6. Introduction to Dynamics

6.1 Introduction

Dynamics in demand and/or supply can be important aspects of competition in oligopoly markets. In many markets demand is dynamic in the sense that (a) consumers' current decisions affect their future utility, and (b) consumers' current decisions depend on expectations about the evolution of future prices (states). Some sources of dynamics in demand are consumer switching costs, habit formation, brand loyalty, learning, and storable or durable products. On the supply side, most firm investment decisions have implications on future profits. Some examples are market entry, investment in capacity, inventories, or equipment, or choice of product characteristics. Firms' production decisions also have dynamic implications if there is learning-by-doing. Similarly, menu costs, or other forms of price adjustment costs, imply that pricing decisions have dynamic effects.

Identifying the factors governing the dynamics is important to understanding competition and the evolution of market structure, and for the evaluation of public policy. To identify and understand these factors, we specify and estimate dynamic structural models of demand and supply in oligopoly industries. A dynamic structural model is a model of individual behavior where agents are forward looking and maximize expected intertemporal payoffs. The parameters are structural in the sense that they describe preferences and technological and institutional constraints. Under the *principle of revealed preference*, these parameters are estimated using longitudinal micro data on individuals' choices and outcomes over time.

We start with some examples and a brief discussion of applications of dynamic structural models of Industrial Organization. These examples illustrate why taking into account forward-looking behavior and dynamics in demand and supply is important for the empirical analysis of competition in oligopoly industries.

Example 1: Demand of storable goods

For a storable product, purchases in a given period (week, month) are not equal to consumption. When the price is low, consumers have incentives to buy a large amount to store the product and consume it in the future. When the price is high, or the household has a large inventory of the product, the consumer does not buy, and instead consumes from her inventory. Dynamics arise because consumers' past purchases and consumption decisions impact their current inventory and therefore the benefits of purchasing today. Furthermore, consumers' expectations about future prices also impact the perceived trade-offs of buying today versus in the future.

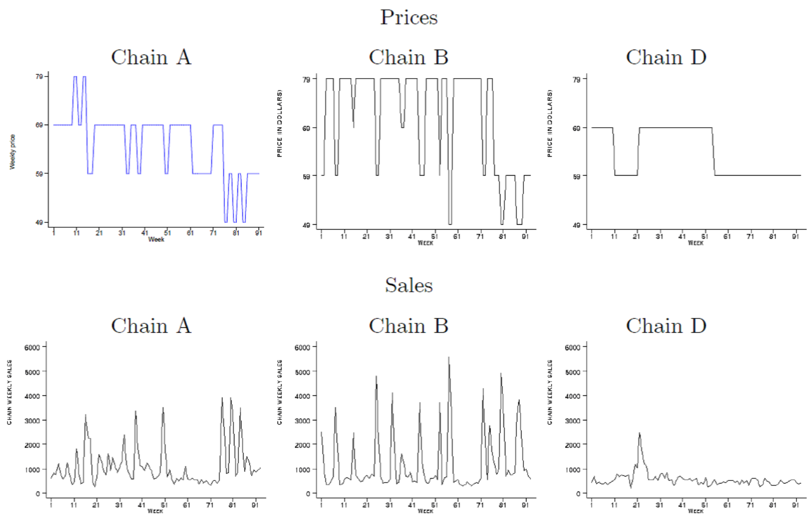

Figure 6.1: Price promotions and sales of a storable good

What are the implications of ignoring consumer dynamic behavior when we estimate the demand of differentiated storable products? An important implication is that we can get serious biases in the estimates of price demand elasticities. In particular, we can incorrectly interpret a short-run intertemporal substitution as a long-run substitution between brands (or stores).

To illustrate this issue, it is useful to consider a specific example. Figure 6.1 presents weekly times series data of prices and sales of canned tuna in a supermarket store. The time series of prices is characterized by "High-Low" pricing, which is quite common in many supermarkets. The price fluctuates between a high regular price and a low promotion price. The promotion price is infrequent and lasts only a few days, after which the price returns to its "regular" level. Sales of this storable product respond to this type of dynamics in prices.

As we can see in figure 6.1, most sales are concentrated in the very few days with low prices. Apparently, the short-run response of sales to these temporary price reductions is very large: the typical discount of a sales promotion is between 10% and 20%, and the increase in sales is around 300%. In a static demand model, this type of response would suggest that the price elasticity of demand of the product is very large. In particular, with these data, the estimation of a static demand model provides estimates of own-price elasticities greater than 8. The static model interprets the large response of sales to a price reduction in terms of consumers' substitution between brands (and to some extent between supermarkets too). Based on these estimates of demand elasticities, a model of competition would imply that price-cost margins are very small and firms (both supermarkets and brand manufacturers) have very little market power. A large degree of substitution between brands implies that product differentiation is small and market power is low.

This interpretation ignores dynamics in consumer purchasing decisions, and can therefore be seriously wrong. Most of the short-run response of sales to a temporary price reduction is not substitution between brands or stores, but intertemporal substitution in households' purchases. The temporary price reduction induces consumers to buy and store today, and to buy less in the future. The long-run substitution effect is much smaller, and it is this long-run effect that is relevant in measuring firms' market power.

In order to distinguish between short-run and long-run responses to price changes, we need to specify and estimate a dynamic model of demand of differentiated products. In this type of models, consumers are forward looking and take into account their expectations about future prices as well as storage costs. Examples of applications estimating dynamic structural models of demand of differentiated storable products are Erdem, Imai, and Keane (2003), Hendel and Nevo (2006), and Wang (2015),

Example 2: Demand of a new durable product.

The price of a new durable product typically declines over time during the months after the introduction of the product. Different factors may explain this price decline, for instance, intertemporal price discrimination, increasing competition, exogenous cost decline, or endogenous cost decline due to learning-by-doing. As in the case of the "high-low" pricing of storable goods, explaining these pricing dynamics requires one to take into account dynamics in supply. For the moment, we concentrate here on demand. If consumers are forward-looking, they expect that the price will be lower in the future, and this generates an incentive to wait and buy the good in the future.

A static model that ignores dynamics in demand of durable goods can introduce two different types of biases in the estimates of the distribution of consumers' willingness to pay, and therefore of demand. The first source of bias comes from the failure to recognize that market size is declining over time. The demand curve moves downwards over time because high willingness-to-pay consumers but the product at early periods and leave the market. A second source of bias comes from ignoring consumers' forward-looking behavior. The estimation of a static model that ignores this forward-looking behavior implies that consumers' willingness-to-pay can be contaminated by consumers' willingness to wait because of the expectation of future lower prices.

We can illustrate the first source of bias using a simple example. Consider a market with an initial mass of 100 consumers and a uniform distribution of willingness to pay over the the unit interval. Consumers are myopic and buy the product if the price is below their willingness to pay. Once consumers buy the product they are out of the market forever. Time is discrete and indexed by $t \in \{1, 2, ...\}$. Every period t, the aggregate demand is:

$$q_t = s_t \ \Pr(v_t \geq p_t) \ = \ s_t \ [1 - F_t(p_t)] \tag{6.1}$$

where q_t and p_t are quantity and price, respectively, s_t is the mass of consumers who remain in market at period t, and F_t is the distribution function of willingness to pay for consumers who remain in the market at period t.

Suppose that we observe a declining sequence of prices equal to $p_1 = 0.9$, $p_2 = 0.8$, $p_3 = 0.7$, and so on. Given this price sequence, it is easy to show that the demand curve changes over as follows:

$$\begin{array}{rcl} \textit{Period}\ \text{t=1:}\ q_1 & = & 100\ (1 - p_1) \\ \\ \textit{Period}\ \text{t=2:}\ q_2 & = & 90\ \left(\frac{0.9 - p_2}{0.9}\right)\ =\ 100\ (0.9 - p_2) \\ \\ \textit{Period}\ \text{t=3:}\ q_3 & = & 80\ \left(\frac{0.8 - p_2}{0.8}\right)\ =\ 100\ (0.8 - p_3) \\ \ldots & & \end{array} \tag{6.2}$$

Therefore, the sequence of realized quantities is constant over time: $q_1 = q_2 = q_3 = \ldots = 10$. A static demand model – with market size constant over time – leads the researcher to conclude that consumers are not sensitive to price, since the same quantity is sold as prices decline. The estimate of the price elasticity would be zero. This example illustrates how ignoring dynamics in demand of durable goods can lead to serious biases in the estimates of the price sensitivity of demand.

We can also illustrate the second source of bias – from ignoring consumer forward-looking behavior – using a simple variation of the previous example. Suppose that consumers are one-period forward-looking: they compare the utility of purchasing at period t versus waiting one period and purchase at $t+1$. Consumers have a time discount factor β and perfect foresight about next period price. Accordingly, a consumer with valuation v purchases the product at period t if $v - p_t \geq \beta(v - p_{t+1})$, or equivalently, if $v \geq (p_t - \beta p_{t+1})/(1-\beta)$. This consumer behavior implies the following aggregate demand at period t:

$$q_t \ = \ s_t \ \left[1 - F_t\left(\frac{p_t - \beta p_{t+1}}{1-\beta}\right)\right] \tag{6.3}$$

A static demand model that ignores forward-looking behavior estimates a demand equation $q_t = s_t\ [1 - F_t\,(p_t)]$, omitting next period price p_{t+1} and discount factor β. This misspecification can lead to substantial biases in the estimation of the probability distribution of consumer valuations.

Examples of empirical applications estimating dynamic structural models of demand of differentiated products are Esteban and Shum (2007), Carranza (2010), Gowrisankaran and Rysman (2012), and Melnikov (2013).

Example 3: Product repositioning in differentiated product markets.

A common assumption in static demand models of differentiated product is that product characteristics, other than prices, are exogenous. However, in many industries, product characteristics are important strategic variables. Firms modify these characteristics in response to changes in demand, costs, regulation, or mergers.

Ignoring the endogeneity of product characteristics has several implications. First, it can bias the estimated demand parameters. A dynamic game that acknowledges the endogeneity of some product characteristics and exploits the dynamic structure of the model to generate valid moment conditions can deal with this problem.

A second important limitation of a static model of firm behavior is that it cannot recover the costs of repositioning product characteristics. As a result, the static model cannot address important empirical questions such as the effect of a merger on product repositioning. That is, the evaluation of the effects of a merger using a static model should assume that the product characteristics (other than prices) of the new merging firm would remain the same as before the merger. This is at odds both with the predictions of theoretical models and with informal empirical evidence. Theoretical models of horizontal mergers show that product repositioning is an important source of value for a merging firm. Informal evidence shows that soon after a merger firms implement significant changes in their product portfolio.

Sweeting (2013) and Aguirregabiria and Ho (2012) are two examples of empirical applications that endogenize product attributes using a dynamic game of competition in a differentiated product industry. Sweeting (2013) estimates a dynamic game of oligopoly competition in the US commercial radio industry. The model endogenizes the choice of radio stations format (genre), and estimates product repositioning costs. Aguirregabiria and Ho (2012) propose and estimate a dynamic game of airline network competition where the number of direct connections that an airline has in an airport is an endogenous product characteristic.

Example 4: Dynamics of market structure

Ryan (2012) and Kasahara (2009) provide excellent examples of how ignoring supply-side dynamics and firms' forward-looking behavior can lead to misleading results.

