Internet Appendix for "Unlocking Local Market Information through Franchising"

Steven Chong Xiao, Jiadi Xu

A Appendix: An Illustrative Model of Franchise Contracting for Learning and Expansion

In this section, we develop a two-period franchising model to analyze how a franchisor can design optimal contracts and investment strategies in the face of demand uncertainty and information asymmetry. We consider a continuum of local markets, each characterized by a demand type p_i, which is privately known to the local franchisee (he) but unobservable to the franchisor (she). This private information creates a screening problem: the franchisor must design contracts that incentivize franchisees to reveal their true demand potential while maximizing her own profits.

To address this problem, we explore three distinct strategies: *Strategy A (Franchising Only)*, where the franchisor never invests in her own store and relies solely on royalties; *Strategy B (Wait to Invest)*, where the franchisor delays investment until period 2, contingent on observing high demand in period 1; and *Strategy C (Immediate Investment)*, where the franchisor invests immediately in period 1 if the inferred demand type p_i exceeds a critical threshold. Each strategy involves a trade-off between attracting franchisees, balancing royalty income with direct operating income, and managing the costs of direct investment. By comparing these strategies, we identify conditions under which each approach is optimal, providing insights into how franchisors can tailor their contracts and investment timing to different market conditions.

A.1 Setup and Assumptions

- **Two Periods and Binary Demand**. There are two periods, $t = 1$ and $t = 2$. There is a *continuum* of local markets $i \in [0, 1]$, each with a type $p_i \sim \text{Uniform}[0, 1]$. Specifically, for each market i, demand D can be either high ($H > 0$) with probability p_i or low ($L > 0$) with probability $1 - p_i$.

- **Private Information**. The probability p_i is the local franchisee's private information; the franchisor only knows $p_i \sim \text{Uniform}[0, 1]$ is possible.

- **Revenues & Encroachment.** One unit of sales yields a margin of 1. If a single outlet operates, it sells D units. If *both* franchisor and franchisee operate, each sells

$$\lambda D \quad \text{with} \quad \lambda \in [0.5, 1),$$

  so total demand expands to $2\lambda D \geq D$. This setup accounts for the possibility that franchisor co-investment can "grow the pie" and expand the customer base, while also inducing some level of encroachment on franchisees' existing customer base. If $\lambda = 0.5$, then it means all franchisor's direct operating revenue comes from cannibalizing franchisee's revenue.

- **Costs**.
  - A one-time *startup cost* $c > 0$ to open a store, paid once per store.
  - A *fixed cost* $f > 0$ per open store in each period it operates.
  - The store remains open into period 2 only if period 1 demand is high ($D = H$).
  - The franchisor and franchisee face the same cost structure.

- **Preferences and Discounting**. Both franchisor and franchisee are risk-neutral, and pay-offs in period 2 are not discounted relative to period 1 (i.e. the discount factor is normalised to one).

- **Contract (Royalty α).** The franchisor offers a take-it-or-leave-it contract with a royalty $\alpha \in [0, 1]$. If the franchisee accepts:
  - He pays startup cost c plus the fixed cost f for each period of operation. He earns $(1-\alpha)$ per unit sold.
  - The franchisor collects α per unit the franchisee sells.
  - The franchisor also has the option to open her own store in period 1 or period 2.
- **Parameter Restrictions.** Throughout the analysis, we assume:

$$L \le f \quad \text{and} \quad \lambda H > (c+f).$$

  The first condition guarantees that a store will close after period 1 if demand is low ($D = L$) because gross revenue cannot cover the fixed cost. The second condition guarantees that, in a high-demand market, total surplus is large enough for two outlets to cover start-up and fixed costs and still earn positive profit.

A.2 Three Contract Strategies

We consider three stylized strategies for the franchisor:

Strategy A (Franchising Only) The franchisor *never* invests in her own store ("franchising only"). She can commit to this by including an anti-encroachment clause in the contract. If the local franchisee accepts, then only the franchisee's store opens, and the franchisor collects royalties α on the franchisee's sales in period 1 (with probability p or $1-p$ of H or L demand), and in period 2 if (and only if) D was high in period 1. Hence the franchisor's objective is to choose α to maximize *royalty income* from all accepting markets.

