The Industry Supply Function and the Long-Run Competitive Equilibrium with Heterogeneous Firms*

Ignacio Esponda  Demian Pouzo
(UC Santa Barbara) (UC Berkeley)

March 12, 2018

Abstract

In developing the theory of long-run competitive equilibrium (LRCE), Marshall (1890) used the notion of a representative firm. The identity of this firm, however, remained unclear, and subsequent theory focused on the case where all firms are identical. Using Hopenhayn’s (1992) model of competitive industry dynamics, we extend the theory of LRCE to account for heterogeneous firms and show that the long-run supply function can indeed be characterized as the solution to the minimization of a representative average cost function. We also highlight that famous principles of competitive markets, such as efficiency of the LRCE allocation, are not robust to heterogeneity.

*We thank Gian Luca Clementi, Ernesto Dal Bó, Daniel Gottlieb, Luis Rayo, and Andrés Rodriguez-Claire for helpful comments. Esponda: Department of Economics, UC Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106, iesponda@ucsb.edu; Pouzo: Department of Economics, UC Berkeley, 530 Evans Hall #3880, Berkeley, CA 94720, dpouzo@econ.berkeley.edu.
Contents

1 Introduction 1

2 Model and illustrative example 4
   2.1 Setup ................................................. 4
   2.2 A simple example .................................. 8

3 Characterization of long-run industry supply 12
   3.1 Proof of Theorem 1 .................................. 13
   3.2 Equilibrium surplus ................................. 18

References 21

A Appendix 22
   A.1 Proof of Lemma 3 ................................. 22
   A.2 Proof of Lemma 5 ................................. 22
Figure 1: Textbook model of long-run competitive equilibrium.

1 Introduction

The theory of long-run competitive equilibrium (LRCE), first developed by Marshall in his *Principles of Economics* (1890), has had a profound influence on our understanding of competitive markets. One distinguishing feature of Marshall’s theory is his conceptualization of the (long-run) industry supply function. Pigou (1928), Viner (1953)[1931] and others subsequently formalized Marshall’s notion of LRCE. The latter author, in particular, is credited for popularizing the typical diagram taught in introductory courses and reproduced in Figure 1.

The figure represents an industry with fixed input prices where all firms are identical and characterized by the marginal (MC) and average (AC) cost functions depicted in the left panel. In an LRCE, price is at the minimum point of the AC function, $p^e$, and aggregate quantity is given by the demand function evaluated at that price, $Q^e_0$. Suppose that there is a shift of the (inverse) demand function from $P^d_0$ to $P^d_1$ in Figure 1. In the short run, the number of firms stays fixed, so price and quantity increase from the original LRCE at point $A$ to the new short-run equilibrium at point $B$, a movement occurring along the short-run supply function $S_0$. But then firms make positive (economic) profits, and these profits attract additional firms into the market. In the long-run, the new LRCE is at point $C$, where all firms make zero profits at price $p^e$ and aggregate production increases to $Q^e_1$. Thus, the (long-run) industry supply function, $S_{LR}$, is horizontal at the minimum of the average cost function, $p^e$.

A distinguishing characteristic of Marshall’s analysis is the notion of a representation...
tative firm. While Marshall recognized that there are different firms in an industry, subsequent developments have focused on the case where all firms are identical in a long-run equilibrium. Viner (1953)[1931], pg. 222, justifies this view:

“If there are particular units of the factors which retain permanently advantages in value productivity over other units of similar factors, these units, if hired, will have to be paid for in the long-run at differential rates proportional to their value productivity, and if employed by their owner should be charged for costing purposes with the rates which could be obtained for them in the open market and should be capitalized accordingly.”

Viner’s argument may justify why firms do not make rents in the presence of markets that bid up the price of advantageous factors, such as exceptional managerial ability. But the argument does not imply that firms with different technologies or productivities cannot coexist in equilibrium. A realistic feature of an industry is that low-productivity firms can potentially become high-productivity firms and vice versa. This feature implies that equilibrium will be characterized both by coexistence of heterogeneous firms and turnover (entry and exit), and it does not seem appropriate to exclude these realistic features from a theory of LRCE.

Our objective in this paper is to go back to Marshall’s original motivation and to extend the classical theory of LRCE to the case of heterogeneous firms. Fortunately, we don’t have to formulate a new model, since Hopenhayn (1992) actually introduced and studied a model of competitive industry dynamics where firms’ productivities evolve over time and exit and entry is an equilibrium phenomenon. We take the steady-state equilibrium in Hopenhayn’s model as the natural extension of the theory of LRCE to the case with heterogeneous firms. Hopenhayn (1992), however, did not link his work to the early theory on LRCE, and our contribution is to fill-in this gap.

Our main result is that the (long-run) industry supply function with heterogeneous firms can indeed be characterized as the solution to the minimization of a representative average cost function, as Marshall originally envisioned. The standard textbook case, depicted in Figure 1, is just a special case where there is no firm heterogeneity.

There are several reasons to care about this result. First, it formalizes Marshall’s original motivation of a representative firm and of the industry supply function in the presence of heterogeneous firms. Second, it provides a connection between the early literature on LRCE and the modern literature on industry dynamics (to be reviewed below). Third, it makes the model of LRCE with heterogeneous firms accessible
to a larger audience (in particular, the example in Section 2.2 conveys much of the intuition and can be taught in introductory courses). Finally, it helps highlight that some famous principles of competitive markets are not robust to the inclusion of firm heterogeneity. We illustrate this last point by showing that aggregate surplus is generally not maximized in an LRCE with heterogeneous firms.

Our paper links the classic theory of LRCE, which does not explicitly model dynamics, with the modern literature on competitive industry dynamics started by Lucas (1967). Lucas and Prescott (1971) developed the first theory of dynamic competitive equilibrium with stochastic demand, costly capital stock adjustments, and correct (i.e., “rational”) expectations about future prices. The theory, however, assumes that firms are identical and that there is no entry and exit. Subsequent developments incorporated both firm heterogeneity and entry and exit, at the expense of no longer studying the dynamics of capital accumulation. Jovanovic (1982) developed the first of such models. Each period, a firm draws a productivity shock from a distribution that depends on an unknown productivity type. Firms have different productivity types and, as they learn their own type, more productive firms stay and less productive firms exit. The objective of these papers was to study the dynamic evolution of a competitive industry, not the steady state. Consequently, all of the interesting action happens outside the steady state and, indeed, there is no entry and exit in the steady state of these models.

Hopenhayn (1992) was the first one to consider a model with both heterogeneous firms and entry and exit in the steady state. In contrast to Jovanovic’s model, firms know their productivity types, but productivity types evolve randomly in such a way that firms that have a low productivity today can have a high productivity tomorrow and vice versa. As mentioned earlier, this is the model that we will use to formalize Marshall’s idea that the LRCE of a competitive industry is characterized by the cost function of a representative firm.¹

For brevity, we focus on the case where input prices are fixed, which implies that the long-run industry supply function is horizontal. The extension to the case of input prices that increase with aggregate quantity was controversial in the early literature—see Opocher and Steedman (2008) for an insightful historical account. The

¹For a model that incorporates capital accumulation to Hopenhayn’s competitive framework, see Clementi and Palazzo (2016). There is also a large literature, beginning with the work of Ericson and Pakes (1995), that studies dynamic equilibrium with capital accumulation, stochastic shocks, heterogeneous firms, and entry and exit under imperfect competition.
initial approach, by Pigou (1928), Viner (1953) [1931], and others, considered a cost function that depends both on individual and aggregate quantity. Subsequent literature (e.g., Kaldor (1934), Allen et al. (1938), and Hicks (1946)) criticized this reduced-form approach because of lack of microfoundations. For either approach, the extension of our result is straightforward: A given aggregate quantity leads to a given equilibrium input price and, fixing this input price, the LRCE price is still the minimum point on a representative average cost function. The long-run industry (inverse) supply function is simply the mapping from aggregate quantities to these minimum points. In particular, the aggregate supply function may be increasing if input prices increase with aggregate quantity.

2 Model and illustrative example

2.1 Setup

We adopt Hopenhayn’s (1992) infinite-horizon model of a competitive industry with a continuum of potential firms, each of which can produce a homogenous product at total cost $C(q, \theta)$ where $q$ is the quantity produced, $\theta \in \Theta = [\theta_L, \theta_H] \subset \mathbb{R}$ is the firm’s type, and $\theta_L < \theta_H$.

Each period $t = 1, 2, ..., \text{product demand is given by } Q^d(p)$, where $p \geq 0$ is the output price. Firms take price as given and choose quantity to maximize profit. There is also an infinite mass of potential entrants with discount factor $\delta \in [0, 1]$ who can decide to enter the market and become a firm. A potential entrant does not know her type, but knows that her type is independently distributed according to the probability measure $\nu \in \Delta(\Theta)$. A firm entering the market pays a one-time entry cost of $\kappa \geq 0$. After paying this cost, a firm immediately learns its own type. Thereafter, types evolve independently across firms according to the probability measure $F(\cdot \mid \theta) \in \Delta(\Theta)$, where $\theta$ is the current type. At the end of the period, each firm makes an exit decision knowing their current, but not future, type. There is also an exogenous exit probability $\rho$. A firm that exits the market (endogenously or exogenously) does so permanently and obtains a payoff of zero.

