A Unified Axiomatic Theory of Microeconomics: Market Structure and Equilibrium

Du, G.

2025

A Unified Axiomatic Theory of Microeconomics: Market Structure and Equilibrium

G. Du

dxspacescience@gmail.com

Abstract

We propose a unified microeconomic theory that takes the behavioral rules of firms in real markets as the endogenous mechanism through which market structures are generated, replacing the standard practice of imposing perfect competition, monopoly, and other market forms as exogenous assumptions. Under a set of minimal axioms (consumer utility maximization, firm profit maximization, firm entry/exit mechanisms, and market clearing) and by introducing realistic cost and demand structures (firm-level cost heterogeneity, fixed costs, and demand elasticity, among others), pricing behavior and market structures long treated as exogenous (such as perfect competition and monopoly) are derived endogenously as equilibrium outcomes of firms' profit-maximizing behavior. Within a single framework, the theory nests traditional cost and marginal analysis, game-theoretic approaches to imperfect competition, and the Walrasian general equilibrium model; different regions of the parameter space naturally yield equilibrium outcomes corresponding to perfect competition, monopoly, monopolistic competition, and competitive-fringe structures. Our analysis shows that perfect competition is only a highly symmetric and intrinsically unstable equilibrium point: even small cost differences suffice to push the system toward more common monopoly or oligopoly configurations, while positive feedback mechanisms render these structures stable, helping to explain why market power is not easily competed away. Under empirically observable premises, the theory coherently derives the principal market forms and, in doing so, clarifies the domains of applicability of various classic models. It can also be used to predict which market structures industries will evolve toward under given conditions, providing a unified and operational theoretical foundation for empirical research, industrial policy, and firms' business and competitive strategy decision-making.

Keywords: Unified Market Structure Theory; Endogenous Structural Evolution; Imperfect Competition; Unified General Equilibrium Framework; Network Externalities and Feedback Mechanisms; Pareto Optimality under Heterogeneity

1 Introduction

In this paper, we propose an economic theory based solely on minimal assumptions, such as consumer utility maximization, firm profit maximization,the mechanism of firm entry and exit, and market clearing (Mas-Colell, Whinston and Green, 1995; Varian, 1992). This theory integrates diverse characteristics of cost and demand curves within a unified framework, explicitly incorporating the interaction between cost structures and product markets into a comprehensive analysis. Consequently, it systematically derives the long-term stability of oligopoly or monopoly structures under a broader set of conditions, while treating perfect competition merely as an extreme special case. This unified framework also positions previously independent models as points along a continuous spectrum, thereby resolving the previously fragmented nature among traditional models.

In conventional microeconomic theory, market structures are often predetermined as modeling assumptions: for instance, firms are assumed to be price-takers under perfect competition, implying $p = MC$, while in monopoly models, firms are assumed to possess pricing power, leading to $p > MC$. Such structural assumptions, however, typically lack behavioral foundations, failing to explain why and how these market structures emerge and remain stable in practice.

By substituting behavioral optimization logic for structural assumptions, and using minimal unified assumptions (with profit maximization as the core axiom), this paper assigns identical decision-making objectives to all firms. By incorporating realistic parameters such as cost heterogeneity, fixed costs, and market elasticity, market structures (price-taking versus price-setting behaviors) are no longer inputs but rather equilibrium outcomes that emerge naturally within the system. The theory presented here does not rely on specific functional forms; instead, it provides a microeconomic theory that derives market structures endogenously from realistic parameters.

Within this theoretical framework, we naturally unify three previously distinct approaches: the fixed cost/ACMC analysis of traditional economics, imperfect competition strategies from game theory, and the Walrasian general equilibrium framework (Tirole, 1988; Arrow and Debreu, 1954; Debreu, 1959). Firms are no longer always passive price-takers, nor are they confined to localized market games or idealized perfectly competitive equilibria; rather, their actions and pricing behaviors are embedded within a comprehensive, global equilibrium incorporating diverse market structures.We explicitly analyze how technological progress, heterogeneity in firm-level costs, network effects, and market elasticity jointly contribute to stable long-term monopoly or oligopoly configurations, and we point out that various forms of market barriers can naturally be represented as differences among firms in their cost and demand curves. Under this theory, we demonstrate that perfect competition emerges merely as a special case within specific parameter combinations, clearly identifying positive feedback mechanisms behind market power, thereby explaining how markets can spontaneously evolve toward stable monopolistic structures. Moreover, this theory provides a systematic method for integrating various independently assumed models into a unified continuum, addressing the apparent isolation or contradictions among different theoretical systems.

Under this perspective, traditional monopoly and perfect competition models are no longer distinct model types determined by external assumptions but rather corner solutions within a unified parameter space. Firms' pricing behaviors and market power are outcomes of their profit structures and competitive environments within the profit maximization framework, not predetermined by the model builder.Based on these distinct configurations, we identify perfect competition as one of several unstable symmetric points within the cost structure. Due to typical differences in cost structures, markets often bypass symmetric states corresponding to perfect competition, oscillating within their respective cost structures or stabilizing at asymmetric points under positive feedback mechanisms. Market structures more commonly evolve from a state dominated by one firm's market power to another firm's dominance, forming an asymmetric yet evolutionarily more sustainable dynamic equilibrium path.

In summary, this work presents a novel attempt at an axiomatic reconstruction of microeconomic theory.We select only a minimal set of basic, observable micro-foundations as axioms, reconstructing the entire theoretical system, and utilize it to uniformly explain diverse market structures, strategic interactions, factor linkages, and real-world industry phenomena.We derive market power, pricing authority, and structural asymmetry directly from first principles rather than from structural assumptions. It naturally derives and encompasses most existing sub-models. The approach presented is not merely a logical derivation; it also provides interfaces for empirical research and industrial policy analysis. It remains predictive and falsifiable, thus retaining the characteristics of scientific rigor.

2 Walrasian General Equilibrium: Theoretical Foundations and Limitations

2.1 Core Assumptions and Potential Limitations of the Walrasian Model

Perfect competition is traditionally viewed as the optimal state for resource allocation. However, this paper argues that within realistic structural mechanisms, perfect competition is merely a special and unstable corner solution. Our primary focus is on understanding structures that are capable of long-term stability, rather than identifying theoretically ideal conditions.

We begin by examining Walrasian equilibrium, a foundational framework in microeconomic theory. Its assumptions and applicability merit careful consideration. This paper systematically analyzes the explanatory power of the Walrasian model in various contexts, as well as its inherent limitations.

The Walrasian equilibrium model, along with the First Welfare Theorem, is built upon a series of stringent assumptions (Arrow and Debreu, 1954; Debreu, 1959; Mas-Colell, Whinston and Green, 1995; Varian, 1992). A critical assumption is the absence of market power, explicitly excluding market structures such as monopolies, oligopolies, and monopolistic competition. The essence of this assumption is that all market agents act as price-takers, and thus individual decisions do not influence market prices, fulfilling the essential criteria of perfect competition. However, specific special cases arise in practice: even when a single firm occupies the entire market share, if it is still regarded as a price-taker, this corner solution remains within the scope of Walrasian equilibrium. Nevertheless, under these circumstances, the traditional assumption of price-taking no longer aligns with realistic market behavior, necessitating a shift towards monopoly models. This illustrates a significant limitation of the Walrasian equilibrium under particular conditions. Building upon this understanding, this paper systematically categorizes multiple solutions that the Walrasian model may generate and clarifies under which specific conditions these solutions become invalid, as well as their economic implications.

2.2 Cost Structures: Analysis of Average and Marginal Costs

Within the Walrasian general equilibrium framework, fixed costs are typically excluded to eliminate the effects of increasing returns to scale. Our analysis primarily concentrates on critical parameters associated with monopoly, such as production volume and market share. Firms determine their output according to the principle of $p = MC$, meaning the market equilibrium price p equals marginal cost MC. The market-clearing condition is thus given by $X = \sum_{i=1}^{n} q_i$, where X denotes total product demand, and q_i represents the output of each firm. We do not restrict the specific form of the production function; one could adopt standard economic forms such as the Cobb-Douglas function $q = AK^{\beta}L^{\alpha}$ or any other form reflecting industry characteristics. The specific functional form chosen does not affect the subsequent conclusions. It is important to emphasize that, for better alignment with reality, the MC and AC functions discussed below do not need explicit analytical forms but can be fitted directly from actual firm-level data. As long as the fitted functions are approximately continuous and satisfy the required shapes, all subsequent results can be derived. Specifying particular functional forms is simply to adhere to standard economic notation.

Let each firm have a cost function denoted as $C(q)$. The marginal cost function can then be expressed as $MC(q) = dC(q)/dq$. Notably, the marginal cost curves can differ across firms. For any $q \in [q_{min}, q_{max}]$ or $q \in D$, we denote differences among firms as $MC_2(q) = MC_1(q) + \Delta(q)$, where $\Delta(q) \geq 0$, or equivalently, we have $MC_1(q) \leq MC_2(q) \leq \cdots \leq MC_n(q)$ for every feasible output level q.

Under the standard equilibrium solution, firms choose their production level based on marginal cost when $p = MC$. If, as output approaches zero, firms still face $p < MC$, they opt to produce zero output. In equilibrium models, firms producing zero output in reality can be viewed as having exited the market or having become bankrupt, corresponding to the conditions $p < \min AVC$ or $p < AC$. Furthermore, Walrasian equilibrium requires that total consumer demand equals the total production of all firms.

The Walrasian model primarily addresses scenarios where fixed costs are negligible. Firms' short-term shutdown condition is $p < \min AVC$, while the long-term market exit (or entry) condition is $p < \min AC$ (or equivalently $p \geq \min AC$). When fixed costs are negligible and there are no increasing returns to scale, this condition equates to p being lower than marginal cost at zero output. For analytical consistency, the subsequent analysis will use the relationship between p and the marginal cost at zero output as a baseline condition. (In cases outside the Walrasian context, involving fixed costs or increasing returns to scale, the criterion $\min AC$ will instead be employed.)

3 Multiple Forms of Marginal Cost under Walrasian Equilibrium: Classification and Analysis

Next, this research will classify and analyze market equilibria based on the characteristics of firms' marginal cost curves, specifically examining cases with different returns to scale. We primarily focus on equilibrium processes within a single market, illustrating how relaxing the price-taker assumption within the Walrasian equilibrium framework (such as utility/profit maximization of consumers and firms, market clearing, etc.) naturally introduces monopolistic or imperfect competition behaviors. The reason for this approach is that imperfect competition itself involves situations where certain firms influence prices rather than passively accepting them. Thus, once we handle the single-market scenario, we can generalize this method to situations involving imperfect competition in other markets, subsequently solving simultaneously the equations for all goods markets.

3.1 Identical Marginal Cost Curves Across Firms

First, consider the scenario where all firms have identical marginal cost curves.

3.1.1 Constant Marginal Cost

1. Constant marginal cost (constant returns to scale): For example, but not limited to, the Cobb-Douglas production function with $\alpha+\beta=1$. In this case, the equilibrium allows firms to freely choose their production quantities q_i. Production may either be entirely by a single firm or arbitrarily allocated among n firms, with the sole constraint that total production equals total consumption. Given the universality of this condition, we will not reiterate it further. Even if one firm supplies the entire market, it cannot independently raise prices since any increase above $p=MC$ allows other firms to enter and capture the whole market. Thus, it can still be viewed as a special form of perfect competition, with arbitrary allocation of market shares due to the indeterminate scale of each firm.

3.1.2 Increasing Marginal Cost

2. Increasing marginal cost (decreasing returns to scale): For instance, when $\alpha+\beta<1$. In the short run, under restrictions preventing free entry and exit, each firm operates at the same finite scale. If entry and exit are free, each firm tends towards minimal production, resulting in nearly infinite firms, aligning precisely with the idealized perfectly competitive state. (In the presence of fixed costs, firms choose the production corresponding to $p=\min AC$ to reach equilibrium.)

3.1.3 Decreasing Marginal Cost

3. Decreasing marginal cost (increasing returns to scale): Such as $\alpha+\beta>1$. Although not consistent with Walrasian equilibrium assumptions, this scenario is still discussed here for comparison. Under this cost structure, if two or more firms compete, even if they initially produce similar quantities, any firm that achieves scale expansion will lower average costs and increase profits, facilitating further expansion, thus squeezing competitors, ultimately

possibly leading to monopoly or oligopoly market formation. Notably, increasing returns to scale usually assume fixed costs, and from a long-term perspective, entry decisions should be based on $\min AC(Q)$.

