin{aligned} y_{t+1} - \beta_l l_{t+1} = \; & \beta_a a_{t+1} + \beta_k k_{t+1} + g\left(P_t, \phi_t - \beta_a a_t - \beta_k k_t\right) \\ & + \xi_{t+1} + \eta_{t+1} \end{aligned}$$

- Again, we can use NLLS to estimate (β_k, β_a).
- Note that $E\left(\xi_{i,t} l_{i,t}\right) \neq 0$ is what creates the need for the first stage.
- For the nonparametric functions, they try polynomial series and normal kernels.

Estimation steps

1. First stage semi-parametric regression:

$$y_{it} = \beta_l l_{it} + \phi_t \left(i_{it}, a_{it}, k_{it} \right) + \eta_{it}$$

2. Estimate propensity scores: $P_t = Pr\left(\chi_{t+1} = 1 | \underline{\omega}_{t+1}\left(k_{t+1}, a_{t+1}\right), J_t\right)$

3. Estimate remaining parameters:

$$\begin{aligned} y_{t+1} - \beta_l l_{t+1} = \quad & \beta_a a_{t+1} + \beta_k k_{t+1} + g\left(P_t, \phi_t - \beta_a a_t - \beta_k k_t\right) \\ & + \xi_{t+1} + \eta_{t+1} \end{aligned}$$

   using fact that innovation term ξ_{t+1} is mean-uncorrelated with variables determined at t, including k_{t+1}.

TABLE VI

ALTERNATIVE ESTIMATES OF PRODUCTION FUNCTION PARAMETERS[a]
(STANDARD ERRORS IN PARENTHESES)

Sample:	Balanced Panel		Full Sample[c, d]						
								Nonparametric F_ω	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation Procedure	Total	Within	Total	Within	OLS	Only P	Only h	Series	Kernel
Labor	.851	.728	.693	.629	.628			.608	
	(.039)	(.049)	(.019)	(.026)	(.020)			(.027)	
Capital	.173	.067	.304	.150	.219	.355	.339	.342	.355
	(.034)	(.049)	(.018)	(.026)	(.018)	(.02)	(.03)	(.035)	(.058)
Age	.002	−.006	−.0046	−.008	−.001	−.003	.000	−.001	.010
	(.003)	(.016)	(.0026)	(.017)	(.002)	(.002)	(.004)	(.004)	(.013)
Time	.024	.042	.016	.026	.012	.034	.011	.044	.020
	(.006)	(.017)	(.004)	(.017)	(.004)	(.005)	(.01)	(.019)	(.046)
Investment	—	—	—	—	.13	—	—	—	—
					(.01)				
Other Variables	—	—	—	—	—	Powers of P	Powers of h	Full Polynomial in P and h	Kernel in P and h
# Obs.[b]	896	896	2592	2592	2592	1758	1758	1758	1758

- Why do within estimators have lower capital coefficients?

TABLE IX

INDUSTRY PRODUCTIVITY GROWTH RATES[a]

Time Period	(1) Full Sample	(2) Balanced Panel
1974–1975	−.279	−.174
1975–1977	.020	−.015
1978–1980	.146	.102
1981–1983	−.087	−.038
1984–1987	.041	.069
1974–1987	.008	.020
1975–1987	.032	.036
1978–1987	.034	.047

[a]The numbers in Table IX are annual averages over the various subperiods.

- Estimate of productivity: $p_{it} = \exp\left(y_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_a a_{it}\right)$
- Plants that eventually exit have low productivity growth
- New entrants have lower productivity than continuing establishments
- Survivors have above-average productivity growth

Productivity decomposition

- Aggregate productivity: $p_t = \sum_{i=1}^{N_t} s_{it} p_{it}$.
- Can be decomposed as follows:

$$\begin{aligned} p_t &= \sum_{i=1}^{N_t} \left(\bar{s}_t + \Delta s_{it}\right)\left(\bar{p}_t + \Delta p_{it}\right) \\ &= N_t \bar{s}_t \bar{p}_t + \sum_{i=1}^{N_t} \Delta s_{it} \Delta p_{it} \\ &= \bar{p}_t + \sum_{i=1}^{N_t} \Delta s_{it} \Delta p_{it} \end{aligned}$$

where $\bar{p}_t$ are unweighted mean productivity and shares in the cross-section.