Ryan (2012) studies the effects of the 1990 Amendments to the Clean Air Act on the US cement industry. This environmental regulation added new categories of regulated emissions, and introduced the requirement of an environmental certification that cement plants have to pass before starting their operation. Ryan estimates a dynamic game of competition where the sources of dynamics are sunk entry costs and adjustment costs associated with changes in installed capacity. The estimated model shows that the new regulation had negligible effects on variable production costs but it increased significantly the sunk cost of opening a new cement plant. A static analysis, that ignores the effects of the policy on firms' entry-exit decisions, would conclude that the regulation had negligible effects on firms profits and consumer welfare. In contrast, the dynamic analysis shows that the increase in sunk-entry costs caused a reduction in the number of plants, which in turn implied higher markups and a decline in consumer welfare.

Kasahara (2009) proposes and estimates a dynamic model of firm investment in equipment and uses it to evaluate the effect of an important increase in import tariffs in Chile during the 1980s. The increase in tariffs had a substantial effect on the price of imported equipment and it may have a significant effect on firms' investment. An important feature of this policy is that the government announced that it was a temporary increase and that tariffs would go back to their original levels after a few years. Kasahara shows that the temporary aspect of this policy exacerbated its negative effects on firm investment. Given that firms anticipated the future decline in both import tariffs and the price of capital, a significant fraction of firms decided not to invest and waited until the reduction of tariffs. This waiting and inaction would not appear if the policy change were perceived as permanent. Kasahara shows that the Chilean economy would have recovered faster from the economic crisis of 1982-83 if the increase in tariffs would have been perceived as permanent.

Example 5: Dynamics of prices in a retail market

The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, charge prices that differ by significant amounts, for instance, 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries.

Interestingly, the price differentials between firms, and the ranking of firms in terms of prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the following week. Using a more graphical description, we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also the construction of price indexes.

Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

For instance, one explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket's ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of placing an order that do not depend on the size of the order, or at least that do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: inventories increase by a large amount up to a maximum threshold when an order is placed, and decline gradually until a minimum value is reached, at which time a new order is placed. Given these dynamics of inventories, it is simple to show that the optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again.

Aguirregabiria (1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. Specifically, I show that these types of inventory-dependence price dynamics can explain more than 20% of the time series variability of prices in the data.

6.2 Firms' investment decisions

Some important firm investment decisions are discrete or at the *extensive margin*. Market entry and exit, machine replacement, or adoption of a new technology are examples of discrete investment investment decisions. Starting with the seminal work by Pakes (1986) on patent renewal and Rust (1987) on machine replacement, models and methods for dynamic discrete choice structural models have been applied to study these investment decisions. In this section, we review models and applications that abstract from dynamic oligopoly competition or assume explicitly that firms operate in either competitive or monopolistic markets.

Let $a_{it} \in \mathscr{A} = \{0, 1, ..., J\}$ be the discrete variable that represents the investment decision of firm i at period t. The profit function is:

$$\Pi_{it} = p_{it}\, Y(a_{it}, k_{it}, \omega_{it}; \theta_y) - C(a_{it}, r_{it}; \theta_c) + \varepsilon_{it}(a_{it}) \tag{6.4}$$

p_{it} represents output price. The term $y_{it} = Y(a_{it}, k_{it}, \omega_{it}; \theta_y)$ is a production function that depends on investment a_{it}, predetermined capital stock k_{it}, total factor productivity ω_{it}, and the structural parameters θ_y. The term $C(a_{it}, r_{it}; \theta_c)$ is the investment cost, where r_{it} is the price of new capital and θ_c is a vector of structural parameters.

The vector of variables $\varepsilon_{it} = \{\varepsilon_{it}(a) : a \in \mathscr{A}\}$ represents a component of the investment cost that is unobservable to the researcher. These unobservables have mean zero and typically they are assumed *i.i.d.* across plants and over time. Capital stock k_{it} depreciates at an exogenous rate δ and increases when new investments are made according to the standard transition rule,

$$k_{it+1} \;=\; (1-\delta)\;(k_{it} + a_{it}). \tag{6.5}$$

Total factor productivity ω_{it} and the price of capital r_{it} are exogenous state variables. They evolve over time according to first order Markov processes.

Every period t, the manager observes the state variables $x_{it} \equiv (k_{it}, p_{it}, r_{it}, \omega_{it})$, and ε_{it}, and decides her investment in order to maximize expected and discounted profits $\mathbb{E}_t\left(\sum_{j=0}^{\infty} \beta^j\, \Pi_{i,t+j}\right)$, where $\beta \in (0,1)$ is the discount factor. The optimal decision rule for investment is:

$$a_{it} = \arg\max_{a \in \mathscr{A}} \{\; v(a, x_{it}) + \varepsilon_{it}(a) \;\}, \tag{6.6}$$

where $v(a, x_{it})$ is the *conditional choice value function*, that is the unique solution to a Bellman equation that we describe in detail in section 6.2.2.

Given the distribution of the unobservables $\varepsilon_{it}\}$, the observable exogenous state variables, and the vector of structural parameters $\theta = (\theta_y, \theta_c, \delta)$, this model implies a probability for the observed path of output and investment of a firm, $\{y_{it}, a_{it} : t = 1, 2, ..., T\}$. A standard approach to estimate the parameters of this model is by (Conditional) Maximum Likelihood. Rust (1987) proposed the Nested Fixed Point algorithm (NFXP) for the computation of this estimator. Hotz and Miller (1993) propose a two-step *Conditional Choice Probabilities* (CCP) estimator that avoids computing a solution of the dynamic programming problem. Aguirregabiria and Mira (2002) propose the Nested Pseudo Likelihood algorithm (NPL) that is a recursive extension of the CCP method that returns the maximum likelihood estimates at a substantially lower computing time than NFXP. Section 6.2.3 describes in detail these estimation methods.

Rust (1987) developed this model and applied it to estimate the costs of bus engine replacement and maintenance in the Madison-Wisconsin metropolitan bus company. Since Rust's seminal work, this model has been applied to different datasets on discrete choice investment problems. Das (1992) studies the decision to operate, hold idle, or retire a kiln by plants in the U.S. cement industry. Kennet (1993) analyzes airlines' replacement decisions of aircraft engines and identifies significant changes in the decision rule after the deregulation of the US airline industry in 1978. Rust and Rothwell (1995) consider the refuelling decisions of nuclear power plants. They use their model to evaluate the impact on maintenance costs, profits, and firm behavior of a change in the safety regulation by the US Nuclear Regulatory Commission. Cooper, Haltiwanger, and Power (1999) and Cooper and Haltiwanger (2006) show that this model provides provides better fit and more plausible expanation for the dynamics of investment in US manufacturing plants. Das, Roberts, and Tybout (2007) investigate why the decision to export by Colombian manufacturing plants is very persistent over time. The authors disentangle the effects of sunk costs, prior exporting experience, and serially correlated unobserved heterogeneity. Kasahara (2009) studies the effect of import tariffs on capital investment decisions by Chilean manufacturing plants. He shows that the temporary feature of a tariff increase in the mid-1980s exacerbated firms' zero-investment response. Rota (2004) and Aguirregabiria and Alonso-Borrego (2014) estimate dynamic discrete choice models of labor demand and use them to measure the magnitude of labor adjustment costs for permanent workers in Italy and Spain, respectively, and the effects of labor market reforms. Holmes (2011) studies the geographic expansion of Wal-Mart stores during the period 1971-2005. He estimates a dynamic model of entry and store location by a multi-store firm that incorporates economies of density and cannibalization between Wal-Mart stores. Holmes finds that Wal-Mart obtains large savings in distribution costs by having a dense store network.

6.2.1 Model

Suppose that we have panel data of N plants operating in the same industry with information on output, investment, and capital stock over T periods of time.

$$\text{Data} = \{ \ y_{it}, a_{it}, k_{it} : i = 1, 2, ..., N \text{ and } t = 1, 2, ..., T \ \} \tag{6.7}$$

Suppose that the investment data is characterized by infrequent and lumpy investments. That is, a_{it} contains a large proportion of zeroes (no investment), and when investment is positive the investment-to-capital ratio a_{it}/k_{it} is quite large. For instance, for some industries and samples we can find that the proportion of zeroes is above 60% (even with annual data!) and the average investment-to-capital ratio conditional on positive investment is above 20%.

A possible explanation for this type of dynamics in firms' investment is that there are significant indivisibilities in the purchases of new capital, and fixed or lump-sum costs associated with purchasing and installing new capital. Machine replacement models are models of investment that emphasize the existence of these indivisibilities and lump-sum costs of investment.

The profit function has the structure in equation 6.4, and capital stock follows the transition rule in equation 6.5. A key feature in models of machine replacement is the indivisibility in the investment decision. In the standard machine replacement model, the firm decides between zero investment ($a_{it} = 0$) and the replacement of the old capital ($k_{it} = 0$) by a "new machine" that implies a fixed amount of new capital k^*. Therefore, $a_{it} \in \{0, \ , k^* - k_{it} \ \}$, and the transition rule for capital stock can be written as:

$$k_{it+1} = (1 - \delta) \ [(1 - a_{it}) \ k_{it} + a_{it} \ k^*] \tag{6.8}$$

This implies that capital stock can take a finite number of possible values: $(1-\delta)k^*$ one period after replacement; $(1-\delta)k^*$ one period after replacement; $(1-\delta)^2k^*$ two periods after replacement; ... $(1-\delta)^d k^*$ d periods after replacement. This implies the following linear relationship between the logarithm of capital stock and variable d_{it} that represents the duration (or number of periods) since the last machine replacement by plant i, or the age of capital:

$$\ln(k_{it}) = \delta_0 \ + \delta_1 \ d_{it} \tag{6.9}$$

with $\delta_0 \equiv \ln(k^*)$ and $\delta_1 \equiv \ln(1-\delta)$. Therefore, capital stock k_{it} and age of capital d_{it} are equivalent state variables. The transition rule for the age of capital is by very simple: d_{it+1} is equal to 1 if there was replacement at period t, and it is equal to $d_{it}+1$ if there was not replacement. That is,

$$d_{it+1} \ = \ (1-a_{it}) \ d_{it} \ +1 \tag{6.10}$$

These assumptions on the values of investment and capital seem natural in applications where the investment decision is actually a machine replacement decision, as in the papers by Rust (1987), Das (1992), Kennet (1994), or Rust and Rothwell (1995), among others. However, this framework may be restrictive when we look at less specific investment decisions, such as investment in equipment as in the papers byCooper, Haltiwanger, and Power (1999), Cooper and Haltiwanger (2006), and Kasahara (2009). In these other papers, investment in the data is very lumpy, which is an implication of a model of machine replacement, but firms in the sample use multiple types of equipment, have very different sizes and their capital stocks are very different. These papers consider that investment is either zero or a constant proportion of the installed capital, that is, $a_{it} \in \{0 \ , q \ k_{it}\}$ where q is a constant, for instance, $q = 25\%$. Here we maintain the most standard assumption of machine replacement models.