3.1.4 U-shaped Marginal Cost Curve

Additionally, there is a typical scenario: scale economies exist at lower outputs, but diseconomies arise beyond a certain threshold (U-shaped MC curve). Equilibrium outcomes depend on where the equilibrium price $p = MC$ falls on the U-shaped MC curve. If on the upward-sloping segment, it corresponds to scenario 2; if near the minimum point, scenario 1; if on the downward-sloping segment, scenario 3. If firms can produce products of various qualities or varieties considered collectively, this curve can approximate reality. For example, when a large firm enters niche segments dominated by smaller firms, it might face increasing marginal costs (the upward-sloping segment), thus allowing coexistence; or, when a monopolist firm expands into new industries due to financial strength but faces rising marginal costs due to unfamiliarity, the expansion may fail, and it remains monopolistic only in its original market. (Of course, precise representation requires the actual cost functions of entering new areas.)

3.2 Heterogeneous Marginal Costs Across Firms

Next, we explore market structures where firms have heterogeneous marginal costs.

For simplicity, assume for all $q \in [q_{min}, q_{max}]$ or $q \in D$, we have $MC_2(q) = MC_1(q) + \Delta(q)$, where $\Delta(q) \geq 0$, or equivalently $MC_1(q) \leq MC_2(q) \leq \cdots \leq MC_n(q)$, representing differences in marginal costs among firms for each potential output q. We do not consider intersecting MC curves between firms here.

3.2.1 Constant Marginal Cost

4. Constant marginal cost (constant returns to scale): Suppose differences in technological capabilities, managerial efficiencies, or decision-making skills result in different marginal costs across firms. As total market output increases, the price p decreases. When the lowest marginal-cost firm satisfies $p = MC_{lowest}$, all other firms necessarily face $p < MC_i$. Because marginal costs do not vary with output and $p < \min(AVC) = MC_i$, the optimal strategy for these higher-cost firms is to exit production, creating a monopoly by the lowest-cost firm. Under Walrasian equilibrium, this situation appears as a corner solution. Formally, under Walrasian equilibrium, this monopolist remains a price-taker, with the price determined by the condition $p = MC_{lowest}$ and market-clearing condition $X = q_{lowest}$. However, effectively, this solution fulfills market-power conditions, constituting a monopoly whose pricing behavior will no longer adhere to Walrasian conditions. This occurs because the monopolist could slightly raise prices, and other firms would still find $p < MC_i$, thus not re-entering production, making the monopolist effectively a price-maker.

Next, we analyze the actual pricing behavior of a monopolist. Considering demand elasticity and profit maximization, we derive the classical monopoly pricing formula:

$$p^* = \left[\frac{\varepsilon(Q)}{\varepsilon(Q)-1}\right] \times MC_{lowest}$$

(Lerner, 1934).

When the monopolist reduces production to increase the price for higher profits, if the price rises to the point of other firms' marginal cost ($P = MC$), these firms re-enter the market, breaking the monopoly. The gap between the lowest marginal cost firm (MC_{lowest}) and the second-lowest marginal cost firm (MC_{2ndLow}) acts as an entry barrier. The monopolist's maximum sustainable price is thus $p = MC_{2ndLow}$ (Bain, 1956). If this price exceeds the profit-maximizing price p^*, the monopolist chooses p^*; if lower than the profit-maximizing price, the monopolist sets a price slightly below MC_{2ndLow} to prevent entry.

If several firms share the same lowest marginal cost while other firms have significantly higher costs, and no collusion exists, scenario 2 applies. Among these few firms, output allocation is arbitrary under Walrasian equilibrium without a unique solution. Continuity suggests current production distribution persists. If these firms collude to jointly control output or prices, the market structure transforms into oligopoly.

3.2.2 Increasing Marginal Costs

5. Increasing marginal costs, corresponding to decreasing returns to scale. Under these circumstances, cost differences among firms may lead to various market structures as follows.

Different Scenarios

5.1. When the lowest marginal-cost firm cannot reduce its marginal cost sufficiently to satisfy total market demand alone, other firms can adjust their output to lower their marginal costs and achieve equilibrium at $P = MC$. This results in a competitive fringe structure composed of one dominant firm, several secondary firms, and numerous small firms. This scenario has two distinct cases:

5.1.1. When the smallest producing firm sets $p = MC(q)$, the marginal cost of non-entering firms at near-zero output is essentially the same as that of the smallest producing firm. This means there is no discontinuity in the marginal cost function at the equilibrium production level. Small firms operate at $p = MC(q)$ with $q \approx 0$, while dominant and secondary firms earn positive economic profits but remain price-takers, as any reduction in output would immediately invite entry by new firms. All active firms earn positive economic profits (the smallest firm's profit can be infinitesimally small or finite, depending on output size; continuity at the equilibrium MC point is required). This scenario is Pareto optimal, and we define it as competitive-fringe price-taker equilibrium.

5.1.2. When the smallest producing firm sets $p = MC(q)$, but the marginal costs at near-zero output for non-entering firms have a finite gap relative to the smallest producing firm, there is a discontinuity at the equilibrium output level. All firms still earn positive economic profits due to this marginal-cost gap.

5.1.2.1. Under specific conditions, firms may reduce their output to elevate prices, thus becoming price makers instead of price takers. For instance, consider the extreme case where firms currently producing have reached their output limits, and any further output would cause their costs to rise steeply (we approximate this scenario by assigning infinite marginal costs beyond the equilibrium output). Thus, total output q_{total} from these firms remains fixed, and one firm can strategically reduce output within residual demand $X(p) - q_{total}$, increasing prices to monopolistic levels and becoming a price maker.

5.2. When the lowest marginal-cost firm is capable of fully occupying the market:

5.2.1. Suppose the lowest marginal-cost firm reduces its marginal cost sufficiently to satisfy total market demand alone, and the resulting equilibrium price p exactly equals the marginal cost at near-zero output of other firms. This means marginal costs remain continuous, with no discontinuity at the equilibrium point. Although the lowest-cost firm dominates the market, it remains a price-taker, yet still earns positive economic profits.

5.2.2. Suppose the lowest marginal-cost firm reduces its marginal cost sufficiently to satisfy total market demand alone, but the resulting equilibrium price p is lower than the marginal cost at near-zero output of other firms, creating a significant gap and naturally resulting in a monopoly. This corner solution superficially appears to satisfy price-taking conditions, but it effectively forms a monopoly. The monopolist does not need to reduce prices to force out all competitors, as slightly increasing prices could yield additional profits that outweigh the loss caused by re-entry of competitors. Thus, the profit-maximizing outcome might involve complete market occupation or allowing some firms to remain in production.

5.2.2.1. If the profit-maximizing price allows partial entry of other firms, the monopolist incorporates the second-lowest-cost firm's cost function into its optimization to satisfy $p = MC$ alongside demand elasticity, determining the monopolist's optimal output and profit. Comparing full monopoly to partial market participation scenarios, the monopolist selects the most profitable strategy. Whether fringe firms exist or not, this structure is not Pareto optimal. It also does not strictly align with the four typical structures distinguishing monopoly from perfect competition (perfect competition, monopolistic competition, oligopoly, monopoly), since it neither fully monopolizes the market nor involves explicit collusion but still sets prices. We thus define it as a monopoly structure with a competitive fringe.

Intermediate Summary

Intermediate summary: From the analysis above, although economics traditionally suggests that firms will enter the market whenever profits exceed the normal profit rate, we see here that if potential entrants cannot achieve normal profit levels due to cost disadvantages, they cannot disrupt the economic profits of established firms. Here we note a distinction between traditional perfect competition definitions and Walrasian definitions. For scenarios 5.1.1 and 5.2.1, firms are considered price-takers, and under Walrasian equilibrium, this might sometimes be termed competitive equilibrium or even perfect competition equilibrium. However, under traditional economics definitions, positive economic profits suggest implicit barriers (such as technological advantages), thus disqualifying these cases as perfectly competitive. Ideally, economics should standardize its definitions to prevent confusion. Here, we adopt the traditional definition of perfect competition, requiring zero economic profit. Traditional definitions argue that, in the long run, due to technology diffusion, no barriers exist, implying firms eventually converge to identical cost functions, corresponding only to scenarios 1 and 2. However, this assumption is questionable, and later analysis will show cost functions may not necessarily converge in the long run. Additionally, we observe that differences in costs inherently generate economic profits.

3.2.3 Decreasing Marginal Costs

6. Decreasing marginal costs, i.e., increasing returns to scale. Although this situation is not permitted within the Walrasian equilibrium framework, the following analysis is provided for comparative purposes. Similar to the previously discussed third scenario, two possibilities arise: First, a single firm achieves a monopoly if it attains the lowest average cost (AC) when its output matches total market demand X. Second, a financially strong firm acquires the lowest-AC firm to secure its technology, thus realizing the minimum achievable AC. Typically, only firms that reach minimum AC at monopolistic scale can form stable equilibria; otherwise, another firm with lower AC could expand production, displacing higher-cost competitors. Given our assumption that $MC_1(Q) \geq MC_2(Q) \geq \cdots \geq MC_n(Q)$ holds across all output levels, the firm with the lowest AC at monopoly output will also have the lowest MC.

3.3 Analysis and Summary

It should be emphasized that the Walrasian general equilibrium theory requires integrating analyses across multiple commodity markets, factor markets, and consumer preferences, whereas partial equilibrium analysis focuses only on a single market. Therefore, conditions within a specific market must be solved simultaneously alongside conditions from other commodity markets.

Based on analyses from scenarios 4 and 5, it is evident that under conditions where the Walrasian model applies, differences in marginal costs often lead to scenarios where a single firm dominates the entire market. Here, the equilibrium solution can first be derived via standard Walrasian equilibrium methods. Subsequently, markets exhibiting single-firm monopolies should replace the price-taking condition with monopoly pricing. Specifically, one must maximize profits of the lowest-MC firm while ensuring the second-lowest-MC firm satisfies $p = MC$, resulting in the final equilibrium solution. For non-monopolized commodities, equilibrium conditions remain consistent with the standard Walrasian approach. These conditions must also be solved simultaneously with the condition that marginal utility per monetary unit spent across goods is equalized, as well as conditions in factor markets, thus obtaining a general market equilibrium. Notably, although the transition from perfectly competitive output to monopolistic output typically involves reduced total demand due to increased prices, such demand shifts do not affect the fundamental establishment of monopoly status. Determining the monopolists optimal pricing strategy, however, requires simultaneous consideration of pricing strategies across all markets. If the demand function is assumed fixed, the analysis can be simplified to partial equilibrium analysis within a single market.

Further consideration reveals that changes in product market outputs impact factor prices. If all firms utilize identical factor input proportions and factor prices change uniformly, marginal costs (MC) across all firms will adjust proportionally, preserving relative MC gaps among firms. Hence, the monopoly conclusion remains unchanged, though specific MC values must still be derived within the general equilibrium framework.

If firms differ in their input proportions or factor prices adjust non-uniformly, explicit simultaneous solving within the general equilibrium framework is required to determine each firms MC and the resulting market structure. Alternatively, if output changes minimally affect factor market equilibrium, factor prices may be treated as exogenously fixed, simplifying the scenario back into partial equilibrium analysis.

4 Extension to the Arrow-Debreu General Equilibrium Model

4.1 Standard Arrow-Debreu General Equilibrium Model Scenario

7. In the standard Arrow-Debreu general equilibrium framework, firms can produce multiple products simultaneously, provided the assumption of convexity in production sets is maintained. This convexity assumption implies the absence of significant "synergies" or "economies of scope," meaning joint production does not significantly lower costs compared to separate independent production processes. Under this convexity condition, production decisions for each good are effectively independent. Therefore, as long as there exist positive market prices, firms set output levels individually for each product, satisfying:

$$p_i = \frac{\partial C}{\partial q_i},$$

which aligns with the Walrasian equilibrium condition $p = MC$ for each product. In particular, a firm may dominate or even monopolize a market if it has a significantly lower marginal cost in product A compared to competitors. Conversely, if its marginal cost in product B is relatively high, it might choose not to produce product B at all. This multi-product scenario can therefore be analyzed within the same single-product Walrasian equilibrium framework described earlier.