- Thus, aggregate productivity decomposes into an unweighted mean and a covariance term.

TABLE XI

DECOMPOSITION OF PRODUCTIVITY[a]
(EQUATION (16))

Year	p_t	$\bar{p}_t$	$\Sigma_i\, \Delta s_{it}\, \Delta p_{it}$	$\rho(p_t, k_t)$
1974	1.00	0.90	0.01	−0.07
1975	0.72	0.66	0.06	−0.11
1976	0.77	0.69	0.07	−0.12
1977	0.75	0.72	0.03	−0.09
1978	0.92	0.80	0.12	−0.05
1979	0.95	0.84	0.12	−0.05
1980	1.12	0.84	0.28	−0.02
1981	1.11	0.76	0.35	0.02
1982	1.08	0.77	0.31	−0.01
1983	0.84	0.76	0.08	−0.07
1984	0.90	0.83	0.07	−0.09
1985	0.99	0.72	0.26	0.02
1986	0.92	0.72	0.20	0.03
1987	0.97	0.66	0.32	0.10

[a] See text for details.

- Covariance term growing $\Rightarrow$ importance of reallocation of production
- No aggregate productivity growth? Note this is revenue productivity

"Estimating Production Functions Using Inputs to Control for Unobservables"

Levinsohn and Petrin (2003)

Main idea

- Same general framework as Olley and Pakes (1996)
- Main idea: rather than use investment to control for unobserved productivity, use materials inputs.
- Two proposed benefits:
  - Investment proxy isn't valid for plants with zero investment. Zero materials inputs typically an issue in the data.
  - Investments may be "lumpy" and not respond to some productivity shocks.

Downsides of investment

- We need to drop observations with zero investment, which can lead to a substantial efficiency loss. Zero investments happen at a non-trivial rate in annual production data.
- Firms might face non-convex capital adjustment costs leading to flat regions in the $i(\omega)$ function even at positive levels of investment.
- What if investment actually happens with only partial information about productivity and then labor is set once the productivity realization is fully observed?

Invertability

- Just as OP require $i_t(\omega_t, k_t)$ be an invertible function of productivity, LP require that input use $m_t(\omega_t, k_t)$ is an invertible function of productivity.
- LP's monotonicity result relies on easily checked properties of the production function, and some may find this more appealing than a result which relies on a Markov perfect equilibrium.
- Unobserved input price variation may be a problem for the LP invertibility condition (but of course it could be for OP, too).

Checking invertibility

- LP claim that
$$\text{sign}\left(\frac{\partial m}{\partial \omega}\right) = \text{sign}\left(f_{ml} f_{l\omega} - f_{ll} f_{m\omega}\right).$$
- To see this, apply the Implicit function theorem to the FOC's to get
$$\begin{pmatrix} \frac{\partial m}{\partial \omega} \\ \frac{\partial l}{\partial \omega} \end{pmatrix} = -\begin{pmatrix} f_{mm} & f_{ml} \\ f_{lm} & f_{ll} \end{pmatrix}^{-1} \begin{pmatrix} f_{m\omega} \\ f_{l\omega} \end{pmatrix}.$$
- Inverting and solving,
$$\Rightarrow \frac{\partial m}{\partial \omega} = \frac{f_{ml} f_{l\omega} - f_{ll} f_{m\omega}}{\begin{vmatrix} f_{mm} & f_{ml} \\ f_{lm} & f_{ll} \end{vmatrix}}.$$
- By the second-order condition for profit maximization, $\begin{pmatrix} f_{mm} & f_{ml} \\ f_{lm} & f_{ll} \end{pmatrix}$ must be negative semidefinite. This means it has exactly two negative eigenvalues, which means its determinant is positive. Therefore, the numerator controls the sign.

Zero inputs

TABLE 2

Per cent of non-zero observations

Industry (ISIC)	Investment	Fuels	Materials	Electricity
Food products (311)	42·7	78·0	99·8	88·3
Metals (381)	44·8	63·1	99·9	96·5
Textiles (321)	41·2	51·2	99·9	97·0
Wood products (331)	35·9	59·3	99·7	93·8

Note: in OP's industry, it was only 8% zeros.