The production function is:

$$y_{it} = Y\left(a_{it}, k_{it}, \omega_{it}\right) = \exp\left\{\theta_0^y + \omega_{it}\right\} \ \left[k_{it} \ + a_{it} \ (k^* - k_{it})\right]^{\theta_1^y} \tag{6.11}$$

where θ_0^y and θ_1^y are parameters, and ω_{it} is log-TFP. The cost function has two components: the investment or replacement cost (when $a_{it} = 1$, which is equal to $\theta^r \ k^* + \varepsilon_{it}$; and the maintenance cost (when $a_{it} = 0$), that increases with the age of capital (or equivalently, declines with the capital stock) according to the function $\theta_1^m \ k_{it} \ + \theta_2^m \ k_{it}^2$. heterogeneity in replacement costs. Therefore, the profit function is:

$$\Pi_{it} = \left\{ \begin{array}{ll} p_{it} \ \exp\left\{\theta_0^y + \omega_{it}\right\} \ [k_{it}]^{\theta_1^y} - \theta_1^m \ k_{it} + \theta_2^m \ k_{it}^2 & \text{if} \quad a_{it} = 0 \\ \\ p_{it} \ \exp\left\{\theta_0^y + \omega_{it}\right\} \ [k^*]^{\theta_1^y} - \theta^r \ k^* - \varepsilon_{it} & \text{if} \quad a_{it} = 1 \end{array} \right. \tag{6.12}$$

Every period t, the firm observes the state variables $s_{it} = (p_{it}, k_{it}, \omega_{it}, \varepsilon_{it})$ and then it decides its investment in order to maximize its expected value:

$$\mathbb{E}_t \left(\sum_{j=0}^{\infty} \beta^j \ \Pi_{i,t+j}\right) \tag{6.13}$$

where $\beta \in (0,1)$ is the discount factor. The main trade-off in this machine replacement decision is simple. On the one hand, the productivity/efficiency of a machine declines over time and therefore the firm prefers younger machines. However, using younger machines requires frequent replacement and replacing a machine is costly.

The firm has uncertainty about future realizations of output price, log-TFP, and the stochastic component of the replacement cost. To complete the model we have to specify the stochastic processes of these state variables. Output price and log-TFP follow a Markov process with transition density function $f_{p\omega}(p_{it+1}, \omega_{it+1}|p_{it}, \omega_{it})$. The the shock in replacement cost ε_{it} is i.i.d. with density function f_{ε}.

Let $V(s_{it}, \varepsilon_{it})$ be the value function. This value function is the solution to the *Bellman equation*:

$$V(s_{it}) = \max_{a_{it}\in\{0,1\}} \left\{ \Pi(a_{it}, s_{it}) + \beta \int V(s_{it+1}) f_s(s_{it+1}|a_{it}, s_{it}) \, ds_{it+1} \right\} \tag{6.14}$$

where $f_s(s_{it+1}|a_{it}, s_{it})$ is the transition probability of the vector of state variables, that based on the previous assumptions has the following form:

$$f_s(s_{it+1}|a_{it}, s_{it}) = 1\{d_{it+1} = (1-a_{it})\, d_{it} + 1\} \;\; f_{p\omega}(p_{it+1}, \omega_{it+1}|p_{it}, \omega_{it}) \;\; f_{\varepsilon}(\varepsilon_{it+1}) \tag{6.15}$$

where $1\{.\}$ is the indicator function. We can also represent the Bellman equation as:

$$V(s_{it}) = \max\{\; v(0, x_{it}) \; ; \; v(1, x_{it}) - \varepsilon_{it} \;\} \tag{6.16}$$

where $x_{it} = (p_{it}, k_{it}, \omega_{it})$ and $v(0, x_{it})$ and $v(1, x_{it})$ are the *conditional choice value functions*. For $a \in \{0, 1\}$:

$$v(a, x_{it}) \equiv \pi(a_{it}, x_{it}) + \beta \int V(s_{it+1}) \; f_s(s_{it+1}|a_{it}, s_{it}) \; ds_{it+1}, \tag{6.17}$$

and $\pi(a_{it}, x_{it})$ being the part of the profit function that does not depend on $varepsilon_{it}$.

Given the value function, $V(.)$, the optimal decision rule gives us the optimal investment decision as a function of the state variables. We use $\alpha(.)$ to represent the optimal decision rule such that $a_{it} = \alpha(s_{it})$. Given the infinite horizon of the model and the time-homogeneity of the profit function and the transition probability functions, Blackwell's Theorem establishes that the value function and the optimal decision rule are time-invariant (Blackwell, 1965).

We use the vector θ to represent all the parameters in this model. It is convenient to distinguish several components in this vector, $\theta = (\theta_{\pi}, \theta_{f_x}, \theta_{\varepsilon}, \beta)$, where: θ_{π} contains the parameters in the profit function, θ_0^y, θ_1^y, θ_r, θ_1^m, and θ_0^y; θ_{f_x} contains the parameters in the transition probability function of the x state variables; and θ_{ε} contains the parameters in the distribution of ε.

It is straightforward to extend this binary choice model to a multinomial choice model. That is, the investment variable a_{it} can take a $J+1$ possible values $0, 1, 2, ..., J$. The profit function, the transition rule of the capital stock, the value function, the Bellman equation, and the conditional choice value functions have the same structure. We only extend the definition of the stochastic component of the investment cost to include J elements: $\varepsilon_{it} = (\varepsilon_{it}(1).\varepsilon_{it}(2), ..., \varepsilon_{it}(J))$.

6.2.2 Solving the dynamic programming problem

The Bellman equation is a functional equation that maps value functions into value functions. That is, given a value function $V : \mathbb{S} \to \mathbb{R}$, the right-hand side of the Bellman equation provides a new value function, say V'. We represent the right-hand side of the Bellman equation for state s as $\Gamma(V)(s)$, and we use $\Gamma(V)$ to represent the mapping for every value of s, that is, $\Gamma(V) \equiv \{\Gamma(V)(s) : s \in \mathbb{S}\}$. Using this notation, we can present the Bellman equation in a compact form as the following functional equation:

$$V \;=\; \Gamma(V) \tag{6.18}$$

Mapping Γ is denoted the *Bellman operator* or *Bellman mapping*. Therefore, the value function V is a fixed point of the Bellman mapping.

Provided $\beta \in (0,1)$, the Bellman mapping is a contraction. That is, for any two functions V and V', the distance between $\Gamma(V')$ and $\Gamma(V)$ is smaller that the distance between V' and V. More precesily,

$$\|V' - V\| \leq \|\Gamma(V') - \Gamma(V)\| \tag{6.19}$$

where the distance operator $\|.\|$ is the supremum or L_∞ distance, and it is defined as $\|f\|$ $= max_{s\in\mathbb{S}}|f(s)|$.

By the *contraction mapping Theoren* (Banach, 1922), this contraction property of the Bellman operator has two important implications for the solution of the dynamic programming problem. First, the solution to the Bellman equation – that is, the value function V and the corresponding optimal decision rule α – is unique. And second, this solution can be obtained using successive iterations in the Bellman operator. That is, given an arbitrary initial value for the value function, say V^0, at every iteration $n \geq 1$, we obtain a new value function as:

$$V^{n+1} \;=\; \Gamma(V^{n+1}) \tag{6.20}$$

Regardless of the initial value V^0, this algorithm always converges to the unique solution: $\lim_{n\to\infty} V^n = V$.

//////////////////////////// *** REVISADO HASTA AQUI *** ////////////////////////////

*** Explain challenges in solution of DP problem with continuous variables: Curse of dimensionality; infinite dimension; and numerical integration. *** Explain three tricks in the literature to deal with these problems: (1) x variables are discrete; (2) use the integrated Bellman equation (integrated over epsilon); and (3) consider a distribution of epsilon that implies a closed form expression for the integrated Bellman operator. ////////////////////////////

For given values of structural parameters and functions, $\{\alpha_0,\ \alpha_1,\ r,\ f_C,\ \sigma_\varepsilon\}$, and of the individual effects η_i^Y and η_i^C, we can solve the DP problem of firm i by simply using successive approximations to the value function, that is, iterations in the Bellman equation.

In models where some of the state variables are not serially correlated, it is computationally very convenient (and also convenient for the estimation of the model) to define versions of the value function and the Bellman equation that are integrated over the non-serially correlated variables. In our model, ε is not serially correlated. The *integrated value function* of firm i is:

$$\bar{V}_i(K_{it}, C_t) \equiv \int V_i(K_{it}, C_t, \varepsilon_{it})\; d f_\varepsilon(\varepsilon_{it})$$

And the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \int \max\{\ v_i(0; K_{it}, C_t)\ ;\ v_i(1; K_{it}, C_t) - \varepsilon_{it}\ \} \, d f_\varepsilon(\varepsilon_{it})$$

The main advantage of using the integrated value function is that it has a lower dimensionality than the original value function.

Given the extreme value distribution of ε_{it}, the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \sigma_\varepsilon \ \ln \left[\exp\left\{ \frac{v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\} + \exp\left\{ \frac{v_i(1; K_{it}, C_t)}{\sigma_\varepsilon} \right\} \right]$$

where

$$v_i(0; K_{it}, C_t) \ \equiv \ \exp\left\{ \alpha_0 + \eta_i^Y \right\} \ K_{it}^{\alpha_1} + \beta \int \bar{V}_i((1-\delta)K_{it}, C_{t+1}) \ f_C(C_{t+1}|C_t)$$

$$v_i(1; K_{it}, C_t) \ \equiv \ \exp\left\{ \alpha_0 + \eta_i^Y \right\} \ K_{it}^{\alpha_1} - C_t \ I^* - r(K_{it}) - \eta_i^C + \beta \int \bar{V}_i((1-\delta)K^*, C_{t+1}) \ f_C(C_{t+1}|C_t)$$

The optimal decision rule of this dynamic programming (DP) problem is:

$$a_{it} = 1\{\ \varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)\ \}$$

Suppose that the price of new capital, C_t, has a discrete and finite range of variation: $C_t \in \{c^1, c^2, ..., c^L\}$. Then, the value function $\bar{V}_i$ can be represented as a $M \times 1$ vector in the Euclidean space, where $M = T * L$ and the T is the number of possible values for the capital stock. Let $\mathbf{V}_i$ be that vector. The integrated Bellman equation in matrix form is:

$$\mathbf{V}_i = \sigma_\varepsilon \ln \left(\exp\left\{ \frac{\Pi_i(0) + \beta\ \mathbf{F}(0)\ \mathbf{V}_i}{\sigma_\varepsilon} \right\} + \exp\left\{ \frac{\Pi_i(1) + \beta\ \mathbf{F}(1)\ \mathbf{V}_i}{\sigma_\varepsilon} \right\} \right)$$

where $\Pi_i(0)$ and $\Pi_i(1)$ are the $M \times 1$ vectors of one-period profits when $a_{it} = 0$ and $a_{it} = 1$, respectively. $\mathbf{F}(0)$ and $\mathbf{F}(1)$ are $M \times M$ transition probability matrices of (K_{it}, C_t) conditional on $a_{it} = 0$ and $a_{it} = 1$, respectively.