4.2 Extension with Synergy Effects

8. When significant "synergy effects" (economies of scope) (Panzar and Willig, 1981; Tirole, 1988) exist in practice, implying that the total cost of jointly producing multiple products is lower than producing them separately, the firm's production set will exhibit non-convex characteristics. Under these circumstances, the classical Arrow-Debreu analytical framework cannot simply decompose multi-product output decisions into independent $p = MC$ decisions. Firms with synergistic production capabilities can simultaneously supply multiple markets at significantly lower costs than competitors. On one hand, their joint production cost advantages are evident; on the other hand, in price-taking or more general dynamic competitive situations, they can strategically lower prices slightly, making it unprofitable for single-product competitors even at minimal output levels, thus forcing their exit. As competitors gradually exit or are acquired, firms with synergistic advantages may simultaneously achieve dominance (or even monopoly) across multiple product markets. This indicates that

once production sets exhibit non-convexity, the Walrasian equilibrium scenario of "multiple firms coexisting" may collapse. Due to the absence of a simple single-product cost concept here, in principle, a multi-product joint analysis is necessary, making the single-product cost somewhat arbitrary. To simplify analysis, we can approximate single-product costs by allocating total costs proportionally based on the existing market prices of each product. Specifically, suppose a firm produces m products, with market prices p_i, and total cost

$$C_{\text{total}}(q_1, \dots, q_m),$$

and corresponding revenues $R_i = p_i q_i$. Then, the allocated cost for product i can be defined as

$$\widetilde{C}_i = \frac{R_i}{\sum_{k=1}^{m} R_k} \times C_{\text{total}}(q_1, \dots, q_m),$$

and the single-product marginal cost approximated as

$$\widetilde{MC}_i = \frac{\partial \widetilde{C}_i}{\partial q_i}.$$

Under the assumption that other product outputs remain unchanged, approximate analysis can be carried out based on the single-product market scenario discussed previously, though this method does not yield the optimal solution.

For a rigorous solution, a simultaneous multi-product analysis is necessary. Here, we must be cautious when assuming one product as a price-taker and another as a price-maker because, once $p_A > MC_A$ and $p_B = MC_B$, the firm can always reallocate costs from the assumed lower-cost product (B) to the higher-cost product (A), effectively making both products price makers. Because there is only one joint cost, the cost allocation among individual products is flexible. Thus, in solving, we must check whether the total existing market prices (or prices of potential producers) $p_A + p_B$ exceed our joint marginal cost MC_{A+B}. If they are equal, products can be sold at market prices. If greater, once we have determined the cost curves of competitors and our joint cost function $C(q_A, q_B)$, we proceed as follows:

1. First, we solve for the unconstrained optimal pricing solution $(p_A^*, p_B^*, \dots)$. Restrictive pricing here refers to competitive market prices prior to entry, or the average cost (AC) of firms we intend to drive out. If all resulting prices comply with their restrictive upper limits (none exceed them), the unconstrained solution is globally optimal. If some product prices (for example, p_A^*) exceed the restrictive upper bound $\bar{p}_A$, the solution becomes infeasible.

2. For prices violating constraints, fix them at their upper limits, then optimize remaining decision variables. If it is found that $p_A^* > \bar{p}_A$, we fix $p_A = \bar{p}_A$ as a known constant in the profit function and allow another price p_B (or other variables) to adjust freely to derive a new optimal solution. If multiple product prices exceed their limits, each is similarly fixed at its upper bound, and optimization is performed on remaining decision variables. This approach yields multiple candidate solutions (e.g., scenarios with one product at the limit or multiple products at their limits), each of which should be compared in terms of profit.

In multi-product scenarios, once one price is fixed at an upper bound, optimal prices for other products must be recalculated and cannot rely on single-product formulas. Ultimately, among these candidate scenarios, the feasible solution yielding the highest total profit must be identified.

Since demand and costs may exhibit mutual influences (cross-elasticities, synergies, joint outputs, etc.), the solution process is more complicated than single-product market analysis, requiring consideration of competitor cost functions across multiple markets to determine joint behavior. Moreover, the joint cost function may exhibit increasing, constant, or decreasing returns to scale, each requiring distinct treatment. Nevertheless, viewed from individual product markets, the outcomes still correspond to the previously classified scenarios.

5 Comprehensive Analysis and Discussion of Various Scenarios

It is noteworthy that due to the presence of fixed costs, firms typically experience an initial stage of increasing returns to scale caused by the dilution of these fixed costs (exhibiting the same trend in terms of total cost changes). Subsequently, firms achieve constant returns to scale by replicating the optimal production structure. Ultimately, as the scale expands, organizational structures may become bloated or firms may enter unfamiliar fields, thus leading to decreasing returns to scale. Therefore, whether a firm with an AC/MC advantage eventually achieves market monopoly depends on the comparison between the target market size and the firms scale return characteristics. Meanwhile, other products, due to cross-elasticities within the Walrasian equilibrium framework, have prices that also influence the market size of this product. As long as the effect of decreasing returns to scale remains insignificant, the equilibrium solution will tend towards monopoly by a single firm.

When a single firm reaches the stage of decreasing returns to scale, the market resembles scenarios 2 or 5 discussed previously, potentially resulting in either multiple firms coexisting in a Pareto-optimal state or the scenario described in 5.2, where a single firm may still wield significant market power. However, there exists a structurally similar yet fundamentally different market structure: differentiated competition or monopolistic competition (Chamberlin, 1933). The core distinction here is the differentiation among firms' products, which are only partially substitutable. Typically, firms achieve monopolies within their segmented markets, exhibiting increasing, constant, or decreasing returns to scale within that segment. However, their average cost (AC) is usually higher than that achievable by other firms if they were to produce identical products. Otherwise, competitors would enter that segment, breaking the differentiated competition pattern. Firms maintain differentiation by continually identifying niches that competitors are less proficient in serving. Under these circumstances, competitors attempting cross-market entry usually face higher AC; if a competitor's AC were lower, it would successfully displace the incumbent, establishing cross-market monopoly. Pricing strategies in differentiated markets thus coincide with either local monopolistic pricing or the lower price derived from AC/MC differences.

It is necessary to emphasize that the scenarios (1-6) discussed previously do not deeply explore cross-market competition. Typically, a firm's experiential advantage in its original market cannot be fully replicated in new markets, resulting in higher AC compared to existing competitors. Such scenarios fundamentally differ from situations 2 or 5, characterized by perfect competition. The core difference between differentiated competition and normal cross-market goods lies in partial substitutability, reflected in significant differences in cross-elasticity of demand. For example, differentiated products A and B compete through price reductions in product A, increasing its demand and consequently decreasing demand for product B, without altering the internal market structures of products A or B (i.e., a monopolistic market structure remains monopolistic, albeit smaller). Thus, whether a market is monopolistic depends solely on the AC/MC differences among firms producing that good, independent of other goods' prices.

In dynamic market scenarios, two typical cases arise: first, all firms initially exhibit similar costs, but subsequently, one firm achieves significantly lower average cost through technological or efficiency improvements; second, a monopolistic incumbent exists, but a new entrant attempts to disrupt the existing market by leveraging lower average costs.

For new entrants experiencing increasing returns to scale, the optimal strategy is rapid expansion or mergers and acquisitions with existing firms to leverage established capacity and capitalize on cost advantages for greater economic profits. If expansion is delayed, new firms risk losses before achieving optimal scale. Until the entrant reaches minimum cost, the incumbent can maintain monopolistic pricing, adjusting only as necessary according to the entrants evolving cost structure. The pricing formula matches that described previously in scenario 4.

If a new entrant exhibits constant returns to scale with costs lower than the incumbent at any output level, but its production capacity is insufficient to dominate the entire market ($X > q_{new,max}$), the incumbent, having previously covered the entire market, effectively possesses infinite production capacity. This scenario resembles case 5.1.2, where the incumbent's optimal strategy involves allowing the entrant to maximize its capacity, then maintaining monopolistic pricing on residual demand, $X(p) - q_{new,max}$. Even if the incumbent lowers prices to its AC, the entrant, with consistently lower costs, will still fully utilize its capacity. Thus, a critical mechanism emerges: despite new firms with lower AC, incumbents retain pricing power by controlling residual demand. This dynamic illustrates market transition, wherein shrinking incumbents eventually experience rising costs due to decreasing output and exit the market. Notably, market power (monopoly) persists throughout this transition, and unusually, the marginalized incumbent retains pricing authority without collusion. This analysis underscores that allowing for explicit market power and clearly defined MC and AC differences within an expanded Walrasian (Arrow-Debreu) framework facilitates a more nuanced parameter-driven analysis, revealing more detailed market power configurations distinct from standard monopolistic models.

6 Typical Scenarios with Non-Negligible Fixed Costs

Scenarios involving non-negligible fixed costs exhibit high similarity to those previously discussed regarding increasing returns to scale. Specifically, fixed costs effectively reduce average costs, mirroring marginal cost curves initially displaying increasing returns, enabling unified analytical treatment. Therefore, the following analysis systematically explores various detailed market scenarios under conditions of non-negligible fixed costs combined with distinct marginal cost curves. Given the similarity between fixed-cost scenarios and initial increasing returns scenarios (as fixed costs and initial increasing returns of the marginal cost curve have analogous effects), further segmentation arises based on differing marginal costs. We will now examine these scenarios in detail.

6.1 Homogeneous Firms: Cases Without Differences in Cost Curves

We first analyze an ideal scenario in which firms have identical marginal costs $MC(q)$ and identical fixed costs, hence no differences in cost structures between firms.

6.1.1 Constant or Increasing Returns to Scale

9. In cases of constant or increasing returns to scale (i.e., constant or decreasing MC), the short-run equilibrium satisfies $p = MC$, but due to fixed cost amortization, this leads to $p < AC$, causing firms to incur losses. In the long run, continuous losses drive firms to exit the market until only a few remain. If these firms continue to compete according to Bertrand competition, further exits occur, eventually leaving only one firm (or a single firm formed through mergers). A typical example is a cloud service provider operating data centers, exhibiting constant returns to scale.

From the classical derivation perspective of $p = MC = AC$, the average cost (AC) curve with fixed costs usually appears U-shaped (corresponding to decreasing returns to scale). The MC curve initially remains below the AC curve and intersects it at its minimum point. However, when market size is small or when the quantity needed to achieve economies of scale is large, actual output is often situated on the downward-sloping or flat portion of the MC curve. Thus, the AC curve effectively resembles the left portion of the U-shaped curve, exhibiting a downward-sloping L-shape (corresponding to constant or increasing returns to scale). In this scenario, the MC curve does not intersect with the minimum point of the AC curve at finite output levels, leading to the situation described earlier. Indeed, under constant returns to scale, MC intersects AC at infinite output levels. However, in reality, since output levels are finite, these situations exhibit similar characteristics and are grouped accordingly.

6.1.2 Decreasing Returns to Scale

10. Under decreasing returns to scale, as firms exit the market, marginal cost (MC) increases, eventually reaching equilibrium at $p = MC = AC$ with market-clearing conditions $X = \sum_{i=1}^{n} q_i$. This results in a finite number of coexisting firms, which can be either one or multiple firms. The exact number depends on the intersection point where each firm's output meets the condition $p = MC = AC$, and the overall market demand at equilibrium.

This scenario demonstrates that, with fixed costs present, the assumption of infinitely many firms inherent to perfect competition is unrealistic. Markets inevitably evolve toward a finite number of coexisting firms. However, due to the absence of cost differences, this can still be interpreted as a perfectly competitive state.

6.2 Heterogeneous Firms: Cases with Differences in Cost Curves

For simplicity, assume for all $q \in [q_{min}, q_{max}]$ or $q \in D$, we have:

$$AC_2(q) = AC_1(q) + \Delta(q), \quad \Delta(q) \geq 0$$

or equivalently,

$$AC_1(q) \le AC_2(q) \le \cdots \le AC_n(q),$$

expressing differences in average costs (AC) among firms at any feasible output level q.