Strategy B (Wait to Invest) The franchisor does not open a store in period 1, but invests at period 2 if period-1 demand was high. In period 1, if the franchisee has accepted, he alone operates, paying $\alpha\, D$ in royalties to the franchisor. The franchisor can contractually commit not to open a nearby company store in period 1, even if the signal conveyed by a franchisee is very strong, but she reserves the right to do so from period 2 onward after observing the actual demand.

Strategy C (Immediate Investment) In this case, the franchisor gives no encroachment protection, but instead *commits* to invest immediately wherever the franchisee accepts the contract. The franchisee understands that acceptance of the contract would trigger the immediate opening of a company-owned outlet next door. [A.1]

A.2.1 Single-Market Payoffs in Each Strategy

We denote $\Pi_{r,A}(p, \alpha)$, $\Pi_{r,B}(p, \alpha)$, and $\Pi_{r,C}(p, \alpha)$ as franchisor's expected single-market profit under Strategies A, B, and C. Likewise, we denote $\Pi_{e,A}(p, \alpha)$, $\Pi_{e,B}(p, \alpha)$, and $\Pi_{e,C}(p, \alpha)$ as franchisees' expected profit under the three strategies.

[A.1] To simplify the analysis, we assume that the franchisor commits to immediately investing in period 1, so that franchisees face no ambiguity as to whether the franchisor would strategically delay their investment when considering the contract. In supplementary analysis Section A.5, we discuss this possible strategy.

Strategy A: Franchising Only

Franchisor's payoff.

If the franchisee accepts, he alone operates:

- Period 1: The franchisor collects royalty $\alpha\, D$, which is $p\alpha\, H + (1-p)\alpha\, L$ in expectation.
- Period 2: The store continues only if period 1 was high, with probability p. So the franchisor collects $\alpha\, H$ again with probability p. Expected value is $p\,(\alpha\, H)$.
- No cost for franchisor.

Hence

$$\Pi_{r,A}(p,\alpha) \;=\; \alpha\big[pH + (1-p)L\big] \;+\; p\,(\alpha\, H). \tag{A.1}$$

If the franchisee rejects, the franchisor's profit is 0 (no store is opened by either side, as Strategy A forbids the franchisor from investing).

Franchisee's payoff.

If he accepts, then:

- Period 1: He sells D at margin $(1-\alpha)$, paying fixed cost f.
- Period 2: Occurs only if $D = H$ in period 1 (prob. p), again selling H at margin $(1-\alpha)$ minus f.
- Startup cost: c at $t = 1$.

Therefore

$$\Pi_{e,A}(p,\alpha) \;=\; p\big[\,H(1-\alpha) - f\,\big] + (1-p)\big[\,L(1-\alpha) - f\,\big] + p\,\big[\,H(1-\alpha) - f\,\big] \;-\; c. \tag{A.2}$$

This simplifies to:

$$\Pi_{e,A}(p,\alpha) = p\Big[\,2\big(H(1-\alpha) - f\big)\Big] \;+\; (1-p)\big[L(1-\alpha) - f\big] \;-\; c.$$

Cutoffs for Strategy A.

A type p franchisee accepts the contract if $\Pi_{e,A}(p,\alpha) \geq 0$. Solving $\Pi_{e,A}(p_0,\alpha) = 0$ for p_0 gives a unique acceptance cutoff $p_0^A(\alpha) \in [0,1]$.

- Solving for $p_0^A(\alpha)$. If we fix α, the lowest type $p_0^A(\alpha)$ who is just indifferent solves $\Pi_{e,A}\big(p_0^A(\alpha),\alpha\big) = 0$. Expanding (A.2) and rearranging yields

$$p_0^A(\alpha) \;=\; \frac{c + f - L(1-\alpha)}{(2H - L)\,(1-\alpha) - f}, \tag{A.3}$$

Strategy B: Wait to Invest

If the franchisor *does not* invest in period 1, the franchisee (if accepting) sells D alone in period 1, paying royalty $\alpha\, D$. The franchisor invests at period 2 only if period-1 demand was high.

Franchisor's payoff.

