Identifying the elasticity of experience and its effect on market structure

Zach Flynn*

April 18, 2019

Abstract

Experience is a factor of production. Practice makes perfect. But when prior production enters the firm's current production function, standard dynamic models of firm production like the Olley and Pakes (1996) proxy model suffer from a simultaneity problem. Furthermore, flexible inputs are chosen dynamically and firms may even choose negative markups temporarily to increase their stock of experience. So identification approaches that are based on assuming zero or positive markups will not work (like Gandhi, Navarro, and Rivers 2019). I develop an identification approach that works for this problem. My approach only requires panel data on firm input and output choice so it can be broadly applied. I use this appoach to study the evolution of learning rates (the rate at which firms learn from experience over time) and how differences in learning rates help explain differences in the firm size distribution across industries.

1. Introduction

A firm's experience producing a product is a key input in the product's production function. This paper is about how to identify the production function when it depends on prior production, firms have flexible inputs, and we have data only on output and input use.

Empirical evidence that firms learn from production abounds. Wright (1936) estimated that the inputs required for airplane production fell by 20% every time cumulative output, a measure of experience, doubled. Benkard (2000) found that input requirements fell 36% when cumulative output doubled for the Lockheed L-1011 TriStar (an airplane) and that the firm forgot some of what they learned via experience as time went on. Thornton and Thompson (2001) found that there were strong experience effects in ship production in World War II and that there were spillover effects from one firm gaining experience on the productivity of other firms. This relationship is so commonly observed in industry that a major consulting firm, the Boston Consulting Group (BCG), claims it as one its "signature concepts" (Reeves, Stalk, and Pasini, 2013) and its founder popularized the idea in industry

*Email: zlflynn@gmail.com.

(Henderson, 1968). While consulting for a semiconductor manufacturer, BCG found unit costs fell by 20-30 percent each time cumulative output doubled.

Experience has implications for industrial organization and market structure. Because firms can control the level of experience they acquire by producing more output, experience functions as an endogenous sunk cost in the sense of Sutton (1991) and Sutton (2007). Firms may produce excessive output from a static perspective and earn negative markups temporarily to obtain lower costs in the long run. Identification methods that identify the production function by assuming zero or positive marginal costs are not applicable.

My goal in this paper is to identify the elasticity of experience and how it affects market structure. But the gross output production function[1] with prior output as an input introduces novel identification problems that are difficult to address without special data. I develop a method that works with only panel data on inputs and output so it can be applied broadly.

Gandhi, Navarro, and Rivers (2019) (GNR) show that the proxy method[2] of estimating the production function is not identified in the presence of *flexible* inputs. Flexible inputs are inputs that affect period t output that are also chosen in period t. The gross output production function will usually depend on some flexible input like "materials" so this identification problem is fundamental.

GNR's solution to the identification problem is to assume that the flexible inputs are also *static*. They have no dynamic consequences. If this is the case and firm's take prices as given, the output elasticities corresponding to the static, flexible inputs can be recovered as the ratio of spending on the inputs to total revenue from the first order conditions of the firm's profit maximization problem. But when experience is an input to the production function, there are no static inputs. Any input that affects output affects the future accumulation of experience so all inputs are dynamic.

In addition, the Levinsohn and Petrin (2003) proxy assumption that flexible input demand is increasing in productivity given inflexible inputs is not true in general when experience is an input in production. Therefore, we cannot rely on it when identifying the production function. Intuitively, the assumption does not hold because greater productivity can *reduce* the marginal dynamic benefit of flexible input use as we reach decreasing marginal returns to experience.

In this paper, I propose a method for estimating structural production functions when experience is an input in the production function. To do so, I build a dynamic programming model of firm production where experience and flexible input choice are treated as dynamic inputs. While the model I propose is clearly in the tradition of the proxy variable models from Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015), it does not use a proxy variable assumption to identify the production function.

[1]As opposed to a value-added production function as in Ackerberg, Caves, and Frazer (2015).

[2]See Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015). The proxy method has been used in a wide variety of empirical applications to estimate production functions. See Syverson (2011) for a survey of applications using this method and others.

