UG Empirical Industrial Organization

Lecture 4: Production Functions: Estimation Methods

Anubhav Jha[1]
Ashoka University

March 11, 2024

[1] This course is based on Victor Aguirregabiria's Empirical IO book. The slides (in my first year of teaching) are extremely close to his slides. These will change in the future iterations of this course.

Outline: Estimation Methods

1. Input prices as IVs
2. Fixed Effects estimator
3. Fixed Effects + Cochrane-Orcutt estimator
4. Arellano-Bond estimator
5. Panel Data: System estimator
6. Control Function: Olley-Pakes estimator
7. Using First Order Conditions

Input prices as IVs

- If input prices r_{it} are observable (wages, cost of capital, fuel and energy prices), then under the assumption that they are not correlated with TFP, $\mathbb{E}(\omega_{it}\, r_{it}) = 0$, we can use them as instruments.
- This approach has several limitations/problems.
- **Problem (1).** Firms in the same industry typically use very similar type of inputs (labor, capital equipment, energy, materials) and they buy these inputs in the same input markets. If these input markets are competitive, the input prices are the same for all the firms in the industry:

$$r_{it} = r_t \quad \text{for every firm } i$$

- If input prices vary only over time, they are perfectly collinear with time-dummies in the PF. No valid instruments.

Input prices as IVs (2)

- **Problem (2).** When input prices have cross-sectional variation, it could be because of endogenous reasons.
  (a) Inputs markets are not competitive and firms with higher productivity pay higher prices. Then,

  $$\text{cov}(\omega_{it}, r_{it}) \neq 0,$$

  making input prices not a valid instrument.

  (b) Firms may be using different types of labor or capital inputs, with different qualities. This difference in the quality of inputs is part of the log-TFP. Then,

  $$\text{cov}(\omega_{it}, r_{it}) \neq 0,$$

  making input prices not a valid instrument.

Input prices as IVs (3)

- An ideal situation for using input prices as IVs is when firms in the same industry produce in different geographic markets where the input markets are competitive.
- The variation in input prices over geographic market is due to different conditions on the supply of inputs (e.g., labor supply, better access to materials) and not to differences in productivity.

Fixed-Effects (FE) estimator

- Consider the PF:

$$y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} \tag{1}$$

- Let's first define the FE (or Within-Groups) estimator and then we will show under which conditions this estimator provides unbiased (consistent) estimates of parameters α_L and α_K.
- If, for each firm i, we average equation (1) considering all the years of observations, we have the equation:

$$\bar{y}_i = \alpha_L \, \bar{\ell}_i + \alpha_K \, \bar{k}_i + \bar{\omega}_i \tag{2}$$

- where:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}; \quad \bar{\ell}_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{it}; \quad \bar{k}_i = \frac{1}{T} \sum_{t=1}^{T} k_{it}; \quad \bar{\omega}_i = \frac{1}{T} \sum_{t=1}^{T} \omega_{it}$$

FE estimator (2)

- If we subtract equation (2) from equation (1), we have:

$$(y_{it} - \bar{y}_i) = \alpha_L(\ell_{it} - \bar{\ell}_i) + \alpha_K(k_{it} - \bar{k}_i) + (\omega_{it} - \bar{\omega}_i) \qquad (3)$$

- This equation is named the Fixed-Effects (or the Within-Groups) transformation of the model.
- The FE estimator is OLS applied to the FE transformed model.
- For instance, if we had only one input, say labor, the FE estimator of α_L would be:

$$\hat{\alpha}_L^{FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(\ell_{it} - \bar{\ell}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(\ell_{it} - \bar{\ell}_i)^2}$$

Consistency of FE estimator

- The FE estimator is OLS in the regression equation:

$$(y_{it} - \bar{y}_i) = \alpha_L(\ell_{it} - \bar{\ell}_i) + \alpha_K(k_{it} - \bar{k}_i) + (\omega_{it} - \bar{\omega}_i) \quad (3)$$

- As any OLS estimator, it is consistent if the error term is not correlated with the regressors. In this case, this implies:

$$\mathbb{E}\left[(\omega_{it} - \bar{\omega}_i)(\ell_{it} - \bar{\ell}_i)\right] = \mathbb{E}\left[(\omega_{it} - \bar{\omega}_i)(k_{it} - \bar{k}_i)\right] = 0$$

- We now present two assumptions on the unobserved log-TFP that imply consistency of the FE estimator (with time dummies).