6.2.1 Increasing Returns to Scale

11. Previously, fixed costs and scale effects were discussed separately due to theoretical rigor. The reduction in AC resulting from spreading fixed costs does not necessarily imply increasing marginal returns at the production-function level. In contrast, genuine increasing returns to scale typically reflect rising marginal output, causing MC to decrease. Particularly from a long-run perspective, fixed costs can be regarded as variable, implying both mechanisms, despite theoretical distinctions, share similar equilibrium implications.

Therefore, the two mechanisms manifest similarly in the cost function structure: both are characterized by a downward-sloping left side of the AC curve, where unit costs decrease continuously with increased output. Thus, for the analysis of market structures and firm behavior (entry, exit, pricing power), both mechanisms can be unified under the previously described sixth category.

6.2.2 Constant Returns to Scale

12. Under constant marginal costs (MC) with non-negligible fixed costs, the scenario corresponds to the bottom of the U-shaped AC curve. If there are differences in AC among firms, the firm with the lowest AC inevitably monopolizes the entire market. This slightly differs from case 4, since here fixed costs ensure economies of scale at low output levels, corresponding to either half of a U-shaped AC curve or an L-shaped curve. In contrast, case 4 involves constant, flat MC and AC curves. However, due to $AC_1(q) \le AC_2(q) \le \cdots \le AC_n(q)$, the firm with the lowest AC clearly can expand production, lower prices, and fully capture the market.

Pricing strategy in this scenario resembles that of case 4. Firms can adopt a secure pricing principle by selecting the lower of either the profit-maximizing monopoly price or the limiting price calculated based on the second-lowest firms minimum AC, to prevent entry and maintain market power. Practically, an even higher limit-pricing strategy is feasible, since the leading firm can always price below the second-lowest firms minimum cost once entry occurs. Hence, potential entrants, recognizing cost curves clearly, will refrain from substantial fixed-cost investment, allowing the dominant firm to set a limit price near the second-lowest firms nearly zero-output-level AC.

6.2.3 Decreasing Returns to Scale

13. In the scenario of decreasing returns to scale (right side of the U-shaped curve), if the AC difference among firms is significanteven if costs rise with output $q$the lowest-AC firm will still monopolize the market provided its $AC(q)$ at total demand $q = X$ remains below the second-lowest firms minimum AC. If differences in costs are smaller, multiple firms coexist in a competitive market structure, becoming price-takers. However, a minimum entry scale arises, ensuring each firm satisfies $p = AC$. The minimum scale depends on the size of fixed

costs. This scenario closely aligns with situation 5 described earlier, thus allowing similar analytical methods and the identification of multiple detailed sub-scenarios.

7 Typical Scenarios of Cross-Market Competition

The situation of cross-market competition has already been discussed within the Arrow-Debreu equilibrium framework, indicating it can be handled under certain conditions by decomposing it into single-market analyses. On the demand side, it's important to note that in cross-market scenarios, product pricing in other markets may expand or shrink the market size for the current product due to cross-elasticity. Thus, when jointly solving with different firms in this market, this might affect the overall market structure. The introduction of fixed costs doesn't significantly change the analytical approach outlined previously.

7.1 Differentiated Competition

14. **Differentiated Competition**

Differentiated competition (Chamberlin, 1933) is one of the most common market structures involving cross-market competition in practice. This type was briefly analyzed earlier; however, due to its unique characteristics compared to single-market competition, it is summarized and analyzed here as a separate category. Under differentiated competition, we assume firms can produce a variety of differentiated products, choosing ultimately to produce the one yielding the highest profits. Consequently, modeling within an Arrow-Debreu equilibrium framework is more appropriate.

The core distinction of this market structure from those discussed previously lies in the differentiated nature of products, which are partially substitutable, making it essentially a multi-product market rather than a single-product market. Although traditional economics often treats differentiated products as a single market, within the Arrow-Debreu framework they should be defined as distinct product markets, with substitutability characterized explicitly through cross-elasticities. As cross-elasticity approaches infinity, the differentiated market returns to a single-product market scenario.

Typically, firms can monopolize their niche markets, exhibiting increasing, constant, or decreasing returns to scale within these niches, though their average cost (AC) generally remains higher than if other firms specialized in the same niche. If a firm's average cost advantage is insufficiently significant, other firms would enter the niche, thus eliminating differentiated competition. Under this scenario, entrants unfamiliar with operations in the niche typically face higher AC upon cross-market entry (unless their AC is lower than the incumbents, enabling them to replace incumbents and achieve cross-market monopolies). Pricing strategies here align with standard monopoly pricing within the niche or limit pricing based on AC/MC cost differentials, whichever is lower. Contrary to the traditional view that entry barriers diminish over the long term, we argue firms, absent significant strategic errors, can sustainably maintain differentiated competition by continually identifying unique market segments, supported by experience, path dependence, and sustained higher fixed-cost investments. This argument will be elaborated upon when analyzing the sources of AC differences subsequently.

7.2 Increasing Returns to Scale on the Demand Side

15. **Demand-Side Increasing Returns to Scale for a Single Product**

Demand-side increasing returns to scale in a single product refer to the self-reinforcing scale effects on the demand side. For such products, demand-side behavior manifests as follows: as demand-side scale variablessuch as installed base, word-of-mouth accumulation, complement richness, or adaptation levelincrease, the willingness to pay (WTP) of the marginal consumer for this product rises. Since this scale effect influences only the demand function of the product itself, the demand function for each firms product is generally specified individually rather than being merged into a unified market demand function. As a result, competition takes place in a differentiated product market. Cross-product competition is primarily reflected through cross-price elasticities, manifesting as the ebb and flow of respective demand scales among products.

A typical example of demand-side increasing returns to scale for a single product is network effects (Katz and Shapiro, 1985). Besides initial differentiation in services, network products often feature strong social characteristics, with distinct user communities that are not completely substitutable. This can be approximated by assuming that other firms face nearly infinite AC when attempting to supply a given firms network product. Each network product effectively constitutes its own monopolistic market structure. Competition occurs through cross-product elasticity, reducing total demand for a product and thus leading firms to exit if revenues fall below costs.

Network effects also have a distinctive characteristic. Unlike standard demand curves (greater supply reducing willingness-to-pay), network effects cause user willingness-to-pay to increase with the size of the user base, resulting in an upward-sloping demand curve within certain regions.This fits the case of single-product demand-side increasing returns to scale.

We now turn to a detailed analysis of single-product demand-side increasing returns to scale.Assuming numerous potential users exist in the market, each user decides whether to purchase (or use) the product by comparing product price p to their own willingness-to-pay W, purchasing only if $W \geq p$. If network effects are present, user willingness-to-pay rises with the total number of current users Q(or installed base / user base / or a parameter capturing word-of-mouth effects). Formally, this relationship can be expressed as:

$$W_i(Q) \;=\; v_i \;+\; \alpha\, Q,$$

where:

- v_i denotes the intrinsic value for user i (personal preference or reservation price for isolated use), varying among users; - $\alpha > 0$ captures the additional value gained per additional user (positive externalities); - Q is the total number of current users of the product.

As Q increases, user is willingness-to-pay $W_i(Q)$ also increases, reflecting the enhanced usefulness of the product or platform due to larger user bases (e.g., more friends on a social network or more buyers/sellers on an e-commerce platform).

Aggregating individual consumer demand yields the overall market demand function. Analytically, this scenario mirrors standard production-side analyses, as firm exit conditions ($p < AC$) remain independent of demand-side functional forms. Thus, firms can maintain monopolies within their product markets. Their exits occur primarily due to decreases in total demand, triggered by price reductions in other firms' products (outside their own product market), leading to prices falling below their own AC.

Profitability depends on expanding the user base to maximize total demand, enabling revenues to cover substantial fixed-cost investments and potentially yield significant monopoly profits. The optimal pricing strategy is determined by the standard monopoly profit-maximization formula:

$$p^* = \left[\frac{\varepsilon(Q)}{\varepsilon(Q) - 1}\right] \times MC.$$

When product A lowers its price, demand and revenue increase, enabling further price reductions to expand its market size. Due to cross-elasticity, this subsequently shrinks the market size and reduces demand for product B, generating effects akin to increasing returns to scale on the production side. Ultimately, this dynamic drives product B out of its market. Once the price-setting firm is identified, its pricing policy follows profit maximization, with the optimal price computed accounting for cross-price elasticities across products. Accordingly, in differentiated-products competition with single-product, demand-side increasing returns to scale and heterogeneous costs, surviving firms earn unequal levels of economic profit.

7.3 Synergies on the Demand Side

16. **Demand-side multi-product synergies**

Similar to multi-product synergies on the production side, the demand side may also exhibit cross-category (or cross-user-group) synergies: when the same firm sells multiple categories simultaneously, the sales of other products can increase consumers' willingness to purchase a given category.

Analogous to economies of scope on the supply side, the demand side likewise features multi-product or multi-dimensional synergies. Suppose a firm has S products/categories (or user-side/functional dimensions), indexed by $s \in \{1, \ldots, S\}$. Define

- the quantity vector $Q = (Q_1, \ldots, Q_S)$;
- the price vector $p = (p_1, \ldots, p_S)$;
- the baseline value (intrinsic willingness to pay) of individual i for dimension s as v_{si}.

Individual utility and participation condition

$$U_{si}(Q) = v_{si} + g_s(Q) - p_s, \qquad \text{participation occurs if and only if } U_{si}(Q) \geq 0.$$

Here $g_s(Q)$ denotes the synergy that dimension s obtains from the overall scale Q: it includes both same-dimension network externalities (increasing with Q_s) and cross-dimensional complementarities/substitutions (changing with other Q_r, $r \neq s$).

For simplicity, suppose the firm sells two products, A and B, simultaneously.The firms users are divided into two groups: side A and side B. The firm charges prices p_A to side A and p_B to side B. If currently Q_A users have purchased (or joined) on side A, the net utility for any user j on side B can be expressed as:

$$U_{Bj}(Q_A) = v_{Bj} + \beta\, Q_A - p_B,$$

where $\beta > 0$ indicates the sensitivity of side B users to the scale of side A. User j purchases (or joins) only if $U_{Bj}(Q_A) \geq 0$. Consequently, the total number of purchasers (or joiners) on side B, Q_B, is an increasing function given by $Q_B = G_B(Q_A,\, p_B)$.

Conversely, for user i on side A, if side B has size Q_B:

$$U_{Ai}(Q_B) = v_{Ai} + \alpha\, Q_B - p_A,$$

where $\alpha > 0$ indicates the dependence of side A users on the scale of side B. Participation on side A, Q_A, is thus given by $Q_A = G_A(Q_B,\, p_A)$.

Aggregating these demands from both sides yields the overall market demand functions. Because demand on side A depends on Q_B and vice versa, the two sides mutually reinforce each other: an increase in Q_A raises the value on side B, further expanding Q_B; similarly, a larger Q_B makes side A more attractive, prompting further growth in Q_A. This positive effect, characteristic of two-sided (Rochet and Tirole, 2003; Armstrong, 2006) and multi-product markets, creates self-reinforcing dynamics on the demand side, i.e., multi-product self-reinforcement. Platforms that appropriately set $(p_A,\, p_B)$ can quickly scale both sides, achieving a market position that is difficult to challenge once critical mass is reached.

Multi-product synergies, as with single-product demand-side increasing returns to scale, are reflected in the demand curve, and the exit condition remains $p < AC$. The demand-side synergies amount to multi-product increasing returns to scale, so that firms with larger scale can still drive rivals out of the market by cutting prices. The actual market structure, however, must be determined jointly with the cost side and calculated using the same method.

7.4 Mixed Demand-Cost Structures

17. **Mixed Demand-Cost Structures**

Previous analysis employed simplified cost curves (constant, increasing, or decreasing marginal costs) for ease of categorization and theoretical derivation. In reality, however, firms' cost structures are usually far more complex. Actual cost curves might simultaneously exhibit increasing and decreasing returns to scale, or even multi-stage, non-unimodal "mixed curve" characteristics. Demand curves, similarly, might simultaneously exhibit features of differentiated competition, network effects, and two-sided market effects. Within the theoretical framework proposed here, such mixed structures can also be analyzed by mapping realistic demand and cost functions across various output ranges, thus enabling firms' profit-maximizing production choices to be determined through unified profit maximization conditions. This yields endogenous market structures consistent with the overall model.

8 Origins of Market Structures and Correspondence with Reality

Extending the Walrasian (Arrow-Debreu) equilibrium framework into imperfect competition allows us to derive various real-world market structures simply by altering total demand X, shapes of marginal cost (MC) and average cost (AC) curves, and cost differences among firms.

