# Production Function Estimation

Paolo Zacchia

Microeconometrics

Lecture 15

# Production function estimation: overview

- Like demand functions, production functions are ubiquitous in economic theory and models. Like demand functions, they are also surprisingly difficult to estimate. The main issue is one of the **omitted variable bias** kind.

- Any decent attempt for a solution shall be based upon **panel data**. Direct panel data approaches are thus reviewed.

- The conventional standard is based upon **control function** methods in the modern formulation by Ackerberg, Caves and Frazer (2015). They are at the center of this lecture.

- As noted by Wooldrdige (2009) these approaches are tightly connected with classical panel data approaches.

- From them, both extensions/applications (De Loecker, 2011) and critiques (Gandhi, Navarro and Rivers, 2020) sprang up.

# Why estimating production functions?

- In many empirical studies, interest falls on estimating **total factor productivity** (TFP). In a Cobb-Douglas setting:

$$\log TFP_i = \log Y_i - \beta_K \log K_i - \beta_L \log L_i$$

thus, consistent estimators $\widehat{\beta}_K$ and $\widehat{\beta}_L$ allow to evaluate (log) TFP as the regression residual. This extends to more general input sets, other functional forms, *et cetera*.

- Production function estimation allows to measure **markups** (De Loecker and Warzynski, 2012). By cost minimization:

$$\eta_{Y_i}^{X_{ki}} = \frac{X_{ki}}{Y_i} \frac{\partial F\left(X_{1i}, \dots, X_{Ki}\right)}{\partial X_{ki}} = \mu_i Z_{ki}$$

where $\mu_i$ is firm $i$'s markup, $F(\cdot)$ is its production function, $Z_{ki}$ is share of the $k$-th input ($X_{ki}$) over revenue, while $\eta_{Y_i}^{X_{ki}}$ is the corresponding output elasticity (within a Cobb-Douglas setting it equals $\beta_k$). Solving for $\mu_i$ requires estimating $\eta_{Y_i}^{X_{ki}}$.

# The transmission bias (1/3)

- Recall the "log-log" production function model motivated on a Cobb-Douglas functional form from Lecture 7.

$$\log Y_i = \alpha + \beta_K \log K_i + \beta_L \log L_i + \omega_i$$

- As discussed in Lecture 12, the regressors are thought to be **endogenous**: $\mathbb{E}\left[\omega_i \middle| \log K_i\right] \neq 0$; $\mathbb{E}\left[\omega_i \middle| \log L_i\right] \neq 0$.

- The motivation is that the error term $\omega_i$ is likely to subsume some **unobserved input**, which is "transmitted" to inputs like capital and labor because of complementarity, as per the First Order Conditions from profit maximization:

$$\log \beta_K + \alpha + (\beta_K - 1) \log K_i + \beta_L \log L_i + \omega_i = \log P_K$$
$$\log \beta_L + \alpha + \beta_K \log K_i + (\beta_L - 1) \log L_i + \omega_i = \log P_L$$

  where $P_K$ is the price of capital while $P_L$ that of labor. This **"transmission bias"** was originally noted by Andrews and Marschack (1944).

# The transmission bias (2/3)

- From a theoretical standpoint, the transmission bias applies only if $\omega_i$ is **observed by firms** when $K_i$ and $L_i$ are chosen. *Timing* is key for production function estimation!

- Error terms of different kind might pose additional problems. For example, $Y_i$ is typically not calculated directly but must obtained by **deflating firm revenues** $R_i$: $Y_i = R_i/P_i$. Here $P_i$ is the price of firm $i$'s goods or services.

- However, typically researchers do not observe $P_i$ but $P_{s(i)}$, a price index for firm $i$'s **industry** $s(i)$. The model becomes:

$$\log R_i - \log P_{s(i)} = \alpha + \beta_K \log K_i + \beta_L \log L_i + \varpi_i + \omega_i$$

  where $\varpi_i = \log P_i - \log P_{s(i)}$ is another error term.

- If $\varpi_i$ is random it poses no problem to estimation. However, there are typically reasons to think that it is *not* random.

# The transmission bias (3/3)

- Issues about deflating variables can also apply to right-hand side regressors (inputs) express in monetary values, thereby leading to measurement error.

- This discussion suggests that information about firm-specific **prices** might help! Unfortunately, this is rarely available or accurate in firm-level data.

- In particular, if $P_K$ and $P_L$ were observable *and* had enough *exogenous* variation they would work as great **instruments**. Unfortunately, those two conditions are hardly satisfied.

- If $P_K$ and $P_L$ are observed with little variation they may still be exploited: in traditional approaches (e.g. McElroy, 1978) they serve *direct estimation of the First Order Conditions*.

- These traditional approaches however can be problematic if some inputs, like capital, are chosen dynamically.

# More general production functions

- The problem can be at least in part mitigated by including other $K$ inputs $(X_{1i}, \ldots, X_{Ki})$ into the model.

$$\log Y_i = \alpha + \beta_K \log K_i + \beta_L \log L_i + \sum_{k=1}^{K} \beta_{X_k} \log X_{ki} + \omega_i$$

  The whole set of inputs is difficult to observe by researchers, but one can often see the total **cost of materials** $M_i$.

- An approach that circumvents the need to observe $P_i$ is:

$$\log V_i = \alpha + \beta_K \log K_i + \beta_L \log L_i + \omega_i$$

  where $V_i$ is a firm's **value added**. Note, however, that this is a model for value added, not for gross output $Y_i$.

- A more general CES specification of the production function (of which the Cobb-Douglas is a special case, see Lecture 11) hardly helps, because the transmission bias still occurs.

# Translog production functions

- To address concerns about the realism of the Cobb-Douglas specification, one can use a **translog** one, which is a better approximation of the (unknown) true production function.

$$\begin{aligned}\log Y_i = \alpha + \beta_K \log K_i + \beta_L \log L_i + \\ + \gamma_{KK} (\log K_i)^2 + \gamma_{LL} (\log L_i)^2 + \\ + \gamma_{KL} (\log K_i)(\log L_i) + \omega_i\end{aligned}$$

- Suitable theory-driven **restrictions** on the parameters may apply, if necessary (example: constant returns to scale).

- There is nothing that prevents OLS estimation of this model. Yet this is about *specification*, not *identification*: a translog model does not prevent the transmission bias.

- With many inputs $X_{ki}$ a *curse of dimensionality* occurs, not unlike in translog models for demand estimation. Here, this is likely to lead to issues of *multicollinearity*.

# Direct panel data approaches (1/5)

- Consider a Cobb-Douglas production function model like the one from Lecture 7, but adapted to panel data:

$$y_{it} = \alpha_i + \beta_K k_{it} + \beta_L \ell_{it} + \omega_{it} + \varepsilon_{it}$$

where:

$$\begin{aligned} y_{it} &\equiv \log Y_{it} \\ k_{it} &\equiv \log K_{it} \\ \ell_{it} &\equiv \log L_{it} \end{aligned}$$

are *logarithms of random variables* and *not* realizations. This is a notational convention typical of production functions.

- Here, the log of "total" productivity $A_{it}$ is split as:

$$\log A_{it} = \alpha_i + \omega_{it} + \varepsilon_{it}$$

that is, between a constant factor $\alpha_i$ and two **time-varying** factors $\omega_{it}$ and $\varepsilon_{it}$ that are discussed next.