"Structural Identification of Production Functions"
Ackerberg, Caves, and Frazer (2015)

Overview

- ACF argue that Olley and Pakes's (1996) and Levinsohn and Petrin's (2003) approach suffer from identification issues, at least in principle.
- They propose a new approach which involves modified assumptions on the timing of input decisions and moves the identification of all coefficients of the production function to the second stage of the estimation.

LP's first stage

- Levinsohn and Petrin's first-stage regression:

$$y_{it} = \beta_l l_{it} + f_t^{-1}\left(m_{it}, k_{it}\right) + \varepsilon_{it}.$$

- LP's approach was based on the premise that materials inputs are a variable input and therefore a function of state variables:

$$m_{it} = m_t\left(\omega_{it}, k_{it}\right),$$

- They also assume that labor is a variable input (or else we would not be able to exclude it from the inversion), so

$$l_{it} = l_t\left(\omega_{it}, k_{it}\right).$$

LP's identification problem

- This means we can write:

$$y_{it} = \beta_l l_t \left(f_t^{-1} \left(m_{it}, k_{it} \right), k_{it} \right) + f_t^{-1} \left(m_{it}, k_{it} \right) + \varepsilon_{it},$$

  and since we're being nonparametric about f_t^{-1},
  it should absorb $\beta_l l_t \left(f_t^{-1} \left(m_{it}, k_{it} \right), k_{it} \right)$.

- There should be no variation in l_{it} left over to identify β_l.
- A similar argument applies when doing the inversion with investment.

Collinearity in practice and in principle

- It could be the case that l_{it} takes different values in the data for the same values of (m_{it}, k_{it}). ACF's argument is about collinearity in principle, given the assumptions of LP.
- Some potential sources of independent variation:
  (*Which one works?*)
  - unobserved variation in firm-specific input prices.
  - measurement error in l_{it} or m_{it}
  - optimization error in l_{it} or m_{it}

Collinearity in practice and in principle

- It could be the case that l_{it} takes different values in the data for the same values of (m_{it}, k_{it}). ACF's argument is about collinearity in principle, given the assumptions of LP.
- Some potential sources of independent variation: (*Which one works?*)
  - unobserved variation in firm-specific input prices.
  - measurement error in l_{it} or m_{it}
  - optimization error in l_{it} or m_{it}
- Unobserved input prices break the inversion; measurement error should have no identifying power.
- While optimization error in l_{it} works econometrically, it's not the most appealing assumption economically.

Another failed solution

- Note that the whole problem comes about because labor and materials are set simultaneously. This means one way to break the collinearity is to assume they are set with respect to different information sets.
- Let's try to break the informational equivalence with timing assumptions. Suppose:
  - m_{it} is set at time t
  - l_{it} is set at time $t - b$ with $0 < b < 1$
  - ω is Markovian in between sub-periods:

$$\begin{aligned} p\left(\omega_{i,t-b}|I_{i,t-1}\right) &= p\left(\omega_{i,t-b}|\omega_{it-1}\right) \\ p\left(\omega_{it}|I_{i,t-b}\right) &= p\left(\omega_{i,t}|\omega_{i,t-b}\right) \end{aligned}$$

- But this doesn't work! And neither does having m_{it} set first. *(Why?)*

An implausible solution

- Let's try again:
  - l_{it} is set at time t
  - m_{it} is set at time $t - b$ with $0 < b < 1$
  - we have a more complicated structure of productivity shocks:

$$y_{it} = \beta_l l_{it} + \beta_m m_{it} + \beta_k k_{it} + \omega_{i,t-b} + \eta_{it},$$

$$p\left(\omega_{i,t-b} | I_{i,t-1}\right) \quad = \quad p\left(\omega_{i,t-b} | \omega_{i,t-1}\right) \ ,$$

  - and there is some unobservable shock to labor prices which is realized between $t - b$ and t. This shock must be i.i.d.

- l_{it} has its own shock to respond to, creating independent variation, and the productivity inversion still works because the new shock is not a state variable.

- This works, but as ACF argue, it's rather ad-hoc and difficult to motivate.

