Production Functions in Empirical IO (EIOBook-3)

Fast, intuitive, and highly coherent review notes

The coherence spine (read this first)

This chapter is one continuous workflow. In empirical IO, we ultimately care about **costs, markups, entry/exit, and counterfactuals**, but these objects depend on firms' underlying **technology** and **productivity**. A production function turns that goal into an estimable equation, but it immediately exposes the central obstacle: firms choose inputs while observing their own productivity, so inputs are correlated with the unobserved productivity component. That correlation makes naive regressions misleading and forces an identification strategy (IV, panel timing, dynamic moments, or control functions derived from firm decisions). Once productivity is credibly recovered, we can interpret why productivity dispersion persists in equilibrium (market structure and cost curvature), and then move from description to mechanism by modeling how productivity evolves—especially through R&D and knowledge accumulation. In short:

IO needs costs/TFP $\Rightarrow$ PF delivers them $\Rightarrow$ endogeneity/selection threatens credibility $\Rightarrow$ methods restore ide

The one mental picture

Production-function estimation = separate technology from productivity. You observe output and inputs, but the firm also has an unobserved productivity term that it knows when choosing inputs. The entire chapter is about recovering technology parameters and productivity in a way that respects that decision-making.

A 60–90 minute study plan

Pass 1 (10 min): What you estimate and why IO cares

- **Technology:** a mapping from inputs into output (the production function).
- **Productivity/TFP:** the residual component shifting output for given inputs.
- **IO link:** once you have technology and productivity, you can build costs (or marginal costs), which feed directly into pricing, markups, competition, and counterfactuals.

Pass 2 (10–15 min): Lock the workhorse form + parameter meaning

A standard log-linear specification is:

$$y_{it} = a_\ell \ell_{it} + a_k k_{it} + a_m m_{it} + \omega_{it} + \varepsilon_{it},$$

where y is (log) output, ℓ labor, k capital, m materials/intermediate inputs, ω_{it} is (log) productivity/TFP observed by the firm, and ε_{it} is measurement error or an unanticipated shock.

- a_ℓ, a_k, a_m are **output elasticities**.
- Returns to scale is $a_\ell + a_k (+a_m)$ (depending on the inputs included).
- **Key object for later:** once a's are known, ω_{it} is interpretable as productivity.

Connection (why the next pass exists). Once the production function is written in logs, it looks like a standard regression with an error term, but that "error" contains productivity ω_{it}, which firms typically observe when choosing inputs. This timing turns inputs into *responses* to productivity. Therefore, the econometric issues are not side details; they are direct consequences of treating the production function as a behavioral object in an optimizing environment.

Pass 3 (15 min): The three landmines (why naive estimation fails)

1. **Simultaneity/endogeneity:** the firm chooses inputs knowing ω_{it}, so inputs correlate with productivity.
2. **Measurement error:** output may be revenue rather than quantity; capital is often constructed; differencing can amplify noise in persistent inputs.
3. **Selection (endogenous exit):** low-ω firms are more likely to exit; the surviving sample is selected.

Pass 4 (20–30 min): Methods as identification stories

Connection (how to read the method menu). All estimators in this chapter solve the same fundamental problem: inputs are correlated with unobserved productivity. They differ only in **which behavioral/timing assumption** they use to break that link. Said differently, each method is a different way to make $\mathbb{E}[\text{unobservables} \mid \text{variation used for estimation}] = 0$ plausible.

A. OLS (baseline)

OLS is a benchmark; it typically fails because $\ell_{it}, k_{it}, m_{it}$ move with ω_{it}.

B. IV using input prices or cost shifters

Use exogenous variation that shifts input choices but is plausibly independent of ω_{it}. Practical limitation: valid price variation is often weak or unavailable.

C. Fixed effects / within estimator

If productivity decomposes as $\omega_{it} = \eta_i + \delta_t + u_{it}$ and u_{it} is realized after input choices (or is otherwise conditionally exogenous), FE can help. Tradeoff: FE/differencing may magnify measurement error and can struggle when productivity is persistent and inputs are dynamic.

D. Dynamic panel GMM (Arellano–Bond; System GMM)

Difference to remove fixed effects, then use lags as instruments:

- **Arellano–Bond:** instruments differenced inputs with deeper lags.
- **System GMM:** adds level equations with additional moment conditions (often improves weak-instrument issues).

Tradeoff: persistence and measurement error can make instruments weak.

E. Control-function methods (OP / LP / ACF)

Core idea: use a firm decision rule (investment or materials) as a **proxy for productivity**:

- **Olley–Pakes (OP):** invert investment (under monotonicity) to control for ω; can also model exit to correct selection.
- **Levinsohn–Petrin (LP):** use intermediate inputs/materials when investment has many zeros.
- **ACF critique/fix:** highlights potential weak identification (e.g., labor absorbed by the proxy) under some timing assumptions; requires careful structure to truly identify coefficients.

F. FOC/share-based identification (for flexible inputs)

If a flexible input (often materials) satisfies a first-order condition, expenditure shares can identify its elasticity. Quasi-fixed inputs (capital) are then identified via dynamics/timing restrictions.

Pass 5 (10 min): Interpreting the productivity dispersion you estimate

Connection (why we do interpretation after estimation). After choosing an estimator, productivity becomes an empirical distribution $\{\omega_{it}\}$. The next step is economic: if productivity differs widely, why doesn't competition instantly eliminate dispersion? The answer links back to IO: equilibrium can sustain dispersion when profit is concave in scale (due to market power/demand curvature or convex costs), so productivity differences do not translate one-for-one into unlimited expansion.

- Dispersion is an equilibrium object shaped by technology curvature and market structure.
- This interpretation prevents treating ω as mere "noise" and connects it to firm dynamics and policy.

Pass 6 (10 min): From residual to mechanism — R&D and knowledge capital

Connection (why R&D naturally comes last). Once dispersion is understood as an equilibrium outcome, the natural question is what moves firms within that distribution over time. R&D provides a mechanism: it makes productivity a state variable with a law of motion, where innovation investment increases the probability or magnitude of future productivity improvements. This connects static measurement (PF) to dynamic counterfactuals (policy/strategy).

- **Knowledge capital approach:** build an R&D stock via perpetual inventory and treat it as an input/state that raises productivity.
- **Structural dynamic approach:** R&D affects productivity through an uncertain, endogenous innovation process identified using FOCs and dynamics.

Exam-ready one-minute summary

- Baseline: $y_{it} = a_\ell \ell_{it} + a_k k_{it} + a_m m_{it} + \omega_{it} + \varepsilon_{it}$.
- Problems: simultaneity (inputs chosen with ω), measurement error, endogenous exit.
- Solutions: IV, FE, dynamic GMM, control functions (OP/LP/ACF), FOC-based identification.
- Output: credible a's and ω; interpret dispersion; model dynamics (R&D).

Five-minute self-test

1. Why is OLS biased in production-function estimation (one sentence)?
2. Why can FE/differencing amplify measurement error for persistent inputs like capital?
3. What is the weak-instrument problem in Arellano–Bond, intuitively?
4. In OP/LP, what is the control function, and what assumption makes inversion possible?
5. Why can endogenous exit bias coefficients, and how does OP address selection?