I apply the identification strategy to study the importance of variation in the elasticity of experience in explaining firm size and its persistence across industries and time. Firms that have already sunk the cost of acquiring experience have a cost advantage relative to current entrants who can only match them with either innately high productivity or by being willing to take on losses to acquire the experience necessary to catch the competition. I bound the counterfactual firm size distribution if firms did not take into account the value of their experience to isolate the influence of experience on the firm size distribution.

Balasubramanian and Lieberman (2011) is the most closely related empirical paper. They apply the Olley and Pakes (1996) version of the proxy method with a *value-added* production function to estimate the output elasticity of cumulative output. Because they assume there are no flexible inputs in the production function, they do not face the identification problem I address in this paper. My contribution is a method to estimate *gross output* production functions where prior output is an input and some inputs are flexible. Gross output production functions are important because the assumptions necessary to justify value added production functions are strong. See GNR for more discussion on the differences between the two.

2. Model

Firms produce output according to the production function[3],

$$q_t = \theta_v v_t + \theta_k k_t + \theta_s s_t + a_t, \tag{1}$$

where q_t is log output, v_t is the log of flexible input use, k_t is log capital, s_t is the log stock of experience (which is acquired by prior production), and a_t is log total factor productivity.

Total factor productivity is a first-order Markov process,

$$a_t = g\left(a_{t-1}\right) + \eta_t, \tag{2}$$

where η_t are innovation shocks to productivity.

Firms acquire knowledge via past production according to the following transition equation (where capital letters represent the levels of the log variables above):

$$S_t = \delta S_{t-1} + Q_{t-1}. \tag{3}$$

The persistence of the stock of knowledge is governed by δ and the stock is increased by production.

[3]Throughout, I use a Cobb Douglas production function. This is not strictly necessary. Identification is nonparametric at the cost of algebra and a few more technical conditions. In the final version of this paper, nonparametric identification results will be included in the text, but the intuition is captured with the Cobb Douglas results.

Firm i chooses its input and output by solving the dynamic programming problem,

$$\begin{aligned} U_{i,t}\left(A,K,S\right) = &\max_{V,X} P_t V^{\theta_v} K^{\theta_k} S^{\theta_s} A - W_{V,t} V - H_i\left(X\right) \\ &+ \beta \int \left[U_{i,t+1}\left(\exp\left(g\left(a\right)+\eta\right), \phi K + X, \delta S + V^{\theta_v} K^{\theta_k} S^{\theta_s} A\right)\right] dM_t\left(\eta\right) \end{aligned} \tag{4}$$

where ϕ is the depreciation rate of capital and M_t is the cumulative distribution function of η_{it}. $H_i\left(\cdot\right)$ is the cost of capital investment which I allow to vary for each firm.

3. Point identification given a known persistence of knowledge and discount factor

The parameters of the model are identified up to (δ, β), given identifying assumptions that are consistent with the model in Section 2. I use assumptions based on the timing and information restrictions embodied in the model as well as the first order conditions to identify the relevant parameters. Although my approach is strongly related to the proxy literature, I do not rely on the "proxy" assumption prominent in that literature that productivity can be controlled for by observables.

The main information assumption of the model is that η_t (or a signal of its value) is not known at time $(t-1)$ so decisions made at that time are uncorrelated with it.

Assumption 1. η_t is uncorrelated with decisions made in prior time periods.

$$\mathbb{E}\left[\begin{pmatrix} k_t \\ s_t \\ q_{t-1} \\ v_{t-1} \\ k_{t-1} \\ s_{t-1} \end{pmatrix} \eta_t\right] = 0. \tag{5}$$