Collinearity in Olley Pakes

- Olley Pakes's control function has the same collinearity issue, but ACF argue it can be avoided with assumptions which "might be a reasonable approximation to the true underlying process."
- Assume that l_{it} is set at $t-b$ with $0<b<1$. ω has a Markovian between subperiods. Then:

$$l_{it} = l_t\left(\omega_{i,t-b}, k_{it}\right),$$

  so we have variation in l_{it} which is independent of (ω_{it}, k_{it}).
- Note that even though l_{it} is set before investment i_{it}, investment won't depend on l_{it} because it is a static input. So the productivity inversion is unchanged.
- Thus, these timing assumptions can save the OP production inversion and first stage.

ACF's alternative procedure I

- Consider value added production function:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it}.$$

- ACF's procedure is based on the same timing assumption that "saves" OP: labor chosen at $t - b$, slightly earlier than when materials are chosen at t.

- Point of first stage is just to get expected output:

$$y_{it} = \Phi_t \left(m_{it}, k_{it}, l_{it} \right) + \epsilon_{it}$$

where

$$\Phi_t \left(m_{it}, k_{it}, l_{it} \right) = \beta_k k_{it} + \beta_l l_{it} + f_{it}^{-1} \left(m_{it}, k_{it}, l_{it} \right)$$

... first stage no longer recovers β_l, avoiding one of our earlier problems.

ACF's alternative procedure II

- After the first stage, we have $\hat{\Phi}_{it}$, expected output.
- We can construct a measure of productivity given coefficients:

$$\hat{\omega}_{it}\left(\beta_k, \beta_l\right) = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it}$$

- Then, non-parametrically regressing $\hat{\omega}_{it}\left(\beta_k, \beta_l\right)$ on $\hat{\omega}_{i,t-1}\left(\beta_k, \beta_l\right)$, we can construct the innovations:

$$\hat{\xi}_{it}\left(\beta_k, \beta_l\right) = \hat{\omega}_{it}\left(\beta_k, \beta_l\right) - E\left(\hat{\omega}_{it}\left(\beta_k, \beta_l\right) | \hat{\omega}_{i,t-1}\left(\beta_k, \beta_l\right)\right)$$

ACF's alternative procedure III

- Estimation relies on the following moments:

$$T^{-1}N^{-1}\sum_t\sum_i \hat{\xi}_{it}\left(\beta_k,\beta_l\right)\begin{pmatrix} k_{it} \\ l_{i,t-1} \end{pmatrix}$$

- In the second stage, these two moments are used to estimate both β_k and β_l.
- In ACF's framework, l_{it} isn't a function of ω_{it} but of $\omega_{i,t-b}$. However, labor will still be correlated with part of the innovation in productivity, so we still need to use lagged labor in the moments.
- The moment with lagged labor is very much in the spirit of OP and LP, and they actually used it as an overidentifying restriction.

Gandhi, Navarro, and Rivers

- Gandhi, Navarro, and Rivers argue that a similar identification concern applies to ACF, at least if materials are on the RHS of the production function.
- OP and ACF's y is the log of revenue minus materials expenditure. This is called a **restricted profit** production function. GNR argue that restricted profit production functions may have problems because they are justified as a local approximation, and the variation in production data is not small.

Structural Value Added

- A structural value added production function starts with a functional form like this:
$$Q = \min\left(\gamma M, F\left(L, K\right)\right)$$
where F could be any form of production function (Cobb-Douglas, Translog).
- Cost minimization implies
$$Q = \gamma M = F\left(L, K\right)$$
- This motivates estimating a production without M on the RHS, but note the LHS is just Q and doesn't subtract materials.
- γ can be recovered from simple ratios.
- De Loecker and Scott (2017) implement this for the brewing industry.

"Reallocation, Firm Turnover, and Efficiency:
Selection on Productivity or Profitability"
Foster, Haltiwanger, and Syverson (2008)

Overview

- They look at some rare industries where quantity data is available, allowing them to separate physical and revenue productivity
- Findings:
  - Physical productivity is inversely correlated with price
  - Young producers charge lower prices than incumbents, meaning the literature understates entrants' productivity advantages

Measurement

- Productivity is measured as follows:

$$tfp_{it} = y_{it} - \alpha_l l_{it} - \alpha_k k_{it} - \alpha_m m_{it} - \alpha_e e_{it}$$

- Coefficients (α) are just taken from input shares by industry.
- Different measures use different output measures y:
  - TFPQ uses physical output
  - TFP uses deflated sales (using standard industry-level deflators from NBER)
  - TFPR are sales deflated by mean prices observed in their data