We maintain the following assumptions.

Assumption 1. Demand: There exists $\nu > 0$ such that $Q^d(\cdot)$ is continuous and decreasing for all $p \in (0, \nu)$, and $Q^d(p) = 0$ for all $p \geq \nu$. 

4
Assumption 2. Costs: For all $\theta \in \Theta$: $C(\cdot, \theta)$ is continuously differentiable, with $C(q, \theta) \geq 0$, $C'(q, \theta) \geq 0$, and $C''(q, \theta) > 0$ for all $q \geq 0$, and $\lim_{q \to \infty} C'(q, \theta) = \infty$; For all $q \geq 0$, $C(q, \cdot)$ is increasing.

Assumption 3. Order over types: For any $\theta_1 < \theta_2$, $F(\theta | \theta_2) < F(\theta | \theta_1)$ for all $\theta \in (\theta_L, \theta_H)$.

Assumption 4. The exogenous probability of exit is positive, i.e., $\rho > 0$.

Assumption 5. Measures over types: (i) $\nu$ has a continuous probability density function (pdf), $f_{\nu}(\cdot)$, with support equal to $\Theta$; (ii) For all $\theta$: $F(\cdot | \theta)$ has a pdf $f(\cdot | \theta)$, with support equal to $\Theta$, and $(\theta', \theta) \mapsto f(\theta' | \theta)$ is jointly continuous.

Assumption 1 implies the existence of a downward sloping inverse demand function, $P^d(\cdot)$. Assumption 2 implies existence and uniqueness of an optimal quantity

$$q(p, \theta) \equiv \arg \max_{q \geq 0} pq - C(q, \theta)$$

The assumption also implies that the profit function

$$\pi(p, \theta) \equiv pq(p, \theta) - C(q(p, \theta), \theta).$$

is nonincreasing in $\theta$, and decreasing for $(p, \theta)$ such that $q(p, \theta) > 0$. Note also that each firm has a fixed cost $C(0, \theta)$ which is sunk once the firm decides to enter or stay in the industry.

Assumption 3 postulates a first-order stochastic dominance relationship across types, so that higher types today are more likely to become higher types tomorrow. Assumption 4 guarantees that the life span of a firm is almost surely finite; in particular, if there is no entry, then there must be zero aggregate production in equilibrium. This assumption is made for simplicity, puts the focus on equilibria with positive entry, and allows us to include the special case where firms’ types are permanent. Assumption 5 lists technical conditions regarding the measures over types.

---

2Hopenhayn (1992) instead assumes that $\rho = 0$ and guarantees finite lifespan with an additional recurrence condition on $F$. He then restricts attention to equilibria with positive entry.
Assumption 6. $\pi(v, \theta_H) > \kappa$.

Assumption 6 is made for simplicity. It rules out equilibria with zero aggregate production by requiring that even the highest-cost firm prefers to enter whenever price equals the maximum willingness to pay, $v$.

The expected net present discounted value of a firm of type $\theta$ who faces (steady-state) price $p$ every period is

$$V(p, \theta) = \pi(p, \theta) + \delta(1 - \rho) \max \left\{ \int_{\Theta} V(p, \theta') F(d\theta' \mid \theta), 0 \right\}. \tag{1}$$

Assumption 3 and the fact that $\pi(p, \cdot)$ is decreasing imply that $\int_{\Theta} V(p, \theta') F(d\theta' \mid \cdot)$ is decreasing. Therefore, the optimal exit decision in steady state is characterized by a marginal type $m \in \Theta$ with the property that all lower types stay and all higher types exit the market.

Let $\mu(n, m)$ denote the steady-state measure of types of firms given the mass of entrants $n \geq 0$ and the marginal type $m \in \Theta$. In particular, for any Borel set $A \subseteq \Theta$,

$$\mu(n, m)(A) = \nu(A)n + (1 - \rho) \int_{\theta_L}^{m} F(A \mid \theta)\mu(n, m)(d\theta). \tag{2}$$

The assumption that $\rho > 0$ guarantees existence of a steady-state measure.

The corresponding aggregate supply at price $p$ is

$$Q^e(p; n, m) \equiv \int_{\Theta} q(p, \theta)\mu(n, m)(d\theta).$$

Definition 1. A tuple $(p^e, n^e, m^e)$ is a long-run competitive equilibrium (LRCE) if the following conditions are satisfied:

(i) Market clearing: $Q^d(p^e) = Q^e(p^e; n^e, m^e)$.

(ii) Unlimited entry: $\int_{\Theta} V(p^e, \theta)\nu(d\theta) \leq \kappa$, with equality if $n^e > 0$.

(iii) Optimal exit: $\int_{\Theta} V(p^e, \theta') F(d\theta' \mid m^e) = 0$ if $m^e \in (\theta_L, \theta_H)$, $\geq 0$ if $m^e = \theta_H$, and $\leq 0$ if $m^e = \theta_L$.

An LRCE captures the steady state of the dynamic competitive industry. The first condition requires market clearing and already incorporates the assumption of

$^{3}$Hopenhayn (1992) called an LRCE a stationary equilibrium and showed that it corresponds to the steady state of a perfect foresight equilibrium of the dynamic environment.
profit maximization. The second condition requires the net present value of entry to equal the entry cost if the mass of entrants is positive. It is known as the “free entry” condition, but we reserve that terminology for the case in which entry is actually free, i.e., $\kappa = 0$. The third condition requires the marginal type to be indifferent between staying or exiting the market, provided it is an interior type.

**Lemma 1.** In any LRCE, both aggregate production and entry must be positive.

**Proof.** Suppose $p^e$ is an LRCE price and $Q^d(p^e) = 0$. By Assumption 1, $p^e \geq v$. By the fact that $\pi(p, \theta)$ is nondecreasing in $p$ and nonincreasing in $\theta$ and by Assumption 6, $\pi(p^e, \theta) \geq \pi(v, \theta) \geq \pi(v, \theta_H) > \kappa$ for all $\theta$. Thus, $V(p^e, \theta) > \kappa$ for all $\theta$, so that $p^e$ does not satisfy the entry condition (ii) in Definition 1, contradicting the fact that $p^e$ is an LRCE price. Therefore, $Q^d(p^e) > 0$ and by condition (i) in Definition 1, $Q^s(p^e; n^e, m^e) > 0$, which then implies, by the assumption that $\rho > 0$, that $n^e > 0$. 

**Definition 2.** The long-run industry (inverse) supply function is a function $Q \mapsto P_{LR}^s(Q)$ with the property that, for any $Q > 0$, $p = P_{LR}^s(Q)$ is the unique price satisfying the following conditions for some $w > 0$, $m \in \Theta$, and $n > 0$:

(i) $Q = Q^s(p; n, m)$.
(ii) $\int_{\Theta} V(p, \theta) \nu(d\theta) = \kappa$.
(iii) $\int_{\Theta} V(p, \theta') F(d\theta' \mid m) = 0$ if $m \in (\theta_L, \theta_H)$, $\geq 0$ if $m = \theta_H$, and $\leq 0$ if $m = \theta_L$.

The next result follows immediately from the definitions and from Lemma 1’s implication that the entry condition in Definition 2 holds with equality in equilibrium.

**Proposition 1.** Suppose that the long-run industry supply function $P_{LR}^s(\cdot)$ exists. Then $p^e$ is part of an LRCE if and only if $p^e = P_{LR}^s(Q^d(p^e))$ and $Q^d(p^e) > 0$.

Proposition 1 simply says that the LRCE price is such that supply equals demand. When firms are identical, it is well known that the long-run industry supply function is horizontal at the minimum point of the average cost function. Our objective is to characterize this function for the environment described in this section, where firms are heterogeneous.
2.2 A simple example

We discuss an example with three objectives in mind: It is simple enough to be taught in introductory courses, it conveys much (but not all) of the intuition behind our results, and it is sufficient to see that standard properties of LRCE with homogeneous firms do not extend to heterogeneous firms.

We assume that: (i) there are only two types, not a continuum, $\theta_H > \theta_L \geq 0$, and each type is equally likely to be drawn by an entrant; (ii) $C(q, \theta) = c(q) + \theta$, so that a firm’s type represents its fixed cost and all firms have the same marginal cost $MC(q) \equiv c'(q)$; (iii) the entry cost is zero, $\kappa = 0$; (iv) types are permanent, so that a firm keeps the type it draws upon entry for its entire lifetime; and (v) firms are impatient, $\delta < 1$. The variable cost function $c(\cdot)$ satisfies the following conditions: $c(0) = 0$, $c'(0) = 0$, $c'(q) > 0$ and $c''(q) > 0$ for all $q > 0$, and $\lim_{q \to \infty} c'(q) = \infty$.