# Direct panel data approaches (2/5)

- Why a distinction between two time-varying components of the error term? Whereas part of the error can be treated as exogenous:
$$\mathbb{E}\left[\left.\varepsilon_{it}\right| k_{it}, \ell_{it}\right] = 0$$
(think about lucky events), the other part may not:
$$\mathbb{E}\left[\alpha_i, \left.\omega_{it}\right| k_{it}, \ell_{it}\right] \neq \mathbf{0}$$
as firm adapt their inputs $k_{it}, \ell_{it}$ to their own circumstances.

- Suppose that $\omega_{it} = 0$ for all $i = 1, \dots, N$ and $t = 1, \dots, T$, as if the only unobserved inputs are constant in time: $\alpha_i$.

- The model can be thus estimated via fixed effects regression. Yet the empirical practice has shown that this typically leads to **unrealistically small** estimates of $\beta_K$; intuitively, $\beta_K$ is identified off insufficient time variation in $k_{it}$.

# Direct panel data approaches (3/5)

- Now reintroduce the time-varying endogenous error $\omega_{it}$, and suppose it follows an AR(1) process:

$$\omega_{it} = \rho\omega_{i(t-1)} + \xi_{it}$$

  where in principle $\rho \in (-1, 1)$, though presumably $\rho \in (0, 1)$.

- The random shock $\xi_{it}$ is called the "innovation" term of the endogenous unobserved productivity. This terminology and notation are shared with more general decompositions of $\omega_{it}$.

- Because $\xi_{it}$ is "new," it is safe to assume:

$$\mathbb{E}\left[\xi_{it} \middle| \Delta k_{i(t-s)}, \Delta \ell_{i(t-s)}\right] = 0$$

  for $s \geq 1$. This also applies to $\upsilon_{it} \equiv \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{i(t-1)}$.

- The lagged main model, multiplied by $\rho$, writes as follows.

$$\rho y_{i(t-1)} = \rho\alpha_i + \beta_K \rho k_{i(t-1)} + \beta_L \rho \ell_{i(t-1)} + \rho\omega_{i(t-1)} + \rho\varepsilon_{it}$$

# Direct panel data approaches (4/5)

- By "ρ-differencing" the original production function, that is by subtracting the previous equation from both sides, it is:

$$y_{it} - \rho y_{i(t-1)} = \alpha_i (1 - \rho) + \beta_K \left( k_{it} - \rho k_{i(t-1)} \right) + \beta_L \left( \ell_{it} - \rho \ell_{i(t-1)} \right) + \upsilon_{it}$$

  and $\omega_{it}$ vanishes. This yields a typical **dynamic model** for panel data, as per the framework outlined in Lecture 12.

- A standard "System GMM" estimation approach is based on **moments in differences** *à la* Blundell and Bond like:

$$\mathbb{E}\left[ \begin{pmatrix} \Delta k_{i(t-s)} \\ \Delta \ell_{i(t-s)} \end{pmatrix} (\alpha_i (1 - \rho) + \upsilon_{it}) \right] = 0$$

  for $s \geq 2$. Observe that this approach is valid if $\mathbb{E}[k_{is}\alpha_i] \neq 0$ and $\mathbb{E}[\ell_{is}\alpha_i] \neq 0$ are **constant in time**, which occurs under the conditions specified by Blundell and Bond (1998).

# Direct panel data approaches (5/5)

- While theoretically sound, even this approach has not stood the test of empirical practice all too well.

- There are two intertwined problems: instruments for high $s$ appear to be **weak**, and overidentification/**exogeneity** tests (along with tests for the **autocorrelation** of the residuals) suggest to select values of $s$ that are even higher than 2.

- In short, $\omega_{it}$ cannot be reduced to an AR(1) process. Taking instruments further back in time to account for that is risky.

- Improvements are obtained by adding to the GMM problem some **moments in levels** *à la* Arellano and Bond (1991):

$$\mathbb{E}\left[\begin{pmatrix} k_{i(t-s)} \\ \ell_{i(t-s)} \end{pmatrix} \Delta \upsilon_{it}\right] = 0$$

  for $s \geq 2$. However, the approach is still not very popular.

# Control function methods: overview

- The so-called **control function** methods for the estimation of production functions are semi-structural methods based on panel data that impose limited assumptions on $\omega_{it}$.

- Estimation is based on semi-parametric, **non-linear** control functions for $\omega_{it}$, *proxied by* some given production inputs.

- They are grounded on assumptions about the **timing** of firm decisions about their production inputs.

- The original method was devised by Olley and Pakes (1996; OP); there, the control function is based on **investment** $I_{it}$.

- Levinsohn and Petrin (2003; LP) proposed an improvement via a control function based on the **cost of materials** $M_{it}$.

- Finding that both methods are flawed, Ackerberg, Caves and Frazer (2015; ACF) developed a suitable **alternative**.

# Proxying unobservables with investment (1/9)

- What follows is an exposition of the OP method that adopts the same notation as in the critical summary by ACF.

- Let the model be as follows:

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + \omega_{it} + \varepsilon_{it}$$

  where $\beta_0$ is constant and where all unobserved heterogeneity is embedded into the so-called **productivity shock** $\omega_{it}$.

- Instead, $\varepsilon_{it}$ is called **transitory shock** because, unlike $\omega_{it}$, it is independent of both its past and future realizations.

- In what follows, denote firm $i$'s investment at time $t$ as $I_{it}$, and let $i_{it} \equiv \log I_{it}$. This choice is somewhat unfortunate ($i$ is duplicated) but is traditional in both OP and ACF.

- It is useful to restate the original OP **assumptions** as ACF also did. The OP procedure supposedly rests on them.

# Proxying unobservables with investment (2/9)

### Assumption 1

**Information set.** The firm's information set at at time $t$, that is $\mathcal{I}_t$, includes current and past productivity shocks $\{\omega_{i\tau}\}_{\tau=0}^{t}$ but does not include future productivity shocks $\{\omega_{i\tau}\}_{\tau=t+1}^{\infty}$. The transitory shocks satisfy $\mathbb{E}\left[\varepsilon_{it} \middle| \mathcal{I}_t\right] = 0$.

### Assumption 2

**First Order Markov.** Productivity shocks evolve according to the probability distribution

$$P\left(\omega_{i(t+1)} \middle| \mathcal{I}_{it}\right) = P\left(\omega_{i(t+1)} \middle| \omega_{it}\right).$$

This distribution is known to firms and stochastically increasing in the conditioned productivity shock $\omega_{it}$.

Both assumptions are commented next, alongside Assumption 3.

# Proxying unobservables with investment (3/9)

**Assumption 3**

**Timing of input choices.** Firms accumulate capital according to

$$k_{it} = \kappa\left(k_{i(t-1)}, i_{i(t-1)}\right)$$

where investment $i_{i(t-1)}$ is chosen in period $t-1$. The labor input $\ell_{it}$ is non-dynamic and chosen at $t$.

Some comments on the assumptions so far are due.

1. Firms cannot foresee the future (short of guessing it).
2. Current productivity $\omega_{it}$ is a sufficient statistic for predicting the future $\omega_{i(t+1)}$.
3. Capital is completely (pre-)determined at time $t$: this is the key assumption (it takes time to buy, install new equipment). Labor is non-dynamic in the sense that today's $\ell_{it}$ does not affect future profits (firms are free to fire workers).