Given this equation, the vector $\mathbf{V}_i$ can be obtained by using value function iterations in the Bellman equation. Let $\mathbf{V}_i^0$ be an arbitrary initial value for the vector $\mathbf{V}_i$. For instance, $\mathbf{V}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at iteration $k = 1, 2, ...$ we obtain:

$$\mathbf{V}_i^k = \sigma_\varepsilon \ln \left(\exp\left\{ \frac{\Pi_i(0) + \beta\ \mathbf{F}(0)\ \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} + \exp\left\{ \frac{\Pi_i(1) + \beta\ \mathbf{F}(1)\ \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} \right)$$

Since the (integrated) Bellman equation is a contraction mapping, this algorithm always converges (regardless of the initial $\mathbf{V}_i^0$) and it converges to the unique fixed point. Exact convergence requires infinite iterations. Therefore, we stop the algorithm when the distance (for instance, Euclidean distance) between $\mathbf{V}_i^k$ and $\mathbf{V}_i^{k-1}$ is smaller than some small constant, for instance, 10^{-6}.

An alternative algorithm to solve the DP problem is the **Policy Iteration algorithm**. Define the *Conditional Choice Probability (CCP) function* $P_i(K_{it}, C_t)$ as:

$$\begin{aligned} P_i(K_{it}, C_t) \ &\equiv \ \Pr(\ \varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)\ ) \\ &= \ \frac{\exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}} \end{aligned}$$

Given that (K_{it}, C_t) are discrete variables, we can describe the CCP function P_i as a $M \times 1$ vector of probabilities $\mathbf{P}_i$. The expression for the CCP in vector form is:

$$\mathbf{P}_i = \frac{\exp\left\{\dfrac{\Pi_i(1)-\Pi_i(0) + \beta\ [\mathbf{F}(1) - \mathbf{F}(0)]\ \mathbf{V}_i}{\sigma_\varepsilon}\right\}}{1 + \exp\left\{\dfrac{\Pi_i(1)-\Pi_i(0) + \beta\ [\mathbf{F}(1) - \mathbf{F}(0)]\ \mathbf{V}_i}{\sigma_\varepsilon}\right\}}$$

Suppose that the firm behaves according to the probabilities in $\mathbf{P}_i$. Let $\mathbf{V}_i^{\mathbf{P}}$ be the vector of values if the firm behaves according to $\mathbf{P}$. That is $\mathbf{V}_i^{\mathbf{P}}$ is the expected discounted sum of current and future profits if the firm behaves according to $\mathbf{P}_i$. Ignoring for the moment the expected future $\varepsilon' s$, we have that:

$$\mathbf{V}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \left[\Pi_i(0) + \beta\ \mathbf{F}(0)\mathbf{V}_i^{\mathbf{P}}\right] + \mathbf{P}_i * \left[\Pi_i(1) + \beta\ \mathbf{F}(1)\mathbf{V}_i^{\mathbf{P}}\right]$$

And solving for $\mathbf{V}_i^{\mathbf{P}}$:

$$\mathbf{V}_i^{\mathbf{P}} = \left(I - \beta\ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} ((1 - \mathbf{P}_i) * \Pi_i(0) + \mathbf{P}_i * \Pi_i(1))$$

where $\mathbf{F}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \mathbf{F}(0) + \mathbf{P}_i * \mathbf{F}(1)$.

Taking into account this expression for $\mathbf{V}_i^{\mathbf{P}}$, we have that the optimal CCP $\mathbf{P}_i$ is such that:

$$\mathbf{P}_i = \frac{\exp\left\{\dfrac{\tilde{\Pi}_i + \beta\ \tilde{\mathbf{F}}\ \left(I - \beta\ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} ((1 - \mathbf{P}_i) * \Pi_i(0) + \mathbf{P}_i * \Pi_i(1))}{\sigma_\varepsilon}\right\}}{1 + \exp\left\{\dfrac{\tilde{\Pi}_i + \beta\ \tilde{\mathbf{F}}\ \left(I - \beta\ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} ((1 - \mathbf{P}_i) * \Pi_i(0) + \mathbf{P}_i * \Pi_i(1))}{\sigma_\varepsilon}\right\}}$$

where $\tilde{\Pi}_i \equiv \Pi_{\mathbf{i}}(1) - \Pi_i(0)$, and $\tilde{\mathbf{F}} \equiv \mathbf{F}(1) - \mathbf{F}(0)$. This equation defines a fixed point mapping in $\mathbf{P}_i$. This fixed point mapping is called the Policy Iteration mapping. This is also a contraction mapping. The optimal $\mathbf{P}_i$ is its unique fixed point.

Therefore we compute $\mathbf{P}_i$ by iterating in this mapping. Let $\mathbf{P}_i^0$ be an arbitrary initial value for the vector $\mathbf{P}_i$. For instance, $\mathbf{P}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at each iteration $k = 1, 2, ...$ we do the following two-step procedure:

Valuation step:

$$\mathbf{V}_i^k = \left(I - \beta\ \mathbf{F}^{\mathbf{P}_i^{k-1}}\right)^{-1} \left((1 - \mathbf{P}_i^{k-1}) * \Pi_i(0) + \mathbf{P}_i^{k-1} * \Pi_i(1)\right)$$

Policy step:

$$\mathbf{P}_i^k = \frac{\exp\left\{\dfrac{\tilde{\Pi}_i + \beta\ \tilde{\mathbf{F}}\ \mathbf{V}_i^k}{\sigma_\varepsilon}\right\}}{1 + \exp\left\{\dfrac{\tilde{\Pi}_i + \beta\ \tilde{\mathbf{F}}\ \mathbf{V}_i^k}{\sigma_\varepsilon}\right\}}$$

Policy iterations are more costly than Value function iterations (especially due to the matrix inversion in the valuation step). However, the policy iteration algorithm requires a much lower number of iterations, especially with β is close to one. Rust (1987,1994) proposes a hybrid algorithm: start with a few value function iterations and then switch to policy iterations.

6.2.3 Estimation

The primitives of the model are: (a) The parameters in the production function; (b) the replacement cost function r; (c) the probability distribution of firm heterogeneity F_η; (d) the dispersion parameter σ_ε; and (e) the discount factor β. Let θ represent the vector of structural parameters. We are interested in the estimation of θ.

Here we describe the Maximum Likelihood estimation of these parameters. Conditional on the observed history of the price of capital and on the initial condition for the capital stock, we have that:

$$\Pr(Data \mid C, K_{i1}, \theta) \quad = \quad \prod_{i=1}^{N} \Pr(a_{i1}, Y_{i1}, ..., a_{iT}, Y_{iT} \mid C, K_{i1}, \theta)$$

The probability $\Pr(a_{i1}, Y_{i1}, ..., a_{iT}, Y_{iT} \mid C, K_{i1}, \theta)$ is the contribution of firm i to the likelihood function. Conditional on the individual heterogeneity, $\eta_i \equiv (\eta_i^Y, \eta_i^C)$, we have that:

$$\begin{aligned}
\Pr(a_{i1}, Y_{i1}, ..., a_{iT}, Y_{iT} \mid C, K_{i1}, \eta_i, \theta) &= \prod_{t=1}^{T} \Pr(a_{it}, Y_{it} \mid C_t, K_{it}, \eta_i, \theta) \\
&= \prod_{t=1}^{T} \Pr(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta) \Pr(a_{it} \mid C_t, K_{it}, \eta_i, \theta)
\end{aligned}$$

where $\Pr(a_{it} \mid C_t, K_{it}, \eta_i, \theta)$ is the CCP function:

$$\Pr(a_{it} \mid C_t, K_{it}, \eta_i, \theta) \quad = \quad P_i(K_{it}, C_t, \theta)^{a_{it}} \; [1 - P_i(K_{it}, C_t, \theta)]^{1-a_{it}}$$

and $\Pr(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta)$ comes from the production function, $Y_{it} = \exp\{\alpha_0 + \eta_i^Y\}$ $[(1-a_{it}) K_{it} + a_{it} K^*]^{\alpha_1}$. In logarithms, the production function is:

$$\ln Y_{it} = \alpha_0 + \alpha_1 \; (1 - a_{it}) \ln K_{it} + \kappa \, a_{it} + \eta_i^Y + e_{it}$$

where κ is a parameter that represents $\alpha_1 \ln K^*$, and e_{it} is a measurement error in output, that we assume i.i.d. $N(0, \sigma_e^2)$, and independent of ε_{it}. Therefore,

$$\Pr(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta) \quad = \quad \phi\left(\frac{\ln Y_{it} - \alpha_0 - \alpha_1 \; (1 - a_{it}) \ln K_{it} - \kappa \, a_{it} - \eta_i^Y}{\sigma_e}\right)$$

where $\phi(.)$ is the density function of the standard normal distribution.

Putting all these pieces together, we have that the log-likelihood function of the model is $\ell(\theta) = \sum_{i=1}^{N} \ell_i(\theta)$ where $\ell_i(\theta) \equiv \ln \Pr(a_{i1}, Y_{i1}, ..., a_{iT}, Y_{iT} \mid C, K_{i1}, \theta)$ such that:

$$\ell_i(\theta) = \ln\left(\sum_{\eta \in \Omega} F_\eta(\eta) \left[\begin{array}{c} \prod_{t=1}^{T} \phi\left(\frac{\ln Y_{it} - \alpha_0 - \alpha_1 \; (1-a_{it}) \ln K_{it} - \kappa \, a_{it} - \eta^Y}{\sigma_e}\right) \\ \\ P_i(K_{it}, C_t, \eta, \theta)^{a_{it}} \; [1 - P_i(K_{it}, C_t, \eta, \theta)]^{1-a_{it}} \end{array}\right]\right) \tag{6.21}$$

Given this likelihood, we can estimate the parameter vector using Maximum Likelihood (ML).

The NFXP algorithm is a gradient iterative search method to obtain the MLE of the structural parameters.

This algorithm nests a BHHH method (outer algorithm), that searches for a root of the likelihood equations, with a value function or policy iteration method (inner algorithm), that solves the DP problem for each trial value of the structural parameters. The algorithm is initialized with an arbitrary vector $\hat{\theta}_0$.

A BHHH iteration is defined as:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left(\sum_{i=1}^{N} \nabla \ell_i(\hat{\theta}_k) \nabla \ell_i(\hat{\theta}_k)' \right)^{-1} \left(\sum_{i=1}^{N} \nabla \ell_i(\hat{\theta}_k) \right)$$

where $\nabla \ell_i(\theta)$ is the gradient in θ of the log-likelihood function for individual i. In a partial likelihood context, the score $\nabla \ell_i(\theta)$ is:

$$\nabla \ell_i(\theta) = \sum_{t=1}^{T_i} \nabla \log P(a_{it}|x_{it}, \theta)$$

To obtain this score we have to solve the DP problem.

In our machine replacement model:

$$\ell(\theta) \quad = \quad \sum_{i=1}^{N} \sum_{t=1}^{T_i} a_{it} \ \log P(x_{it}, \theta) + (1 - a_{it}) \ \log(1 - P(x_{it}, \theta))$$

with:

$$\mathbf{P}(\theta) \quad = \quad F_{\tilde{\varepsilon}} \begin{pmatrix} [\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta \ \mathbf{F}_x(0)\mathbf{V}(\theta)] \\ -[\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta \ \mathbf{F}_x(1)\mathbf{V}(\theta)] \end{pmatrix}$$

The NFXP algorithm works as follows. At each iteration we can distinguish three main tasks or steps.

Step 1: Inner iteration: DP solution. Given $\hat{\theta}_0$, we obtain the vector $\bar{\mathbf{V}}(\hat{\theta}_0)$ by using either successive iterations or policy iterations.