**Steady-state measure of types.** It is easy to see that type $\theta_L$ will stay and type $\theta_H$ will exit in equilibrium; in particular, we will drop $m$ from the notation. The steady-state mass of firms of type $\theta_L$, denoted by $\mu_L$, is determined by the steady-state mass of entrants, $n$, as follows:

$$\mu_L = \frac{n}{2} + \mu_L(1 - \rho). \quad (3)$$

The RHS of equation $(3)$ is the sum of the mass of entrants of type $\theta_L$, $n/2$, and the mass of firms of type $\theta_L$ that were already present and did not exit exogenously, $\mu_L(1 - \rho)$. The equation implies that, in steady state, the mass of type $\theta_L$ remains constant. For firms of type $\theta_H$, who never stay for more than one period, their mass is half the mass of entrants. Thus, the steady-state masses of firms of each type as a function of the mass of entrants, $n$, are

$$\mu_L(n) = \frac{n}{2(2\rho)} \quad \text{and} \quad \mu_H(n) = \frac{n}{2}.$$

**Long-run industry supply function.** The conditions in the definition of the long-run supply function become:

(i) $Q = (\mu_L(n) + \mu_H(n))q(p) > 0$.

(ii) (Free entry) $NPV(p) \equiv \frac{1}{2} \pi(p, \theta_L)/(1 - \delta(1 - \rho)) + \frac{1}{2} \pi(p, \theta_H) = 0$.

$^4$For the free entry condition to hold for $p > 0$, the profit of type $\theta_H$ must be negative. Because types are permanent, type $\theta_H$ will find it optimal to exit.
Condition (i) requires aggregate output supply to equal $Q$. Condition (ii) requires that the net present value of an entrant is zero. With probability $1/2$, a firm is of type $\theta_L$ and remains in the market until it has to exogenously exit, thus expecting a net present value of $\alpha(p, \theta_L)/(1 - \delta(1 - \rho))$. With probability $1/2$, a firm is of type $\theta_H$, makes profit $\pi(p, \theta_H)$, and exits the market.

The weights on the profit functions of each type in the free-entry condition have an intuitive interpretation. The weight $\Lambda_L\equiv 1/(2(1 - \delta(1 - \rho)))$ on $\pi(p, \theta_L)$ is equal to the steady-state mass of type $\theta_L$, normalized by the mass of entrants $n$, in a hypothetical world where firms, instead of exiting with probability $\delta$, exit with probability $1 - \delta(1 - \rho)$.

The hypothetical and actual probabilities of exit coincide as $\delta \to 1$, and so the weight asymptotically equals the actual, normalized steady-state mass of type $\theta_L$. Similarly, the weight $\Lambda_H\equiv 1/2$ on $\pi(p, \theta_L)$ is equal to the normalized steady-state mass of firms of type $\theta_H$ (here, $\delta$ is irrelevant because type $\theta_H$ exits with probability 1). Thus, the net present value of entry can be written as

$$NPV(p) = \Lambda_L \pi(p, \theta_L) + \Lambda_H \pi(p, \theta_H)$$

$$= pq(p)(\Lambda_L + \Lambda_H) - (\Lambda_L C(q(p), \theta_L) + \Lambda_H C(q(p), \theta_H)).$$

(4)

By equation (4), the solution $p^e$ to $NPV(p^e) = 0$ satisfies

$$p^e = AC^e(q(p^e), \Lambda) \equiv \frac{\Lambda_L AC(q(p^e), \theta_L) + \Lambda_H AC(q(p^e), \theta_H)}{\Lambda_L + \Lambda_H},$$

(5)

where $\Lambda \equiv (\Lambda_L, \Lambda_H)$, $AC(q, \theta) \equiv C(q, \theta)/q$ is the average cost of type $\theta$, and $q \mapsto AC^e(q, \Lambda)$ is a weighted average cost function.

By profit maximization, $p^e = MC(q(p^e)))$, and so (5) implies that $p^e$ equalizes marginal and weighted average cost,

$$p^e = MC(q(p^e)) = AC^e(q(p^e), \Lambda).$$

(6)

The left panel of Figure 2 illustrates how to find $p^e$. The figure plots the marginal cost function common to all types, $MC(\cdot)$, the average cost function for each type, $AC(\cdot, \theta)$, and the weighted average cost function $AC^e(\cdot, \Lambda)$. The zero-profit price $p^e$ is given by the intersection of the marginal cost and weighted average cost functions, and

$^5$Formally, $\Lambda_L \equiv \mu_E(n, \delta)(\theta_L)/n$, where $\mu_E(n, \delta)(\theta_L)$ solves $\mu_E(n, \delta)(\theta_L) = n/2 + \mu_E(n, \delta)(\theta_L)\delta L(1 - \rho)$. 
Figure 2: Long-run competitive equilibrium in the example.

this intersection occurs at the minimum point on the weighted average cost function.\(^6\)

Therefore, \( q(p^e) = q_{\text{min}}^e \equiv \arg \min_q AC^e(q, \Lambda) \) and the zero-profit price \( p^e \) is

\[
p^e = AC^e(q_{\text{min}}^e, \Lambda) = \min_q AC^e(q, \Lambda).
\]

Finally, it is straightforward to check that, since \( p^e > 0 \), there exists \( n(Q) > 0 \) satisfying condition (i) in Definition 2, i.e., \( Q = (\mu_L(n(Q)) + \mu_H(n(Q)))q(p^e) \). Therefore, the long-run supply function exists and is horizontal at the price that minimizes the weighted average cost function \( AC^e(\cdot) \). Thus, provided that \( P^d(0) > p^e \), there exists a unique LRCE where price is \( p^e \) and the mass of entrants \( n^e \) is such that the product market clears, i.e., \( Q^d(p^e) = (\mu_L(n^e) + \mu_H(n^e))q_{\text{min}}^e \).

Figure 2 also illustrates that aggregate profits are strictly positive in an LRCE.

The equilibrium profit of the average firm is \( \pi^e \equiv (p^e - AC^*(q_{\text{min}}^e))q_{\text{min}}^e > 0 \), where \( q_{\text{min}}^e \) is the quantity produced by each firm and

\[
AC^*(\cdot) = \frac{(1/(2\rho))AC(\cdot, \theta_L) + (1/2)AC(\cdot, \theta_H)}{((1/(2\rho)) + 1/2)}
\]

is the per-unit cost function of the average firm producing in equilibrium. The weights in \( AC^*(\cdot) \) correspond to the steady-state proportion of firms of each type. While these

\(^6\)For a proof that the intersection occurs at the minimum point of \( AC^e(\cdot, \Lambda) \), note that the first order condition for the problem \( \min_q AC^e(q, \Lambda) \) is precisely the condition \( MC(q) = AC^e(q, \Lambda) \). Moreover, the second order condition is satisfied because \( c''(q) > 0 \) for all \( q > 0 \).

\(^7\)The solution is unique and given by \( n^e = Q^d(p^e)/((1/2\rho + 1/2)q_{\text{min}}^e) \).
weights converge to $\Lambda$ as $\delta \to 1$, for the case $\delta < 1$, $AC^*(\cdot)$ puts more weight on the low cost type relative to $AC^e(\cdot, \Lambda)$. Intuitively, the selection in exit implies that the steady-state composition of firms is tilted towards low-cost firms relative to the ex-ante perception of a potential entrant who discounts the future. Thus, while potential entrants make zero profits ex-ante, the actual firms operating in the steady state make strictly positive profits.

The result that profits are strictly positive in an LRCE changes many of the implications of the textbook model of LRCE, such as the idea that 100% of the incidence from tax policy must fall on the demand side, or the idea that benefits from a subsidies must accrue exclusively to the owner of a fixed input factor, or the idea that aggregate surplus is maximized in equilibrium. In this example, a planner who wishes to maximize steady-state surplus prefers a higher aggregate quantity $Q^*$, a lower quantity per firm $q^*_{\min}$, and a higher mass of entrants $n^*$ compared to the LRCE quantities $Q^e$, $q^e_{\min}$, and $n^e$. Of course, the planner’s preferred outcome is not an equilibrium outcome, because the net present value of entry would be negative and firms would not enter to begin with.

In the special case where there is a single type, $\theta_L = \theta_H$, the standard textbook results hold: The industry supply function is horizontal at the price that equals the minimum of the average cost function (all firms have the same cost function), each firm makes zero profits, and aggregate surplus is maximized in an LRCE (irrespective of the value of the discount factor $\delta$). Alternatively, we can interpret the standard textbook model as a case where firms are of different types but know their types before entering the market. In that case, only firms of type $\theta_L$ will operate in the market in an LRCE.

**Beyond the simple example.** We extend the logic in the example in several directions. First, marginal costs may differ by type. We will tackle this case by expressing the average cost function in terms of price, not quantity. Second, types may be non-permanent. When types follow a more general Markov process, optimal entry decisions are the solution to a non-trivial dynamic optimization problem. We will use results from the theory of bounded linear operators to show that, nevertheless, ex-ante expected profits can still be expressed as the weighted average of the profits of each type. Third, there may be a continuum of types. In this case, exit decisions will no longer be trivially characterized and we will see, for example, that exit decisions are also inefficient from the perspective of a planner who wants to maximize aggregate
equilibrium surplus. Fourth, strictly positive entry costs need to be incorporated into the definition of average cost.