Notably, most realistic scenarios tend to result in monopolistic or other forms of imperfect competition; only under very specific conditions does the market achieve the special case of perfect competition.

The emergence of diverse market structures hinges primarily on differences in marginal costs (MC), average costs (AC), and shapes of demand curves among firms. To ensure theoretical relevance, it is crucial to align these differences with real-world conditions. Various factors lead to these differences in reality, and the following explains how to incorporate them uniformly into the extended general equilibrium model of unified imperfect and perfect competition presented in this paper:

- **Network Effects and Two-sided Markets**: Treated as "increasing returns to scale" and "synergies" on the demand side. The handling of these effects was addressed in scenarios 15 and 16, respectively.
- **Brand Effects**: Represent additional demand-side benefits under differentiated competition, enhancing perceived product value or customer loyalty.
- **Firm-held Patents**:
  - If the patent primarily impacts production, model it as a reduction in the firms costs (its MC/AC becomes lower than competitors).
  - If the patent affects consumption or demand, model it as enhanced demand, similar to brand effects.
- **Government Licensing or Administrative Monopoly**: Equivalent to raising entry costs. Can be modeled as an additional fixed cost (including both explicit and implicit costs). If a firm is unable to obtain a license, its entry cost is effectively infinite, preventing market entry.
- **Other Sources of Cost or Demand Heterogeneity**: All other factors causing differences in firms' cost or demand structures (such as information acquisition, financing channels, organizational management, path dependency, and geographic advantages) can be handled similarlyby mapping them into parameter differences within firms' respective cost or demand functions.

Through this method, diverse real-world barriers can be systematically embedded within an extended general equilibrium model, seamlessly bridging theoretical analysis of perfect-imperfect competition with the practical sources of monopoly power observed in real markets.

In practice, market structure and competition dynamics can be analyzed using publicly available market data, such as financial reports from listed companies (assuming their authenticity can be confirmed). However, our model requires knowledge of potential entrants' cost structures, which is typically difficult to acquire, as these scenarios involve hypothetical situations. Nonetheless, given antitrust regulations, dominant firms in reality often permit other companies to enter the market at their higher average cost (AC), thereby avoiding stricter regulatory scrutiny. Consequently, we can use the costs of marginal, zero-profit firms as proxies for the costs of potential entrants, providing a practical basis for analyzing competitive structures or monopoly conditions in various markets.

This approach, however, necessitates translating accounting-based revenues and costs into economically relevant counterparts. Generally, this involves aggregating multi-period data since the firm's inception (to appropriately amortize R&D expenses, fixed costs, etc.). At the outset of this study, we deliberately did not restrict the functional form of cost functions, focusing instead on simplified characterizations of marginal cost (MC) and average cost (AC) curves, and the differences between firms. The rationale is that real-world cost functions should directly inform the analysiseven if mathematically less elegant. Once this translation is complete, the initially abstract "toy model" of Walrasian equilibrium can effectively integrate with real market contexts, becoming a significantly more operational analytical tool.

9 Fundamental Axioms and Market Structure

In the preceding sections, we have classified and analyzed, case by case, a variety of canonical market structures (perfect competition, monopoly, competitive fringe configurations, network-effect-dominated structures, and so on). It should be emphasized that these cases are not a disconnected patchwork of models, but equilibrium solutions on different parameter ranges that can be rigorously derived in formal (in particular, mathematical) terms from a single axiomatic system and solution procedure set out below.

Put differently, the case-by-case verbal analyses and conclusions above are precisely the derivation schemes and computational outcomes of the axiomatic framework developed in this section under different specifications of cost functions and demand structures. The choice to present them in verbal form rather than through full mathematical derivations is made purely for ease of understanding.

This section sets out the unified axioms and structure, and the Appendix provides a formal mathematical derivation example that illustrates how, starting from these axioms, specific cost and demand specifications can be mapped into the corresponding market structures.

9.1 Fundamental Axioms

1. **Finite Firms and Profit Maximization**
   Each firm i maximizes its profit:

   $$\Pi_i = p \cdot q_i - C_i(q_i),$$

   where p denotes the market price or the equilibrium price function derived from network effects, two-sided market effects, or multi-firm profit-maximizing strategies; $C_i(\cdot)$ is the firm's cost function (including fixed costs, production costs, synergies, etc.). This research explores various firm entry, exit, and output decisions across multiple input parameters, but the logic uniformly derives from the same profit-maximization conditions.

2. **market clearing and the mechanisms of firm entry/exit and capacity expansion/reduction**

There exists a demand or network-effect mechanism satisfying multi-firm equilibrium:

$$\sum_i q_i = X(p),$$

or more general game-theoretic interactions among multiple firms. Throughout this process, firms' decisions about entry, exit, and capacity expansion or reduction are unified under profit maximization. Network externalities and differentiated demand can be incorporated through adjustments to $X(p)$. Despite potential variations in functional forms of the demand side across scenarios (e.g., differentiation levels, returns to scale), the core remains the determination of a unified equilibrium price p and an allocation of output $\{q_i\}$ that matches supply and demand.

3. **Consumer Utility Maximization Axiom**
   To determine the equilibrium price and quantity, the following consumer-side condition is introduced: assuming equivalent utility structures, consumers always prefer goods with lower unit prices (or, in multi-product markets, combinations of products providing the highest total utility per unit cost). Together with market clearing, this axiom constrains the set of feasible equilibria.

4. **Self-Consistency and Verifiability**
   Based on the above principles of profit maximization and market clearing, we can empirically fit firms' cost and demand functions using observed data to derive equilibrium solutions. When key parameters, such as fixed costs or factor prices, change, the model's predictions regarding market structure or firm entry/exit decisions can be verified against observations. Regardless of whether scenarios involve price-takers or allow for one or more price-makers, the research consistently relies on the same profit-maximization equations and market-clearing conditions, thus ensuring logical consistency, practical applicability, and empirical verifiability.

10 Structural Unification, Endogenous Behavior, and Scientific Modeling

10.1 Structural Unification and Partitioning of Solution Spaces

The central idea of the theory is that, once the heterogeneity of cost and demand functions and the mechanisms of firm entry/exit and capacity expansion/reduction are given, then, under consumer utility maximization and firm profit maximization, firms expansion, shutdown, entry, and exit decisions, together with market clearing in all goods markets, naturally generate all market structures and the corresponding equilibrium prices.

Although this paper formally employs a case-by-case categorization of market outcomes (such as monopoly, oligopoly, perfect competition, or network-effect dominance), these scenarios are not independent or fragmented models. Instead, they uniformly emerge from the general system of equationscomprising multi-firm profit maximization, market clearing, and network effect/cost parametersunder a coherent market logic, yielding solutions corresponding to different parameter configurations. When critical parameters (such as price relative to average cost, network externality thresholds, or intervals of increasing returns) cross threshold values, firms' optimal decisions exhibit discontinuous jumps: moving from zero output to positive production or from multiple coexisting firms converging to a single dominant player. These threshold-based solutions do not represent patched-together models; rather, they are spontaneously emerging corner solutions derived from the same fundamental profit-maximizing conditions under varying demand-cost parameter intervals. Thus, the seemingly discrete market structures categorized by "case analysis" are, at their essence, traceable to equilibria within different parameter ranges of a unified axiomatic system. This approach is rigorously formalized, anchored in standardized equations derived from first principles.

10.2 Endogeneity of Behavioral Mechanisms

Notably, conditions such as price-taking or price-setting (market power) are not prior assumptions of this study. Instead, they emerge naturally within this model from heterogeneous firms' profit-maximizing choices. Specifically, when a firm's optimal profit-maximizing output decision satisfies $p = MC$, the firm acts as a price-taker at equilibrium. Conversely, if a firm's optimal decision leads it to produce at $p > MC$, it becomes a price-maker. These behaviors result from firms' autonomous choices based on actual demand-cost conditions. Consequently, such "behavioral patterns" are not presupposed but arise endogenously as discrete jump solutions from the same underlying profit-maximization equations under varying circumstances. Through this process, the model attains closure, endogeneity, operationalizability, and empirical testabilitythe four fundamental principles of economic modeling.

10.3 Scientificity: Measurability and Falsifiability

Because various market structuresfrom perfect competition to imperfect competitionare systematically derived using the same underlying demand-cost framework, the model effectively addresses traditional microeconomic ambiguities arising from experiential assumptions like "sufficiently large number of firms," "absence of market power," "absence of hidden barriers," or "free entry and exit." These traditional assumptions often lack clear quantitative benchmarks, making formal representation and empirical application challenging and limiting their validity and testability. In contrast, by partitioning the parameter intervals of the cost function, this paper transforms the classification of market structures from an empirical premise into a result derived from theoretical deduction. This enables theoretical predictions to directly connect with firms' actual cost data and undergo empirical validation. This transformationfrom abstract modeling toward enhanced falsifiability with measurable, predictable, and replicable attributescan be viewed as a valuable exploration into the scientific rigor of microeconomics.

11 Reasons Why Differences in MC/AC Are Difficult to Eliminate Through Technological Diffusion

The classical perfect-competition assumptionthat cost differences among firms diminish over time due to free technological diffusionis overly simplistic and differs significantly from reality.

We provide a detailed analysis below, examining both demand-side and supply-side factors.

11.1 Demand Side: Brand Lock-In and Consumer Inertia

Brand effects on the demand side create enduring advantages within specific market segments, enabling dominant firms (as exemplified by the OEM business model) to secure consumer loyalty through sustained marketing efforts and high-quality services. Once consumer loyalty is established, incumbent firms benefit from entrenched customer bases. Even if new entrants approach incumbents' technological levels, overcoming existing demand-side advantages proves challenging. Establishing a brand often requires substantial advertising investments, and incumbents' financial strength allows them to invest more effectively in advertising and brand building compared to disadvantaged competitors.

11.2 Cost Side: Entrenched Division-of-Labor Advantages and Reinforcement from Economic Profits

On the cost side, crucial factors include division-of-labor mechanisms that create enduring cost disparities (AC/MC differences) and economic profits that reinforce and amplify these disparities.

Specialization and division of labor often yield persistent cost advantages difficult to replicate, significantly contributing to market concentration. The conventional "technological diffusion" assumption typically suggests that, given enough time or through open technology licensing, later entrants or other firms can gradually achieve parity with pioneers in production techniques and costs. However, theories of division-of-labor, combined with factors such as tacit knowledge and organizational specialization, result in persistent cost gaps. These enduring cost differences, in turn, yield persistent economic profit disparities. Consequently, dominant firms attain significantly higher unit profits compared to weaker competitors. Although traditional economics suggests that late entrants can bridge funding gaps through financing and borrowing, dominant firms higher profitability enables superior financing capabilities, further exacerbating capital disparities. These dynamics imply that cost convergence is not guaranteed, specifically manifesting as follows:

11.2.1 Deep Experience and Knowledge Accumulation through Specialization (Differentiated Products)

When firms consistently specialize in particular products, processes, or market operations, they accumulate substantial *tacit knowledge* spanning production methods, management processes, supply-chain integration, and human capital. Such knowledge encompasses firm-specific production nuances, team collaboration models, internal decision-making efficiency, brand-building, and customer relationships. Even if external firms replicate publicly available technical documentation or production equipment, they struggle in the short run to acquire these embedded organizational experiences and skills. Due to human learning costs and experiential accumulation requiring extensive hands-on operations, a firm's specialized expertise developed by executives over sustained commitment to a niche market is difficult to transfer directly into other segments, making it challenging for new entrants to match incumbents' Average Cost (AC) or Marginal Cost (MC).

11.2.2 Advantages from "Hard-to-Imitate" Management and Branding on the Production Side

Competitive advantages under specialized division of labor arise not only from production technologies but also from supply-chain optimization, brand recognition, stable sales networks, and upstream-downstream relationships. The entrenched status of long-established firms enables them to maintain negotiation advantages in raw material procurement, talent acquisition, and market channels, sustaining lower per-unit costs. Such advantages are generally difficult to replicate through publicly accessible market information or short-term investments, ensuring that incumbent firms maintain enduring leadership in MC/AC levels.