Consistency of FE estimator (2)

- **Assumption FE-1:** Log-TFP has the following structure:

$$\omega_{it} = \eta_i + \delta_t + u_{it}$$

- η_i is interpreted as managerial ability, or a different technology that is constant over time.
- δ_t represents productivity that affects in the same way all the firms in the industry.
- u_{it} is a firm-specific transitory shock.

Consistency of FE estimator (3)

- **Assumption FE-2:** The firm-specific transitory shock, u_{it}, is not correlated over time and it is realized after the firm chooses the amount of inputs at period t.
- u_{it} is a *surprise* that is realized after the firm has chosen inputs. For any two periods t and s, u_{it} is not correlated with inputs ℓ_{is} and k_{is}.

Consistency of FE estimator (4)

- Under assumptions FE-1 we have that:

$$\bar{\omega}_i = \eta_i + \bar{\delta} + \bar{u}_i$$

such that:

$$\omega_{it} - \bar{\omega}_i = \delta_t - \bar{\delta} + u_{it} - \bar{u}_i$$

- If we control for $\delta_t - \bar{\delta}$ using time dummies, the remaining error term is $u_{it} - \bar{u}_i$.
- Under Assumption FE-2, the error term $u_{it} - \bar{u}_i$ is not correlated with the regressors $\ell_{it} - \bar{\ell}_i$ and $k_{it} - \bar{k}_i$ because, for any two periods t and s, u_{it} is not correlated with inputs ℓ_{is} and k_{is}.
- Under FE-1 and FE-2, the FE estimator is unbiased / consistent.

Cochrane-Orcutt estimator

- The assumption that the firm-specific transitory shock is not serially correlated (and fully unknown to the firm at period t) is quite strong.
- This assumption is testable (Arellano-Bond test for serial correlation). If rejected, this assumption can be relaxed.
- Suppose that we maintain assumption FE-1 but we replace assumption FE-2 with the following.
- **Assumption FE-CO.** The firm-specific transitory shock, u_{it}, follows an Autoregressive-1 process, AR(1):

$$u_{it} = \rho \, u_{i,t-1} + a_{it}$$

  where ρ is a parameter, and a_{it} is not correlated over time and it is realized after the firm chooses the amount of inputs at period t.

Cochrane-Orcutt estimator (2)

- In this model where u_{it} is serially correlated, the standard FE estimator is inconsistent (biased) because $u_{it} - \bar{u}_i$ is correlated with the regressors.
- However, we can define a new version of the FE estimator (Cochrane-Orcutt FE) that is consistent under these conditions and the additional condition that the number of periods T is large.

Cochrane-Orcutt estimator (3)

- Consider the PF at periods t and $t-1$ under assumption FE-1:

$$y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \eta_i + \delta_t + u_{it}$$
$$y_{it-1} = \alpha_L \, \ell_{it-1} + \alpha_K \, k_{it-1} + \eta_i + \delta_{t-1} + u_{it-1}$$

- Multiplying the equation at $t-1$ by ρ and subtracting it from the equation at period t, we get:

$$\begin{aligned} y_{it} - \rho \, y_{it-1} = & \alpha_L [\ell_{it} - \rho \, \ell_{it-1}] + \alpha_K [k_{it} - \rho \, k_{it-1}] \\ & + [1-\rho]\eta_i + [\delta_t - \rho \, \delta_{t-1}] + a_{it} \end{aligned}$$

- because $u_{it} - \rho \, u_{it-1} = a_{it}$.
- This is called a quasi-difference transformation.

Cochrane-Orcutt estimator (4)

- The quasi-difference transformation can be written as:

$$y_{it} = \beta_1 y_{i,t-1} + \beta_2 \ell_{it} + \beta_3 \ell_{it-1} + \beta_4 k_{it} + \beta_5 k_{it-1} + \eta_i^* + \delta_t^* + a_{it}$$

- with

$$\beta_1 = \rho, \quad \beta_2 = \alpha_L, \quad \beta_3 = -\rho\alpha_L, \quad \beta_4 = \alpha_K, \quad \beta_5 = -\rho\alpha_K,$$

and

$$\eta_i^* = (1-\rho)\eta_i, \quad \delta_t^* = \delta_t - \rho\delta_{t-1}$$

- Note that given the β parameters we can obtain the parameters ρ, α_L, and α_K. In fact, there are additional (over-identifying) restrictions:

$$\rho = \beta_1 = -\frac{\beta_3}{\beta_2} = -\frac{\beta_5}{\beta_4}$$

$$\alpha_L = \beta_2 = -\frac{\beta_3}{\beta_1}, \quad \alpha_K = \beta_4 = -\frac{\beta_5}{\beta_1}$$

- Now, under assumption FE-CO, in equation (2), the transitory shock a_{it} is not correlated with the inputs.