Correlations

TABLE 1—SUMMARY STATISTICS FOR OUTPUT, PRICE, AND PRODUCTIVITY MEASURES

Correlations								
Variables	Trad'l. output	Revenue output	Physical output	Price	Trad'l. TFP	Revenue TFP	Physical TFP	Capital
Traditional output	1.00							
Revenue output	0.99	1.00						
Physical output	0.98	0.99	1.00					
Price	−0.03	−0.03	−0.19	1.00				
Traditional TFP	0.19	0.18	0.15	0.13	1.00			
Revenue TFP	0.17	0.21	0.18	0.16	0.86	1.00		
Physical TFP	0.17	0.20	0.28	−0.54	0.64	0.75	1.00	
Capital	0.86	0.85	0.84	−0.04	0.00	−0.00	0.03	1.00
Standard deviations								
	1.03	1.03	1.05	0.18	0.21	0.22	0.26	1.14

Notes: This table shows correlations and standard deviations for plant-level variables for our pooled sample of 17,669 plant-year observations. We remove product-year fixed effects from each variable before computing the statistics. All variables are in logs. See the text for definitions of the variables.

- correlation between traditional TFP, physical TFP is substantial, but imperfect
- negative correlation between price and physical TFP

Demand

- They estimate a demand system for each industry:

$$\ln q_{it} = \alpha_0 + \alpha_1 p_{it} + \sum_t \alpha_t YEAR_t + \alpha_2 \ln (INCOME_{mt}) + \eta_{it}$$

  where $INCOME_{mt}$ is the income in a firm's local market m

- They use the residuals from these regressions as a measure of demand shocks.

Persistence

TABLE 3—PERSISTENCE OF PRODUCTIVITY, PRICES AND DEMAND SHOCKS

	Five-year horizon		Implied one-year persistence rates	
Dependent variable	Unweighted regression	Weighted regression	Unweighted regression	Weighted regression
Traditional TFP	0.249	0.316	0.757	0.794
	(0.017)	(0.042)		
Revenue TFP	0.277	0.316	0.774	0.794
	(0.021)	(0.042)		
Physical TFP	0.312	0.358	0.792	0.814
	(0.019)	(0.049)		
Price	0.365	0.384	0.817	0.826
	(0.025)	(0.066)		
Demand shock	0.619	0.843	0.909	0.966
	(0.013)	(0.021)		

TABLE 6—SELECTION ON PRODUCTIVITY OR PROFITABILITY?

Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Traditional TFP	−0.073						
	(0.015)						
Revenue TFP		−0.063					
		(0.014)					
Physical TFP			−0.040			−0.062	−0.034
			(0.012)			(0.014)	(0.012)
Prices				−0.021		−0.069	
				(0.018)		(0.021)	
Demand shock					−0.047		−0.047
					(0.003)		(0.003)
Controlling for plant capital stock							
Traditional TFP	−0.069						
	(0.015)						
Revenue TFP		−0.061					
		(0.013)					
Physical TFP			−0.035			−0.059	−0.034
			(0.012)			(0.014)	(0.012)
Prices				−0.030		−0.076	
				(0.018)		(0.021)	
Demand shock					−0.030		−0.029
					(0.004)		(0.004)
Capital stock	−0.046	−0.046	−0.046	−0.046	−0.023	−0.046	−0.023
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)

- one-standard deviation increases physical TFP and prices seem to have similar impacts on exit probabilities

Production and Markups

"Markups and Firm-Level Export Status"
De Loecker and Warzynski (2012)

Overview

- Demonstrates how production function can be used to make inferences about markups
- Applied question: how do markups of exporters differ from non-exporters, and how does a firm's productivity change when it becomes an exporter?
- Findings:
  - Exporters have higher markups than importers
  - Markups increase when a firm becomes an exporter
  - Note similarity to De Loecker (2011), but focus is now on exporter status rather than trade liberalization

Sketch of main idea I

- Definition of markup: $\mu = P/MC$
- Let P_{it}^{v} represent the price of input v and let P_{it} represent the price of output.
- Production function:

$$Q_{it} = Q_{it}\left(X_{it}^{1}, \ldots, X_{it}^{V}, K_{it}, \omega_{it}\right)$$

  where $v = 1, 2, \ldots, V$ indexes variable inputs.
- Assumption: variable inputs are set each period to minimize costs.