Assumption 1 allows us to identify (θ_k, θ_s) for a given θ_v. Because, if θ_v is known,

$$q_t = \theta_v v_t + \theta_k k_t + \theta_s s_t + a_t \tag{6}$$

$$\implies q_t - \theta_v v_t = \theta_k k_t + \theta_s s_t + g\left(q_{t-1} - \theta_v v_{t-1} - \theta_k k_{t-1} - \theta_s s_{t-1}\right) + \eta_t \tag{7}$$

$$\implies q_t - \theta_v v_t = \theta_k k_t + \theta_s s_t + \widetilde{g}\left(q_{t-1}, v_{t-1}, k_{t-1}, s_{t-1}\right) + \eta_t, \tag{8}$$

and by Assumption 1, all the right hand side covariates are uncorrelated with η_t. So regressing $(q_t - \theta_v v_t)$ on $(k_t, s_t, q_{t-1}, v_{t-1}, k_{t-1}, s_{t-1})$ is sufficient to identify (θ_k, θ_s) so long as $\delta \neq 0$. If $\delta = 0$, then $s_t = q_{t-1}$ and the two are perfectly collinear.

To identify θ_v, I use the first order conditions of the firm's problem. The first order condition with respect to V is,

$$P_t Q \frac{\theta_v}{V} - W_{V,t} + \beta \int \frac{\partial U_{i,t+1}}{\partial S} \times \frac{\theta_v}{V} \times Q dM_t\left(\eta\right) = 0 \tag{9}$$

I use the envelope theorem to recover the derivative of U with respect to S.

$$\frac{\partial U_{i,t+1}}{\partial S} = P_{t+1} Q_{i,t+1} \frac{\theta_s}{S_{i,t+1}} + \beta \int \frac{\partial U_{i,t+2}}{\partial S} \times \left[\delta + \theta_s \frac{Q_{i,t+1}}{S_{i,t+1}} \right] dM_{t+1}(\eta) \tag{10}$$

We can use the first order condition for the $(t+1)$ decision about $V_{i,t+1}$ to recover the value of the second term.

$$\theta_v P_{t+1} Q_{i,t+1} - W_{V,t+1} V_{i,t+1} + \beta \theta_v Q_{i,t+1} \int \frac{\partial U_{i,t+2}}{\partial S} dM_{t+1}(\eta) = 0 \tag{11}$$

$$\implies \int \frac{\partial U_{i,t+2}}{\partial S} dM_{t+1}(\eta) = \frac{W_{V,t+1} V_{i,t+1} - \theta_v P_{t+1} Q_{i,t+1}}{\beta \theta_v Q_{i,t+1}} \tag{12}$$

It is useful to write the firm's information set at time t as $I_{i,t}$, a vector containing all the variables known to the firm at time t. Taking the expression for $(\partial U_{i,t+1}/\partial S)$ from above, the first order condition with respect to $V_{i,t}$ is then,

$$\beta \mathbb{E} \left[(-\theta_s) \times \underbrace{\frac{P_{t+1} Q_{i,t+1}}{S_{i,t+1}}}_{\text{Experience productivity}} + \left[\delta + \theta_s \frac{Q_{i,t+1}}{S_{i,t+1}} \right] \times \underbrace{\left(P_{t+1} - \frac{W_{V,t+1} V_{i,t+1}}{Q_{i,t+1} \theta_v} \right)}_{\text{Future static markups}} | I_{i,t} \right] \tag{13}$$

$$= \underbrace{P_t - \frac{W_{V,t} V_{i,t}}{Q_{i,t} \theta_v}}_{\text{Present static markups}} \tag{14}$$

Static markups are what markups would be if firms were static cost minimizers (they are not). From the envelope theorem, in static cost minimization problems,

$$\text{MC} = \frac{W_V V}{Q} \times \frac{1}{\theta_v}. \tag{15}$$

So the first order condition says that, holding future static markups fixed, greater future experience productivity (weighted by the experience elasticity) incentivizes the firm to take on lower present static markups. If the experience productivity term is sufficiently large relative to expected future markups, current static markups will be negative. In fact, the first order condition says that if $\theta_s \neq 0$, static markups cannot always be zero. To see this last result, replace the future and present static markup terms above with zero to find the first order condition cannot be satisfied without $\theta_s = 0$.