Cochrane-Orcutt estimator with Fixed Effects (5)

- Consider equation (4) in firm-specific means:

$$\bar{y}_i = \beta_1 \, \bar{y}_{i(-1)} + \beta_2 \, \bar{\ell}_i + \beta_3 \, \bar{\ell}_{i(-1)} + \beta_4 \, \bar{k}_i + \beta_5 \, \bar{k}_{i(-1)} + \eta_i^* + \bar{\delta}^* + \bar{a}_i$$

- And in deviations with respect to firm-specific means:

$$\begin{aligned} y_{it} - \bar{y}_i = \; & \beta_1(y_{it-1} - \bar{y}_{i(-1)}) + \beta_2(\ell_{it} - \bar{\ell}_i) + \beta_3(\ell_{it-1} - \bar{\ell}_{i(-1)}) \\ & + \beta_4(k_{it} - \bar{k}_i) + \beta_5(k_{it-1} - \bar{k}_{i(-1)}) \\ & + (\delta_t^* - \bar{\delta}^*) + (a_{it} - \bar{a}_i) \end{aligned}$$

- The FE–Cochrane-Orcutt estimator is applying OLS to this equation.

Cochrane-Orcutt estimator: Large T condition

- **IMPORTANT NOTE:** The FE–Cochrane–Orcutt is consistent (asymptotically unbiased) only when T is large, e.g., larger than 30 or 40 periods.
- Note that under condition FE–CO, we have that $(a_{it} - \bar{a}_i)$ is not correlated with regressors

$$\ell_{it} - \bar{\ell}_i, \quad \ell_{it-1} - \bar{\ell}_{i(-1)}, \quad k_{it} - \bar{k}_i, \quad k_{it-1} - \bar{k}_{i(-1)}.$$

- However, even under this condition, we have that $(a_{it} - \bar{a}_i)$ is correlated with regressor $y_{it-1} - \bar{y}_{i(-1)}$.
- Note that y_{it-1} depends on $a_{i,t-1}$ and that $a_{i,t-1}$ is part of $\bar{a}_i$.
- This correlation goes to zero as T becomes large.

Implementation of FE and FE-CO (1)

- Run the following

$$y_{it} = \alpha_L \ell_{it} + \alpha_K k_{it} + \eta_i + \delta_t + u_{it} \quad \text{FE}$$

$$y_{it} = \beta_1 y_{i,t-1} + \beta_2 \ell_{it} + \beta_3 \ell_{it-1} + \beta_4 k_{it} + \beta_5 k_{it-1} + \eta_i^* + \delta_t^* + a_{it} \quad \text{FE-CO}$$

```
library(fixest)
## FIRM ID and Year is generally given
fe_ols<-feols(y~ k + l| firm_id +year,
                   data=df)
## Lagged output and lagged inputs need to be
   created
df<-df %>%
group_by(firm_id)  %>%
arrange(year, .by_group = TRUE) %>%
mutate(lag_y =  dplyr::lag(y,1),
        lag_k = dplyr::lag(k,1),
        lag_l =  dplyr::lag(l,1)) %>%
ungroup()
```

Implementation of FE and FE-CO (2)

- Run the following

$$y_{it} = \alpha_L \ell_{it} + \alpha_K k_{it} + \eta_i + \delta_t + u_{it} \quad \text{FE}$$
$$y_{it} = \beta_1 y_{i,t-1} + \beta_2 \ell_{it} + \beta_3 \ell_{it-1} + \beta_4 k_{it} + \beta_5 k_{it-1} + \eta_i^* + \delta_t^* + a_{it} \quad \text{FE-CO}$$

```
library(fixest)
## FIRM ID and Year is generally given
feco_ols<-feols(y~lag_y + l + lag_l +
                  k + lag_k | firm_id + year,
                  data=df)
```

FE-CO tables

- FE-CO1 and FE-CO2 are Cochrane-Orcutt estimator. Assume that the FE-CO assumptions hold, can you fill in the "?" in the table?