Sketch of main idea II

- Lagrangian for cost minimization problem:

$$L\left(X_{it}^{1},\ldots,X_{it}^{V},K_{it},\lambda_{it}\right)=\sum_{v=1}^{V}P_{it}^{v}X_{it}^{v}+r_{it}K_{it}+\lambda_{it}\left(Q_{it}-Q_{it}\left(\cdot\right)\right)$$

- First-order condition:

$$P_{it}^{v}-\lambda_{it}\frac{\partial Q_{it}\left(\cdot\right)}{\partial X_{it}^{v}}=0,$$

where λ_{it} is the marginal cost of production at production level Q_{it}.

Sketch of main idea III

- First-order condition:

$$P_{it}^{v} - \lambda_{it} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}} = 0.$$

- Multiplying by X_{it}^{v}/Q_{it}:

$$\frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}} \frac{X_{it}^{v}}{Q_{it}} = \frac{1}{\lambda} \frac{P_{it}^{v} X_{it}^{v}}{Q_{it}}.$$

- With $\mu_{it} \equiv P_{it}/\lambda_{it}$,

$$\frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}} \frac{X_{it}^{v}}{Q_{it}} = \mu_{it} \frac{P_{it}^{v} X_{it}^{v}}{P_{it} Q_{it}}$$

where we have multiplied and divided by P_{it} on the RHS.

The markup formula

This leads to a simple expression:

$$\mu_{it} = \theta_{it}^{v} \left(\alpha_{it}^{v}\right)^{-1}$$

where θ_{it}^{v} is the output elasticity with respect to input v, and α_{it}^{v} is expenditures on input v as a share of revenues.

- On its own, this formula is nothing new
- What's new about DLW is how flexible they are about estimating θ_{it}^{v} and how they base their inferences about markups on careful production function estimation.

The demand-based approach

- Recall the formula for monopoly pricing:

$$\frac{p}{mc} = \frac{1}{1 + \mathcal{E}_D^{-1}}$$

  where $\mathcal{E}_D^{-1}$ is the inverse elasticity of demand.

- In more complicated settings (e.g., differentiated products), we can still solve for markups as a function of demand elasticities.

- Demand-based approach has been the standard, but notice the many assumptions involved:
  - Typically static Nash-Bertrand competition (or at least some imperfect competition game where we can easily solve for the equilibrium)
  - Instruments to identify demand
  - Functional form assumptions on demand system, model of consumer heterogeneity

CD: example

- Assume labor is a flexible input.
- With Cobb-Douglas production function,

$$Q_{it} = \exp\left(\omega_{it}\right) L^{\beta_L} K^{\beta_K},$$

output elasticity of labor is just a constant:

$$\theta_{it}^{L} = \frac{\partial Q_{it}}{\partial L_{it}} \frac{L_{it}}{Q_{it}} = \beta_L.$$

- Markup:

$$\mu_{it} = \frac{\beta_L}{\alpha_{it}^{L}}$$

CD: concerns

Cobb-Douglas markup:

$$\mu_{it} = \frac{\beta_L}{\alpha_{it}^L}$$

Some things we might worry about:

- Bias in estimating β_L without appropriate econometric strategy (always a concern in production function estimation)
- Cobb-Douglas is very restrictive, imposing output elasticity which does not depend on Q nor the relative levels of inputs. Variation in expenditure shares will be only source of variation in markups.
- If we assume variation of input share is independent of output elasticity, then any variation in productivity which affects the input share is being treated as variation in markups.

Translog production function

- DLW's main results are based on a translog production function:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{lk} l_{it} k_{it} + \omega_{it} + \varepsilon_{it}.$$

- Translog output elasticities:

$$\hat{\theta}_{it}^L = \hat{\beta}_l + 2\hat{\beta}_{ll} l_{it} + \hat{\beta}_{lk} k_{it},$$

so translog production is flexible enough to allow for a first-order approximation to how output elasticities vary with input use.