3 Characterization of long-run industry supply

To state the main result, we first define an average weighted cost function. Letting $\mathcal{M}(\Theta)$ be the space of finite Borel measures that are absolutely continuous with respect to Lebesgue, we define $\bar{C} : [0, \infty) \times \mathcal{M}(\Theta) \to [0, \infty)$ as

$$\bar{C}(p, \eta) = \int_{\Theta} C(q(p, \theta), \theta) \eta(d\theta) + \kappa$$

for all $p \geq 0$ and $\eta \in \mathcal{M}(\Theta)$. This is the weighted cost with respect to a measure $\eta$. Similarly, let $\bar{q} : [0, \infty) \times \mathcal{M}(\Theta) \to [0, \infty)$ be defined by

$$\bar{q}(p, \eta) = \int_{\Theta} q(p, \theta) \eta(d\theta).$$

The corresponding average weighted cost function is then defined by

$$\bar{AC}(p, \eta) \equiv \bar{C}(p, \eta)/\bar{q}(p, \eta),$$

provided that $\bar{q}(p, \eta) > 0$. In the case where marginal costs are identical, the average weighted cost coincides with the weighted average cost, as in the example, but this is not true in general.

Next, for each $n$, $m$, and $\delta$, we define $\mu_E(n, m, \delta) \in \mathcal{M}(\Theta)$ to be the steady-state measure of types of firms when the mass of entrants is $n$, firms survive with exogenous probability $\delta(1 - \rho)$, and surviving firms exit endogenously if their type is lower than $m \in \Theta$, i.e., for any Borel set $A \subseteq \Theta$,

$$\mu_E(n, m, \delta)(A) = \nu(A)n + \delta(1 - \rho) \int_{\theta_L}^{m} F(A \mid \theta) \mu_E(n, m, \delta)(d\theta).$$

For the special case of $\delta = 1$, $\mu_E(n, m, 1) = \mu(n, m)$ is the actual steady-state measure of types defined in equation (2), because in the model firms survive with exogenous probability $1 - \rho$, not $\delta(1 - \rho)$.

---

8If $\bar{q}(p, \eta) = 0$, we define $\bar{AC}(p, \eta) = \infty$. 

12
Finally, since $\mu_E$ is linear in $n$, we define the normalized mass

$$\Lambda(m, \delta) \equiv \mu_E(n, m, \delta)/n \in \mathcal{M}(\Theta).$$

We now state the main result.

**Theorem 1.** The long-run industry supply function exists and it is given, for any $Q > 0$, by

$$P^S_{LR}(Q) = \min_{p, m} \bar{AC}(p, \Lambda(m, \delta)).$$

Theorem 1 extends the textbook characterization of the long-run supply function to a setting with heterogeneous firms. The long-run supply function is horizontal at a price that minimizes the average weighted cost function, where the minimum is with respect to both price and the marginal type. The average weight cost function is constructed using the measure $\Lambda(m, \delta)$, which can be viewed as the normalized steady-state cross-sectional distribution of firm types in a hypothetical world where firms survive with exogenous probability $\delta(1-\rho)$ and surviving firms exit endogenously if their type is lower than $m$.

In particular, Theorem 1 formalizes Marshall’s notion of a representative firm as a hypothetical firm with average cost function $\bar{AC}$. In the special case where all firms have identical marginal cost functions (as in the example), the average cost function of the representative firm, $\bar{AC}$, corresponds to a weighted average of the average cost functions.

**Corollary 1.** There exists a unique LRCE and it is characterized by positive entry and positive aggregate production.

*Proof.* Follows immediately from Proposition 1, Theorem 1, and the fact that assumption 6 and monotonicity of $\pi(\cdot, \theta)$ imply that $\min_{p, m} \bar{AC}(p, \Lambda(m, \delta)) < v$. 

### 3.1 Proof of Theorem 1

We will show that there is a unique solution $(p^e, m^e)$ to conditions

(ii) $\int_\Theta V(p, \theta)\nu(d\theta) = \kappa$, and

(iii) $\int_{\Theta} V(p, \theta')F(d\theta' | m) = 0$ if $m \in (\theta_L, \theta_H)$, $\geq 0$ if $m = \theta_H$, and $\leq 0$ if $m = \theta_L$. 

13
in Definition 2, and that this solution satisfies

\[(p^c, m^c) = \min_{p, m} AC(p, \Lambda(m, \delta)).\]

The proof has three steps. Throughout the proof, we let \(\varrho \equiv \delta(1 - \rho).\)

**STEP 1.** For any \((p, \theta) \in \mathbb{R}_+ \times \Theta\) and \(m \in \Theta,\) let

\[V_m(p, \theta) = \pi(p, \theta) + \varrho T_m[V_m(\cdot)](\theta)\]  

where \(T_m[g](\theta) = 1\{\theta \leq m\} \int_\Theta g(\theta')F(d\theta' \mid \theta).\) In words, \(V_m\) differs from the value function \(V\) defined in equation (1) in that it forces a possibly suboptimal exit decision threshold \(m.\)

Consider the system of equations:

\[\text{(ii')} \int_\Theta V_m(p, \theta) \nu(d\theta) = \kappa, \quad \text{and}\]
\[\text{(iii')} \int_\Theta V_m(p, \theta')F(d\theta' \mid m) = 0 \text{ if } m \in (\theta_L, \theta_H), \geq 0 \text{ if } m = \theta_H, \text{ and } \leq 0 \text{ if } m = \theta_L.\]

We will show that we can work with the system of equations (ii’)-(iii’) rather than (ii)-(iii).

**Lemma 2.** If \((p, m)\) is the unique solution to (ii’)-(iii’), then \((p, m)\) must also be the unique solution to (ii)-(iii).

**Proof.** Let \((p, m)\) be the unique solution to (ii’)-(iii’). In particular, \(m\) is the unique solution to (iii’) given \(p.\) Let \(m_0\) be the optimal exit threshold given \(p.\) In particular, \(\int_\Theta V_{m_0}(p, \theta')F(d\theta' \mid m) = 0 \text{ if } m \in (\theta_L, \theta_H), \geq 0 \text{ if } m = \theta_H, \text{ and } \leq 0 \text{ if } m = \theta_L.\)

Since \(m\) is the unique solution to (iii’) given \(p,\) it follows that \(m = m_0\) and, therefore, \(V_m = V_{m_0}.\) In addition, by optimality of \(m_0\) and the one-shot deviation principle, \(V_{m_0} = V.\) Therefore, \((p, m)\) solves (ii)-(iii). To show uniqueness, suppose that \((p', m')\) solves (ii)-(iii). Then \(V = V_{m'}\) and so \((p', m')\) must also solve (ii’)-(iii’). But since \((p, m)\) is the unique solution to (ii’)-(iii’), it must be that \((p', m') = (p, m).\)

**STEP 2.** In this step, we will show that conditions (ii’)-(iii’) can be equivalently expressed using weighted profit functions. This is one of the main insights of the proof and it relies on the concept of the adjoint of a bounded operator to identify the appropriate weight over profit functions.
For each $m \in \Theta$, define an operator $\Phi_m : \mathcal{M}(\Theta) \to \mathcal{M}(\Theta)$ such that, for all $A \subseteq \Theta$ Borel,
\[ \Phi_m[\eta](A) = \int_{\theta \in \Theta} F(A \mid \tilde{\theta}) \eta(\tilde{\theta}) \, d\tilde{\theta}. \]
$\Phi_m[\eta]$ gives the measure of types that results from applying the Markov operator $F$ to current types that are below the marginal type $m$, when the measure of current types is $\eta$.

The next result collects two useful properties of the operator $\Phi_m$.

**Lemma 3.** (i) For any $\varrho \in [0, 1)$ and $m \in \Theta$, $\sum_{j=0}^{\infty} \varrho^j \Phi^j m = (I - \varrho \Phi_m)^{-1}$ is a bounded operator from $\mathcal{M}(\Theta)$ to itself, where $I$ is the identity operator; (ii) For all $j$, $\Phi^j m$ is the adjoint operator of $T^j m$.