After some algebra, I get θ_v as a linear function of θ_s where the coefficients are data,

$$\implies \beta\theta_s Q_{i,t}\mathbb{E}\left[\frac{P_{t+1}Q_{i,t+1}}{S_{i,t}}|I_t\right] + \beta\delta\frac{Q_{i,t}}{\theta_v}\times\mathbb{E}\left[\frac{W_{V,t}V_{i,t+1}}{Q_{i,t+1}}|I_{i,t}\right] - \beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] \quad (16)$$

$$+\beta Q_{i,t}\times\frac{\theta_s}{\theta_v}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right] - \beta Q_{i,t}\theta_s\mathbb{E}\left[\frac{P_{t+1}Q_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right] = \quad (17)$$

$$\beta\delta\frac{Q_{i,t}}{\theta_v}\times\mathbb{E}\left[\frac{W_{V,t}V_{i,t+1}}{Q_{i,t+1}}|I_{i,t}\right] - \beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] + \beta Q_{i,t}\times\frac{\theta_s}{\theta_v}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right] = \quad (18)$$

$$\frac{W_{V,t}V_{i,t}}{\theta_v} - P_tQ_{i,t} \quad (19)$$

I then multiply through by θ_v to recover θ_v as a linear function of θ_s,

$$\beta\delta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t}V_{i,t+1}}{Q_{i,t+1}}|I_{i,t}\right] - \beta\theta_v\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] + \theta_s\times\beta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right] \quad (20)$$

$$= W_{V,t}V_{i,t} - P_tQ_{i,t}\theta_v \quad (21)$$

$$\implies \frac{\beta\delta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t}V_{i,t+1}}{Q_{i,t+1}}|I_{i,t}\right] - W_{V,t}V_{i,t}}{\left(\beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] - P_tQ_{i,t}\right)} + \theta_s\times\frac{\beta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right]}{\left(\beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] - P_tQ_{i,t}\right)} \quad (22)$$

$$= \theta_v. \quad (23)$$

I can then write the production function equation as,

$$q_{i,t} - \frac{\beta\delta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t}V_{i,t+1}}{Q_{i,t+1}}|I_{i,t}\right] - W_{V,t}V_{i,t}}{\left(\beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] - P_tQ_{i,t}\right)}\times v_{i,t} = \quad (24)$$

$$\left[\frac{\beta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right]}{\left(\beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] - P_tQ_{i,t}\right)}v_{i,t} + s_{i,t}\right]\theta_s + \theta_k k_{i,t} + \widetilde{g}\left(q_{i,t-1}, v_{i,t-1}, k_{i,t-1}, s_{i,t-1}\right) + \eta_{i,t} \quad (25)$$

From Assumption 1, I have sufficient instruments to identify (θ_s, θ_k) from the above equation for known (δ, β). To identify (θ_s, θ_k), I only need to assume that $s_{i,t}$ has power as an instrument for the θ_s term.

Assumption 2. $s_{i,t}$ has power as an instrument.

$$\mathbb{E}\left[s_{i,t}\left\{\frac{\beta Q_{i,t}\times\mathbb{E}\left[\frac{W_{V,t+1}V_{i,t+1}}{S_{i,t+1}}|I_{i,t}\right]}{\left(\beta\delta Q_{i,t}\mathbb{E}\left[P_{t+1}|I_{i,t}\right] - P_tQ_{i,t}\right)}v_{i,t} + s_{i,t}\right\}\right] \neq 0. \quad (26)$$

Assumption 2 is highly credible because, of course, $s_{i,t}$ is correlated with $s_{i,t}$. So all that the assumption requires is that this correlation is not undone by a negative correlation with the first term exactly equal in magnitude to the second moment of $s_{i,t}$.

So, given Assumption 1 and 2, the production function $(\theta_v, \theta_k, \theta_s)$ is identified for given (δ, β) because θ_v can be recovered after identifying θ_s.

4. Partial identification across the persistence of experience parameter and discount factors

There are two paths to identifying the production function unconditional on (δ, β):

1. I could assume the (δ, β) parameters are known. This assumption is not as strong as it might seem at least relative to the current state of the literature. A vast empirical literature in macroeconomics and industrial organization that uses dynamic programming methods assumes a known time discount factor β. Balasubramanian and Lieberman (2011) and other work on the "experience curve" (Henderson, 1968), implicitly assume $\delta = 1$.