Dependent Variable:		y	
Model:	(FE)	(FE-CO1)	(FE-CO2)
Variables			
l	0.6244***	?	0.0208
	(0.1331)	(0.11)	(0.1162)
k	0.1712***	?	0.1038*
	(0.0587)	(0.082)	(0.0573)
lag_y		0.9118***	0.8361***
		(0.0199)	(0.031)
lag_l		-0.0323	?
		(0.1231)	(0.15)
lag_k		-0.0949*	?
		(0.0525)	(0.08)
Fixed-effects			
IndustryCode	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	1,134	1,107	871
R^2	0.97323	0.99588	0.7528
Within R^2	0.44341	0.91291	0.7013

Arellano-Bond estimator

- Assumption FE-2 (or for that matter FE-CO) has two parts:
  - FE-2(a): u_{it} is not serially correlated.
  - FE-2(b): u_{it} is not known to the firm when it decides the amounts of inputs.
- In most applications, the stronger of the two assumptions is FE-2(b).
- We now present a panel data estimator that relaxes assumption FE-2(b).

Arellano-Bond estimator (2)

- We maintain assumptions FE-1, $\omega_{it} = \eta_i + \delta_t + u_{it}$, and FE-2(a), u_{it} is not serially correlated.
- Define the variables in first differences:

$$\Delta y_{it} = y_{it} - y_{it-1}; \quad \Delta \ell_{it} = \ell_{it} - \ell_{it-1}; \quad \text{etc.}$$

- And consider the PF in first differences (equation at period t minus equation at period $t-1$):

$$\Delta y_{it} = \alpha_L \, \Delta \ell_{it} + \alpha_K \, \Delta k_{it} + \Delta \delta_t + \Delta u_{it}$$

- We have removed the term η_i from the error term, and we can control for the term $\Delta \delta_t$ by including time-dummies.
- But we still have the term Δu_{it} that is correlated with the regressors $\Delta \ell_{it}$ and Δk_{it}.

Arellano-Bond estimator (3)

$$\Delta y_{it} = \alpha_L \, \Delta \ell_{it} + \alpha_K \, \Delta k_{it} + \Delta \delta_t + \Delta u_{it}$$

- Consider the following general models for demand of capital and labor inputs:

$$\begin{aligned} \text{(LD)} \quad \ell_{it} &= \beta_1^{LD} \ell_{i,t-1} + \beta_2^{LD} k_{i,t-1} + \beta_3^{LD} \omega_{it} + \beta_4^{LD} r_{it} \\ \text{(KD)} \quad k_{it} &= \beta_1^{KD} \ell_{i,t-1} + \beta_2^{KD} k_{i,t-1} + \beta_3^{KD} \omega_{it} + \beta_4^{KD} r_{it} \end{aligned}$$

- This means that ℓ_{it} and k_{it} depend on the current and the past histories of the transitory shocks:

$$u_{it}, u_{i,t-1}, u_{i,t-2}, \ldots$$

- But not on future shocks: $u_{it+1}, u_{i,t+1}, \ldots$

Arellano-Bond estimator (4)

$$\Delta y_{it} = \alpha_L \, \Delta \ell_{it} + \alpha_K \, \Delta k_{it} + \Delta \delta_t + \Delta u_{it}$$

- This implies that $\ell_{i,t-2}$ and $k_{i,t-2}$ are valid instruments in this equation.
- They are **Relevant**: $\Delta \ell_{it}$ and Δk_{it} are correlated with $\ell_{i,t-2}$ and $k_{i,t-2}$.
- They are not correlated with the error term:

$$\mathbb{E}[\ell_{i,t-2} \, \Delta u_{it}] = \mathbb{E}[\ell_{i,t-2} \, u_{it}] - \mathbb{E}[\ell_{i,t-2} \, u_{i,t-1}] = 0$$

Arellano-Bond estimator (5)

- This idea implies many moment restrictions that can be used to estimate α_L, α_K, and $\Delta\delta_t$:

$$\mathbb{E}(\ell_{i,t-j}\,\Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T; \quad \text{and } j := 2$$

$$\mathbb{E}(k_{i,t-j}\,\Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T; \quad \text{and } j := 2$$

$$\mathbb{E}(y_{i,t-j}\,\Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T; \quad \text{and } j := 2$$

- The Arellano-Bond estimator exploits all these restrictions optimally: optimal weighting; optimal Generalized Method of Moments (GMM) estimator.

System GMM

- When labor and capital inputs are strongly correlated, the Arellano–Bond estimator suffers from a weak instruments problem: low correlation between instruments and endogenous variables, and imprecise estimates.
- Note that if ℓ_{it} and k_{it} follow *random walks*, then $\Delta\ell_{it}$ and Δk_{it} are not serially correlated and therefore they are not correlated with the instruments $\ell_{i,t-2}$ and $k_{i,t-2}$.
- For these cases, Blundell–Bond derive additional restrictions that help to identify the PF.