- Period 1: $\alpha\big[pH + (1-p)L\big]$ in expectation, since franchisor has no cost.
- Period 2: With probability p, the store remains (since period 1 must have been high), and the franchisor now invests, paying $c+f$ to earn $\lambda H\alpha$ in royalties and λH in direct operating revenue in that final period. Hence

$$p\left[\lambda H(1+\alpha) - f - c\right].$$

- Franchisor's total expected profit:

$$\Pi_{r,B}(p,\alpha) \;=\; \alpha\big[pH + (1-p)L\big] \;+\; p\left[\lambda H(1+\alpha) - f - c\right]. \tag{A.4}$$

Franchisee's payoff.

He alone operates in period 1, then share some demand with franchisor in period 2 only if $D = H$ in period 1:

$$\Pi_{e,B}(p,\alpha) \;=\; p\left[\,H(1-\alpha) - f\right] + (1-p)\left[\,L(1-\alpha) - f\right] + p\left[\lambda H(1-\alpha) - f\right] \;-\; c. \tag{A.5}$$

This simplifies to:

$$\Pi_{e,B}(p,\alpha) = p\left[(1+\lambda)H(1-\alpha) - 2f\right] \;+\; (1-p)\left[L(1-\alpha) - f\right] \;-\; c.$$

Cutoffs for Strategy B.

- Solving for the cutoff $p_0^B(\alpha)$. If α is given, $\Pi_{e,B}\big(p_0^B(\alpha),\alpha\big) = 0$. The result can be shown to be

$$p_0^B(\alpha) = \frac{c + f - L(1-\alpha)}{\big[(1+\lambda)H - L\big](1-\alpha) - f}. \tag{A.6}$$

Strategy C: Immediate Investment

Under Strategy C, the franchisor commits to investing immediately in period 1 if the inferred demand type p exceeds a critical threshold $p_0^C(\alpha)$.

Franchisor's payoff.

If both operate from the start, each period's revenue split is λD per store. Summing over two periods (the second occurs only if period-1 was high) and subtracting c:

$$\Pi_{r,C}(p,\alpha) \;=\; p\Big[2\lambda H(1+\alpha) - 2f\Big] + (1-p)\big[\lambda L(1+\alpha) - f\big] - c, \tag{A.7}$$

Franchisee's payoff.

$$\Pi_{e,C}(p,\alpha) = p\Big[2\lambda H(1-\alpha) - 2f\Big] + (1-p)\big[\lambda L(1-\alpha) - f\big] - c. \tag{A.8}$$

Cutoffs for Strategy C. A franchisee of type p accepts the contract only if $\Pi_{e,C}(p,\alpha) \geq 0$. Solving $\Pi_{e,C}(p,\alpha) = 0$ with (A.8) gives

$$p_0^C(\alpha) \;=\; \frac{c + f - \lambda L(1-\alpha)}{\lambda(2H-L)(1-\alpha) - f}. \tag{A.9}$$

the lowest demand–probability type willing to sign for a given royalty α.

A.2.2 Franchisor's Expected Payoff Across Market Under Each Strategy

Consider a uniform distribution $p \sim U[0,1]$ across market. Let $(\alpha_A^*, p_{0,A}^*)$, $(\alpha_B^*, p_{0,B}^*)$, and $(\alpha_C^*, p_{0,C}^*)$ be the equilibrium for Strategies A, B, and C, respectively. The expected profit across market:

- Strategy A (Franchising Only).

$$\begin{aligned} \mathrm{E}[\Pi_r]_A(\alpha) \;&=\; \int_{p_{0,A}}^{1} \Pi_{r,A}\big(p,\alpha\big)\,\mathrm{d}p \\ &=\; \alpha L\left(1 - p_0^A(\alpha)\right) \;+\; \frac{\alpha(2H-L)}{2}\left(1 - \left(p_0^A(\alpha)\right)^2\right), \end{aligned} \tag{A.10}$$

- Strategy B (Wait to Invest).