**Proof.** See the Appendix. 

Using the operator $\Phi_m$, $\mu_E$ can be alternatively written as
\[ \mu_E(n, m, \delta) = \nu n + \varrho \Phi_m[\mu_E(n, m, \delta)]. \]
Analogously, we can define $\mu_X(n, m, \delta) \in \mathcal{M}(\Theta)$ as the same measure, except that the distribution of entrants is the one facing the marginal exit type, $F(\cdot \mid m)$, i.e.,
\[ \mu_X(n, m, \delta) = F(\cdot \mid m)n + \varrho \Phi_m[\mu_X(n, m, \delta)]. \]

By Lemma 3(i),
\[ \Lambda(m, \delta) = \mu_E(n, m, \delta)/n = (I - \varrho \Phi_m)^{-1}[\nu] \quad (8) \]
and
\[ \Lambda_X(m, \delta) \equiv \mu_X(n, m, \delta)/n \equiv (I - \varrho \Phi_m)^{-1}[F(\cdot \mid m)]. \]

Our goal is to show that we can express the value functions in terms of weighted profit functions, with weights $\Lambda$ and $\Lambda_X$ for the entry and exit conditions, respectively. For this purpose, we define the weighted profit function $\bar{\pi} : [0, \infty) \times \mathcal{M}(\Theta) \to \mathbb{R}$, where
\[ \bar{\pi}(p, \eta) = \int \pi(p, \theta) \eta(d\theta) \]
for all \( p \geq 0 \) and \( \eta \in \mathcal{M}(\Theta) \). We then state the following two conditions, which the next lemma will show to be equivalent to conditions (ii')-(iii').

Condition (ii'). \( \pi(p, \Lambda(m, \delta)) = \kappa. \)

Condition (iii'). \( \pi(p, \Lambda_X(m, \delta)) = 0 \) if \( m \in (\theta_L, \theta_H) \), \( \geq 0 \) if \( m = \theta_H \), and \( \leq 0 \) if \( m = \theta_L \).

Lemma 4. \((p, m)\) solves (ii')-(iii') if and only if it solves (ii'')-(iii'').

Proof. By repeatedly applying equation (7), it follows that

\[
V_m(p, \theta) = \sum_{j=0}^{\infty} \rho^j T_m^j [\pi(p, \theta)](\theta).
\]

Then

\[
\int V_m(p, \theta) \nu(d\theta) = \int \sum_{j=0}^{\infty} \rho^j T_m^j [\pi(p, \cdot)](\theta)
\]

\[
= \int \pi(p, \theta) \left( \sum_{j=0}^{\infty} \rho^j \Phi^j_m [\nu](d\theta) \right)
\]

\[
= \int \pi(p, \theta) (1 - \rho \Phi_m)^{-1} [\nu](d\theta)
\]

\[
= \int \pi(p, \theta) \Lambda(m, \delta)(d\theta) = \bar{\pi}(p, \Lambda(m, \delta)),
\]

where the second line follows because \( \Phi^j_m \) is the adjoint operator of \( T_m^j \) (see Lemma 3(ii)) and the last line follows by definition of \( \Lambda \) in equation (8). A similar argument establishes \( \int_\Theta V_m(p, \theta') F(d\theta' \mid m) = \bar{\pi}(p, \Lambda_X(m, \delta)) \).

Step 3. We conclude the proof by showing that the solution to (ii'')-(iii'') is unique and minimizes the average weighted cost function.

Lemma 5. There is a unique \((p^e, m^e)\) satisfying conditions (ii'')-(iii''), and it is characterized by

\[
\{(p^e, m^e)\} = \arg \min_{p', m'} \bar{AC}(p', \Lambda(m', \delta)).
\]
Figure 3: Characterization of entry and exit conditions.

Proof. See the Appendix.

Figure 3 describes the intuition behind Lemma 5. The pair \((p^e, m^e)\) that solves (ii\(^\prime\))-\((iii\(^\prime\))\) is given by the intersection of the zero entry-profit schedule \(\bar{\pi}(p, w, \Lambda(m, \delta)) = \kappa\) and the zero exit-profit schedule \(\bar{\pi}(p, w, \Lambda_X(m, \delta)) = 0\) in the \((p, m)\) space. By a simple generalization of the textbook model, the former equation is equivalent to the condition that \(p = \bar{AC}(p, w, \Lambda(m, \delta)) = \min_{p'} \bar{AC}(p', \Lambda(m, \delta))\); denote the solution to this equation by \(\hat{p}(m)\). As illustrated by the figure, it is also the case that the zero exit-profit schedule intersects the zero entry-profit schedule at the minimum point of the latter. Thus, \(m^e\) minimizes \(\bar{AC}(\hat{p}(m), \Lambda(m, \delta))\). In other words, \((p^e, m^e)\) jointly minimize \(\bar{AC}\), as stated in Lemma 5.

The reason why the two schedules in Figure 3 intersect at the minimum point of the zero entry-profit schedule is as follows. Consider a point \((\tilde{p}, \tilde{m})\) on the zero entry-profit schedule such that \(\tilde{m} < m^e\). This point lies above the zero exit-profit schedule; that is, \(\bar{\pi}(\tilde{p}, \Lambda_X(\tilde{m}, \delta)) > 0\), and so the marginal type \(\tilde{m}\) makes a strictly positive profit. If the marginal type were slightly increased from \(\tilde{m}\) to \(\tilde{m} + \varepsilon\), then a potential entrant would get to stay whenever drawing a type in \((\tilde{m}, \tilde{m} + \varepsilon)\). By continuity, its profit from having a type in the interval would be positive, and so the firm’s ex-ante profit would increase from zero to a strictly positive number. The price would then need to fall in order to remain on the zero entry-profit schedule. Thus, the zero entry-profit schedule is decreasing whenever it is above the zero exit-profit schedule. By a similar
argument, the zero entry-profit is increasing whenever it is below the zero exit-profit schedule.

### 3.2 Equilibrium surplus

We conclude by comparing the equilibrium allocation with the allocation that maximizes steady-state aggregate surplus. We focus on the interesting case where \( m_e \neq \theta_L \), so that not all firms exit in equilibrium.

**Proposition 2.** Consider an LRCE such that \( m_e \neq \theta_L \). Then the equilibrium allocation maximizes steady-state surplus if and only if \( \delta = 1 \). Moreover, if \( \delta < 1 \), aggregate quantity is strictly lower and each firm’s individual quantity is weakly higher in the LRCE, compared to the surplus-maximizing, allocation.

**Proof.** A planner who wishes to maximize steady-state surplus must equalize marginal costs across all firms. Letting \( p \) be this common marginal cost, the planner’s problem becomes

\[
\max_{(Q, p, m)} \int_0^Q P^d(\bar{Q})d\bar{Q} - \int_{\Theta} C(q(p, \theta), \theta)\mu(m, n)(d\theta) - \kappa n
\]

subject to \( Q = \bar{q}(p, \Lambda(m, 1))n \). Substituting the constraint and using the definition of \( AC \), the problem becomes

\[
\max_{(Q, p, m)} \int_0^Q P^d(\bar{Q})d\bar{Q} - Q\bar{AC}(p, \Lambda(m, 1)).
\]

It is immediate that the planner’s solution is \( (p^*, m^*) = \min_{p, m} \bar{AC}(p, \Lambda(m, 1)) \) and \( P^d(Q) = \bar{AC}(p^*, \Lambda(m^*, 1)) \). By Theorem 1, the planner’s solution coincides with the LRCE allocation for \( \delta = 1 \).

For the second part, let \( \langle p_1^e, n_1^e, m_1^e \rangle \) and \( \langle p_2^e, n_2^e, m_2^e \rangle \) be LRCE for \( \delta_1 \) and \( \delta_2 \), respectively, where \( \delta_1 < \delta_2 \) and \( m_1^e > \theta_L \). Let \( V_1 \) and \( V_2 \) be the corresponding value functions. Then

\[
\int V_2(p_1^e, \theta)\nu(d\theta) > \int V_1(p_1^e, \theta)\nu(d\theta) = \kappa = \int V_2(p_2^e, \theta)\nu(d\theta),
\]

where the strict inequality follows from the fact that future payoffs are strictly positive (since \( m_1^e > \theta_L \) and \( \nu \) has full support) and the equality follows from the zero-profit
equilibrium condition. Since $V_2(\cdot, \theta)$ is nondecreasing, it follows that $p_2^\delta < p_1^\delta$. The second part of the claim then follows by setting $\delta_2 = 1$ and applying the first part. \[\square\]

Intuitively, the firms’ entry problem is a problem of experimentation. Firms don’t know their types ex-ante and, once they enter, they have the potential to become more productive over time. A long-lived planner who is more patient than the firms puts a higher value on entry because lower productivity firms exit and, in the long-run, only the more productive firms will remain in the industry. As Proposition 2 shows, the case of a planner who cares about long-run surplus is an extreme case where the planner is essentially infinitely patient. The divergence in discount factors implies that, under such planner, the steady state will be characterized by more productive firms, higher aggregate production, and higher consumer surplus, resulting in a higher overall aggregate steady-state surplus. This is not true, however, in the standard textbook model where firms are homogeneous and there is free entry.\(^9\)

We are not necessarily advocating for a planner who cares only about the steady state, although we note that, as Proposition 2 highlights, this is the implicit assumption whenever researchers focus on long-run equilibrium outcomes, which is a common approach.\(^10\) More generally, it does make sense for a long-lived planner, such as an antitrust authority, to have a higher discount factor than the firms. In that case, Proposition 2 continues to hold, as illustrated by Figure 4. As the discount factor increases from $\delta$ to $\delta' > \delta$, both the zero entry-profit and the zero exit-profit schedules characterized in the proof of Theorem 1 move to the right. The former does so strictly and the latter does so weakly in the special case of permanent types, and strictly in all other cases. Consequently, a planner who is more patient than the firms always wants the price to be lower than the LRCE price, meaning that he wants higher aggregate production. The effect on exit is, however, ambiguous, as it depends on how much each of the schedules shifts relative to the other one. It is always the case, however,

\(^9\)Notably, the LRCE allocation also fails to maximize surplus whenever entry costs are positive, even with homogeneous firms. The reason is that discounting once again becomes relevant with positive entry costs, since less patient firms discount future profits more relative to the initial entry cost.