# Proxying unobservables with investment (4/9)

**Assumption 4**

**Scalar unobservable.** Firms' investment decisions are given by

$$i_{it} = f_t\left(k_{it}, \omega_{it}\right).$$

**Assumption 5**

**Strict monotonicity.** $f_t\left(k_{it}, \omega_{it}\right)$ is strictly increasing in $\omega_{it}$.

Here are brief comments for these two additional assumptions.

4. Investment depends on capital and productivity as they are the only *state variables* (labor is not since it is non-dynamic).
5. Monotonicity is implied by Assumption 2 and the underlying dynamic optimization problem.

Note: all firms have the same $f_t\left(\cdot\right)$, though it varies over time.

# Proxying unobservables with investment (5/9)

- These assumptions motivate the OP **estimation** approach, which proceeds in **two stages**.

- The key idea is to "invert" the monotonic $f_t(k_{it}, \omega_{it})$ for $\omega_{it}$:

$$\omega_{it} = f_t^{-1}(k_{it}, i_{it})$$

so as to obtain a control function for the productivity shock.

- This delivers a so-called **first stage** that identifies $\beta_L$:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + f_t^{-1}(k_{it}, i_{it}) + \varepsilon_{it} \\ &= \beta_L \ell_{it} + \Phi_t(k_{it}, i_{it}) + \varepsilon_{it} \end{aligned}$$

where $\Phi_t(k_{it}, i_{it})$ is a composite function that is treated **non-parametrically**. This is framed via a **moment condition**.

$$\mathbb{E}\left[\varepsilon_{it} | \mathcal{I}_t\right] = \mathbb{E}\left[y_{it} - \beta_L \ell_{it} - \Phi_t(k_{it}, i_{it}) | \mathcal{I}_t\right] = 0$$

# Proxying unobservables with investment (6/9)

- The **second stage** identifies $\beta_K$. First, by Assumption 2:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \middle| \mathcal{I}_{t-1}\right] + \xi_{it} = g\left(\omega_{i(t-1)}\right) + \xi_{it}$$

where $\mathbb{E}\left[\omega_{it} \middle| \mathcal{I}_{t-1}\right] = \mathbb{E}\left[\omega_{it} \middle| \omega_{i(t-1)}\right]$ and $\mathbb{E}\left[\xi_{it} \middle| \mathcal{I}_{t-1}\right] = 0$.

- Substituting this into the model gives:

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + g\left(\omega_{i(t-1)}\right) + \xi_{it} + \varepsilon_{it}$$

where $\omega_{i(t-1)} = \Phi_{t-1}\left(k_{i(t-1)}, i_{i(t-1)}\right) - \beta_0 - \beta_K k_{i(t-1)}$ as per the previous definition of the composite function $\Phi_t(\cdot)$. Here $g(\cdot)$ is also treated **non-parametrically**.

- This yields another **moment condition**.

$$\mathbb{E}\left[\xi_{it} + \varepsilon_{it} \middle| \mathcal{I}_{t-1}\right] = \mathbb{E}\Big[y_{it} - \beta_0 - \beta_K k_{it} - \beta_L \ell_{it} -$$

$$- g\left(\Phi_{t-1}\left(k_{i(t-1)}, i_{i(t-1)}\right) - \beta_0 - \beta_K k_{i(t-1)}\right) \Big| \mathcal{I}_{t-1}\Big] = 0$$

# Proxying unobservables with investment (7/9)

- By expressing $\mathcal{I}_t$ as a set of **instruments**: typically, suitable lags of $k_{i(t-s)}$, $i_{i(t-s)}$ and $\ell_{i(t-s)}$ for $s = 0, 1, \dots, t-1$, one can easily recast the moment conditions in a way amenable to GMM estimation (via the Law of Iterated Expectations).

- The non-parametric functions $\mathit{\Phi}_t(\cdot)$ and $g(\cdot)$ are expressed in the empirical model via **polynomial series** (typically of third or fourth degree) of their arguments.

- Ideally, both sets of moments shall be **jointly** estimated (Ai and Chen, 2003; Wooldridge, 2009), but the presence of the two non-parametric functions can make this cumbersome.

- The popular approach is thus to estimate the two stages **in sequence**. In the second stage, $\mathit{\Phi}_{t-1}(\cdot)$ is *substituted* by:

$$\widehat{\varphi}_{i(t-1)} = \widehat{\mathit{\Phi}}_{t-1}\left(k_{i(t-1)}, i_{i(t-1)}\right)$$

  as predicted by the first stage (a "plug-in" approach).

# Proxying unobservables with investment (8/9)

- In their original paper, OP applied their method to estimate production functions in the US telecommunications industry of their time (1963-1987).

- They also included a firm's *age* $a_{it}$ in their control functions, but this is not common nowadays.

- Since they worked with an unbalanced sample drawn from an evolving industry, all their theoretical results were obtained *conditional on firm survival* (not "exiting"). They computed estimates $\widehat{P}_{it}$ of a firm's *survival probability* at time $t$.

- Their first stage was as follows:

$$y_{it} = \beta_L \ell_{it} + \sum_{l=0}^{4-m-n} \sum_{m=0}^{4-n} \sum_{n=0}^{4} \phi_{lmn} i_{it}^{l} k_{it}^{m} a_{it}^{n} + \varepsilon_{it}$$

giving $\widehat{\varphi}_{it} = \sum_{l=0}^{4-m-n} \sum_{m=0}^{4-n} \sum_{n=0}^{4} \widehat{\phi}_{lmn} i_{it}^{l} k_{it}^{m} a_{it}^{n}$ for later use.

# Proxying unobservables with investment (9/9)

- Their second stage was instead as follows, given $\widehat{\beta}_L$ (the first stage estimate of $\beta_L$) and $\widehat{\varphi}_{it}$. They estimated it via NLLS.

$$
\begin{aligned}
y_{it} - \widehat{\beta}_L \ell_{it} = \beta_0^* + \beta_A a_{it} + \beta_K k_{it} + \\
+ \sum_{m=0}^{4-n} \sum_{n=0}^{4} \gamma_{mn} \widehat{P}_{it}^{m} \left( \widehat{\varphi}_{i(t-1)} - \beta_A a_{i(t-1)} - \beta_K k_{i(t-1)} \right)^n + \\
+ \xi_{it} + \varepsilon_{it}
\end{aligned}
$$

- They experimented with a **kernel estimator** of the second stage as well, regressing $y_{it} - \widehat{\beta}_L \ell_{it} - \beta_A a_{it} - \beta_K k_{it}$ on $\widehat{P}_{it}$ and on the first lag of the composite term $\widehat{\varphi}_{it} - \beta_A a_{it} - \beta_K k_{it}$ for given $(\beta_A, \beta_K)$ fully non-parametrically, then searching for the pair $(\beta_A, \beta_K)$ that minimizes the squared residuals.

- Their procedure delivers realistic estimates, yet very close to baseline OLS. There is little/no gain from kernel estimators.

# Proxying unobservables with materials (1/4)

- Some drawbacks of the OP approach were noted quite soon.
- First, Assumption 5 is hard to verify, because it depends on a difficult dynamic programming problem.
- Relatedly, it invalidates the approach for those quite frequent observations where investment data is "lumpy" ($i_{it} = 0$).
- Second, Assumption 4 is too stringent: it rules out any other *dynamic* factors affecting investment $i_{it}$ – yet function $f_t(\cdot)$ is constant across firms (Griliches and Mairesse, 1998).
- To circumvent this, LP proposed to base the control function on the (logarithmic) cost of materials: $m_{it} = \log M_{it}$ (often available in the data). Their baseline model is as follows.