2. I could allow for a range of "reasonable" values for (δ, β), leading to partial identification of the production function.

I take the second path, but I also highlight the results for specific, commonly chosen (δ, β) parameters. In particular, $(\delta = 1, \beta = 0.90, 0.95, 0.98)$.

I allow δ to be between $[\underline{\delta}, 1]$ where $\underline{\delta}$ is small, representing rapid depreciation of experience. $\underline{\delta} > 0$ to avoid collinearity between (s_t, q_{t-1}). Technically, if v_t and s_t are correlated, I can identify the Cobb Douglas production function with $\delta = 0$, but that results depends on the parametric form.

Assumption 3. A firm's past experience weakly depreciates in value over time,

$$0 < \underline{\delta} \leq \delta \leq 1. \tag{27}$$

I then consider reasonable values for β. Similarly to δ, I assume β lies between two known values. I start with the weakest possible assumption, but I also present results for ranges of β that are commonly assumed in the empirical dynamic programming literature: $0.88 \leq \beta \leq 0.98$.

Assumption 4. The firm cares more about today than tomorrow,

$$0 \leq \beta \leq 1. \tag{28}$$

Lastly, I assume the production function is increasing which may rule out some choices for δ and β that imply negative output elasticities.

Assumption 5. The production function is increasing in all inputs,

$$\theta_v \geq 0, \theta_k \geq 0, \theta_s \geq 0. \tag{29}$$

The set of identified production functions are then,

$$\Theta = \left\{ (\theta_v, \theta_k, \theta_s) \in \mathbb{R}^3_+ : \theta = \theta(\beta, \delta), 0 \leq \beta \leq 1, \underline{\delta} \leq \delta \leq 1 \right\}, \tag{30}$$

where $\theta(\beta, \delta)$ is the unique θ identified for a given (β, δ).

My objects of interest are scalar functions of the data and the production function parameters. I write the general parameter of interest as a function of θ and the distribution of the relevant variables (y) in the data $\mathcal{F}_y$.

$$T\left(\theta, \mathcal{F}_y\right). \tag{31}$$

The bounds on T are the smallest and largest value of T given the restrictions on (β, δ, θ).

$$\begin{aligned} \left\{\underline{T}, \overline{T}\right\} = \max \text{ or } \min_{\beta,\delta} \quad & T\left(\theta\left(\beta, \delta; \mathcal{F}_y\right), \mathcal{F}_y\right) \\ \text{ST:} \quad & 0 \leq \beta \leq 1, \underline{\delta} \leq \delta \leq 1, \theta\left(\beta, \delta; \mathcal{F}_y\right) \geq 0. \end{aligned} \tag{32}$$

5. Inference on functions of the production function

But I do not known the distribution of the data $\mathcal{F}_y$. I have to estimate it, and so, I need to make inference on the bounds on T given the data.

I use a Bayesian approach. Bayesian methods have several practical advantages over frequentist methods in partially identified models. The bounds on T are maximums and minimums of a number of estimated parameters. Because maximization and minimization are non-differentiable functions, standard delta-method arguments do not work. See Chernozhukov, Lee, and Rosen (2013), Hsieh, Shi, and Shum (2017), and other work on frequentist inference in partially identified models for a discussion of the problem. But nondifferentiability does not pose a challenge for Bayesian methods. Bayesian inference is conditional on the data so we do not need to account for how the distribution of the estimator changes as sample size increases which is the key problem in making inference on non-smooth functions of moments in frequentist inference.

I assume that $y = (q_t, v_t, k_t, s_t, q_{t-1}, v_{t-1}, k_{t-1}, s_{t-1}, \dots)$ is distributed as a finite mixture of multivariate normal distributions or another flexible, parametric specification. From my posterior beliefs about the distribution of y, I can recover the posterior of both the upper and lower bounds on T.