\(^10\)An alternative is to take a explicit stand on the dynamics leading to equilibrium. For example, Hopenhayn (1992) studies a perfect foresight equilibrium of the dynamic environment, where an equilibrium entry condition holds every period. He shows that a perfect foresight equilibrium is efficient and that, if it converges, it converges to an LRCE. Thus, under the assumption of perfect foresight and convergence of equilibrium, one could interpret an LRCE as being efficient from the point of view of a planner who discounts the future at the same rate as the firms.
that the planner wants to encourage firms to stay more than desired at the planner’s optimal price $p^*$. 

Figure 4: Comparison of equilibrium vs. planner’s allocation ($\delta < \delta'$).
References

Allen, Roy G. D., London School of Economics, and Political Science, “Mathematical analysis for economists,” Technical Report, Macmillan London 1938.

Clementi, Gian Luca and Berardino Palazzo, “Entry, Exit, Firm Dynamics, and Aggregate Fluctuations,” American Economic Journal: Macroeconomics, July 2016, 8 (3), 1–41.

Ericson, Richard and Ariel Pakes, “Markov-perfect industry dynamics: A framework for empirical work,” The Review of Economic Studies, 1995, 62 (1), 53–82.

Hicks, John R., Value and Capital: An Inquiry Into Some Fundamental Principles of Economic Theory, Oxford: Clarendon Press, 1946.

Hopenhayn, Hugo A, “Entry, exit, and firm dynamics in long run equilibrium,” Econometrica, 1992, pp. 1127–1150.

Jovanovic, Boyan, “Selection and the Evolution of Industry,” Econometrica, 1982, pp. 649–670.

Kaldor, Nicholas, “The equilibrium of the firm,” The Economic Journal, 1934, 44 (173), 60–76.

Lucas, Robert E, “Adjustment costs and the theory of supply,” The Journal of Political Economy, 1967, pp. 321–334.

Lucas, Robert E. and Edward C. Prescott, “Investment under uncertainty,” Econometrica, 1971, pp. 659–681.

Marshall, Alfred, “Principles of economics,” 1890.

Opocher, Arrigo and Ian Steedman, “The industry supply curve: Two different traditions,” The European Journal of the History of Economic Thought, 2008, 15 (2), 247–274.

Pigou, Arthur C, “An analysis of supply,” The economic journal, 1928, 38 (150), 238–257.

Viner, Jacob, “Cost curves and supply curves,” in Readings in price theory, G. J. Stigler and K. E. Boulding, eds., 1953, pp. 198–232.
A Appendix

A.1 Proof of Lemma 3

Let \( L(\mathcal{M}(\Theta)) \) denote the space of linear bounded operators mapping \( \mathcal{M}(\Theta) \) to itself. (i) Since \( \|\Phi_m\| < 1 \) (here \( \|\cdot\| \) is the operator norm\(^{11}\)), it is easy to see that the sequence \( \left( \sum_{j=0}^{n} \phi^j \Phi_m \right) \) is Cauchy (under the operator norm). Because \( L(\mathcal{M}(\Theta)) \) is complete, then \( S = \sum_{j=0}^{\infty} \phi^j \Phi_m \in L(\mathcal{M}(\Theta)) \). It is easy to see that \( \phi \Phi_m S = S - I \) or, equivalently, \( (I - \phi \Phi_m)S = I \); similarly \( S(I - \phi \Phi_m) = I \). Therefore, \( S \) is the inverse of \( (I - \phi \Phi_m) \), denoted by \( (I - \phi \Phi_m)^{-1} \). (ii) Let \( g \in L^\infty(\Theta) \) and let \( \eta \) be any Borel measure of \( \Theta \). By Fubini’s Theorem,

\[
\int_{\Theta} T_m[g](\theta) \eta(d\theta) = \int_{\Theta} g(\theta') \left\{ \int_{1\{\theta \leq m\}} F(d\theta' | \theta) \eta(d\theta) \right\} = \int_{\Theta} g(\theta') \Phi_m[\eta](d\theta').
\]

Expression (9) can be equivalently be cast as \( \langle T_m[g], \eta \rangle = \langle g, \Phi_m[\eta] \rangle \), where \( \langle \cdot, \cdot \rangle \) denotes the integral operation. Using this notation, it is easy to see that, for any \( j \),

\[
\langle T_m^j[g], \eta \rangle = \langle T_m[T_m^{j-1}[g]], \eta \rangle = \langle [T_m^{j-1}[g]], \Phi_m[\eta] \rangle = \ldots = \langle g, \Phi_m^j[\eta] \rangle.
\]

\( \square \)

A.2 Proof of Lemma 5

Throughout this proof, we use the following properties for \( V_m \). The proof of these properties follow from standard fixed point arguments and are thus omitted: (1) For any \( m \in \Theta \), \( p \mapsto M[V_m(p, \cdot)](m) \) is nondecreasing and increasing over \( p \) such that \( q(p, m) > 0 \); (2) For any \( m \in \Theta \), \( \theta \mapsto M[V_m(p, \cdot)](\theta) \) is decreasing; (3) For any \( m \in \Theta \), \( p \mapsto M[V_m(p, \cdot)](m) \) is continuous.

Before proving Lemma 5, we state and prove two preliminary results.

Lemma 6. For any \( p > 0 \) and any \( m \in \Theta \) such that \( \bar{\pi}(p, \Lambda_X(m)) = 0 \), \( M[V_m'(p, \cdot) - V_m'(p, .)](\theta) < 0 \) for all \( m' \neq m \) and \( \theta \in \Theta \).

---

\(^{11}\)The space \( \mathcal{M}(\Theta) \) is equipped with the total variation norm and the operator norm \( \|\Phi_m\| \equiv \sup_{\eta \neq 0} \frac{\|\Phi_m[\eta]\|_{TV}}{\|\eta\|_{TV}} \leq 1 \) where \( \|\eta\|_{TV} \equiv 0.5 \int_{\Theta} |f_{\eta}(\theta)|d\theta \) where \( f_{\eta} \) is the Radon-Nikodym derivative of \( \eta \) with respect to Lebesgue.
Proof. Fix any $\theta \in \Theta$. We first show the result for $m' < m$. By definition of $V_m(p, \cdot)$,  

$$M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) = \varrho \int \{ 1\{ \theta' \leq m' \} M[V_{m'}(p, \cdot)] - 1\{ \theta' \leq m \} M[V_m(p, \cdot)] \}(\theta') \ F(d\theta' | \theta)$$

$$= \varrho \int 1\{ m \leq \theta' \leq m' \} M[V_m(p, \cdot)](\theta') F(d\theta' | \theta)$$

$$+ \varrho \int 1\{ \theta' \leq m' \} M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta') F(d\theta' | \theta)$$

$$\equiv A_{m',m}(\theta) + \varrho K_{m'} [M[V_{m'}(p, \cdot) - V_m(p, \cdot)]](\theta),$$

where $K_{m'} : L^\infty(\Theta) \to L^\infty(\Theta)$.

Observe that $M[V_m(p, \cdot)](m) = \bar{\pi}(p, \Lambda_X(m)) = 0$ and also $\theta \mapsto M[V_m(p, \cdot)](\theta)$ is decreasing, hence $1\{ m \leq \theta \leq m' \} M[V_m(p, \cdot)](\theta) < 0$, which implies $A_{m',m}(\cdot) < 0$ (note that $F(\cdot | \theta)$ has full support for all $\theta \in \Theta$ by Assumption 5(ii)). Since $\varrho||K_{m'}|| = \varrho \sup_{g \in L^\infty(\Theta)} \frac{||K_{m'}[g]||_{L^\infty(\Theta)}}{||g||_{L^\infty(\Theta)}} \leq \varrho < 1$, by the analogous arguments in the proof of Lemma 3,

$$M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) = (I - \varrho K_{m'})^{-1} [A_{m',m}](\theta) = \sum_{j=0}^\infty \varrho^j K_{m'}^j [A_{m',m}](\theta).$$

We note that for any $g(\cdot) < 0$, $K_{m'}[g](\cdot) = \int 1\{ \theta' \leq m' \} g(\theta') F(d\theta' | \cdot) < 0$. Hence, from this fact and the fact that $A_{m',m}(\cdot) < 0$, we can show inductively that for each $j$, $\varrho^j K_{m'}^j [A_{m',m}](\cdot)$ and thus $M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) < 0$.