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + \beta_M m_{it} + \omega_{it} + \varepsilon_{it}$$

# Proxying unobservables with materials (2/4)

LP replace OP's Assumptions 4 and 5 with the following ones.

**Assumption 4b**

**Scalar unobservable.** The intermediate input demand of firms is given by

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

**Assumption 5b**

**Strict monotonicity.** $f_t\left(k_{it}, \omega_{it}\right)$ is strictly increasing in $\omega_{it}$.

These two assumptions still allow inversion of $f_t\left(\cdot\right)$ for $\omega_{it}$:

$$\omega_{it} = f_t^{-1}\left(k_{it}, m_{it}\right)$$

yet evade the Griliches-Mairesse critique. Since $m_{it}$ is a variable (non-dynamic) input, heterogeneous dynamics is not a concern.

# Proxying unobservables with materials (3/4)

- The LP **first stage** identifies $\beta_L$, like in OP.

$$\mathbb{E}\left[\varepsilon_{it} \middle| \mathcal{I}_t\right] = \mathbb{E}\left[y_{it} - \beta_L \ell_{it} - \Phi_t\left(k_{it}, m_{it}\right) \middle| \mathcal{I}_t\right] = 0$$

- The LP **second stage** identifies both $\beta_K$ and $\beta_M$ instead.

$$\begin{aligned}
&\mathbb{E}\left[\xi_{it} + \varepsilon_{it} \middle| \mathcal{I}_{t-1}\right] = \\
&\quad = \mathbb{E}\Big[y_{it} - \beta_0 - \beta_K k_{it} - \beta_L \ell_{it} - \beta_M m_{it} - \\
&\quad\quad - g\Big(\Phi_{t-1}\left(k_{i(t-1)}, m_{i(t-1)}\right) - \beta_0 - \\
&\quad\quad\quad - \beta_K k_{i(t-1)} - \beta_M m_{i(t-1)}\Big)\Big| \mathcal{I}_{t-1}\Big] = 0
\end{aligned}$$

- Apart from this, estimation is implemented pretty much like in OP, as polynomial series approximate the non-parametric components of the moment conditions.

# Proxying unobservables with materials (4/4)

- LP originally applied their extension of the OP method on a Chilean manufacturing census panel dataset for 1979-1986 (which was quite popular) focusing on four large industries.
- They further add two more inputs $\log X_{kit}$ to their estimated model: *fuel* and *electricity*, observed in their Chilean dataset. Yet they mainly use log-materials $m_{it}$ in the control function.
- While OP calculate their standard errors analytically, using results from a separate paper (Pakes and Olley, 1995), LP circumvent this "difficult task" (*ibidem*) by bootstrapping.
- They provide nice *specification tests* about the choice of the proxy and the monotonicity assumption.
- They show that their empirical estimates differ from baseline OLS in a more marked way than OP's estimates do.

# The functional dependence problem (1/3)

- The key contribution by ACF was to show that both OP and LP suffer from a so-called "functional dependence problem" that invalidates their first stages: $\beta_L$ is not really identified.

- This clearly implies that also their second stage is flawed.

- The problem is best illustrated in the LP setting. Consider the profit maximization First Order Condition for $M_{it}$:

$$\beta_M K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M - 1} \exp\left(\beta_0 + \omega_{it}\right) = \frac{P_M}{P_i}$$

  where $P_M$ is the price of $M_{it}$. This implicitly gives $f_t\left(\cdot\right)$.

- *Inverting* for $\omega_{it}$ and substituting back into the production function yields a "first stage" that does not depend on $\beta_L$.

$$y_{it} = \log\left(\frac{1}{\beta_M}\right) + \log\left(\frac{P_M}{P_i}\right) + m_{it} + \varepsilon_{it}$$

# The functional dependence problem (2/3)

- This result comes from a fully parametric treatment of $f_t(\cdot)$, but it can be generalized. Suppose the labor input follows:

$$\ell_{it} = h_t(k_{it}, \omega_{it})$$

similarly to $m_{it}$. Then, the "inversion" step gives:

$$\ell_{it} = h_t\left(k_{it}, f_t^{-1}(k_{it}, m_{it})\right)$$

hence, $\ell_{it}$ cannot be *non-parametrically identified* separately from $m_{it}$ (as $\ell_{it}$ is a function of $m_{it}$).

- Formally, this implies that the following random matrix:

$$\boldsymbol{H}_L = \mathbb{E}\left[[\ell_{it} - \mathbb{E}(\ell_{it}|\, k_{it}, m_{it})](\ell_{it} - \mathbb{E}[\ell_{it}|\, k_{it}, m_{it}])^{\mathrm{T}}\right]$$

is *not* positive definite, implying non-identification of $\beta_L$ in the "partially linear" LP first stage (Robinson, 1988).

# The functional dependence problem (3/3)

A similar discussion also applies to the OP model. Adding prices to $f_t(\cdot)$ and $h_t(\cdot)$ would not break functional dependence (prices work best as IVs) neither in OP nor in LP.

How to break it, then? ACF discussed three theoretical options.

1. There is some exogenous **"optimization error"** in $\ell_{it}$ (e.g. workers fall sick) but similar optimization error in $m_{it}$ would re-introduce the problem, and violate Assumption 4.

2. The information set $\mathcal{I}_t$ that informs input choices is different for $\ell_{it}$ and $m_{it}$: this occurs for example if $m_{it}$ is chosen *before* $\ell_{it}$ and **new information** becomes available in between (but not the reverse).

3. Only in OP, $\ell_{it}$ is **non-dynamic** and chosen before $i_{it}$.

These are all unlikely scenarios. Ultimately, one needs a **shifter** of the control function **external** to the production function.

# The modern control function approach (1/5)

- ACF suggest a more conservative approach that accounts for the functional dependence problem.

- Their analysis is restricted to a "value added" specification:

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + \omega_{it} + \varepsilon_{it}$$

  where $y_{it} = \log V_{it}$ is now the logarithm of **value added** $V_{it}$. No attempt is made at identifying a coefficient for $m_{it}$.

- Materials still enter the grand production function for gross output $Y_{it}$, but in a way that breaks functional dependence.

- This occurs for example in a **Leontiev** specification in value added and materials (this can be generalized).

$$Y_{it} = \min \left\{ K_{it}^{\beta_K} L_{it}^{\beta_L} \exp\left(\beta_0 + \omega_{it}\right), \beta_M M_{it} \right\}$$

- ACF provide updated versions of the OP-LP assumptions.

# The modern control function approach (2/5)

### Assumption 3c

**Timing of input choices.** Firms accumulate capital according to

$$k_{it} = \kappa\left(k_{i(t-1)}, i_{i(t-1)}\right)$$

where investment $i_{i(t-1)}$ is chosen in period $t-1$. The labor input $\ell_{it}$ has potentially dynamic implication and it is chosen at period $t$, $t-1$ or $t-b$, with $0 < b < 1$.

### Assumption 4c

**Scalar unobservable.** The intermediate input demand of firms is given by

$$m_{it} = \widetilde{f}_t\left(k_{it}, \ell_{it}, \omega_{it}\right).$$

### Assumption 5c

**Strict monotonicity.** $\widetilde{f}_t\left(k_{it}, \ell_{it}, \omega_{it}\right)$ is strictly increasing in $\omega_{it}$.