Step 2: Construction of scores. Then, given $\hat{\theta}_0$ and $\bar{\mathbf{V}}(\hat{\theta}_0)$ we construct the choice probabilities

$$\mathbf{P}(\hat{\theta}_0) \quad = \quad F_{\tilde{\varepsilon}} \begin{pmatrix} \left[\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta \ \mathbf{F}_x(0)\mathbf{V}(\hat{\theta}_0)\right] \\ -\left[\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta \ \mathbf{F}_x(1)\mathbf{V}(\hat{\theta}_0)\right] \end{pmatrix}$$

the Jacobian $\dfrac{\partial \bar{\mathbf{V}}(\hat{\theta}_0)'}{\partial \theta}$ and the scores $\nabla \ell_i(\hat{\theta}_0)$

Step 3: BHHH iteration. We use the scores $\nabla \ell_i(\hat{\theta}_0)$ to make a new BHHH iteration in order to obtain $\hat{\theta}_1$.

$$\hat{\theta}_1 = \hat{\theta}_0 + \left(\sum_{i=1}^{N} \nabla \ell_i(\hat{\theta}_0) \nabla \ell_i(\hat{\theta}_0)' \right) \left(\sum_{i=1}^{N} \nabla \ell_i(\hat{\theta}_0) \right)$$

Then, we replace $\hat{\theta}_0$ by $\hat{\theta}_1$ and go back to step 1.
* We repeat steps 1 to 3 until convergence: that is, until the distance between $\hat{\theta}_1$ and $\hat{\theta}_0$ is smaller than a pre-specified convergence constant.

The main advantages of the NFXP algorithm are its conceptual simplicity and, more importantly, that it provides the MLE which is the most efficient estimator asymptotically under the assumptions of the model.

The main limitation of this algorithm is its computational cost. In particular, the DP problem should be solved for each trial value of the structural parameters.

6.3 Patent Renewal Models

• **What is the value of a patent? How to measure it?**
• The valuation of patents is very important for: merger and acquisition decisions; using patents as collateral for loans; value of innovations; value of patent protection.
• Very few patents are traded, and there is substantial selection. We cannot use a "hedonic" approach.
• The number of citations of a patent is a very imperfect measure of patent value.
• Multiple patents are used in the production of multiple products, and in generating new patents. A "production function approach" seems also unfeasible.

6.3.1 Pakes (1986)

• Pakes (1986) proposes to use information on patent renewal fees together with a *Revealed Preference approach* to estimate the value of a patent.
• Every year, a patent holder should pay a renewal fee to keep her patent.
• If the patent holder decides to renew, it is because her expected value of holding the patent is greater than the renewal fee (that is publicly known).
• Therefore, observed decisions on patent renewal / non renewal contain information on the value of a patent.

Model: Basic Framework
• Consider a patent holder who has to decide whether to renew her patent or not. We index patents by i.
• This decision should be taken at ages $t = 1, 2, ..., T$ where $T < \infty$ is the regulated term of a patent (for instance, 20 years in US, Europe, or Canada).
• Patent regulation also establishes a sequence of **Maintenance Fees** $\{c_t : t = 1, 2, ..., T\}$. This sequence of renewal fees is deterministic such that a patent owner knows with certainty future renewal fees.
• The schedule $\{c_t : t = 1, 2, ..., T\}$ is typically increasing in patent age t, and it may increase from a few hundred dollars to a few thousand dollars.

• A patent generates a sequence of profits $\{\pi_{it} : t = 1, 2, ..., T\}$.
• At age t, a patent holder knows current profit π_{it} but has uncertainty about future profits $\pi_{i,t+1}$, $\pi_{i,t+2}$, ...
• The evolution of profits depends on the following factors:
(1) the initial "quality" of the idea/patent;

(2) innovations (new patents) which are substitutes for the patent and therefore, depreciate its value or even make it obsolete;

(3) innovations (new patents) which are complements of the patent and therefore, increase its value.

Stochastic process of patent profits

• Pakes proposes the following stochastic process for profits, that tries to capture the three forces mentioned above.

• A patent profit at the first period is a random draw from a log-normal distribution with parameters μ_1 and σ_1:

$$\ln(\pi_{i1}) \sim N(\mu_1, \sigma_1^2)$$

• After the first year, profit evolves according to the following formula:

$$\pi_{i,t+1} = \tau_{i,t+1} \ \max\left\{ \ \delta \ \pi_{it} \ ; \ \xi_{i,t+1} \right\}$$

• $\delta \in (0,1)$ is the depreciation rate. In the absence of unexpected shocks, the value of the patent depreciates according to the rule: $\pi_{i,t+1} = \delta \ \pi_{it}$.

• $\tau_{i,t+1} \in \{0,1\}$ is a binary variable that represents the patent becoming obsolete (that is, zero value) due to competing innovations. The probability of this event is a decreasing function of profit at the previous year:

$$\Pr(\tau_{i,t+1} = 0 \mid \pi_{it}, t) = \exp\{-\lambda \ \pi_{it}\}$$

• The larger the profit of the patent at age t, the smaller the probability of it becoming obsolete.

• Variable $\xi_{i,t+1}$ represents innovations which are complements of the patent and increase its profitability.

• $\xi_{i,t+1}$ has an exponential distribution with mean γ and standard deviation $\phi^t \sigma$:

$$p(\xi_{i,t+1} \mid \pi_{it}, t) = \frac{1}{\phi^t \sigma} \exp\left\{ -\frac{\gamma + \xi_{i,t+1}}{\phi^t \sigma} \right\}$$

• If $\phi < 1$, the variance of $\xi_{i,t+1}$ declines over time (and the $\mathbb{E}(\max\left\{ \ x \ ; \ \xi_{i,t+1} \right\})$ value declines as well).

• If $\phi > 1$, the variance of $\xi_{i,t+1}$ increases over time (and the $\mathbb{E}(\max\left\{ \ x \ ; \ \xi_{i,t+1} \right\})$ value increases as well).

• Under this specification, profits $\{\pi_{it}\}$ follow a non-homogeneous Markov process with initial density $\pi_{i1} \sim \ln N(\mu_1, \sigma_1^2)$, and transition density function:

$$f_\varepsilon\left(\pi_{it+1} | \pi_{it}, t\right) = \begin{cases} \exp\{-\lambda \ \pi_{it}\} & \text{if} \quad \pi_{it+1} = 0 \\ \\ \Pr\left(\xi_{it+1} < \delta \pi_{it} \mid \pi_{it}, t\right) & \text{if} \quad \pi_{it+1} = \delta \pi_{it} \\ \\ \dfrac{1}{\phi^t \sigma} \exp\left\{ -\dfrac{\gamma + \pi_{it+1}}{\phi^t \sigma} \right\} & \text{if} \quad \pi_{it+1} > \delta \pi_{it} \end{cases}$$

• The vector of structural parameters is $\theta = (\lambda, \delta, \gamma, \phi, \sigma, \mu_1, \sigma_1)$.

Model: Dynamic Decision Model

• $V_t(\pi)$ is the value of an active patent of age t and current profit π.
• Let $a_{it} \in \{0,1\}$ be the decision variable that represents the event "the patent owner decides to renew the patent at age t".
• The value function is implicitly defined by the Bellman equation:

$$V_t(\pi_{it}) = \max\left\{ 0 \; ; \; \pi_{it} - c_t + \beta \int V_{t+1}(\pi_{i,t+1}) \; f_\varepsilon(d\pi_{i,t+1} \mid \pi_{it}, t) \right\}$$

with $V_t(\pi_{it}) = 0$ for any $t \geq T+1$.
• The value of not renewal ($a_{it} = 0$) is zero. The value of renewal ($a_{it} = 1$) is the current profit $\pi_{it} - c_t$ plus the expected and discounted future value.

Model: Solution (Backwards induction)
• We can use backwards induction to solve for the sequence of value functions $\{V_t\}$ and optimal decision rules $\{\alpha_t\}$:
• Starting at age $t = T$, for any profit π:

$$V_T(\pi) = \max\{\, 0 \; ; \; \pi - c_T \}$$

and

$$\alpha_T(\pi) = 1\{\, \pi - c_T \geq 0 \,\}$$

• Then, for age $t < T$, and for any profit π:

$$V_t(\pi) = \max\left\{ 0 \; ; \; \pi - c_t + \beta \int V_{t+1}(\pi') \; f_\varepsilon(d\pi'|\pi, t) \right\}$$

and

$$\alpha_t(\pi) = 1\left\{ \pi - c_t + \beta \int V_{t+1}(\pi') \; f_\varepsilon(d\pi'|\pi, t) \geq 0 \right\}$$

Solution - A useful result.
• Given the form of $f_\varepsilon(\pi'|\pi,t)$, the future and discounted expected value, $\beta \int V_{t+1}(\pi')$ $f_\varepsilon(d\pi'|\pi,t)$, is increasing in current π.
• This implies that the solution of the DP problem can be described as a **sequence of threshold values for profits** $\{\pi_t^* : t = 1,2,...,T\}$ such that the optimal decision rule is:

$$\alpha_t(\pi) = 1\{\, \pi \geq \pi_t^* \,\}$$

• π_t^* is the level of current profits that leaves the owner indifferent between renewing the patent or not: $V_t(\pi_t^*) = 0$.
• These threshold values are obtained using backwards induction:
• At period $t = T$:

$$\pi_T^* = c_T$$

• At period $t < T$, π_t^* is the unique solution to the equation:

$$\pi_t^* - c_t + \mathbb{E}\left(\sum_{s=t+1}^{T} \beta^{s-t} \; \max\{\, 0 \; ; \; \pi_{t+1} - \pi_{t+1}^* \,\} \mid \pi_t = \pi_t^* \right) = 0$$

• Solving for a sequence of threshold values is much simpler that solving for a sequence of value functions.