11.2.3 Organizational Gaps Difficult to Bridge by Traditional Technology Diffusion

Classic economics often assumes that "knowledge is a public good and diffuses easily," suggesting that long-term free entry reduces cost differences. However, in highly specialized industries (such as high-tech, biomedicine, and precision manufacturing), key factors involve soft power like organizational structures, team capabilities, and market reputation, which require years of accumulation and path-dependence to form. Even with equivalent equipment, new entrants often fail to quickly build internal managerial efficiency and customer loyalty comparable to incumbents. Consequently, cost differences persist or even widen over the long term.

11.2.4 Technological Diffusion Speed Unable to Match the Pace of Technology Advancement Reinforced by Financial Resources

For monopoly firms, sustained high economic profits and mature R&D infrastructures enable them to innovate technologically faster than industry averages. Even if certain technological aspects partially diffuse to competitors, latecomers find it difficult to match the continual upgrade and iteration of leading firms. Leaders consistently invest substantial resources to enhance products and processes, leaving competitors constantly behind at the diffusion "starting point," widening the technological and cost gap. Only disruptive innovations by latecomerswhich would then place them in a cost-advantageous positioncan significantly alter this dynamic.

11.2.5 Higher Risk Tolerance Enabled by Greater Economic Profits

Firms earning high economic profits possess superior financial and resource buffers, enhancing their resilience against potential failures or risks in strategic operations and R&D. Here we distinguish two risk types: per-unit-product-based risks (losses scale with sales volume) and firm-level risks (losses approximately equal per firm). Leading firms demonstrate advantages in both scenarios (higher per-unit profits enhance resistance to per-unit-product risks, and larger production volumes increase total profits to better absorb firm-level risks).

Specifically:

Per-Unit-Product-Based Risks

1 External Risks External economic fluctuations or demand downturns may cause price declines after product production, or rising input costs may elevate expenses. Transaction prices might fall below the average cost for marginal firms yet remain above average costs for advantageous firms.

2 Internal Risks Product recalls and compensations arising from quality issues, or strategic marketing missteps causing broad-scale product sales failures, represent batch-wide losses proportional to company sales volume. Larger firms incur greater absolute losses in such scenarios, classifying these as per-unit-product-based risks.

Thus, advantageous firms demonstrate stronger resilience against external and internal shocks. High profit margins enhance financial flexibility, allowing firms to absorb losses from strategic mistakes or economic volatility without immediate financial distressunless the entire market disappears. Such risk tolerance starkly contrasts weaker competitors with limited financial buffers.

Firm-Level Risks

1 External Risks Examples include one-off expenditures from regulatory changes affecting market entry, purchasing proprietary technology for exclusive leadership, or rapidly investing substantial resources in technology replication.

2 Internal Risks Internal risks involve trial-and-error R&D via multiple pathways or shifting R&D directions after initial setbacks.

11.2.6 Enhanced Ability to Absorb Technological Progress and R&D Risks through Higher Overall Profits

Leading firms' substantial overall profits equip them with superior capacity to withstand technological advancements and R&D risks. In cases of external technological progress, they can sustain technological leadership through exclusive, high-priced acquisitions or quickly replicate competitors' innovations, thereby maintaining technological parity.

When independently engaging in R&D for technological leadership, these firms possess ample profits enabling investment across multiple R&D pipelines or technological projects simultaneously. Even if certain projects fail, their overall financial stability remains unaffected. This multi-project parallel strategy significantly increases the probability of technological breakthroughs and allows exploration of more frontier or potentially promising yet riskier fields. Such firms can repeatedly experiment, rapidly iterate technologies, and achieve accelerated technological advancement.

External competitors, constrained by limited funds or narrow profit margins, must cautiously strategize and selectively choose R&D projects, making them incapable of competing with leading firms aggressive experimentation and expansion. Moreover, when new technologies emerge, these competitors lack sufficient financial resources to match leading firms' acquisition offers, thereby perpetuating technological inferiority. Consequently, leading firms continuously consolidate their cost advantages or technological barriers, further enhancing market concentration.

11.2.7 High Economic Profits Enable Firms to Artificially Generate Operational Risks for Competitors

Monopoly firms can strategically lower prices, relinquishing part of their profits to consumers while still pricing above production costs, thereby avoiding accusations of dumping. However, such pricing wars force competitors' prices below their cost thresholds, increasing their risk exposure and discouraging them from substantial investment. Competitors thus lack sufficient funds for technological and managerial advancements. Particularly, the combined effects discussed in Sections 11.2.5 and 11.2.7 result in shorter lifecycles for non-leading firms. Consequently, these firms rarely accumulate sufficient technological and managerial competencies before their demise, maintaining persistent gaps in their AC and MC relative to industry leaders.

11.3 Evolutionary Path from Minor Differences to Market Monopoly

We can explicitly illustrate a simple, complete dynamic process by which minor initial differences inevitably evolve toward market monopolization. Initially, differences in division of labor may produce initial cost disparities driven by tacit knowledge or other factors, solidifying long-term differences in costs (or demand). Persistent cost disparities then further widen gaps in capital accumulation speeds, granting leading firms higher economic profits than disadvantaged competitors. As capital accumulation disparities expand, differences in branding, market experience, technological capability, and risk resilience intensify, enlarging the gap with disadvantaged firms. Leading firms thus continue expanding until reaching the boundaries of their niche marketsmarket monopolies within niche segments often remain unrecognized by antitrust laws.

Upon cross-segment expansion, firms lose their division-of-labor advantage, thus halting growth. Alternatively, to comply with antitrust regulations, firms proactively halt further expansion within their original markets, conceding some market share. Accumulated excess economic profits beyond immediate operational needs subsequently flow into other industries through corporate investments or shareholder dividends reinvested elsewhere. Therefore, the process acts as a self-reinforcing positive feedback system, spontaneously driving markets toward various monopoly structures until equilibrium boundaries are reached.

Barring significant market shifts, catastrophic managerial errors, high-level organizational turbulence undermining strategic decision-making advantages, or the emergence of disruptive, non-acquirable entrants offering substantially lower costs or increased demand, market dominance persists, often widening further relative to subsequent entrants. Even if adverse events dismantle incumbent leaders, these self-reinforcing mechanisms persist, allowing new dominant firms to emerge rather than reverting markets to perfect competition. Hence, monopolies replacing monopolies constitute the typical real-world market structure.

Specifically:

(1) Initial minor differences (technology, capital, connections, entrepreneurial talent, etc., leading to initial cost discrepancies)

(2) Division of labor and tacit knowledge → sustain differences, resisting technological diffusion; mature R&D experience expands technological advantages

(3) Profit disparities, financing capacities, and capital accumulation → further reinforce branding/technology/management/risk resilience advantages

(4) Self-reinforcing system reaches market boundaries → eventually stabilizing into monopolistic or oligopolistic structures

(5) Even if a dominant firm collapses, new dominant firms fill the vacuum; markets do not revert to multi-firm perfect competition, initiating another cyclical phase

Refutation of the "Long-Term Equalization" Hypothesis

Traditional textbooks typically argue that with sufficiently open market entry, prices ultimately approach minimum average costs, eliminating excess profits. However, under conditions of brand advantages, specialized division of labor, advanced R&D systems, and superior financial resources, this reasoning assumes rapid, complete technological imitation or diffusion. When proprietary management expertise, brand foundations, R&D capabilities, and financial strengthimplicit resourcesprove difficult to replicate, firms achieving minimum AC dominate industry leadership.

Thus, minor initial disparities may amplify rather than diminish through competitive evolution, establishing self-reinforcing effects continuously elevating market concentration toward monopoly or oligopoly outcomes.

Reinforcing Core Conclusions of this Study

Previously, we stated that "minor disparities in marginal or average costs easily drive market equilibria toward single-firm dominance." Division-of-labor theory illustrates that such initial disparities likely persist over time rather than diminish, reinforced by tacit knowledge, organizational specialization, and brand loyalty.

These persistent disparities remain insurmountable by traditional "technology diffusion" approaches. Within the expanded Walrasian (or ArrowDebreu) general equilibrium framework incorporating imperfect competition, explicitly incorporating "irreversible division-of-labor advantages" or directly embedding brand, division-of-labor, R&D capabilities, and capital disparities, further demonstrates: perfectly competitive equilibria represent theoretical extremes; realistically, enduring MC/AC differencessustained by branding, specialized division of labor, R&D capabilities, and capital strengthsnaturally drive markets toward monopolistic or highly concentrated structures.

12 Assumptions and Conditions of the Perfect Competition Model

Next, we explore the fundamental assumptions and conditions underlying the traditional perfect competition model. The model assumes that technological diffusion leads to uniform marginal cost (MC) curves, the existence of an infinitely large number of firms, and decreasing returns to scale.

12.1 Conditions for the Assumption of Infinite Firms

Regarding the assumption of infinitely many firms in a market, as analyzed previously under Walrasian equilibrium (scenario 2), its validity relies on the absence or negligibility of fixed costs and an increasing marginal cost. Under these conditions, the average cost (AC) curve corresponds to the right half of a U-shaped curve, with its lowest point at zero output, meaning firms approach infinitesimal scale. As fixed costs increase, the lowest point of the U-shaped curve shifts rightward. According to the previous scenario 8, this yields a market with finitely many perfectly competitive firms. Once the lowest-point output surpasses total market demand, the AC curve transitions to the left half of the U-shape, causing market concentration. Theoretically, the number of firms a market can accommodate depends on the relationship between market demand scale and the minimum efficient scale of production.

12.2 Re-examining Mechanisms Explaining Decreasing Returns to Scale

Traditionally, three explanations are proposed for decreasing returns to scale: firstly, scarcity-driven price increases of critical resources with scale expansion; secondly, organizational inefficiencies and rising coordination costs as management complexity grows; and thirdly, diminishing marginal returns from single-factor input expansions. Below, we critically revisit these mechanisms.

12.2.1 Scarcity-induced Resource Price Increases

We first examine resource scarcity. Scarcity implies rising unit costs as production expands and key inputs (labor, raw materials, land, water, electricity, etc.) become limited or difficult to procure. In long-run cost curve analysis, scarcity often explains the upward-sloping segment of the AC curve. However, resource scarcity is fundamentally an industry-wide phenomenon. If the entire industry faces scarcity, costs uniformly rise for all firms, not creating a firm-specific right-side segment on the AC curve. Particularly, the perfect competition assumption implies each firms production is small, so individual firm output changes rarely trigger resource shortages.

For instance, assuming no scarcity-driven increase in MC/AC, marginal cost remains constant. Introducing scarcity uniformly shifts all firms' MC curves upward. Suppose variable costs vary linearly with output. A rise in input prices merely alters the slope or intercept without inducing additional decreasing/increasing returns. Consider the total cost function:

$$TC(Q) = FC + w \cdot Q,$$

where w is constant unit variable cost. If input prices rise, w increases yet remains independent of output levels. Thus, marginal cost $MC = w$ remains constant (or uniformly shifts up to a new constant w'). The average cost is:

$$AC(Q) = \frac{FC}{Q} + w.$$

For any finite Q, since $\frac{FC}{Q} > 0$,

$$AC(Q) = \frac{FC}{Q} + w > w = MC.$$

Only when $Q \to \infty$ does $\frac{FC}{Q} \to 0$, causing AC to converge toward MC. At finite output levels, AC never equals MC; they only intersect at the infinite output limit.

Under such scenarios, the intersection point of MC and AC may shift rightward or occur in another output range. In extreme models (e.g., fixed costs plus constant marginal cost with factor prices linearly rising with total output), MC and AC intersect only asymptotically at infinite output. The classical intersection of $MC = AC$ at the minimum of AC assumes constant factor prices and marginal returns initially increasing then decreasing. However, if factor-price-induced cost increases initially low and later high, this aligns with standard rising MC, preserving the minimum-point intersection. Industry-wide input-price increases thus do not create a single firm's right-side U-shaped segment; hence, this explanation can generally be disregarded.

Nevertheless, if a single large firm dominates a region and does not share scarce resources with others, expansion may locally elevate costs due to resource shortages. This situation might create the right half of a U-shaped AC curve, but firms can mitigate this by establishing facilities in other locations.