# The modern control function approach (3/5)

- Their revised Assumption 3 allows labor to be dynamic.

- More crucially, their revised Assumptions 4 and 5 formulate "conditional" input demand functions that fully account for functional dependence even between non-dynamic inputs.

- The **first stage** proceeds similarly as in OP and LP:

$$\omega_{it} = \widetilde{f}_t^{-1}\left(k_{it}, \ell_{it}, m_{it}\right).$$

Let $\widetilde{\Phi_t}\left(k_{it}, \ell_{it}, m_{it}\right) = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + \widetilde{f}_t^{-1}\left(k_{it}, \ell_{it}, m_{it}\right)$ so as to construct a proper **moment condition**.

$$\mathbb{E}\left[\varepsilon_{it} \middle| \mathcal{I}_t\right] = \mathbb{E}\left[y_{it} - \widetilde{\Phi_t}\left(k_{it}, \ell_{it}, m_{it}\right) \middle| \mathcal{I}_t\right] = 0$$

- The first stage is similar to LP's, but it does not feature the term $\beta_L \ell_{it}$ which is embedded in the control function.

- Hence, this yields a first stage **estimate** $\widehat{\widetilde{\varphi}}_t = \widehat{\widetilde{\Phi_t}}\left(k_{it}, \ell_{it}, i_{it}\right)$.

# The modern control function approach (4/5)

- It is the ACF **second stage** that identifies both $\beta_K$ and $\beta_L$. The relative **moment condition** is as follows.

$$\mathbb{E}\left[\xi_{it}+\varepsilon_{it}\middle|\mathcal{I}_{t-1}\right]=\mathbb{E}\Big[y_{it}-\beta_0-\beta_K k_{it}-\beta_L \ell_{it}- \\ -g\Big(\Phi_{t-1}\left(k_{i(t-1)},\ell_{i(t-1)},m_{i(t-1)}\right)- \\ -\beta_0-\beta_K k_{i(t-1)}-\beta_L \ell_{i(t-1)}\Big)\Big|\mathcal{I}_{t-1}\Big]=0$$

- This is estimated by replacing $\Phi_{t-1}\left(k_{i(t-1)},\ell_{i(t-1)},m_{i(t-1)}\right)$ with $\widehat{\widehat{\varphi}}_{t-1}$, as in OP and LP.

- Relative to OP and LP, the second stage needs at least one additional instrument in $\mathcal{I}_{t-1}$ in order to identify $\beta_L$ (which is not identified in the first stage).

- Both $\ell_{it}$ and $\ell_{i(t-1)}$ are good candidates: the choice depends on the **timing assumptions** about labor demand.

# The modern control function approach (5/5)

- ACF provide some Monte Carlo experiments that show how *under their favorite Leontiev functional form* their procedure delivers consistent estimates, unlike LP's.

- Symmetrically (and unsurprisingly) LP's works better in the ACF experiments under assumptions favorable to it.

- The method by ACF is currently the **standard approach** in production function estimation. Occasionally the method by LP (and to a lesser extent that by OP) is still used.

- Their method, like OP's and LP's, can be extended to more general specifications, like the translog production function.

- In their paper, ACF also make a very important point: their method is comparable to direct panel data approaches. This connection is best understood through Wooldridge (2009).

# A unified panel data approach (1/5)

Wooldridge (2009) provides a unified framework for OP, LP and ACF. He considers the following more general model:

$$y_{it} = \alpha + \mathbf{w}_{it}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{x}_{it}^{\mathrm{T}} \boldsymbol{\gamma} + \omega_{it} + \varepsilon_{it}$$

with:

$$\omega_{it} = f^{-1} \left( \mathbf{x}_{it}, \mathbf{m}_{it} \right)$$

and where:

- $\mathbf{w}_{it}$ are the **variable inputs** (e.g. $\ell_{it}$);
- $\mathbf{x}_{it}$ are the **state variables** (e.g. $k_{it}$);
- $\mathbf{m}_{it}$ are the **proxy variables** (e.g. $i_{it}$ or $m_{it}$).

Wooldridge allows $f^{-1}(\cdot)$ to be time-varying and acknowledges the functional dependence problem; this does not fundamentally affect his analysis.

# A unified panel data approach (2/5)

Wooldridge poses the following **sets of moment conditions**:

$$\mathbb{E}\left[\varepsilon_{it}\middle|\left\{\mathbf{w}_{i(t-s)},\mathbf{x}_{i(t-s)},\mathbf{m}_{i(t-s)}\right\}_{s=0}^{t-1}\right]=0$$

for $t=1,\dots,T$; and:

$$\mathbb{E}\left[\varepsilon_{it}+\xi_{it}\middle|\mathbf{x}_{it},\left\{\mathbf{w}_{i(t-s)},\mathbf{x}_{i(t-s)},\mathbf{m}_{i(t-s)}\right\}_{s=1}^{t-1}\right]=0$$

for $t=2,\dots,T$ and given $\xi_{it}\equiv\omega_{it}-\mathbb{E}\left[\omega_{it}\middle|\,\omega_{i(t-1)}\right]$.

They evidently correspond to the "first stage" and "second stage" moment conditions by OP and LP, respectively.

Wooldridge claims that these moment conditions **jointly** identify both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, even in ACF: "$\mathbf{x}_{it}$, $\mathbf{x}_{i(t-1)}$ and $\mathbf{m}_{i(t-1)}$ act as their own instruments, and $\mathbf{w}_{i(t-1)}$ acts as an instrument for $\mathbf{w}_{it}$."

# A unified panel data approach (3/5)

Wooldridge illustrates this with **polynomial approximations**. He writes $\boldsymbol{c}(\mathbf{x}_{it}, \mathbf{m}_{it})$ as a $Q$-long vector of polynomial functions of its arguments (which contains $\mathbf{x}_{it}$ and $\mathbf{m}_{it}$ "separately"), and:

$$f^{-1}(\mathbf{x}_{it}, \mathbf{m}_{it}) = \lambda_0 + [\boldsymbol{c}(\mathbf{x}_{it}, \mathbf{m}_{it})]^{\mathrm{T}} \boldsymbol{\lambda}$$

where $\mathbf{c}_{it}$ can be used as shorthand for $\boldsymbol{c}(\mathbf{x}_{it}, \mathbf{m}_{it})$. Furthermore, Wooldridge posits the following.

$$\mathbb{E}\left[\omega_{it} \middle| \omega_{i(t-1)}\right] = \rho_0 + \rho_1 \omega_{i(t-1)} + \cdots + \rho_G \omega_{i(t-1)}^G$$

Substituting, the model can be written, for $\alpha_0 \equiv \alpha + \lambda_0$, as:

$$y_{it} = \alpha_0 + \mathbf{w}_{it}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{x}_{it}^{\mathrm{T}} \boldsymbol{\gamma} + \mathbf{c}_{it}^{\mathrm{T}} \boldsymbol{\lambda} + \varepsilon_{it}$$

and, for $\eta_0 \equiv \alpha + \rho_0$ and $\upsilon_{it} = \varepsilon_{it} + \xi_{it}$, as follows.