We now show the case for $m' > m$. Following the same steps as those above one obtains

$$M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) = -A_{m,m'}(\theta) + \varrho K_{m'} [M[V_{m'}(p, \cdot) - V_m(p, \cdot)]](\theta).$$

Since $A_{m,m'}(\theta) = 1\{ m' \leq \theta \leq m \} M[V_m(p, \cdot)](\theta)$, it follows that $A_{m,m'}(\theta) > 0$. This observation and analogous derivations to the ones for $m' < m$ imply that $M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) < 0$. $\square$

Lemma 7. $(p, m) \mapsto \bar{\pi}(p, \Lambda(m))$ and $(p, m) \mapsto \bar{\pi}(p, \Lambda_X(m))$ are continuous.

Proof. We only prove continuity of $(p, m) \mapsto \bar{\pi}(p, \Lambda_X(m))$ since continuity of $(p, m) \mapsto \bar{\pi}(p, \Lambda(m))$ is obtained by an analogous argument. By definition of $V_m$, we want to show that $(p, m) \mapsto M[V_m(p, \cdot)](m)$ is continuous. Let $(p_n, m_n) \to (p, m)$ and note
that, for sufficiently large $n$,

\[
|M[V_{m_n}(p_n, \cdot)](m_n) - M[V_m(p, \cdot)](m)| \leq |M[V_{m_n}(p_n, \cdot)](m_n) - M[V_m(p_n, \cdot)](m_n)| \\
+ |M[V_m(p_n, \cdot)](m_n) - M[V_m(p, \cdot)](m)| \\
\leq \sup_{p \in C} ||M[V_{m_n}(p, \cdot) - V_m(p, \cdot)]||_{L^\infty} \\
+ |M[V_m(p_n, \cdot)](m_n) - M[V_m(p, \cdot)](m)|
\]

where $C$ is some compact neighborhood of $p$. The second term in the RHS vanishes because $(p, t) \mapsto M[V_m(p, \cdot)](t)$ is continuous (the proof follows from standard contraction mapping arguments and is omitted). Thus, the desired result follows by showing that, for sufficiently large $n$,

\[
|M[V_m(p, \cdot) - V_m(p, \cdot)](\theta)| \leq \varrho \int (1\{\theta \leq m_n\} - 1\{\theta \leq m\}) M[V_{m_n}(p, \cdot)](\theta') f(\theta' \mid \theta) d\theta' \\
+ \varrho \int (1\{\theta \leq m\}) M[V_{m_n}(p, \cdot) - V_m(p, \cdot)](\theta') f(\theta' \mid \theta) d\theta' \\
\leq \varrho |B_{m_n,m,p}(\theta)| + \varrho ||M[V_{m}(p, \cdot) - V_m(p, \cdot)]||_{L^\infty}.
\]

where $B_{m_n,m}(\theta) \equiv \int (1\{\theta \leq m_n\} - 1\{\theta \leq m\}) M[V_{m'}(p, \cdot)](\theta') f(\theta' \mid \theta) d\theta'$. Therefore, since $\varrho < 1$, it suffices to show that there exists a $\delta > 0$ such that $\limsup_{n \to \infty} \sup_{p \in C} ||B_{m_n,m,p}||_{L^\infty} = 0$. To do this, we first show that for each $\theta$, $\limsup_{n \to \infty} \sup_{p \in C} |B_{m_n,m,p}(\theta)| = 0$.

It is easy to show that there exists a $K < \infty$ such that $\sup_{p \in C} \sup_{m \in \Theta} ||V_m(p, \cdot)||_{L^\infty} \leq K$. So, for any $\theta' \in \Theta$,

\[
\sup_{p \in C} |(1\{\theta \leq m_n\} - 1\{\theta \leq m\}) M[V_{m_n}(p, \cdot)](\theta') f(\theta' \mid \theta)| \leq K |(1\{\theta \leq m_n\} - 1\{\theta \leq m\}) f(\theta' \mid \theta)|.
\]

Thus, for any $\theta' \neq m$, $\limsup_{n \to \infty} \sup_{p \in C} |(1\{\theta \leq m_n\} - 1\{\theta \leq m\}) M[V_{m}(p, \cdot)](\theta') f(\theta' \mid \theta)| = 0$. By the DCT, this readily implies that for any $\theta \in \Theta$, $\limsup_{n \to \infty} \sup_{p \in C} |B_{m_n,m,p}(\theta)| = 0$.

We now show that $\limsup_{n \to \infty} \sup_{p \in C} \sup_{\theta \in \Theta} |B_{m_n,m,p}(\theta)| = 0$. Since $\Theta$ is compact and we already established pointwise convergence, by the Arzela-Ascoli theorem it suffices to show that the family $\{\sup_{p \in C} |B_{m_n,m,p}(\cdot)|\}_{n \in \mathbb{N}}$ is equi-continuous. To do
this, note that for any $\theta$ and $\theta'$,
\[
\sup_{p \in C} |B_{m_n, m, p}(\theta') - B_{m_n, m, p}(\theta)| \leq \sup_{p \in C} \{ |B_{m_n, m, p}(\theta')| - |B_{m_n, m, p}(\theta)| \} \\
\leq \sup_{p \in C} \left| \int (1\{\theta \leq m_n\} - 1\{\theta \leq m\}) M[V_{m_n}(p, \cdot)](t) \left( f(t \mid \theta) - f(t \mid \theta') \right) dt \right| \\
\leq K \times |(f(t \mid \theta) - f(t \mid \theta')) dt|.
\]

The RHS is continuous by Assumption 5(ii), and its “modulus of continuity” does not depend on $m'$. Hence, $\{\sup_{p \in C} |B_{m', m, p}(\cdot)|\}_{n \in \mathbb{N}}$ is equi-continuous.

Proof of Lemma 5. Throughout the proof, we fix $\delta$ and omit it from the notation. We now define certain mappings that will be used throughout the proof.

Let $m \mapsto p_E(m) \equiv \{ p : \bar{\pi}(p, \Lambda(m)) = \kappa \}$, and $p \mapsto m_X(p) \equiv \{ m : \bar{\pi}(p, \Lambda_X(m)) = 0 \}$ and $m \mapsto p_X(m) = \{ p : m_X(p) = m \}$. For the mapping $m_X$, it is implicit that if $\bar{\pi}(p, \Lambda_X(m)) < 0$ then $m_X(p) = \theta_L$ and if $\bar{\pi}(p, \Lambda_X(m)) > 0$ then $m_X(p) = \theta_H$.

STEP 1. We now show that a solution to the system (ii’)-(iii’) exists and is unique and, moreover, we show that for any $(m, p)$ such that $\bar{\pi}(p, \Lambda_X(m)) = 0$ and $\bar{\pi}(p, \Lambda(m)) = \kappa$, then $p < p_E(m')$ for all $m' \neq m$, i.e., $m$ is a global minimizer of the function $p_E$.

Observe that by Assumption 5(i), $\nu(\{C(0, \theta) > 0\}) > 0$. Also, $\text{supp}(\Lambda(m)) \supset \text{supp}(\nu)$ for all $m$, so $\int C(0, \theta) \Lambda(m)(d\theta) > 0$. This implies that if $\bar{q}(p, \Lambda(m)) = 0$, then $\bar{\pi}(p, \Lambda(m)) < 0 \leq \kappa$, so a $(p, m)$ such that $\bar{q}(p, \Lambda(m)) = 0$ can never be a solution to $\bar{\pi}(p, \Lambda(m)) = \kappa$ (if it exists). Therefore, if the solution exists it would be such that $\bar{q}(p, \Lambda(m)) > 0$, in particular, this implies that $\bar{p} = 0$ cannot be part of a solution. Therefore, henceforth we focus on $(p, m)$ such that $\Lambda(m)(\{\theta : q(p, w, \theta) > 0\}) > 0$, in particular, we only consider $m \in M \equiv \{ m \in \Theta : \exists p : \Lambda(m)(\{\theta : q(p, \theta) > 0\}) > 0 \}$.

One of the following cases occurs: (a) $p_E - p_X < 0$; (b) $p_E - p_X > 0$ or (c) neither (a) nor (b) occurs (i.e., $p_E - p_X$ changes signs at least once in $\Theta$). If (a) occurs, then the solution to (ii’)-(iii’) exists and is given by $m = \theta_L$ and $p$ such that $\bar{\pi}(p, \Lambda(m)) = \kappa$ and $\bar{\pi}(m, \Lambda_X(m)) < 0$. Similarly, if (b) occurs, then the solution to (ii’)-(iii’) exists and is given by $m = \theta_H$ and $p$ such that $\bar{\pi}(p, \Lambda(m)) = \kappa$ and $\bar{\pi}(m, \Lambda_X(m)) > 0$. Therefore, if either (a) or (b) occurs a solution exists and is unique.