Arellano–Bond Setup

We want to estimate a Cobb–Douglas production function:

$$\log(\text{sales}_{it}) = \beta_L \log(\text{labor}_{it}) + \beta_K \log(\text{capital}_{it}) + u_{it}$$

using Arellano–Bond GMM (difference GMM).

Data: balanced panel `blundell_bond`

- `id`: firm identifier
- `year`: year
- `sales`: firm revenue
- `labor`, `capital`

Data Preparation in R

```
df2 <- blundell_bond %>%
  group_by(id) %>%
  arrange(year, .by_group = TRUE) %>%
  mutate(y = log(sales),
         l = log(labor),
         k = log(capital)) %>%
  ungroup()
```

Data Preparation in R

```
# DO INSTALL plm BEFORE LOADING IT BLINDLY.
library(plm)

pdf <- pdata.frame(df2, index = c("id","year"))

pgmm(
  y ~ k + l  |
    lag(y, 2:2) + lag(k, 2:2) + lag(l, 2:2),
  data = pdf,
  effect = "individual",
  model = "onestep",
  transformation = "d"
)
```

Interpreting Output

Typical pgmm output:

- **"Oneway (individual) effect One-step model Difference GMM"** $\Rightarrow$ First-difference GMM estimator, one-step weighting, removing firm FE.
- **Sargan test** Tests validity of over-identifying restrictions (instruments). High p-value $\Rightarrow$ instruments valid. Low p-value $\Rightarrow$ overfit/misspecified.
- **Autocorrelation test (1)** Tests AR(1) in differenced residuals (should exist).
- **Autocorrelation test (2)** Tests AR(2) in differenced residuals (should NOT exist). High p-value desired.
- **Wald test for coefficients** Tests joint significance of regressors.

Output

Table: Arellano–Bond Difference GMM Estimates

	Estimate	Std. Error	z-value
log(capital) (k)	0.426	0.029	14.82***
log(labor) (l)	0.530	0.093	5.72***
Model diagnostics			
Sargan test (df = 16)	$\chi^2 = 131.32 \quad p < 0.001$		
AR(1) test	$z = -5.32 \quad p < 0.001$		
AR(2) test	$z = -3.88 \quad p < 0.001$		
Wald test (df = 2)	$\chi^2 = 630.40 \quad p < 0.001$		
Observations	3054		
Firms (n)	509		
Time periods (T)	8		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Linear Hypothesis : For feols Regressions

- Suppose you wish to test if the estimated production function is CRS.

```
library(fixest)
library(car) #### INSTALL THIS PACKAGE
library(sandwich) #### INSTALL THIS PACKAGE
fe_olsobj<-feols(y~ k + l|industry+year,
                 data=df2)
coef(fe_olsobj)
# Displaying output now.
# DO NOT BLINDLY COPY
[output]
      k           l
0.003620979 0.033905935
# BACK TO CODE
linearHypothesis(fe_olsobj, "k + l = 1", vcov =
    vcovCL(fe_tw))
# linearHypothesis() CANNOT automatically
# extract vcov from feols output. Provide it
```

Linear Hypothesis : For pgmm Regressions (1)

- Suppose you wish to test if the estimated production function is CRS.

```
library(plm)

pdf <- pdata.frame(df2, index = c("id","year"))

pgmm(
  y ~ k + l  |
   lag(k, 2:2) + lag(l, 2:2),
  data = pdf,
  effect = "individual",
  model = "onestep",
  transformation = "d"
)
```

Linear Hypothesis : For pgmm Regressions (2)

- Suppose you wish to test if the estimated production function is CRS.

```
# Displaying output now.
# DO NOT BLINDLY COPY
[output]
      k           l
0.01635895  0.40369588
# BACK TO CODE
linearHypothesis(
  ab1,
  hypothesis.matrix = "k + l  = 1",
  test = "Chisq" )
# linearHypothesis() CAN automatically
# extract vcov from pgmm output.
# no need to Provide it
```

Advance Control Function Methods

- Olley & Pakes (1996; OP) and Levinsohn & Petrin (2003; LP) are control function methods.
- Instead of looking for instruments for K and L, we look for observable variables that can control (or proxy) unobserved TFP.
- The control variables should come from a model of firm behavior.
- **Note:** Both OP and LP assume that labor is perfectly flexible input. This assumption is completely innocuous for their results. To emphasize this point, I present here versions of OP and LP that treat labor as a potentially dynamic input.