Data

• Sample of N patents with complete (uncensored) durations $\{d_i : i = 1, 2, ...N\}$, where $d_i \in \{1, 2, ..., T + 1\}$ is patent i's duration or age at its last renewal period.
• The information in this sample can be summarized by the empirical distribution of $\{d_i\}$:

$$\widehat{p}(t) = \frac{1}{N} \sum_{i=1}^{N} 1\{d_i = t\}$$

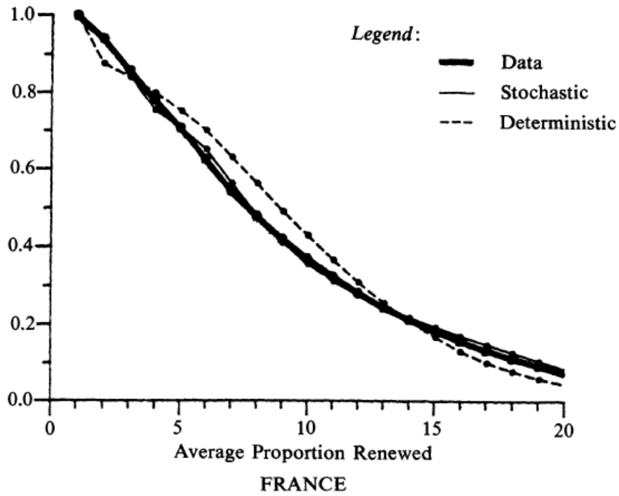

Figure 6.2: Pakes (1986)- Empirical Distribution - France

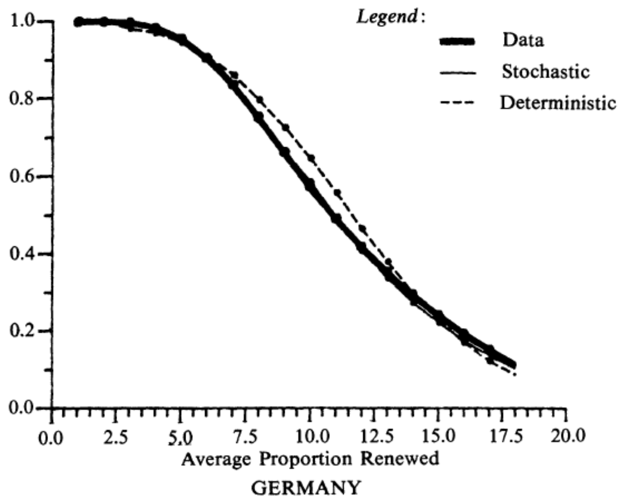

Figure 6.3: Pakes (1986)- Empirical Distribution - Germany

Estimation: Likelihood • The log-likelihood function of this model and data is:

$$
\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{N}\sum_{t=1}^{T+1} 1\{d_i = t\} \ \ln \Pr(d_i = t|\theta) \\
&= N \sum_{t=1}^{T+1} \widehat{p}(t) \ \ln P(t|\theta)
\end{aligned}
$$

where:

$$
\begin{aligned}
P(t|\theta) &= \Pr\left(\pi_s \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_t < \pi_t^* \mid \theta\right) \\
&= \int_{\pi_1^*}^{\infty} \ldots \int_{\pi_{t-1}^*}^{\infty} \int_{0}^{\pi_t^*} dF\left(\pi_1, \ldots, \pi_{t-1}, \pi_t\right)
\end{aligned}
$$

• Computing $P(t|\theta)$ involves solving an integral of dimension t. For t greater than 4 or 5, it is computationally very costly to obtain the exact value of these probabilities. Instead, we approximate these probabilities using Monte Carlo simulation.

Estimation: Simulation of Probabilities

• For a given value of θ, let $\{\pi_t^{sim}(\theta) : t = 1,2,...,T\}$ be a simulated history of profits for patent i.
• Suppose that, for a given value of θ, we simulate R **independent** profit histories. Let $\{\pi_{rt}^{sim}(\theta) : t = 1,2,...,T;\, r = 1,2,...,R\}$ be these histories.
• Then, we can approximate the probability $P(t|\theta)$ using the following simulator:

$$\tilde{P}_R(t|\theta) = \frac{1}{R}\sum_{r=1}^{R} 1\{\pi_{rs}^{sim}(\theta) \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_{rt}^{sim} < \pi_t^*\}$$

• $\tilde{P}_R(t|\theta)$ is a *raw frequency simulator*. It has the following properties (Note that these are properties of a simulator, not of an estimator. $\tilde{P}_R(t|\theta)$ does not depend on the data).
(1) Unbiased: $\mathbb{E}\left(\tilde{P}_R(t|\theta)\right) = P(t|\theta)$
(2) $Var(\tilde{P}_R(t|\theta)) = P(t|\theta)(1-P(t|\theta))/R$
(3) Consistent as $R \to \infty$.
• It is possible to obtain better simulators (with lower variance) by using importance-sampling simulation. This is relevant because the bias and variance of simulated-based estimators depend on the variance (and bias) of the simulator.
• Furthermore, when $P(t|\theta)$ is small, the simulator $\tilde{P}_R(t|\theta)$ can be zero even when R is large, and this creates problems for ML estimation.
• A simple solution to this problem is to consider the following simulator which is based on the raw-frequency simulated probabilities $\tilde{P}_R(1|\theta)$, $\tilde{P}_R(2|\theta)$, ... $\tilde{P}_R(T+1|\theta)$:

$$P_R^*(t|\theta) = \frac{\exp\left\{\dfrac{\tilde{P}_R(t|\theta)}{\eta}\right\}}{\sum_{s=1}^{T+1}\exp\left\{\dfrac{\tilde{P}_R(s|\theta)}{\eta}\right\}}$$

where $\eta > 0$ is a smoothing parameter.
• The simulator P_R^* is biased. However, if $\eta \to 0$ as $R \to \infty$, then P_R^* is consistent, it has lower variance than $\tilde{P}_R$, and it is always strictly positive.

Simulation-Based Estimation.

• The estimator of θ (Simulated Method of Moments estimator) is the value that solves the system of T equations: for $t = 1,2,...T$:

$$\frac{1}{N}\sum_{i=1}^{N}\left[1\{d_i = t\} - \tilde{P}_{R,i}(t|\theta)\right] = 0$$

where the subindex i in the simulator $\tilde{P}_{R,i}(t|\theta)$ indicates that for each patent i in the sample we draw R independent histories and compute independent simulators.
• **Effect of simulation error.** Note that $\tilde{P}_{R,i}(t|\theta)$ is unbiased such that $\tilde{P}_{R,i}(t|\theta) = P(t|\theta) + e_i(t,\theta)$, where $e_i(t,\theta)$ is the simulation error. Since the simulation errors are

independent random draws:

$$\frac{1}{N}\sum_{i=1}^{N} e_i(t,\theta) \rightarrow_p 0 \quad \text{and} \quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_i(t,\theta) \rightarrow_d N(0,V_R)$$

The estimator is consistent an asymptotically normal for any R. The variance of the estimator declines with R.

Identification

• Since there are only 20 different values for the renewal fees $\{c_t\}$ we can at most identify 20 different points in the probability distribution of patent values.
• The estimated distribution at other points is the result of interpolation or extrapolation based on the functional form assumptions on the stochastic process for profits.
• It is important to note that the identification of the distribution of patent values is NOT up to scale but in dollar values.
• For a given patent of age t, all we can say is that: if $a_{it} = 0$, then $V_{it} < V(\pi_t^*)$; and if $a_{it} = 1$, then $V_{it} \geq V(\pi_t^*)$.

Empirical Questions • The estimated model can be used to address important empirical questions.
• **Valuation of the stock of patents.** Pakes uses the estimated model to obtain the value of the stock of patents in a country.
• According to the estimated model, the value of the stock of patents in 1963 was \$315 million in France, \$385 million in UK, and \$511 million in Germany.
• Combining these figures with data on R&D investments in these countries, Pakes calculates rates of return of 15.6%, 11.0% and 13.8%, which look quite reasonable.

Empirical Questions.
• **Factual policies.** The estimated model shows that a very important part of the observed between-country differences in patent renewal can be explained by differences in policy parameters (that is, renewal fees and maximum length).
• **Counterfactual policy experiments.** The estimated model can be used to evaluate the effects of policy changes (in renewal fees and/or in maximum length) that are not observed in the data.

6.3.2 lanjow_1999 (lanjow_1999)

Estimates the private value of patent protection for four technology areas—computers, textiles, combustion engines, and pharmaceuticals - using new patent data for West Germany, 1953-1988. The model takes into account that patentees must pay not only renewal fees to keep their patents but also legal expenses to enforce them. The dynamic structural model takes into account the potential need to prosecute infringement. Results show that the aggregate value of protection generated per year is in the order of 10% of related R&D expenditure.

6.3.3 Trade of patents: Serrano (2018)

The sale of patents is an incentive to invest in R&D, especially for small firms. This market can generate social gains by reallocating patent rights from innovators to firms that may be more effective in using, commercializing, or enforcing these rights. There are also potential social costs, if the acquiring firms can exercise more market power. Serrano (2018) investigates the value of trading patents by estimating a structural model that includes renewal and trading decisions.

Data: Panel of patents granted to U.S. small firms (no more than 500 employees) in the period 1988-1997 (15% of patents granted to firms). In the U.S. patent system, the patent holder needs to pay renewal fees to maintain the patent only at ages 5, 9, and 13 years. Fee increases with age: $c_{13} > c_9 > c_5$. **serrano_2000** (**serrano_2000**) constructs the dataset with renewals and transfers/sales. Working sample: 54,840 patents from 10 granting cohorts (1988 to 1997), followed from granting period until 2001 or until non-renewal.

Renewal and trading frequencies. Probability that a patent is traded (between renewal dates): - higher if previously untraded. - decreases with age. Probability of patent expiration (at renewal dates): - lower for previously traded. - increase over time.

Renewal and trading frequencies

Model: Key features. The transfer/sale of a patent involves a transaction cost. This transaction cost creates a selection effect: patents with higher per period returns are more likely to be traded. This selection effect explains the observed pattern that previously traded patents are: - more likely to be traded; - less likely to expire.

Returns. At age t, a patent has: - an **internal return** for the current patent owner, x_t; - a potential **external return** for the best alternative user, y_t. There is an "improvement factor", g_t^e, that relates external and internal returns: $y_t = g_t^e\ x_t$, where g_t^e is *i.i.d.* with a truncated (at zero) exponential distribution: $\gamma^e \equiv \Pr(g_t^e = 0)$, and σ^e is the mean of the exponential. Initial (internal) returns: $\log(x_1) \sim N(\mu, \sigma_R^2)$. Next period returns:

$$x_{t+1} = \begin{cases} g_t^i\ x_t & \text{if not traded at age } t \\ g_t^i\ y_t & \text{if traded at age } t \end{cases}$$

g_t^i is a random variable with a truncated (at zero) exponential distribution: $\gamma^i \equiv \Pr(g_t^i = 0)$, and σ_t^i is the mean of this exponential, and $\sigma_t^i = \phi^t\ \sigma_0^i$, with $\phi \in (0,1)$. This implies that x_{t+1} follows a first order Markov process. Remember that there is a lump-sum transaction cost, τ. It is assumed to be paid by the buyer.

Model: Renewal and Sale decisions. Let $V_t(x_t, y_t)$ be the value of a patent with age t, current internal and external returns x_t and y_t, resp.

$$V_t(x_t, y_t) = \max\left\{0,\ V_t^K(x_t, y_t),\ V_t^S(x_t, y_t)\right\}$$

$V_t^K(x_t, y_t) =$ value of keeping; $V_t^S(x_t, y_t) =$ value of selling. And for $t \leq T = 17$:

$$\begin{aligned} V_t^K(x_t, y_t) &= x_t - c_t + \beta\ \mathbb{E}\left[V_{t+1}(x_{t+1}, y_{t+1}) \mid x_t, y_t, a_t = K\right] \\ V_t^S(x_t, y_t) &= x_t - c_t - \tau + \beta\ \mathbb{E}\left[V_{t+1}(x_{t+1}, y_{t+1}) \mid x_t, y_t, a_t = S\right] \end{aligned}$$

with $V_{T+1}^K = V_{T+1}^S = 0$.

Model: Optimal decision rule.