$$\begin{aligned} \mathrm{E}[\Pi_r]_B(\alpha) \;&=\; \int_{p_{0,B}}^{1} \Pi_{r,B}\big(p,\alpha\big)\,\mathrm{d}p \\ &=\; \alpha L\big(1 - p_0^B(\alpha)\big) + \frac{\alpha\big[(1+\lambda)H - L\big] + \lambda H - f - c}{2}\big(1 - (p_0^B(\alpha))^2\big). \end{aligned} \tag{A.11}$$

- Strategy C (Immediate Investment).

$$\begin{aligned} \mathrm{E}[\Pi_r]_C(\alpha) \;&=\; \int_{p_{0,C}}^{1} \Pi_{r,C}\big(p,\alpha\big)\,\mathrm{d}p \\ &=\; \big[\lambda L(1+\alpha) - f - c\big]\big(1 - p_0^C(\alpha)\big) \\ &\quad+ \frac{\lambda(2H-L)(1+\alpha) - f}{2}\big(1 - (p_0^C(\alpha))^2\big). \end{aligned} \tag{A.12}$$

where $p_0^A(\alpha)$, $p_0^B(\alpha)$, and $p_0^C(\alpha)$ are the franchisee break-even thresholds defined in Equations (A.3), (A.6), and (A.9), respectively.

A.3 Equilibrium Conditions

The franchisor's problem across the three strategies comes down to selecting an optimal royalty rate α. The franchisee break-even thresholds are strictly decreasing in α under the parameter restrictions $L \leq f$ and $\lambda H > (c+f)$, so higher royalties attract fewer franchisees. All three franchisors' ex-ante profit functions are continuous on $(0,1)$ and twice differentiable wherever the denominators of the cut-offs are non–zero, and vanish at the upper bound $\alpha = 1$ because no franchisee signs when the royalty is 100%.

Here are the equilibrium rules for α under the three strategies:

Strategy A (Franchising Only). The franchisor chooses

$$\alpha_A^* \;=\; \arg\max_{\alpha\in(0,1)} \; \mathrm{E}[\Pi_r]_A(\alpha).$$

Strategy B (Wait to Invest).

$$\alpha_B^* = \arg\max_{\alpha\in(0,1)} \mathrm{E}[\Pi_r]_B(\alpha),$$

Strategy C (Immediate Investment).

$$\alpha_C^* \;=\; \arg\max_{\alpha\in(0,1)} \; \mathrm{E}[\Pi_r]_C\,(\alpha), \tag{A.13}$$

where $\mathrm{E}[\Pi_r]_C(\alpha)$ is given in (A.12).

A.4 Numerical Illustration

To visualise how the optimal royalty and the franchisor's expected profit vary with demand uncertainty, we fix the start-up cost $c = 20$, the per-period fixed cost $f = 5$, the low–demand level $L = 5$, and let the high–demand level H range from the minimal value consistent with our parameter restriction, $H = 55 = 2(c + f) + 5$, up to $H = 200$. For each value of H, we solve the franchisor's optimization problem under the three contract strategies described in Sections A.2 and A.3.

We consider the following two scenarios: fixed markets ($\lambda = 0.5$), and growing markets ($\lambda > 0.5$).

A.4.1 Fixed Markets (Complete Cannibalization)

When $\lambda = 0.5$, meaning that franchisors' direct operating revenue completely comes from cannibalizing franchisees' revenue. The resulting equilibrium royalties α^*, franchisee break-even threshold p_0, and expected profits $E[\Pi_r]^*$ are plotted in Figures A.1, A.2 and A.3.

Based on these figures, we have the following observations:

Observation 1 (Optimal royalty ordering) *The equilibrium royalty rates under the three strategies follow the order:* $\alpha_A^* > \alpha_B^* > \alpha_C^*$.

We also provide an analytical proof of this observation in Section A.6.

The ordering $\alpha_A^* > \alpha_B^* > \alpha_C^*$ emerges naturally from the strategic interaction between franchisors and franchisees under varying degrees of encroachment. Under Strategy A, the franchisor does not directly compete with franchisees, making franchisees relatively less sensitive to changes in royalty rates. Consequently, franchisors find it profitable to increase royalties to maximize returns, resulting in the highest equilibrium royalty, α_A^*.