Olley and Pakes (OP)

- Consider the following model of simultaneous equations:

$$\begin{aligned} \text{(PF)} \quad & y_{it} = \alpha_L \ell_{it} + \alpha_K k_{it} + \omega_{it} + \varepsilon_{it} \\ \text{(LD)} \quad & \ell_{it} = f_L(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}) \\ \text{(ID)} \quad & i_{it} = f_K(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}) \end{aligned}$$

- *(LD)* & *(ID)*: firms' optimal labor and investment given state variables $(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it})$; r_{it} = input prices.
- OP consider the following assumptions:

(OP - 1) $f_K(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it})$ is invertible in ω_{it}

(OP - 2) No cross-sectional variation in r_{it}: $r_{it} = r_t$

(OP - 3) ω_{it} follows a first order Markov process.

(OP - 4) k_{it} is decided at $t-1$: $k_{it} = (1-\delta)k_{i,t-1} + i_{i,t-1}$

Olley and Pakes (2)

- OP method deals both with the simultaneity problem and with the selection problem due to endogenous exit.
- It doesn't deal with potential measurement error in inputs.
- OP method proceeds in two stages.
- First stage: estimates α_L [Assumptions (OP-1) and (OP-2) are key];
  and the second stage estimates α_K [Assumptions (OP-3) and (OP-4) are key].

- Assumptions (OP-1) and (OP-2) imply that the investment equation is invertible in ω_{it}:

$$\omega_{it} = f_K^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t)$$

- Solving this equation in the PF we have:

$$\begin{aligned} y_{it} &= \alpha_L \ell_{it} + \alpha_K k_{it} + f_K^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t) + e_{it} \\ &= \alpha_L \ell_{it} + \varphi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it} \end{aligned}$$

- This is a partially linear model. Parameter α_L and functions $\varphi_1(.)$, ..., $\varphi_T(.)$ can be estimated using semiparametric methods.
- A possible method is Robinson's method (199ss). OP use an n-th order polynomial to approximate the φ_t functions.

Olley and Pakes First Stage

- This first stage is a **Control Function** method: instead of instrumenting the endogenous regressors, we include additional regressors that capture the endogenous part of the error term.
- We are controlling for endogeneity by including $(\ell_{i,t-1}, k_{it}, I_{it})$ as proxies of ω_{it}.
- Key assumptions for the identification of α_L:
  (a) Invertibility of $f^k(\ell_{i,t-1}, k_{it}, \omega_{it}, r_t)$ w.r.t. ω_{it}.
  (b) $r_{it} = r_t$, i.e., no cross-sectional variability in unobservables, other than ω_{it}, affecting investment.
  (c) Given $(\ell_{i,t-1}, k_{it}, I_{it}, r_t)$, labor l_{it} still has sample variability.

- Example (with parametric linear investment func.):

$$\begin{aligned} \text{(PF)} \quad & y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} + \varepsilon_{it} \\ \text{(Inverse ID)} \quad & \omega_{it} = \gamma_1 \, i_{it} + \gamma_2 \, \ell_{i,t-1} + \gamma_3 \, k_{it} + \gamma_4 \, r_t \end{aligned}$$

- Then,

$$y_{it} = \alpha_L \, \ell_{it} + (\alpha_K + \gamma_3) \, k_{it} + \gamma_1 \, i_{it} + \gamma_2 \, \ell_{i,t-1} + \gamma_4 \, r_t + \varepsilon_{it} \quad (1)$$

$$\Rightarrow y_{it} = \alpha_L \, \ell_{it} + \underbrace{(\alpha_K + \gamma_3) \, k_{it} + \gamma_1 \, i_{it} + \gamma_2 \, \ell_{i,t-1} + \delta_t}_{\varphi_t(\ell_{i,t-1}, k_{it}, i_{it})} + \varepsilon_{it} \quad (2)$$

- Note that ℓ_{it} is correlated with r_t. Therefore, we need $r_{it} = r_t$ and include time dummies to control for r_t in order to have consistency of the OLS estimator in this regression.
- Note also that to identify ℓ_{it} with enough precision we need to **not** have high collinearity between this variable and $(k_{it}, i_{it}, \ell_{i,t-1})$.

Olley and Pakes Second Stage

- **Estimation of α_K**: It is based on the other two assumptions:

(OP-3) ω_{it} follows a first order Markov process.
(OP-4) k_{it} is decided at $t-1$: $k_{it} = (1-\delta)k_{i,t-1} + i_{i,t-1}$

- Since ω_{it} is first order Markov, we can write:

$$\omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{i,t-1}] + \xi_{it} = h(\omega_{i,t-1}) + \xi_{it}$$

  where ξ_{it} is an innovation which is mean independent of any information at $t-1$ or before. And $h(.)$ is some unknown function.