12.2.2 Increases in Management Costs

The traditional assumption of escalating management costs with larger scale warrants reassessment in modern contexts. Advances in contemporary management, particularly automation and digitalization, suggest firms rarely experience sudden, sharp management cost increases at global operational scales. Modern management practices may have shifted the "critical scale" at which unit management costs markedly increase beyond the typical global market size. Consequently, firms may maintain relatively stable management expense ratios even at substantial scales, as observable by examining management cost ratios across firms of different sizes.

12.2.3 Single-factor Expansion and Diminishing Marginal Returns

Diminishing marginal returns from single-factor input expansion likely persist in smaller firms or historically common cottage industries. Constrained by management capability or limited capital, these firms are unable to proportionally expand all factors optimally, thus encountering diminishing returns from single-factor expansion. Yet, in large modern corporations benefiting from robust financing options, this scenario is uncommon.

12.3 Revisiting Technological Diffusion and Cost Convergence

Perfect competition assumes homogeneous firms, implying technological diffusion leads to convergent marginal and average cost curves (MC/AC). Such an assumption may hold in industries with low technological intensity or slow technological progress. However, as previously argued in the specialization section, technology diffusion is challenging in high-tech sectors. Particularly, leading firms with first-mover advantages enjoy substantial economic profits for continuous R&D investment. Without major strategic mistakes, MC or AC differences not only persist but potentially widen over time, solidifying market dominance structures.

12.4 Conclusion: Perfect Competition as a Special Case, not Mainstream

12.4.1 Equivalence of Perfect and Imperfect Competition

As demonstrated above, the traditional perfect competition model applies only to limited special circumstances characterized by small scale, low technological content, slow technological advancements, and an absence of sophisticated management systems. It represents a special case rather than the foundational economic model often assumed. Conversely, other market structuresincluding various monopolistic formsare not deviations from the ideal model, but outcomes equally valid under different realistic parameter conditions. When incorporating demand and cost conditions along with firm heterogeneity, both general equilibrium and partial equilibrium frameworks naturally yield diverse competitive or monopolistic structures.

12.4.2 Historical Basis and Limited Applicability of the Perfect Competition Model

We propose that the perfect competition model emerged and gained widespread acceptance primarily due to specific historical and technological constraints of its era, which limited the formation of large-scale enterprises and naturally produced two core assumptions: "price-taking behavior" and "marginal cost pricing":

Firstly, limitations in management and communication significantly shaped market structures:

1. Firms were unable to expand geographically or efficiently organize large-scale, cross-regional production and marketing activities;
2. Absence of modern management practices led to declining efficiency with increasing scale;

3. Predominance of single-factor expansion in agriculture and artisanal production mistakenly interpreted diminishing marginal returns of single-factor inputs as decreasing returns to scale (modern economics corrects this, applying it only in short-term analyses).

Secondly, geographical limitations of firms shaped market characteristics:

1. Constraints in transportation and information flow fragmented markets, restricting firms to local markets with limited output;
2. Firms operated at similar scales, unable to influence market prices, thus becoming "price takers".

Thirdly, low entry barriers due to modest initial investments and fixed costs:

1. Agricultural and artisanal production required minimal upfront investment (tools, simple equipment), allowing numerous small producers;
2. Firms could sustain operations through marginal cost pricing without incurring substantial fixed costs.

Finally, the lack of technology and management methods to achieve economies of scale:

1. Absence of automation, supply chain integration, and modern management techniques restricted firms from lowering unit costs through scale expansion;
2. Resulted in coexistence of numerous small-scale producers satisfying local demand via marginal cost pricing.

Under these historical conditions, the market structure characterized by "countless small firms + marginal cost pricing + price-taking behavior + negligible fixed costs" emerged, forming the practical basis for the perfect competition model. Coupled with its status as Pareto optimal, the perfect competition model long served as the baseline in economics.

Nevertheless, with significant advancements in modern management and communication technologies, increased fixed costs, pronounced economies of scale, and persistent cost differences due to specialization and capital advantages, the applicability of the perfect competition model has greatly diminished. It no longer accurately approximates most modern enterprises, with industries naturally stabilizing into differentiated monopolistic competition, competitive fringe structures, oligopolies, monopolies, or natural monopolies.

12.4.3 Instability of Perfect Competition

Moreover, synthesizing previous analyses, we observe that if perfect competition is adopted as a standard model, it lacks stabilityanalogous to a ball balanced atop a hill, susceptible to small perturbations from cost heterogeneity, quickly shifting to imperfect competitive structures. Additionally, since cost differences possess self-reinforcing characteristics, markets rarely revert to perfect competition. Thus, given that most market structures deviate from perfect competition, welfare economic theorems built upon this model become largely inapplicable to realistic economic scenarios.

12.5 From Optimal Assumptions to One of Several Unstable Symmetric Structures: Repositioning Perfect Competition in Reality

In traditional microeconomic theory, “perfect competition" is viewed as the ideal market structure and endowed with high normative status due to its characteristics of price equaling marginal cost, firms being price takers, zero economic profits, and Pareto-optimal resource allocation. However, it should be noted that numerous imperfectly competitive structures exist in reality, and even if these generate positive economic profits, they may still satisfy the marginal-efficiency condition $p = MC$ under certain circumstances. For example, as illustrated in scenario 5.1.1 (or certain subdivisions of scenario 13, as well as other qualifying cases), firms may enjoy positive economic profits due to cost advantages and increasing marginal costs, traditionally interpreted as evidence of implicit entry barriers. In fact, we need to distinguish between replicable and non-replicable sources of advantage. As long as a firm possesses unique, non-replicable endowments enabling it to earn higher profitsadvantages not derived from institutional barriersthis condition alone is sufficient to achieve first-best Pareto optimality. This viewpoint differs from the traditional assertion that “only perfect competition can yield frictionless first-best Pareto optimality.” Here, we provide this analysis solely to discuss the canonical status of perfect competition. Whether the firm’s cost or profit advantage stems from replicable or non-replicable sources only affects the long-run equilibrium conditions and the attainment of first-best Pareto optimality, without altering our analysis of market structures.

Meanwhile, scenario 1 violates the traditional conditions of perfect competition regarding market sizespecifically, even if a single firm holds significant or even complete market share, it may still be unable to influence market prices. Whether a firm can influence prices depends fundamentally on the AC/MC structure of incumbent and potential entrant firms.

Within the theoretical structure proposed here, perfect competition retains its role as an idealized limiting case. By strictly deriving from cost functions, we clearly define specific parametric conditions necessary for perfect competition, thereby moving away from the descriptive assumptions typically found in textbooks, such as numerous buyers and sellers, firms as price takers, homogeneous products, complete information, free resource mobility, and sufficiently small firm sizes. Similarly, the same cost-function derivation method allows systematic deduction of descriptive conditions for various imperfectly competitive market structures, representing a significant theoretical advance. Thus, theoretical rigor is enhanced, market structure conditions become continuous and unified, and these conditions become measurable and testable, effectively overcoming ambiguities inherent in traditional descriptive assumptions. Within our parametric space, perfect competition corresponds to the extreme combination of no cost differences between firms ($MC_i \approx MC_j$), zero fixed costs ($F_i = 0$), decreasing returns to scale, perfectly homogeneous products, and no positive feedback or scale synergies, namely scenario 2.

If we relax the assumptions, allowing for a finite number of firms, scenario 10 could also represent perfect competition. However, once firms have heterogeneous costs, scenarios 2 and 10 naturally shift toward scenarios 5 and 13, respectively. Thus, perfect competition lacks stability, easily transitioning toward other market structures. Traditional economics generally suggests that technological diffusion eliminates cost heterogeneity among firms, but it must be noted that technological diffusion typically reduces unit costs (the height of the cost curve) without changing the overall shape of the MC/AC curve. Consequently, market structures oscillate between certain closed curves, such as between scenarios 2 and 5 or between scenarios 10 and 13. Due to the presence of fixed costs, scenario 13 does not evolve into scenario 2. More generally, because other market structures possess cost curves of differing shapes (considering differences in demand shapes, such as network effects, also influences market structures; hence, rigorous analysis should employ profit curves, which encompass a wider range of scenarios than cost curves alone), technological diffusion does not necessarily revert these market structures back to perfect competition. Instead, each structure oscillates within market conditions corresponding to homogeneous or heterogeneous firm cost structures. Thus, perfect competition corresponds specifically to increasing marginal cost conditions. Equally fundamental structures with homogeneous firms include decreasing and constant marginal costs (scenarios 1, 3, and 9; each can be considered a symmetric equilibrium point within its respective cost structure, possessing an equivalent evolutionary status to perfect competition along their trajectories). Different market structures evolve along their respective trajectories without necessarily intersecting with the perfect competition point. Thus, perfect competition is one of several unstable symmetric points (referring to homogeneous firms) on oscillating trajectories, and the evolution of market structures does not necessarily pass through it. Unless a market mechanism transforms firms' cost structures from another shape into an increasing marginal cost structure and simultaneously homogenizes firms, the perfect competition point will never be traversed. Since complete homogeneity among all firms is practically unattainable, heterogeneity inevitably exists. Consequently, any unstable symmetric equilibrium requiring firm homogeneity is difficult to reach in practice, and markets merely transition from one imperfectly competitive point to another. This explains why, in reality, we typically observe one monopoly replaced by another monopoly rather than the emergence of numerous small firms. Thus, we assert that perfect competition is merely a special theoretical caseelegant but not exceptional. Even if a market temporarily satisfies conditions of perfect competition with increasing marginal costs, any slight relaxation of the cost homogeneity condition or activation of positive feedback mechanisms will rapidly push the market toward stable imperfect competition structures represented by scenarios 5 or 13.

Furthermore, analyzing from a practical perspective, we observe that real-world markets commonly exhibit:

1. Fixed costs and economies of scale;
2. Heterogeneity in cost structures;
3. Product differentiation;
4. Positive feedback mechanisms and path dependencies;

5. Differences in investors capabilities, experience, and information-processing abilities.

These realistic mechanisms imply that although perfect competition remains one of many Pareto-optimal structures, in actual economic evolution it frequently exists only as an unstable boundary equilibrium. Market structures typically drift from this equilibrium due to minor asymmetries, transitioning towards more evolutionarily stable local equilibria such as monopolies, oligopolies, competitive-fringe structures, differentiated competition, or other imperfectly competitive structures.

Thus, within this theoretical framework, perfect competition is not invalidated but repositioned as a special, minimally stable corner solution within the parametric space. This reinterpretation enables us to simultaneously explain:

1. Why perfect competition represents a theoretical optimum;
2. The existence of other Pareto-optimal structures beyond perfect competition;
3. The rarity of perfect competition in actual industries;
4. Why non-optimal structures persist long-term and exhibit evolutionary stability.

This also suggests that market structure analysis should avoid overemphasizing the perfect competition model or relying solely on optimality criteria. Instead, it should prioritize evolutionary stability and compatibility with realistic economic mechanisms. Correspondingly, welfare economic theorems predicated upon perfect competition assumptions require reevaluation in light of the widespread existence of imperfect competition.

In other words, real-world market structures should not be seen as distortions or deviations from perfect competition, but rather as natural responses to structural conditions in the pursuit of profit maximization.

References

Arrow, K. J., and G. Debreu (1954). Existence of an equilibrium for a competitive economy. *Econometrica* 22(3), 265–290.

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics* 37(3), 668–691.

Bain, J. S. (1956). *Barriers to New Competition*. Cambridge, MA: Harvard University Press.

Chamberlin, E. H. (1933). *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. New York: Wiley.

Katz, M. L., and C. Shapiro (1985). Network externalities, competition, and compatibility. *American Economic Review* 75(3), 424–440.

Lerner, A. P. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies* 1(3), 157–175.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Panzar, J. C., and R. D. Willig (1981). Economies of scope. *American Economic Review* (Papers and Proceedings) 71(2), 268–272.

Rochet, J.-C., and J. Tirole (2003). Platform competition in two-sided markets. *Journal of the European Economic Association* 1(4), 990–1029.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Varian, H. R. (1992). *Microeconomic Analysis* (3rd ed.). New York: W. W. Norton.

Appendix : Example of Mathematical Formalization

This appendix supplements the formal derivation discussed in the main text, illustrating how imperfect competition can be rigorously incorporated into the Walrasian (Arrow-Debreu) general equilibrium framework. Although the main body primarily presents results through classification, each scenario is derived from a unified mathematical approach. Here, we provide a simplified formal example and a general solution strategy, outlining the methods underlying the analytical results presented above.