We now show that the same holds if (c) occurs. Clearly, for existence of a solution in this case it suffices that $m \mapsto p_X(m)$ is continuous (i.e., for any $(m_n)_n$ and $(p_n)_n$
such that \(m_n \to m\) and \(p_n \in p_X(m_n)\) with \(p_n \to p\) then \(p \in p_X(m)\) and closed-and convex-valued; and that \(m \mapsto p_E(m)\) is single-valued and continuous. Continuity of \(m \mapsto p_X(m)\) follows from Lemma 7; and by continuity and monotonicity of \(p \mapsto M[V_m(p, \cdot)](m)\), it follows that, for each \(m \in \Theta, p_X(m)\) is a closed interval. Since \(p \mapsto \pi(p, \theta)\) is nondecreasing and increasing over \(p\) such that \(q(p, \theta) > 0\) and \(\text{supp}(\Lambda(m)) \supseteq \text{supp}(\nu)\) for all \(m\), it follows that for any \(m \in M, p \mapsto \pi(p, \Lambda(m))\) is increasing. Hence, \(p_E(m)\) has at most one element. Moreover, since \(\pi(0, \Lambda(m)) \leq 0\) and \(\lim \inf_{p \to \infty} \pi(p, \Lambda(m)) = \infty\), continuity of \(p \mapsto \pi(p, \Lambda(m))\) ensures that \(p_E(m)\) is non-empty. Finally, continuity of \(m \mapsto p_E(m)\) follows from Lemma 7.

It thus remains to show that the solution in case (c) is unique. To do this, it suffices to show that for any \((m, p)\) such that \(\pi(p, \Lambda_X(m)) = 0\) and \(\pi(p, \Lambda(m)) = \kappa\), then \(p < p_E(m')\) for all \(m' \neq m\), i.e., \(m\) is a global minimizer of the function \(p_E\). Since \(p \mapsto \pi(p, \Lambda(m))\) is increasing, it suffices to show that for any \(m' \neq m\), \(\pi(p, \Lambda(m')) < \pi(p, \Lambda(m)) = \kappa\).

For any \(m_1 \leq m_2\), let \(\theta \mapsto A_{m_1,m_2}(\theta) \equiv 1\{m_1 \leq \theta \leq m_2\}M[V_m(p, \cdot)](\theta)\). Note that \(M[V_m(p, \cdot)](m) = \pi(p, \Lambda_X(m)) = 0\) and also \(\theta \mapsto M[V_m(p, \cdot)](\theta)\) is decreasing, so \(M[V_m(p, \cdot)](\cdot < (>)0\) for all \(\theta > (\cdot)m\). This, in turn, implies that \(A_{m_1,m}(. > 0\) and \(A_{m,m_2}(. < 0\).

By definition of \(V_m\), it follows that: If \(m' > m\),

\[
\pi(p, \Lambda(m')) - \pi(p, \Lambda(m)) = \int A_{m,m'}(\theta)\nu(d\theta) + \int 1\{\theta \leq m'\}M[V_m(p, \cdot) - V_m(p, \cdot)](\theta)\nu(d\theta)
\]

and if \(m' < m\),

\[
\pi(p, \Lambda(m')) - \pi(p, \Lambda(m)) = -\int A_{m',m}(\theta)\nu(d\theta) + \int 1\{\theta \leq m'\}M[V_m(p, \cdot) - V_m(p, \cdot)](\theta)\nu(d\theta).
\]

By our previous observations, \(\int A_{m,m'}(\theta)\nu(d\theta) < 0\) and \(-\int A_{m',m}(\theta)\nu(d\theta) < 0\). By Lemma 6 \(M[V_{m'}(p, \cdot) - V_m(p, \cdot)](\theta) < 0\) for any \(m' \neq m\) and any \(\theta \in \Theta\). So, \(\pi(p, \Lambda(m')) - \pi(p, \Lambda(m)) = \pi(p, \Lambda(m')) - \kappa < 0\) as desired.

**STEP 2.** We now show that the solution \((p^\epsilon, m^\epsilon)\) for (ii')-(iii'), which is unique (see Step 1), satisfies

\[
(p^\epsilon, m^\epsilon) = \arg \min_{p',m'} AC(p', \Lambda(m', \delta)).
\]
To show this, we first show that for any \( m, \bar{\pi}(p, \Lambda(m)) = \kappa \) iff \( p = \bar{A}C(p, \Lambda(m)) \) iff \( p = \min_{\nu > 0} AC(p', \Lambda(m)) \). The first ‘iff’ follows from simple algebra. To show the second ‘iff’, let \( p_m \equiv \inf_{\nu \in [p, \Lambda(m)]} p < \infty \), and that implies that \( \bar{A}C(p_m, \Lambda(m)) = \infty \) for all \( p \leq p_m \). Suppose for now (we show it below) that the following holds: (I) If \( p < \bar{A}C(p, \Lambda(m)) \), then \( \bar{A}C(p', \Lambda(m)) < \bar{A}C(p, \Lambda(m)) \) for all \( p' \) such that \( p < p' < \bar{A}C(p, \Lambda(m)) \); (II) If \( p > \bar{A}C(p, \Lambda(m)) \), then \( \bar{A}C(p', \Lambda(m)) > \bar{A}C(p, \Lambda(m)) \) for all \( p' > p \); and (III) There is at most one solution \( p \) to \( p = \bar{A}C(p, \Lambda(m)) \).

We claim that by (I) and the facts that \( \bar{A}C(p_m, \Lambda(m)) = \infty \) for all \( p \leq p_m \) and continuity of \( \bar{A}C(\cdot, \Lambda(m)) \) over \( p > p_m \), there exists a solution \( p \) to \( p = \bar{A}C(p, \Lambda(m)) \) and \( p > p_m \). To show this, suppose not, i.e., \( p < \bar{A}C(p, \Lambda(m)) \) for all \( p' \). This implies that there exists a \( p' \) such that \( \bar{A}C(p', \Lambda(m)) < \bar{A}C(p''', \Lambda(m)) \) for all \( p'' = \bar{A}C(p', \Lambda(m)) \) and this contradicts (I). By (I) and (II), this solution minimizes \( \bar{A}C(\cdot, \Lambda(m)) \), and, by (III), this is the unique solution.

We now prove (I)-(III). Let \( p, p' > p_m \) and \( p' > p \). By definition of optimality, \( pq(p, \bar{\pi}) - C(q(p, \bar{\pi}), \bar{\pi}) \geq \bar{q}(p', \Lambda(m)) - C(q(p', \bar{\pi}), \bar{\pi}) \). By simple algebra, integrating over \( \Theta \) using \( \Lambda(m)(p') \), and the fact that \( \bar{q}(p', \Lambda(m)) - \bar{q}(p, \Lambda(m)) > 0 \) (by the assumption that \( p' > p > p_m \)) and the fact that \( p \mapsto q(p, \bar{\pi}) \) is increasing for over \( p \) such that \( q(p, \bar{\pi}) > 0 \),

\[
p < \frac{\bar{C}(p', \Lambda(m)) - \bar{C}(p, \Lambda(m))}{\bar{q}(p', \Lambda(m)) - \bar{q}(p, \Lambda(m))} \leq p'.
\]

First, suppose that \( p < \bar{A}C(p, \Lambda(m)) \). Then (10) implies that, for all \( p' \) such that \( p < p' < \bar{A}C(p, \Lambda(m)) \),

\[
\bar{A}C(p', \Lambda(m)) \equiv \frac{\bar{C}(p', \Lambda(m))}{\bar{q}(p', \Lambda(m))} < \frac{\bar{C}(p, \Lambda(m))}{\bar{q}(p, \Lambda(m))} \equiv \bar{A}C(p, \Lambda(m)).
\]

Thus, (I) is proven. Next, let \( p > \bar{A}C(p, \Lambda(m)) \). Then (10) implies that \( \bar{A}C(p', \Lambda(m)) > \bar{A}C(p, \Lambda(m)) \) for all \( p' > p \); thus, (II) is proven. Finally, suppose \( p = \bar{A}C(p, \Lambda(m)) \) and \( p' = \bar{A}C(p', \Lambda(m)) \) with \( p' > p \). Putting together the two inequalities in (10), \( p = \bar{A}C(p, \Lambda(m)) = \bar{A}C(p', \Lambda(m)) = p' \), which contradicts \( p' > p \). A similar contradiction obtains if we assume \( p' < p \). Therefore, \( p' = p \), and so (III) is proven.

Note that \( p_E(m) = \arg \min_{\nu > 0} \bar{A}C(p', \Lambda(m)) \). Moreover, if \((m^e, p^e)\) solves (ii')-(iii'), \( p^e = p_E(m^e) \). So in order to show the desired result it suffices to show that \( m^e = \arg \min_{m \in \Theta} AC(p_E(m), \Lambda(m)) \), or equivalently, \( AC(p_E(m^e), \Lambda(m^e)) < AC(p_E(m), \Lambda(m)) \)
for all $m \neq m^c$. By step 1,
\[
p^c = p_E(m^c) < p_E(m')
\]
for all $m' \neq m$. Since, by our previous calculations in this step, \(AC(p_E(m), \Lambda(m)) = p_E(m)\) for all $m$, the desired result follows. \(\square\)