- $\hat{\varphi}_{it}$ is identified from 1st step as

$$\text{general setup } \hat{\varphi}_{it} = \hat{\varphi}_t(\ell_{i,t-1}, k_{it}, i_{it}) \quad (3)$$

$$\text{linear example: } \hat{\varphi}_{it} = (\alpha_K + \gamma_3)\, k_{it} + \gamma_1\, i_{it} + \gamma_2\, \ell_{i,t-1} + \delta_t \quad (4)$$

  where $\hat{\varphi}_t$ is denoting the estimated version of φ_t which you may estimate using a polynomial.

- Recall $\varphi_t(\ell_{i,t-1}, k_{it}, i_{it}) = \alpha_K \cdot k_{it} + \omega_{it} \Rightarrow \omega_{it} = \hat{\varphi}_{it} - \alpha_K \cdot k_{it}$. Then:

$$\omega_{it} = h(\omega_{it-1}) + \xi_{it} \Rightarrow \hat{\varphi}_{it} - \alpha_K \cdot k_{it} = h(\hat{\varphi}_{it} - \alpha_K \cdot k_{it}) + \xi_{it} \quad (5)$$

$$\hat{\varphi}_{it} = \alpha_K\, k_{it} + h\left(\hat{\varphi}_{i,t-1} - \alpha_K \cdot k_{i,t-1}\right) + \xi_{it} \quad (6)$$

- We estimate $h(.)$ and α_K by applying recursively the same type of semiparametric method as in the first stage of OP.

$$\hat{\varphi}_{it} = \alpha_K k_{it} + h\left(\hat{\varphi}_{i,t-1} - \alpha_K \cdot k_{i,t-1}\right) + \xi_{it}$$

- Suppose that we consider a quadratic function for $h(.)$: i.e., $h(\omega) = \pi_0 + \pi_1\omega + \pi_2\omega^2$. Then:

$$\begin{aligned}\hat{\varphi}_{it} =& \pi_0 + \alpha_K k_{it} + \pi_1 \cdot (\hat{\varphi}_{i,t-1} - \alpha_K \cdot k_{i,t-1}) \qquad (7) \\ &+ \pi_2 \cdot (\hat{\varphi}_{i,t-1} - \alpha_K \cdot k_{i,t-1})^2 + \xi_{it} \qquad (8)\end{aligned}$$

- It is clear that π_0, α_K, π_1, and π_2 are identified in this equation.

- Time-to-build is a key assumption for the consistency of this method. If investment at period t is productive, then the equation becomes:

$$\varphi_{it} = \alpha_K k_{i,t+1} + h\left(\alpha_K \cdot k_{it} - \varphi_{i,t}\right) + \xi_{it}$$

- $k_{i,t+1}$ depends on investment at period t and therefore it is correlated with the innovation ξ_{it}.
- To relax this assumption, you need an instrument for k_{it+1}.

OP-Implentation (1)

```
library(tidyverse)
library(fixest)
deg <- 3
dat <- df_op %>%
  mutate(across(c(ly1, ll, lk, li), as.numeric))
      %>%
  group_by(ccn) %>% arrange(year, .by_group =
     TRUE) %>%
  mutate(ll_lag = dplyr::lag(ll,1),
         lk_lag = dplyr::lag(lk,1)) %>%
  ungroup() %>%
  filter(!is.na(ly1), !is.na(ll), !is.na(lk), !
     is.na(li), !is.na(ll_lag))
# Stage 1
m1 <- feols(
  ly1 ~ ll + I(year):poly(ll_lag, lk, li, degree
      = deg, raw = TRUE),
  data = dat
)
```

OP-Implentation (2)

```
alpha_L_hat <- unname(coef(m1)["ll"])
# creating phi_hat, lag logk, lag phi_hat
dat$phi_hat<- fitted(m1) - alpha_L_hat * dat$ll
dat2<- dat %>%
  group_by(ccn) %>%
  arrange(year, .by_group = TRUE) %>%
  mutate(lag_phi_hat=dplyr::lag(phi_hat,1),
         lag_lk=dplyr::lag(lk,1)) %>%
  ungroup() %>%
  na.omit()
#Stage - 2
m2<-feols(phi_hat ~ lk + poly(lag_phi_hat,lag_lk
   ,degree = 3,raw=TRUE),data=dat2)
```