$$y_{it} = \eta_0 + \mathbf{w}_{it}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{x}_{it}^{\mathrm{T}} \boldsymbol{\gamma} + \rho_1 \left(\mathbf{c}_{i(t-1)}^{\mathrm{T}} \boldsymbol{\lambda}\right) + \cdots + \rho_G \left(\mathbf{c}_{i(t-1)}^{\mathrm{T}} \boldsymbol{\lambda}\right)^G + \upsilon_{it}$$

# A unified panel data approach (4/5)

Wooldridge argues that it is easy to verify that all the parameters $\boldsymbol{\theta} = (\alpha_0, \eta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \rho_1, \ldots, \rho_G)$ are **identified**. Write:

$$\mathbf{z}_{it} \equiv \begin{pmatrix} 1 & \mathbf{x}_{it}^{\mathrm{T}} & \mathbf{w}_{i(t-1)}^{\mathrm{T}} & \mathbf{c}_{i(t-1)}^{\mathrm{T}} & \mathbf{q}_{i(t-1)}^{\mathrm{T}} \end{pmatrix}$$

where $\mathbf{q}_{i(t-1)}$ is a set of at least $G$ non-linear functions of $\mathbf{c}_{i(t-1)}$. Then, the **instruments matrix** for this system of equations is:

$$\mathbf{Z}_{it} = \begin{pmatrix} \mathbf{w}_{it}^{\mathrm{T}} & \mathbf{c}_{it}^{\mathrm{T}} & \mathbf{z}_{it}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & \mathbf{z}_{it}^{\mathrm{T}} \end{pmatrix}$$

for $t = 2, \ldots, T$. The system can be expressed as follows.

$$\begin{aligned} \mathbf{r}_{it}(\boldsymbol{\theta}) &= \begin{pmatrix} r_{1it}(\boldsymbol{\theta}) \\ r_{2it}(\boldsymbol{\theta}) \end{pmatrix} \\ &= \begin{pmatrix} y_{it} - \alpha_0 - \mathbf{w}_{it}^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{c}_{it}^{\mathrm{T}}\boldsymbol{\lambda} \\ y_{it} - \eta_0 - \mathbf{w}_{it}^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\gamma} - \sum_{g=1}^{G} \rho_g \left(\mathbf{c}_{i(t-1)}^{\mathrm{T}}\boldsymbol{\lambda}\right)^g \end{pmatrix} \end{aligned}$$

# A unified panel data approach (5/5)

Hence, the moment conditions can be expressed succinctly as:

$$\mathbb{E}\left[\mathbf{Z}_{it}^{\mathrm{T}}\mathbf{r}_{it}\left(\theta\right)\right]=\mathbf{0}$$

for $t=2,\ldots,T$. As Wooldridge suggests, this enables easy **joint estimation** via standard GMM.

Wooldridge further suggests that one particular case is especially illustrative: when $\omega_{it}$ follows a random walk with drift – that is, $G=1$ and $\omega_{it}=\rho_0+\omega_{i(t-1)}+\xi_{it}$. Thus the system writes as:

$$\mathbf{r}_{it}\left(\theta\right)=\begin{pmatrix} y_{it} \\ y_{it} \end{pmatrix}-\begin{pmatrix} 1 & 0 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{it} \\ 0 & 1 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{i(t-1)} \end{pmatrix}\begin{pmatrix} \alpha_0 \\ \eta_0 \\ \beta \\ \gamma \\ \lambda \end{pmatrix}$$

and estimation is straightforward; also, including $\mathbf{q}_{i(t-1)}$ into $\mathbf{Z}_{it}$ is unnecessary but it provides overidentifying restrictions.

# Control functions and panel data: a summary

- The approach by Wooldridge dispenses details on structural assumptions and provides a more transparent econometrics.

- Yet one needs to make sense of the differences between ACF and LP in light of it. The ACF approach corresponds to:

$$f^{-1}\left(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}\right) = \lambda_0 + \left[\boldsymbol{c}\left(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}\right)\right]^{\mathrm{T}} \boldsymbol{\lambda}$$

  which is more general than the one outlined by Wooldridge, based on LP (and that can be seen as a *restriction* of ACF, if functional dependence is not a problem).

- Ultimately however, the identifying moments are **similar**.

- There are also similarities with the Blundell-Bond approach, where $\omega_{it} = \rho\omega_{i(t-1)} + \xi_{it}$ but $f_t\left(\cdot\right)$ is **unrestricted**. There, identification is also based on a similar set of *lagged* inputs.

- Joint estimation is ideal, but it is still often impractical.

# Incorporating demand into the model (1/6)

- Control function approaches enabled substantial progress in the estimation of production functions. However they ignore the **demand side** altogether, which can be problematic.

- This is illustrated in the contribution by De Loecker (2011), which studies the impact of trade liberalization (the removal of tariffs and similar barriers) on productivity.

- Traditional approaches to this question typically pose that:

$$\omega_{it} = \lambda_0 + \lambda_1 qr_{it} + \zeta_{it}$$

  where $qr_{it} \in [0, 1]$ is a variable that measures the extent to which firm $i$'s products are "protected" by quotas that apply to foreign countries: at the extremes, $qr_{it} = 0$ if no product is protected and $qr_{it} = 1$ if all products are protected.

- Clearly, $\zeta_{it}$ here is a residual error of the productivity shock.

# Incorporating demand into the model (2/6)

- It is hypothesized that $\lambda_1 \leq 0$ because of **competition**, but to what extent is this empirically true?

- One could estimate $\lambda_1$ by specifying $qr_{it}$ into the production function, or by regressing the estimated residual $\widehat{\omega}_{it}$ on it.

- Both approaches fail even if OP/LP/ACF are used, because of the confounding effect of **demand** changes. In fact, trade liberalization is likely to affect sale prices!

- Recall that researchers estimate production functions using deflated sales $\log R_{it} - \log P_{s(i)t}$ as their dependent variable, unless actual *physical* output $Y_{it}$ is observed (which is rare).

- Naturally, $qr_{it}$ correlates with $\varpi_{it} = \log P_{it} - \log P_{s(i)t}$: that is, unobserved error in firm $i$'s own price.

- Thus, naïve estimation of $\lambda_1$ likely leads to **overstate** it.

# Incorporating demand into the model (3/6)

- The key contribution by De Loecker was to incorporate the following **demand function** in the estimation.

$$Y_{it} = Y_{s(i)t} \left( \frac{P_{it}}{P_{s(i)t}} \right)^{\sigma_{s(i)}} \exp\left(\eta_{it}\right)$$

- Above, $Y_{s(i)t}$ is a **demand shifter**, $\eta_{it}$ is a **demand shock** (unobserved), and $\sigma_{s(i)}$ is the **demand elasticity**.

- This demand function follows directly from CES preferences, a classical ingredient of many economic models.

- In logarithms (represented by lower-case variables), it reads:

$$y_{it} = y_{s(i)t} + \sigma_{s(i)} \left( p_{it} - p_{s(i)t} \right) + \eta_{it}$$

which can also obtain from a random utility model of choice as in Berry (1994), with a different interpretation for $\sigma_{s(i)}$.