Lemma 1: $V_t(x_t, y_t)$ is weakly increasing in x_t and y_t, and weakly decreasing in t. Proposition 1. There are two threshold values: $x_t^*(\theta)$ that depends on age and structural

A-1: Percentage of Active Small Business Patents Traded and

Age	All	Not Previously Traded	Previously Traded (Years since last trade)	
			Any Year	One year
A. Probability that an active patent is traded				
2	2.99	2.85	7.47	7.47
7	2.81	2.46	4.79	6.63
11	2.51	2.13	3.77	2.55
B. Probability that an active patents is allowed to expire				
5	17.2	17.7	12.7	6.2
9	25.6	26.6	21.4	11.6
13	25.5	26.6	22.5	14.1

Figure 6.4: Serrano - Table of Frequencies

parameters, and $g_t^*(x,\theta)$, that depends on age, internal return, and parameters, such that the optimal decision rule a_t is:

$$a_t = \begin{cases} S & \text{if} \quad g_t^e \geq g_t^*(x_t,\theta) \\ K & \text{if} \quad g_t^e < g_t^*(x_t,\theta) \ \text{ and } \ x_t \geq x_t^*(\theta) \\ 0 & \text{if} \quad g_t^e < g_t^*(x_t,\theta) \ \text{ and } \ x_t < x_t^*(\theta) \end{cases}$$

Identification and Estimation. Method: Simulated method of moments. Moments describing the history of trading and renewal decisions of patent owners. (1) probability that an active patent is traded at different ages conditional on having been previously traded, and conditional on not having been previously traded. (2) probability that an active patent is allowed to expire at different renewal dates conditional on having been

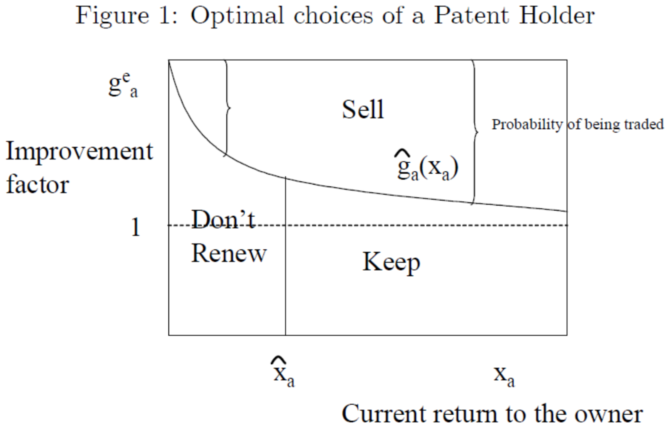

Figure 6.5: Serrano - Decision Rule

previously traded, and conditional on not having been previously traded. A total of 186 moments.

Parameter estimates. Transaction cost: \$5,850, about one-third of the average return at age 1 (8% of the average value at age 1). On average, internal growth of returns is greater than external.

Evaluating the value of the market for patents. The possibility of trading patents has two types of effects on the value of the pool of patents: - a direct causal effect due to the reallocation to an owner with higher returns; - a selection effect, through the renewal decisions (renewal decision is different with and without the possibility of trading).

Serrano measures these two sources of value.

Evaluating the value of the market for patents. (1) Total effect on the value of patents: - 50% of the total value of patents. - Only 23% of patents are sold, but the value of a

Table 1: Parameter Estimates

Description (Parameter)	Estimate[a]
A. Patent initial returns	
Mean parameter of the Lognormal Initial Distribution (μ)	8.4179 (4.0*10^{-2})
Std. Deviation parameter of the Lognormal Initial Distribution (σ_R)	1.6911 (1.2*10^{-2})
B. Internal growth of returns	
Depreciation factor (δ)	0.8917 (5.2*10^{-3})
Not obsolescence (γ^i)	0.9673 (6.5*10^{-3})
Internal Growth of Returns (σ^i)	0.4450 (3.6*10^{-3})
Upside opportunities (ϕ)	0.5941 (6.1*10^{-3})
C. Market for patents and transaction costs	
Transaction cost (τ)	5,850.1 (50.49)
Mean External Growth of Returns (σ^e)	0.3745 (2.9*10^{-3})
Proportion of unsuccessful transfers (γ^e)	0.0385 (4*10^{-4})
Random transfers (ε)	0.0059 (4*10^{-4})
Size of sample	54,840
Simulations in the estimation	164,520
MSE[b]	3.784*10^{-4}

Figure 6.6: Serrano - Parameter Estimates

traded patent is 3 times higher than untraded patent ($173,668 vs. $54,960). (2) Direct gains from trade (from reallocation) - accounts for 10% of the total value of the traded patents. - The distribution of the gains from trade is very skewed.

Counterfactual: Reducing transaction cost . Lowering transaction cost by 50% (from $5,850 to $2,925) raises the proportion of patents traded by 6 percentage points: from 23.1% to 29.6%. It also boosts the gains from trade (reallocation) by an additional 8.7%. Overall, it increases the total value of the patent market by 3%.

6.4 Dynamic pricing

Retail firms selling the same product and operating in the same narrowly defined geographic market can charge prices that differ by significant amounts. Cross-sectional dispersion of prices has been well established in different retail markets such as gas stations or supermarkets, among others. Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in the U.S. (see **hosken_reiffen_2004** ,**hosken_reiffen_2004**, **nakamura_ steinsson_2008** ,**nakamura_ steinsson_2008**, or **midrigan_2011** ,**midrigan_2011**). Sales promotions can also account for a substantial part of cross-sectional price dispersion. Therefore, understanding the determinants of temporary sales is important to understand price stickiness, price dispersion, and firms' market power and competition.

varian_1980 (**varian_1980**) presents a model of price competition in an homogeneous product market, with two types of consumers: consumers that are informed about the price, and consumer that are not. Informed customers always buy in the store with the lowest price. Uninformed consumers choose a store at random and buy there as long as the price of the store is not above their reservation price. The model does not have an equilibrium in pure strategies. In mixed strategies, there is a unique symmetric equilibrium characterized by a U-shape density function on the price interval between the marginal cost and the reservation price of uninformed consumers. According to this equilibrium, the price charged by a store changes randomly over time between a "low" and a "high" price.

Though Varian's model can explain some important empirical features in the cross-section and time series of prices in retail markets, it cannot explain the time dependence of sales promotions that have been reported in many empirical studies (for instance, Slade ,1998, Aguirregabiria ,1999, or Pesendorfer ,2002, among others). The probability of a sales promotion increases with the duration since the last sale. Several studies have proposed and estimated dynamic structural models of retail pricing that can explain price dispersion, sales promotions and their state dependence. These studies also provide estimates of the magnitude and structure of firms' price adjustments costs.

Slade (1998) proposes a model where the demand of a product in a store depends on a stock of goodwill that accumulates over time when the store charges low prices, and erodes when the price is high. The model also incorporates menu costs of changing prices. The optimal pricing policy consists of a cycle between a low price (or sales promotion) and a high price. Slade estimates this model using weekly scanner data of prices and quantities of saltine crackers in four supermarket chains. The estimated model fits well the joint dynamics of prices and quantities. Her estimates of the cost of adjusting prices are approximately 4% of revenue.

Aguirregabiria (1999) studies the relationship between inventories and prices in supermarkets. The cost of placing orders to wholesalers has a fixed component. Retailers have also menu costs of changing prices, face substantial demand uncertainty, and have stockouts. Aguirregabiria proposes a model of price and inventory decisions that incorporates these features. In the optimal decision rule of this model, inventories follow an (S,s) cycle, and prices have a "high-low" cyclical pattern. When a new order is placed, the probability of a stockout declines, expected demand becomes more elastic, and the optimal price drops to a minimum. When inventories decline between two orders, the probability of a stockout increases, expected sales become more inelastic, and the optimal price eventually increases and stays high until the next order. Aguirregabiria estimates this model using data on inventories, prices, and sales from the warehouse of a supermarket chain. The estimated model fits well the joint cyclical pattern of prices and inventories in the data and can explain temporary sales. The estimated values for the fixed ordering cost and the menu cost are 3.1% and 0.7% of monthly revenue, respectively. According to the estimated model, almost 50% of sales promotions are associated to the dynamics of inventories.

Pesendorfer (2002) studies the dynamics of consumer demand as a factor explaining sales promotions of storable products. He proposes a model of demand of a storable product with two types of consumers: store-loyal consumers and shoppers. The equilibrium of the model predicts that the probability that a store has a sale increases with the duration since the last sale both in that store and in other stores. The implied pattern of prices consists of an extended period of high prices followed by a short period of low prices. He tests the predictions of the model using supermarket scanner data for ketchup products. The effects of the duration variables are significant and have the predicted sign. Though this evidence suggests that demand accumulation could be important in the decision to conduct a sale, it is also consistent with models in Slade (1998) and Aguirregabiria (1999). As far as we know, there is no empirical study that has tried to disentangle the relative contribution of consumer inventories, firm inventories, and goodwill to explain temporary sales promotions.

kano_2013 (**kano_2013**) makes an interesting point on the estimation of menu costs in oligopoly markets. Dynamic price competition in oligopoly markets implies a positive strategic interaction between the prices of different firms. This strategic interaction may be an important source of price inertia even when menu costs are small. If a firm experiences an idiosyncratic increase in its marginal cost, it may prefer not to increase its price if the competitor maintains a constant price. A model of monopolistic competition that ignores strategic interactions among firms may spuriously overestimate menu costs. Kano estimates a dynamic pricing model that accounts for these strategic interactions and finds that they account for a substantial part of price rigidity.

6.4.1 Aguirregabiria (1999)

The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, charge prices that differ by significant amounts, for instance, 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries. Interestingly, the price differentials between firms, and the ranking of firms in terms of prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the next week. Using a more graphical description we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also for the construction of price indexes.

Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

For instance, one explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket's ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of making an order that do not depend on the size of the order, or at least that do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: inventories increase by a large amount up to a maximum threshold when an order is placed, and decline gradually until a minimum value is reached, at which time a new order is placed. Given these dynamics of inventories, it is simple to show that the optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again. Aguirregabiria (1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. Specifically, I show that these types of inventory-dependence price dynamics can explain more than 20% of the time series variability of prices in the data.

Temporary sales and inventories. Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in the U.S.: **hosken_reiffen_2004** (**hosken_reiffen_2004**); **nakamura_steinsson_2008** (**nakamura_steinsson_2008**); **midrigan_2011** (**midrigan_2011**). Understanding the determinants of temporary sales is important to understand price stickiness and price dispersion, and it has important implications on the effects of monetary policy. It has also important implications in the study of firms' market power and competition. Different empirical models of sales promotions: Slade (1998) [Endogenous consumer loyalty], Aguirregabiria (1999) [Inventories], Pesendorfer (2002) [Intertemporal price discrimination], and **kano_2013** (**kano_2013**).

This paper studies how retail inventories, and in particular (S,s) inventory behavior, can explain both price dispersion and sales promotions in retail markets. Three factors are key for the explanation provided in this paper:

(1) Fixed (lump-sum) ordering costs, that generates (S,s) inventory behavior.

(2) Demand uncertainty.

(3) Sticky prices (Menu costs) that, together with demand uncertainty, creates a positive probability of excess demand (stockout).

Model: Basic framework

Consider a retail firm selling a product. We index products by i. Every period (month) t the firm chooses the retail price and the quantity of the product to order to manufacturers/wholesalers. **Monthly sales** are the minimum of supply and demand:

$$y_{it} = \min\{\ d_{it}\ ;\ s_{it} + q_{it}\ \}$$

y_{it} = sales in physical units; d_{it} = demand; s_{it} = inventories at the beginning of month t; q_{it} = orders (and deliveries) during month t.