OP-Implentation (3)

```
etable(m1, m2,
  keep= c("%ll$", "%lk"),                # only
      show alpha_L and alpha_K
  drop = c("\\(Intercept\\)", "^year::", "poly
     \\("),
  dict = c("ll"="$\\alpha_L$",
           "lk"="$\\alpha_K$"),
  se.below    = TRUE,# estimates with SE in ( )
  fitstat     = ~ n + r2 + ar2 + f,
  digits      = 3,
  signif.code = c("***"=0.01,"**"=0.05,"*"=0.1),
  tex = TRUE
)
```

OP-Implementation (4)

Dependent Variables:	ly1	phi_hat
Model:	(1)	(2)
Variables		
α_L	0.201***	
	(0.030)	
α_K		0.013***
		(0.0008)
Fit statistics		
Observations	8,287	4,251
R^2	0.43253	0.59275
Adjusted R^2	0.43116	0.59179
F-test	315.02	617.13

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Stage-2 SEs are conventional OLS and ignore that $\hat{\psi}$ and its lag were estimated in Stage 1.

OP: Empirical Application

- US Telecom. equipment industry: 1974–1987.
- Technological change and deregulation.
  - Elimination of barriers to entry;
  - Antitrust decisions against AT&T: The Consent Decree (implemented in 1984) $\rightarrow$ divestiture of AT&T.
  - Substantial entry/exit of plants.
- Data: US Census of manufacturers.

OP: Empirical Application

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

OP: Empirical Application

- Going from OLS balanced panel to OLS full sample almost doubles α_K and reduces α_L by 20%. [Importance of endogenous exit].
- Controlling for simultaneity further increases α_K and reduces α_L.

Total Factor Productivity (TFP)

- Production function:

$$Y_{it} = A_{it} F(K_{it}, L_{it}, M_{it})$$

- A_{it} is denoted as Total Factor Productivity (TFP).
- It is a factor-neutral shifter that captures variations in output not explained by observable inputs.
- TFP is a residual.

Large & Persistent Differences in TFP Across Firms

- **Ubiquitous:** Observed even within narrowly defined industries and products.
- **Large differences:** 90th to 10th percentile TFP ratios $\frac{A_{90th}}{A_{10th}}$
  - U.S. manufacturing, average within 4-digit SIC industries = 1.92
  - Denmark: average = 3.75
  - China or India, average > 5
- **Persistent:** AR(1) of log-TFP with annual frequency:
  - Autoregressive coefficients between 0.6 to 0.8.
- **It matters:** Higher TFP producers are more likely to survive, innovate, invest.

Why Do Firms Differ in Their Productivity Levels?

- What supports such large productivity differences in equilibrium?
- Can producers control the factors that influence productivity, or are they purely external effects of the environment?
- If firms can partly control their TFP, what type of choices increase it?

Why TFP Dispersion Is Possible in Equilibrium?

- Because the profit function is concave in output and the optimal amount of profit for a monopolist (or duopolist, ...) is smaller than total demand.
- Let the profit of a firm be:

$$\pi_i = P_i(Y_i)Y_i - C(Y_i, A_i)$$

  $P_i(Y_i) =$ Inverse demand function; $C(Y_i, A_i) =$ Cost function.
- Key condition: Either $P_i(Y_i)Y_i$ is strictly concave in Y_i, or $C(\cdot)$ is strictly convex in Y_i.
  - The profit function is strictly concave.
- **Example: Decreasing Returns to Scale (DRS) even with perfect competition:** PY_i is linear in Y_i, but $C(\cdot)$ is strictly convex due to DRS.
- **Example: Oligopoly competition even with Constant Returns to Scale (CRS):** $C(\cdot)$ is linear, but $P_i(Y_i)Y_i$ is strictly concave in Y_i if demand is downward sloping.

Why TFP Dispersion Is Possible in Equilibrium? [2]

- Equilibrium implies the marginal condition for optimal output:

$$MR_i \equiv \frac{\partial[P(Y_i)Y_i]}{\partial Y_i} = \frac{\partial C(Y_i, A_i)}{\partial Y_i} \equiv MC_i$$

- If variable profit is strictly concave, this equilibrium can support firms with different TFPs, A_i.
- It is not optimal for the firm with the highest TFP to provide all the output in the industry.
- Firms with different TFPs (above a certain threshold value) operate in the same market.

How Can a Firm Affect Its TFP?

- Human resources and managerial practices.
- Learning-by-doing.
- Organizational structure:
  - Vertical integration vs outsourcing.
- Higher quality of labor and capital inputs.
- Adoption of new technologies.
- Investment in R&D.
- Innovation:
  - Process innovation.
  - Product innovation.