# Incorporating demand into the model (4/6)

Substituting the demand function into the LP model with $\beta_0 = 0$:

$$\widetilde{r}_{it} = \gamma_K k_{it} + \gamma_L \ell_{it} + \gamma_M m_{it} + \gamma_{s(i)} y_{s(i)t} + \omega^*_{it} + \eta^*_{it} + \varepsilon^*_{it}$$

where:

- $\widetilde{r}_{it} = r_{it} - p_{s(i)t}$ is the actually used **deflated revenue**;
- $\gamma_H = \left(\frac{\sigma_{s(i)} + 1}{\sigma_{s(i)}}\right) \beta_H$ for $H = K, L, M$;
- $\gamma_{s(i)} = \frac{1}{\left|\sigma_{s(i)}\right|}$;
- $\omega^*_{it} = \left(\frac{\sigma_{s(i)} + 1}{\sigma_{s(i)}}\right) \omega_{it}$ and $\varepsilon^*_{it} = \left(\frac{\sigma_{s(i)} + 1}{\sigma_{s(i)}}\right) \varepsilon_{it}$;
- $\eta^*_{it} = \frac{\eta_{it}}{\left|\sigma_{s(i)}\right|}$.

# Incorporating demand into the model (5/6)

- De Loecker also specifies, under some assumptions, a version of this equation for multi-product firms.

- Estimating this equation consistently would allow to recover *both* production functions parameters and $\sigma_{s(i)}$: the demand elasticity. This entails tackling the **endogenous** $\omega^*_{it}$ and $\eta^*_{it}$.

- De Loecker specifies $\eta^*_{it}$ as:

$$\eta^*_{it} = \mathbf{d}_{it}^{\mathrm{T}} \boldsymbol{\delta} + \tau q r_{it} + \widetilde{\eta}_{it}$$

  where $\mathbf{d}_{it}$ is a vector of **product dummies** (to account for firm $i$'s products), $\boldsymbol{\tau}$ is a parameter that introduces a demand channel for quotas, and $\widetilde{\eta}_{it}$ is a residual *orthogonal* shock.

- Instead, De Loecker specifies $\omega^*_{it}$ as in LP, but with a twist: the law of motion of productivity is affected by trade quotas.

$$\omega_{it} = g_t \left( \omega_{i(t-1)}, q r_{it} \right) + \xi_{it}$$

# Incorporating demand into the model (6/6)

De Loecker's **final model** is thus as follows, for $\varepsilon_{it}^{**} = \varepsilon_{it}^{*} + \widetilde{\eta}_{it}$.

$$\widetilde{r}_{it} = \gamma_K k_{it} + \gamma_L \ell_{it} + \gamma_M m_{it} + \gamma_{s(i)} y_{s(i)t} + \mathbf{d}_{it}^{\mathrm{T}} \boldsymbol{\delta} + \tau q r_{it} + \omega_{it}^{*} + \varepsilon_{it}^{**}$$

In performing estimation, De Loecker attempts all of OP, LP and ACF to tackle $\omega_{it}^{*}$ (although $\gamma_M$ is dubiously *always* estimated). De Loecker then estimates **productivity-per-input** $\omega_{it}$ as:

$$\widehat{\omega}_{it} = \left(\widetilde{r}_{it} - \widehat{\gamma}_K k_{it} - \widehat{\gamma}_L \ell_{it} - \widehat{\gamma}_M m_{it} - \widehat{\gamma}_{s(i)} y_{s(i)t} - \widehat{\tau} q r_{it}\right) \left(\frac{\widehat{\sigma}_{s(i)}}{\widehat{\sigma}_{s(i)} + 1}\right)$$

and regresses this measure on $qr_{it}$ in order to estimate $\lambda_1$.

In summary, his **results** are as follows:

- $(\beta_K, \beta_L, \beta_M)$ are estimated similarly as in OLS, and OP/LP;
- the resulting estimate of $\lambda_1$ is hardly significant (both in the statistical and economic sense).

# A modern non-parametric treatment (1/9)

- A recent contribution by Gandhi, Navarro and Rivers (2020, GNR) revisits the econometrics of production functions from a fully non-parametric perspective.

- Their starting observation is that the literature culminating with ACF provides what is essentially a negative result about the identification of *gross output* production functions, with more positive prospects reserved to models for *value added*.

- Yet interest typically falls on gross output, not value added. The starting point of GNR is a model for gross output $y_{it}$:

$$y_{it} = \log F\left(k_{it}, \ell_{it}, m_{it}\right) + \omega_{it} + \varepsilon_{it}$$

  where $F(\cdot)$ is flexibly treated non-parametrically.

- GNR develop a method for the non-parametric identification of $F(\cdot)$ using information about input prices.

# A modern non-parametric treatment (2/9)

- GNR first revisit the functional dependence problem. Their Theorem 1 proves that under Assumptions 1-3 by ACF and if firms take all prices as given, function $F(\cdot)$ is not identified separately from $g(\cdot)$: the law of motion of $\omega_{it}$.

- Their main result (Theorem 2) proves that also allowing for Assumptions 4-5 by ACF, the **elasticity** of $F(\cdot)$ to a given input is identified off variation in **input prices**. This result exploits the First Order Conditions, and echoes traditional literature (most notably Griliches and Ringstad, 1971).

- Write the First Order Condition for $m_{it}$ as:

$$P_t^M = P_{it} \frac{\partial F(k_{it}, \ell_{it}, m_{it})}{\partial \exp(m_{it})} \exp(\omega_{it}) \mathcal{E}$$

where $P_t^M$ is the price of materials whereas $\mathcal{E} \equiv \mathbb{E}[\exp(\varepsilon_{it})]$. Unlike in ACF, firms "expect" $\varepsilon_{it}$, but with uncertainty.

# A modern non-parametric treatment (3/9)

By taking the logarithm of the First Order Conditions:

$$\log P_t^M = \log P_{it} - \log M_{it} + \log\left[\frac{\partial F\left(k_{it}, \ell_{it}, m_{it}\right)}{\partial m_{it}}\right] + \omega_{it} + \log \mathcal{E}$$

and substituting $\omega_{it} = \log Y_{it} - \log F\left(k_{it}, \ell_{it}, m_{it}\right) - \varepsilon_{it}$, one gets:

$$z_{it}^M = \log D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right) - \varepsilon_{it}$$

where:

$$z_{it}^M \equiv \log\left(P_t^M M_{it}\right) - \log\left(P_{it} Y_{it}\right)$$

is the logarithmic *share* of the *cost of materials* on total revenue (typically observed in the data) and $D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right)$ is as follows.

$$\begin{aligned} D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right) &\equiv \mathcal{E}\left[\frac{\partial \log F\left(k_{it}, \ell_{it}, m_{it}\right)}{\partial m_{it}}\right] \\ &= \mathcal{E}\left[\frac{1}{F\left(k_{it}, \ell_{it}, m_{it}\right)} \frac{\partial F\left(k_{it}, \ell_{it}, m_{it}\right)}{\partial m_{it}}\right] \end{aligned}$$

# A modern non-parametric treatment (4/9)

Theorem 2 by GNR proceeds as follows. Starting from equation

$$z_{it}^{M} = \log D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right) - \varepsilon_{it}$$

they observe that function $D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right)$ is non-parametrically identified since the transitory shock is exogenous.

$$\mathbb{E}\left[\varepsilon_{it} \middle| k_{it}, \ell_{it}, m_{it}\right] = 0$$

In addition, the constant term

$$\mathcal{E} = \mathbb{E}\left[\exp\left(\log D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right) - z_{it}^{M}\right)\right]$$

is also obviously identified. Therefore, the elasticity of interest is identified residually.

$$\frac{\partial \log F\left(k_{it}, \ell_{it}, m_{it}\right)}{\partial m_{it}} = \frac{D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right)}{\mathcal{E}}$$

For example, in the Cobb-Douglas case $D^{\mathcal{E}}\left(k_{it}, \ell_{it}, m_{it}\right)/\mathcal{E} = \beta_M$.