In Strategy B, franchisees anticipate potential future franchisor competition if market outcomes are favorable. This anticipated encroachment heightens franchisees' sensitivity to royalty adjustments (i.e., higher $p_0'(\alpha)$) compared to Strategy A. Therefore, the franchisor optimally chooses a lower

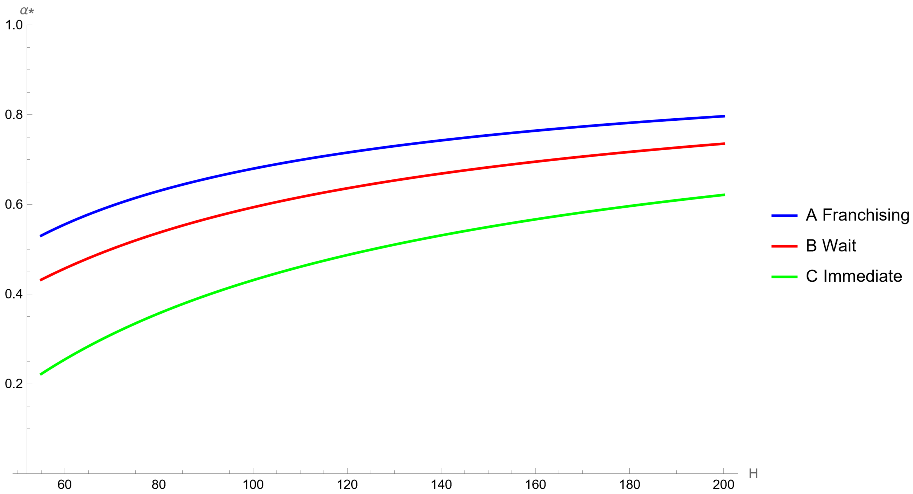

Figure A.1: Optimal royalty α^* as a function of H under $\lambda = 0.5$.

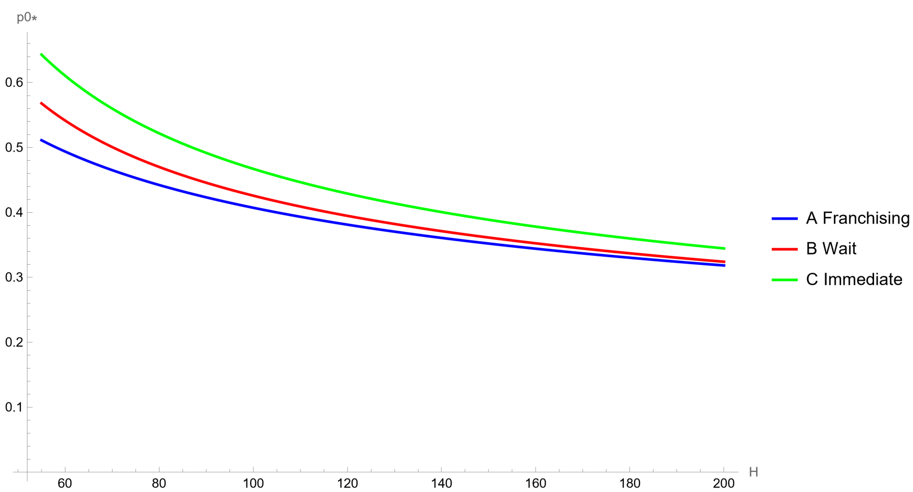

Figure A.2: Franchisee break-even threshold p_0 as a function of H under $\lambda = 0.5$.