Empirical framework

- Consistent with production function estimation literature, they assume Hicks-neutral productivity shocks:

$$Q_{it} = F\left(X_{it}^{1}, \ldots, X_{it}^{V}, K_{it}; \beta\right) \exp\left(\omega_{it}\right).$$

- Also allow for some measurement error in production:

$$y_{it} = \ln Q_{it} + \varepsilon_{it}$$

$$y_{it} = f\left(x_{it}, k_{it}; \beta\right) + \omega_{it} + \varepsilon_{it}$$

The control function

- Following Levinsohn and Petrin, use materials to proxy for productivity

$$m_{it} = m_t\left(k_{it}, \omega_{it}, \mathbf{z}_{it}\right)$$

  where $\mathbf{z}_{it}$ are controls.

- Note: a big claim of the paper is estimating "markups without specifying how firms compete in the product market"
- But here, $\mathbf{z}_{it}$ must control for everything which shifts input demand choices or else there will be variation in productivity they're not controlling for (and hence some of the variation in their inferred markups may actually come from variation in productivity). Related: identification critique of Doraszelski and Jaumandreu (2019)
- In the appendix, they explain that $\mathbf{z}_{it}$ includes input prices, lagged inputs (meant to capture variation in input prices), and exporter status.

Physical output vs. sales

- Note that the theory is developed in terms of outputs, but DLW only have sales (as usual).
- For a price-taking firm, there's no problem rewriting the formula in terms of sales:

$$\frac{\partial R_{it}(\cdot)}{\partial X_{it}^{v}} \frac{X_{it}^{v}}{R_{it}} = \frac{P_t \partial Q_{it}(\cdot)}{\partial X_{it}^{v}} \frac{X_{it}^{v}}{P_t Q_{it}} = \mu_{it} \frac{P_{it}^{v} X_{it}^{v}}{P_{it} Q_{it}}$$

  because $\frac{\partial R_{it}(\cdot)}{\partial X_{it}^{v}} = \frac{P_t \partial Q_{it}(\cdot)}{\partial X_{it}^{v}}$.
- However, if the firm has market power,

$$\frac{\partial R_{it}(\cdot)}{\partial X_{it}^{v}} = \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}} \left(P_{it} + \frac{\partial P_{it}}{\partial Q_{it}} \right).$$

- Potential for analogous issue in input prices...

Production vs. Demand approach

- De Loecker and Scott (2017) compare the two approaches for US brewing industry and find agreement between the two strategies.
- Are they substitutes? For measurement of markups and retrospective studies, arguably they are. However, production approach cannot be used for merger analysis. Counterfactuals require an understanding of demand.
- Note that since the two approaches rely on different assumptions, they can be used together to test and/or relax certain assumptions.

Topic: Rise of Market Power

- De Loecker, Eeckhout, and Unger (2020) argue that mean markups have risen since 1980, driven by fattening of upper tail.
- Raval documents differences in markup trends depending on choice of variable input (labor v. materials).
- Raval and Doraszelski and Jaumandreu (2013, 2019) point to the importance of non-neutral productivity differences.

Some intuition

- Consider static Cobb-Douglas production with

$$Q_i = L^{\beta_{it}^L} M^{\beta_{it}^M} K^{\beta^K}$$

with $\forall i: \quad \beta_{it}^L + \beta_{it}^M + \beta^K = 1$

- Suppose there is no heterogeneity in or changes in markups
- Recall that input shares will be proportional to β's.
- If β_{it} is increasing and β_{it} is decreasing over time, and if we estimate β^M and β^L without allowing for heterogeneity, then $\frac{\hat{\beta}^L}{s_{it}^L}$ and $\frac{\hat{\beta}^M}{s_{it}^M}$ will mechanically have opposite trends.
- Looking within the cross section, if we aggregate too much, then firms with higher materials shares will have lower labor shares, and their markups will be negatively correlated.

Topic: Misallocation

- Hsieh and Klenow (2009) dispersion in marginal product can be interpreted as evidence of distortions and/or misallocation. There appears to be much more dispersion in China and India than the US.
- Asker, Collard-Wexler, and De Loecker (2014): volatility in productivity combined with non-transferability of capital is naturally going to lead to dispersion in marginal product of capital. Cross-industry volatility in productivity can largely explain Hsieh and Klenow's cross-country differences.
- Haltiwanger, Kulick, and Syverson (2018): revisit Hsieh and Klenow with more flexible framework.