# A modern non-parametric treatment (5/9)

The last important result by GNR (Theorem 3) is that the whole production function $F(\cdot)$ is non-parametrically identified. Write:

$$\mathcal{D}(k_{it}, \ell_{it}, m_{it}) \equiv \int_{\mathbb{R}} \frac{\partial \log F(k_{it}, \ell_{it}, m_{it})}{\partial m_{it}} dm_{it}$$
$$= \log F(k_{it}, \ell_{it}, m_{it}) + \mathcal{C}(k_{it}, \ell_{it})$$

by the Fundamental Theorem of Calculus (notice that knowledge of $\mathcal{C}(k_{it}, \ell_{it})$ identifies the production function). Also define:

$$\mathcal{Y}_{it} \equiv y_{it} - \varepsilon_{it} - \mathcal{D}(k_{it}, \ell_{it}, m_{it}) = -\mathcal{C}(k_{it}, \ell_{it}) + \omega_{it}$$

where $\mathcal{Y}_{it}$ is a random variable that can be "recovered" from the data. The final step exploits $\omega_{it} = g\left(\omega_{i(t-1)}\right) + \xi_{it}$:

$$\mathcal{Y}_{it} = -\mathcal{C}(k_{it}, \ell_{it}) + g\left(\mathcal{Y}_{i(t-1)} + \mathcal{C}\left(k_{i(t-1)}, \ell_{i(t-1)}\right)\right) + \xi_{it}$$

hence $\mathcal{C}(k_{it}, \ell_{it})$ is non-parametrically identified if $\mathbb{E}[\xi_{it} | \mathcal{I}_t] = 0$, similarly to both Blundell-Bond and OP/LP/ACF.

# A modern non-parametric treatment (6/9)

Practical implementation of such a non-parametric identification result requires a **two-step** estimation procedure.

The **first step** seeks those coefficients $\boldsymbol{\gamma} = \{\gamma_{r_k, r_\ell, r_m}\}_{r_k + r_\ell + r_m \leq R}$ that solve (via NLLS):

$$\min_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ z_{it} - \log \left( \sum_{r_k=1}^{R-r_\ell-r_m} \sum_{r_\ell=1}^{R-r_m} \sum_{r_m=1}^{R} \gamma_{r_k, r_\ell, r_m} k_{it}^{r_k} \ell_{it}^{r_\ell} m_{it}^{r_m} \right) \right]^2$$

that is, the parameter estimates of an approximating polynomial of degree $R$ for $D^{\mathcal{E}}(k_{it}, \ell_{it}, m_{it})$. This writes as follows.

$$\widehat{D}^{\mathcal{E}}(k_{it}, \ell_{it}, m_{it}) = \sum_{r_k=1}^{R-r_\ell-r_m} \sum_{r_\ell=1}^{R-r_m} \sum_{r_m=1}^{R} \widehat{\gamma}_{r_k, r_\ell, r_m} k_{it}^{r_k} \ell_{it}^{r_\ell} m_{it}^{r_m}$$

Notice that this is enough to identify the elasticities of interests, since $\mathcal{E}$ is easily estimated as the empirical average of $\exp(\widehat{\varepsilon}_{it})$, where $\widehat{\varepsilon}_{it}$ are the residuals of the least squares problem.

# A modern non-parametric treatment (7/9)

Given an estimate $\widehat{\mathcal{E}}$ for $\mathcal{E}$, one can calculate:

$$\widehat{\mathcal{D}}\left(k_{it},\ell_{it},m_{it}\right)=\left(\widehat{\mathcal{E}}\right)^{-1}\sum_{r_k=1}^{R-r_\ell-r_m}\sum_{r_\ell=1}^{R-r_m}\sum_{r_m=1}^{R}\frac{\widehat{\gamma}_{r_k,r_\ell,r_m}}{r_m+1}k_{it}^{r_k}\ell_{it}^{r_\ell}m_{it}^{r_m+1}$$

as well as $\widehat{\mathcal{Y}}_{it}=y_{it}-\widehat{\varepsilon}_{it}-\widehat{\mathcal{D}}\left(k_{it},\ell_{it},m_{it}\right)$.

The **second stage** is based on more polynomial approximations:

$$\mathcal{C}\left(k_{it},\ell_{it}\right)=\sum_{s_k=1}^{S-s_\ell}\sum_{s_\ell=1}^{S}\delta_{s_k,s_\ell}k_{it}^{s_k}\ell_{it}^{s_\ell}$$

and:

$$g\left(\omega_{i(t-1)}\right)=\sum_{a=1}^{A}\alpha_a\omega_{i(t-1)}^{a}$$

for some degrees $S$ and $A$.

# A modern non-parametric treatment (8/9)

This yields an equation estimable via NLLS, very much like the OP/LP/ACF second stage.

$$\widehat{\mathcal{Y}}_{it} = -\sum_{s_k=1}^{S-s_\ell}\sum_{s_\ell=1}^{S} \delta_{s_k,s_\ell} k_{it}^{s_k}\ell_{it}^{s_\ell} + \\ + \sum_{a=1}^{A}\alpha_a\left(\widehat{\mathcal{Y}}_{i(t-1)} + \sum_{s_k=1}^{S-s_\ell}\sum_{s_\ell=1}^{S} \delta_{s_k,s_\ell} k_{i(t-1)}^{s_k}\ell_{i(t-1)}^{s_\ell}\right)^a + \xi_{it}$$

The non-parametric estimate of $F(\cdot)$ is thus recovered as follows.

$$\log \widehat{F}(k_{it},\ell_{it},m_{it}) = \widehat{\mathcal{D}}(k_{it},\ell_{it},m_{it}) - \widehat{\mathcal{C}}(k_{it},\ell_{it}) = \\ = \left(\widehat{\mathcal{E}}\right)^{-1}\sum_{r_k=1}^{R-r_\ell-r_m}\sum_{r_\ell=1}^{R-r_m}\sum_{r_m=1}^{R}\frac{\widehat{\gamma}_{r_k,r_\ell,r_m}}{r_m+1}k_{it}^{r_k}\ell_{it}^{r_\ell}m_{it}^{r_m+1} - \\ - \sum_{s_k=1}^{S-s_\ell}\sum_{s_\ell=1}^{S}\widehat{\delta}_{s_k,s_\ell}k_{it}^{s_k}\ell_{it}^{s_\ell}$$

# A modern non-parametric treatment (9/9)

- GNR rely upon recent non-parametric econometric literature (Hahn, Liao and Ridder, 2018) for the asymptotic properties of their estimator. They bootstrap the standard errors of key functionals, such as the elasticities.

- In Monte Carlo simulations with $R = S = 2$ and $A = 3$, they show that their method retrieves input elasticities quite well for Cobb-Douglas, translog and CES production functions.

- In an application on Colombian and Chilean data, GNR find that their method yields estimates that differ markedly from those by OLS, and they find their own more realistic.

- GNR acknowledge that conceptually, their method is not too different from panel data and control function methods. Yet their contribution is important as it highlights the role of $z_{it}$. Their approach is likely to replace ACF as the standard.