Because T is only partially identified, my posterior beliefs about T itself always depends on my prior beliefs about T no matter how much data I have. But the bounds on T are based purely on the joint distribution of the data. So with sufficient evidence my beliefs about the bounds on T will converge to the truth. I make inference on T without specifying a prior for T itself by focusing on the likelihood a given value for T belongs to the identified set instead of on the likelihood T lies in a certain region.

For any given T_0, my posterior belief about the likelihood it belongs to the identified set is,

$$\pi\left(\underline{T} \leq T_0 \leq \overline{T}\right), \tag{33}$$

where π is my posterior belief about the distribution of the data which determines $\underline{T}$ and $\overline{T}$. I use these probabilities to construct "hypothesis intervals" which I also used in Flynn

(2019). Hypothesis intervals are the set of all parameters T_0 for which our posterior belief that they are included in the identified set exceeds a certain threshold,

$$\mathrm{HI}_\alpha = \left\{ T_0 : \pi\left(\underline{T} \leq T_0 \leq \overline{T}\right) \geq \alpha \right\}. \tag{34}$$

Hypothesis intervals are uniquely defined and depend only on the posterior of point-identified parameters unlike Bayesian credible intervals. Credible intervals are any range of the posterior distribution which is more than α likely to contain the parameter. In partially identified models, such an interval will always depend on the prior we have for T no matter how much data we collect, and there are many credible intervals.

I use simulation methods to construct the posterior and hypothesis interval in the following way:

1. Obtain draws from the posterior of the parameters (γ) governing the joint distribution of the data, $\mathcal{F}_y$. I use NUTS, a Hamiltonian Monte Carlo method via STAN, a programming language for specifying Bayesian models that can be called from R. See Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddel (2017).

2. For each of the L draws from the posterior $(\gamma_1, \ldots, \gamma_L)$, simulate a synthetic dataset of size R, as large as is feasible.

3. Compute $\underline{T}_\ell$ and $\overline{T}_\ell$ for each simulated dataset $\ell = 1, \ldots, L$.

4. For a given T_0, compute

$$\pi\left(\underline{T} \leq T_0 \leq \overline{T}\right) \approx \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{1}\left(\underline{T}_\ell \leq T_0 \leq \overline{T}_\ell\right), \tag{35}$$

   if this statistic exceeds α, then T_0 is included in the hypothesis interval. Otherwise, it is left out.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*, 2411–2451.

Balasubramanian, N. and M. Lieberman (2011). Learning-by-doing and market structure. *The Journal of Industrial Economics 59*(2), 177–198.

Benkard, C. L. (2000). Learning and forgetting: the dynamics of aircraft production. *American Economic Review 90*(4), 1034–1054.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddel (2017). Stan: a probabilistic programming language. *Journal of Statistical Software 76*(1).

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: estimation and inference. *Econometrica 81*(2), 667–737.

Flynn, Z. (2019). Identifying productivity when it is a choice. *Working Paper*.

Gandhi, A., S. Navarro, and D. Rivers (2019). On the identification of production functions: how heterogeneous is productivity? *Working Paper*.

Henderson, B. (1968). The experience curve. *Boston Consulting Group Perspectives*.

Hsieh, Y.-W., X. Shi, and M. Shum (2017). Inference on estimators defined by mathematical programming. *Working Paper*.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*, 1263–1297.

Reeves, M., G. Stalk, and F. S. Pasini (2013). Bcg classics revisited: Experience curve. https://www.bcg.com/en-us/publications/2013/growth-business-unit-strategy-experience-curve-bcg-classics-revisited.aspx. Accessed: 2019-03-13.

Sutton, J. (1991). *Sunk Costs and Market Structure*. MIT Press.

Sutton, J. (2007). Market structure: Theory and evidence. *The Handbook of Industrial Organization*, 2303–2364.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

Thornton, R. and P. Thompson (2001). Learning from experience and learning from others: an exploration of learning and spillovers in wartime shipbuilding. *American Economic Review 91*(5), 1350–1368.

Wright, T. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences 3*(4), 122–128.