Demand and Expected sales. The firm has uncertainty about **current demand**:

$$d_{it} = d_{it}^{e}\ \exp(\xi_{it})$$

d_{it}^{e} = expected demand; ξ_{it} = zero mean demand shock unknown to the firm at t. Therefore, **expected sales** are:

$$y_{it}^{e} = \int \min\{d_{it}^{e}\ \exp(\xi)\ ;\ s_{it} + q_{it}\ \} dF_{\xi}(\xi)$$

Assume monopolistic competition. **Expected Demand** depends on the own price, p_{it}, and a demand shock ω_{it}. The functional form is isoelastic:

$$d_{it}^{e} = \exp\{\gamma_0 - \gamma_1 \ \ln(p_{it}) + \omega_{it} \ \}$$

where γ_0 and $\gamma_1 > 0$ are parameters.

Price elasticity of expected sales. **Demand uncertainty** has important implications for the relationship between prices and inventories. The price elasticity of expected sales is a function of the **supply-to-expected-demand ratio** $(s_{it} + q_{it})/d_{it}^{e}$:

$$\begin{aligned} \eta_{y^e|p} \equiv \frac{-\partial y^e}{\partial p}\frac{p}{y^e} &= -\left[\int I\{d^e \exp(\xi) \ ; \ s+q \ \} dF_\xi(\xi)\right] \frac{\partial d^e}{\partial p}\frac{p}{y^e} \\ &= \gamma_1 \ F_\xi\left(\log\left[\frac{s+q}{d^e}\right]\right)\frac{d^e}{y^e} \end{aligned}$$

And we have that:

$$\eta_{y^e|p} \longrightarrow \begin{cases} \gamma_1 & \text{as} \quad (s+q)/d^e \longrightarrow \infty \\ \\ 0 & \text{as} \quad (s+q)/d^e \longrightarrow 0 \end{cases}$$

Price elasticity of expected sales

$$\eta_{y^e|p} = \gamma_1 \ F_\xi\left(\log\left[\frac{s+q}{d^e}\right]\right)\frac{d^e}{y^e}$$

[FIGURE: $\eta_{y^e|p}$ increasing in $\dfrac{s+q}{d^e}$, with asymptote at γ_1]

When the supply-to-expected-demand ratio is large, the probability of a stockout is very small and $y^e \simeq d^e$, so the elasticity of expected sales is just the elasticity of demand. However, when the supply-to-expected-demand ratio is small, the probability of a stockout is large and the elasticity of expected sales can be much lower than the elasticity of demand.

Markup and inventories (myopic case). This has potentially important implications for the optimal price of an oligopolistic firm. To give some intuition, consider the pricing decision of the monopolistic firm without forward-looking behavior. That optimal price is:

$$\frac{p-c}{p} = \frac{1}{\eta_{y^e|p}}$$

OR

$$\frac{p-c}{c} = \frac{1}{\eta_{y^e|p} - 1}$$

Variability over time in the supply-to-expected-demand ratio can generate significant fluctuations in price-cost margins. It can also explain temporary sales promotions. That can be the case under (S,s) inventory behavior.

Empirical Application

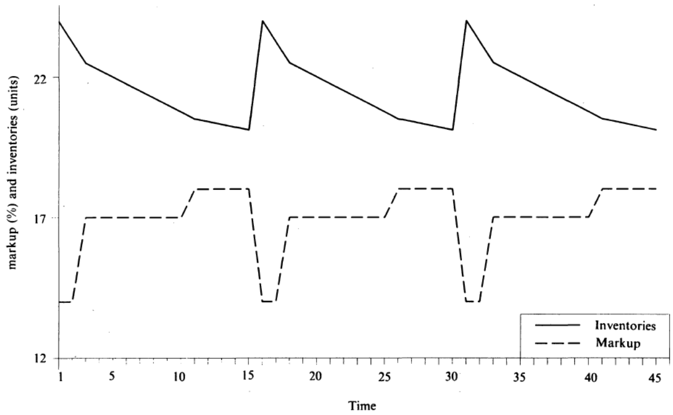

Figure 6.7: Cyclical Inventories and Prices

The paper investigates this hypothesis using data from a supermarket chain, with rich information on prices, sales, inventories, orders, and wholesale prices for many different products. Reduced form estimations present evidence that supports the hypothesis:

(1) Prices depend negatively and very significantly on the level of inventories.

(2) Inventories of many products follow (S,s) cycles.

(3) Price cost margins increase at the beginning of an (S,s) cycle, and decline monotonically during the cycle.

I estimate the parameters in the profit function (demand parameters, ordering costs, inventory holding costs) and use the estimated model to analyze how much of price variation and temporary sales promotions can be explained by firm inventories.

Profit function. **Expected current profits** are equal to expected revenue, minus ordering costs, inventory holding costs and price adjustment costs:

$$\pi_{it} = p_{it}\; y_{it}^{e} - OC_{it} - IC_{it} - PAC_{it}$$

OC_{it} = ordering costs; IC_{it} = inventory holding costs; PAC_{it} = price adjustment (menu) costs. **Ordering costs:**

$$OC_{it} = \begin{cases} 0 & \text{if} \quad q_{it} = 0 \\ \\ F_{oc} + \varepsilon_{it}^{oc} - c_{it}\; q_{it} & \text{if} \quad q_{it} > 0 \end{cases}$$

F_{oc} = fixed (lump-sum) ordering cost. Parameter; ε_{it}^{oc} = zero mean shock in the fixed ordering cost; c_{it} = wholesale price. **Inventory holding costs:**

$$IC_{it} = \alpha\; s_{it}$$

Menu costs:

$$PAC_{it} = \begin{cases} 0 & \text{if} \quad p_{it} = p_{i,t-1} \\ \\ F_{mc}^{(+)} + \varepsilon_{it}^{mc(+)} & \text{if} \quad p_{it} > p_{i,t-1} \\ \\ F_{mc}^{(-)} + \varepsilon_{it}^{mc(-)} & \text{if} \quad p_{it} < p_{i,t-1} \end{cases}$$

$F_{mc}^{(+)}$ and $F_{mc}^{(-)}$ are price adjustment cost parameters; $\varepsilon_{it}^{mc(+)}$ and $\varepsilon_{it}^{mc(-)}$ are zero mean shocks in menu costs

State variables. The state variables of this DP problem are:

$$\left\{ \underbrace{s_{it},\, c_{it},\, p_{i,t-1},\, \omega_{it},}_{x_{it}}\;\; \underbrace{\varepsilon_{it}^{oc},\, \varepsilon_{it}^{mc(+)},\, \varepsilon_{it}^{mc(+)}}_{\varepsilon_{it}} \right\}$$

The decision variables are q_{it} and $\Delta p_{it} \equiv p_{it} - p_{i,t-1}$. We use a_{it} to denote $(q_{it},\Delta p_{it})$. Let $V(x_{it},\varepsilon_{it})$ be the value of the firm associated with product i. This value function solves the Bellman equation:

$$V(x_{it},\varepsilon_{it}) = \max_{a_{it}} \left\{ \begin{array}{c} \pi(a_{it},x_{it},\varepsilon_{it}) \\ \\ +\beta\; \int V(x_{i,t+1},\varepsilon_{i,t+1})\; dF(x_{i,t+1},\varepsilon_{i,t+1}|a_{it},x_{it},\varepsilon_{it}) \end{array} \right\}$$

Discrete Decision variables. Most of the variability of q_{it} and Δp_{it} in the data is discrete. For simplicity, we assume that these variables have a discrete support.

$$\begin{array}{rcl} q_{it} & \in & \{0\;,\; \kappa_i\} \\ \\ \Delta p_{it} & \in & \{0\;,\; \delta_i^{(+)},\; \delta_i^{(-)}\} \end{array}$$

where $\kappa_i > 0$, $\delta_i^{(+)} > 0$, and $\delta_i^{(-)} < 0$ are parameters. Therefore, the set of choice alternatives at every period t is:

$$a_{it} \in A = \left\{\; (0,0),\; (0,\delta_i^{(+)}),\; (0,\delta_i^{(-)}),\; (\kappa_i,0),\; (\kappa_i,\delta_i^{(+)}),\; (\kappa_i,\delta_i^{(-)})\; \right\}$$

The transition rules for the state variables are:

$$\begin{array}{rcl} s_{i,t+1} & = & s_{it} + q_{it} - y_{it} \\ p_{it} & = & p_{i,t-1} + \Delta p_{it} \\ c_{i,t+1} & \sim & AR(1) \\ \omega_{i,t+1} & \sim & AR(1) \\ \varepsilon_{it} & \sim & i.i.d. \end{array}$$

Integrated Bellman Equation. The components of ε_{it} are independently and extreme value distributed with dispersion parameter σ_ε. Therefore, as in Rust (1987), the integrated value function $\bar{V}(x_{it})$ is the unique fixed point of the integrated Bellman equation:

$$\bar{V}(x_{it}) = \sigma_\varepsilon \ln \left(\sum_{a \in A} \exp \left\{ \frac{v(a, x_{it})}{\sigma_\varepsilon} \right\} \right)$$

where:

$$v(a, x_{it}) = \bar{\pi}(a, x_{it}) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) \; f_x(x_{i,t+1} | a, x_{it})$$

Discrete choice profit function

• $\bar{\pi}(a, x_{it})$ is the part of current profit which does not depend on ε_{it}:

$$\bar{\pi}(a, x_{it}) = \begin{cases} R_{it}(0,0) - \alpha \; s_{it} & \text{if} \quad a = (0,0) \\ R_{it}(0, \delta_i^{(+)}) - \alpha \; s_{it} - F_{mc}^{(+)} & \text{if} \quad a = (0, \delta_i^{(+)}) \\ R_{it}(0, \delta_i^{(-)}) - \alpha \; s_{it} - F_{mc}^{(-)} & \text{if} \quad a = (0, \delta_i^{(-)}) \\ R_{it}(\kappa_i, 0) - \alpha \; s_{it} - F_{oc} - c_{it} \kappa_i & \text{if} \quad a = (\kappa_i, 0) \\ R_{it}(\kappa_i, \delta_i^{(+)}) - \alpha \; s_{it} - F_{oc} - c_{it} \kappa_i - F_{mc}^{(+)} & \text{if} \quad a = (\kappa_i, \delta_i^{(+)}) \\ R_{it}(\kappa_i, \delta_i^{(-)}) - \alpha \; s_{it} - F_{oc} - c_{it} \kappa_i - F_{mc}^{(-)} & \text{if} \quad a = (\kappa_i, \delta_i^{(-)}) \end{cases}$$

where $R_{it}(.,.)$ is the expected revenue function.

Some predictions of the model. Fixed ordering cost F_{oc} generates infrequent orders: $(\mathbf{S}, \mathbf{s})$ **inventory policy**. (S, s) inventory behavior, together with demand uncertainty (that is, optimal prices depend on the supply-to-expected demand ratio) generate a cyclical pattern in the price elasticity of sales. Prices decline significantly when an order is placed (sales promotion). This price decline and the consequent inventory reduction generate a price increase. Then, as inventories decline between two orders, prices tend to increase.

Data. Data from the central warehouse of a supermarket chain in the Basque Country (Spain). Monthly data: period January 1990 to May 1992. Estimation of Structural Parameters. Counterfactual Experiments