1. Basic Model Setup

This appendix considers only a relatively simple single-good setting with two types of cost structures and presents a representative derivation, in order to illustrate the mathematical implementation of the general solution procedure.

Consider a single-commodity market with multiple consumers and firms. Let p denote price, and q_i represent firm i's output.

1.1 Consumers and Demand Function

Assume the aggregate demand function is given by:

$$D(p) = \sum_{h=1}^{H} d_h(p),$$

where $d_h(p)$ represents consumer h's optimal demand at price p. Typically, $D(p)$ is continuous and decreases monotonically with respect to p.

1.2 Producers and Cost Functions

There are n firms in the market, indexed by $i = 1, 2, \ldots, n$, each producing quantity $q_i \geq 0$. Firm i's total cost function is:

$$C_i(q_i).$$

The marginal cost function for firm i is defined as:

$$MC_i(q_i) = \frac{dC_i(q_i)}{dq_i}.$$

If fixed costs F_i and variable costs per unit c_i exist, we may write:

$$C_i(q_i) = F_i + c_i \cdot q_i,$$

or more generally,

$$C_i(q_i) = F_i + \int_0^{q_i} c_i(s)\, ds.$$

Increasing returns to scale correspond to total costs increasing at a decelerating rate, whereas decreasing returns correspond to total costs increasing at an accelerating rate.

2. Walrasian Equilibrium Conditions

In the simplest "single-market + price-taker" scenario, Walrasian equilibrium requires:

$$\sum_{i=1}^{n} q_i^* = X(p^*),$$

with each active firm i ($q_i^* > 0$) satisfying:

$$p^* = MC_i(q_i^*).$$

If for some firm i, the marginal cost exceeds p^* at all feasible output levels, then equilibrium implies $q_i^* = 0$.

3. Corner Solution with Single-Firm Monopoly

Consider a case with heterogeneous marginal costs:

$$C_1(q_1) = F_1 + c_1\, q_1, \quad C_2(q_2) = F_2 + c_2\, q_2,$$

with $0 \leq c_1 < c_2$. If at price $p^* = c_1$, market demand $D(c_1) > 0$, only firm 1 will supply the market, as other firms cannot profitably produce since $c_2 > p^*$. Thus, a corner solution emerges where one firm dominates.

However, the assumption that firm 1 passively accepts $p^* = c_1$ contradicts the fact that it can influence price through its output decision. This implies that, when cost heterogeneity exists, Walrasian equilibrium naturally gravitates towards a "natural monopoly" outcome.

4. Modified Monopoly Pricing

If firm 1 maximizes profits actively, rather than passively accepting price, we adopt the classic single-product monopoly framework. The firm solves:

$$\max_{q \geq 0} \Pi(q) = p(q) \cdot q \;-\; C_1(q),$$

where $p(q)$ is the inverse demand function, and $C_1(q)$ is the firm's cost. Ignoring entry threats from other firms, the classic monopoly solution yields:

$$p^* \;=\; \left[\frac{\varepsilon(q)}{\varepsilon(q) - 1}\right] \times MC_1(q),$$

or simply,

$$p^* = \frac{\varepsilon}{\varepsilon - 1} \times MC_{lowest},$$

where elasticity ε and marginal cost determine the equilibrium price.

In reality, the threat of entry from the firm with the second-lowest marginal cost (MC_{2ndLow}) constrains pricing. If the monopolist raises price to $p \geq MC_{2ndLow}$, other firms could profitably enter the market, breaking the monopoly. Therefore, the highest feasible price for firm 1 becomes:

$$p = MC_{2ndLow},$$

slightly below the second-lowest marginal cost, ensuring no entry occurs.

Consequently, the monopolist must choose a price balancing two considerations:

- If $p^* \leq MC_{2ndLow}$, select p^*, as this avoids attracting entry and maximizes profit. - If $p^* > MC_{2ndLow}$, set the price slightly below MC_{2ndLow}, effectively employing limit pricing to prevent potential entrants.

Thus, the feasible monopoly pricing rule can be summarized as:

$$p_{\text{actual}} = \min\{ p^*, MC_{2ndLow} \}.$$

Here, p^* is the static monopoly solution absent entry threats, and MC_{2ndLow} represents the entry threshold of competitors. This modified approach realistically captures imperfect competition pricing dynamics by incorporating both monopolistic influence and competitive entry constraints.

5. Extension to Multi-Product and Factor Markets

Under the Arrow-Debreu general equilibrium framework, we must simultaneously consider:

- Prices and equilibria in multiple factor markets (labor, capital, etc.).
- Supply-demand equilibria across multiple product markets.
- Firms potentially exhibiting non-convex technologies, economies of scope, or other cross-product interactions.

When firms have significant cost advantages in certain product markets, monopolistic or oligopolistic structures naturally emerge, deviating substantially from the traditional "multiple firms, price-taker" equilibrium.

6. Framework Extensions and Pricing Mechanism Discussion

The previous example illustrates that, whenever marginal cost differences or significant fixed costs exist, Walrasian equilibria easily produce corner solutions with a single firm dominating the market. If we then consider the active production decisions of this dominant firm, the model naturally transitions to monopoly or imperfect competition analysis rather than simply assuming firms as price takers.

All classification outcomes in the main text can follow this methodology: we first solve the Walrasian equilibrium using the standard $p = MC = AC$ approach. Upon identifying violations of the price-taker condition, we analyze the firms responsible for these violations. Specifically, when only one firm emerges as a price-maker, we determine its profit-maximizing price as:

$$p_{\text{actual}} = \min\{ p^*, MC_{\text{2ndLow}} \}.$$

Other firms remain price takers. The primary complexity beyond the single-market scenario lies in the necessity to jointly solve the multi-product demand equations and multi-factor cost equations. If aligning with reality, one should directly use empirically derived shapes of these demand and cost curves, bypassing complex game-theoretic formulations and directly yielding the distinct classifications described earlier.

When considering competitive fringe monopoly structureswhere permitting entry by smaller firms actually increases total profitswe eliminate restrictive pricing conditions that would otherwise block entry (such as setting prices below the lowest marginal cost of potential entrants). Instead, we retain only the traditional monopoly profit-maximizing condition but subtract the output of potential entrants from total demand when calculating the monopolist's revenue.

If considering regulatory requirements from antitrust laws that enforce market-share constraints resembling a competitive fringe scenario, we directly incorporate rivals' cost functions to calculate the equilibrium price ensuring required market-share thresholds, thus deriving the monopolist's optimal pricing strategy.

7. Multiple Price-makers and Game-Theoretic Considerations

If the market contains multiple price-setting firms, if the existence or identity of price-makers is unknown, or if more accurate characterization of interactions (including collusion) between multiple price-makers is desiredas discussed in scenario 5.1.2then explicit interaction or reaction functions between firms must be specified. In such cases, employing appropriate game-theoretic models (Bertrand, Cournot, Stackelberg, dynamic games, repeated games with collusion) becomes essential.

Different game-theoretic models primarily differ in their mechanisms: whether the decision variables are quantities or prices, existence of capacity constraints, or possibility of collusion. These differences emerge directly from realistic market environments and actual institutional arrangements. At this point we can, under the maintained assumption that all firms are price-makers, specify the interaction/response functions among them, substitute these into the extended Walrasian system, and solve in one shot for the market structure and prices in all goods markets; even under this price-maker assumption, the firms profit-maximization conditions can still imply that for some firms the optimal choice is to behave as price-takers. The advantage of this approach is that it yields the equilibrium in a single computation, without first imposing a Walrasian perfectly competitive equilibrium, then inferring which firms are price-makers, and then performing multiple iterations of that procedure.

Consequently, the appropriate game-theoretic framework must incorporate:

- Explicit, heterogeneous, firm-specific cost functions (covering increasing, constant, or decreasing returns to scale, and differences among firms, including potential entrants).
- Demand functions capturing inter-product market interactions through cross-price elasticities.
- Cost functions that simultaneously account for factor-market equilibrium conditions.

Models overly simplified by excluding these elements cannot effectively depict actual market conditions. Moreover, the choice of short-term versus long-term analysis simply reflects approximations intended for computational convenience regarding temporal dynamics. Thus, to accurately capture firm interactions, models must comprehensively specify these three categories of conditions.

1. Model rules derived from market rules and the production and sales characteristics of firms within the market.
2. Specification of all potential participating firms and their internal heterogeneity.
3. Approximations of firms' behavior based on short-run or long-run analysis.

A straightforward example is the basic Bertrand model, where, in the absence of knowledge about the existence of a price-maker, we initially assume all firms in the market (two or n firms) behave as price-makers. However, through Bertrand's competition logic, we conclude that none of them act as true price-makersthey all become price-takers. This illustrates how a game-theoretic model, initially allowing firms to set prices (or influence price through output), ultimately converges to the classical perfectly competitive framework where firms are price-takers. Of course, game theory can also yield scenarios consistent with traditional economics that involve a single price-maker (imperfect competition), though details are omitted here. Traditional perfectly competitive frameworks, notably, cannot depict interactions among multiple price-makers, thus necessitating the introduction of explicit response functions between firms, as provided by theories (e.g., game theory) linked to Walrasian or ArrowDebreu frameworks. In other words, the price-taking behavior in traditional competitive economics emerges as a special simplification of more general game-theoretic treatments.

In reality, firms do not precisely know their competitors' cost functions, nor can they exactly anticipate demand curves or fully understand factor markets, resulting in decision-making fluctuations around our theoretical ideal solution. Especially when the market involves more than two price-setting firms, we require game-theoretic models to characterize the interactions between firms accurately, although these interactions can be quite complex in reality. Existing game-theoretic models may not precisely capture firm behaviors, especially as firm decisions often involve significant environmental or personal factors. Hence, we must be fully aware of the potential errors in these models. Nevertheless, we can at least assert (and prove) that total oligopoly profits must lie somewhere between average cost (AC) and the profits of a single monopolist. Even if firms engage in price wars where the short-term price falls below AC, the present value of expected future profits must justify temporary losses for future strategic benefits.

8. Existence, Uniqueness, and Real-World Compatibility of Model Solutions

Existence and uniqueness are central concerns in economics. Because our framework takes a parametric form covering common cases, there is no universal existenceuniqueness theorem. If we are to provide such results, then under different maintained assumptions one obtains equilibria corresponding to different market structures. Taking those structures as primitives, one can carry over the standard analyses of existence, uniqueness, or multiplicity from the associated traditional models.

Crucially, traditional models treat market structure as an ex ante assumption, whereas in our framework it is endogenous. Hence, when a traditional model with a preset structure has no equilibrium under given parameters, this only shows that no equilibrium consistent with that particular structure existsnot that no equilibrium exists at all. In our system, an equilibrium may well exist under a different market structure.

Based on this reasoning, we therefore adopt the following stance. Because our model aims to depict real-world scenarios and uses different firm-interaction response functions for different realities, the issues of existence and uniqueness transform into examining solutions' existence and uniqueness under the specific response functions implied by realistic parameters. Under certain conditions, these solutions converge to single-firm monopoly or perfect competition. Thus, the existence and uniqueness of solutions arise naturally from the specific real-world conditions we explore. In reality, we also know that solutions are often not unique or may not exist.

Therefore, we should adopt the following viewpoint: if our model lacks solutions under certain conditions, and if our theory is correct, such conditions should not exist in reality either. Conversely, if such conditions do occur in reality, yet our model has no solutions, this implies a contradiction that necessitates modifying the model. If multiple solutions arise under certain conditions, it indicates that multiple real-world scenarios fit our model's conditions, meaning our model is incomplete and some conditions remain uncharacterized. This typically occurs because reality is inherently complex; simplifications inevitably lead to non-uniqueness. If detailed investigation into these multiple solutions is required, we simply add more refined conditions based on reality. Once we introduce new conditions to differentiate scenarios, we typically obtain clarity on which solution emerges under specific conditions. Occasionally, we might simply treat multiple solutions as random outcomes (which requires embracing a non-deterministic view, such as accepting random, unexplained human decisions).

Thus, under the criterion of falsifiability, we may continually expand our theoretical explanatory power by adding further conditions. The existence and uniqueness of solutions are tools to assist our understanding of the mathematical characteristics of the model. Ultimately, whether our models accurately correspond with reality is the crucial factor, rather than merely mathematical properties.