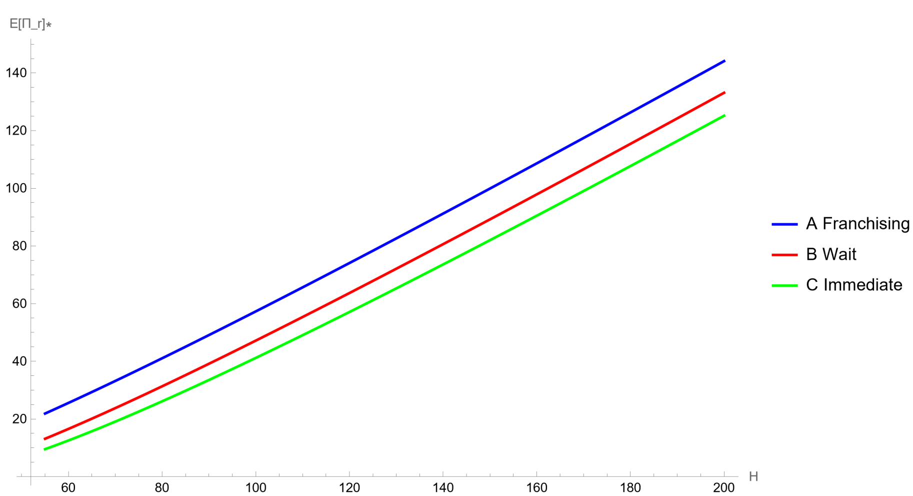

Figure A.3: Franchisor optimal expected profit $E[\Pi_r]^*$ as a function of H under $\lambda = 0.5$.

royalty rate (α_B^*) relative to Strategy A, balancing immediate royalty revenues against future direct investment opportunities.

Under Strategy C, the royalty serves a critical screening function for franchisors to identify high-potential markets. Moreover, immediate encroachment sharply increases franchisees' sensitivity to royalty changes relative to the other strategies (see Lemma 1 in Section A.6 for this result). Consequently, as Figures A.1 and A.2 show, Strategy C has the highest equilibrium royalty rate (α_C^*), but the highest equilibrium cut off probability (p_0^C), consistent with franchisees' heightened sensitivity to royalty under the threat of direct competition and the screening role of royalty rate.

Observation 2 (Profit ordering under complete cannibalization) *Under complete cannibalization ($\lambda = 0.5$) and model assumptions ($L \leq f$, $\lambda H > c + f$), the franchisor's optimal expected profits satisfy:*

$$\mathrm{E}[\Pi_r]_A^* > \mathrm{E}[\Pi_r]_B^* > \mathrm{E}[\Pi_r]_C^*.$$

In Section A.6, we also have provided an analytical proof for this observation (see Proposition 2).

When $\lambda = 0.5$, franchisor investment exclusively cannibalizes the franchisee's existing customer base without expanding overall market size. In this scenario, direct investment by the franchisor duplicates both fixed and startup costs without increasing total market revenue, thereby diminishing the aggregate surplus available for distribution. Additionally, the immediate or anticipated encroachment from franchisor investment significantly increases franchisees' sensitivity to changes in royalty rates. This heightened sensitivity is reflected in $\frac{dp_0^A}{d\alpha} < \frac{dp_0^B}{d\alpha} < \frac{dp_0^C}{d\alpha}$ (see Lemma 1 in Section A.6). Franchisees therefore require substantially lower royalty rates (α) to justify entry, which leads to higher franchisee demand thresholds (p_0) and consequently reduces franchise participation. Collectively, these dynamics reduce total surplus and shrink the franchisee base, leading to Strategy A consistently outperforming Strategies B and C under conditions of complete cannibalization.

A.4.2 Growing Markets (Partial Cannibalization)

In this case, we set $\lambda = 0.6 > 0.5$, meaning that franchisors' direct operating revenue partly comes from cannibalizing franchisees' revenue, and partly comes from market expansion. With $\lambda = 0.6$, franchisor direct investment increases the size of the local market by 20%. In the resulting equilibrium royalties α^*, franchisee break-even threshold p_0, and expected profits $E[\Pi_r]^*$ are plotted in Figures A.4, A.5 and A.6.

We have the following observations:

Observation 3 (Profit ordering under partial cannibalization) *Under $\lambda > 0.5$ and model assumptions ($L \leq f$, $\lambda H > c + f$), there exists a threshold $H^* > 0$ such that for all $H > H^*$:*

$$\mathrm{E}[\Pi_r]_C^* > \mathrm{E}[\Pi_r]_B^* > \mathrm{E}[\Pi_r]_A^*,$$

When a franchisor's investment expands the market, Strategy C results in market expansion. Moreover, higher demand potential (H) amplifies the market expansion effect. This multiplicative market growth allows the franchisor to achieve higher profit through direct investment.

When $\lambda > 0.5$, franchisor entry partially expands the market, significantly altering strategic implications. Franchisor entry generates incremental revenues beyond mere cannibalization, increasing

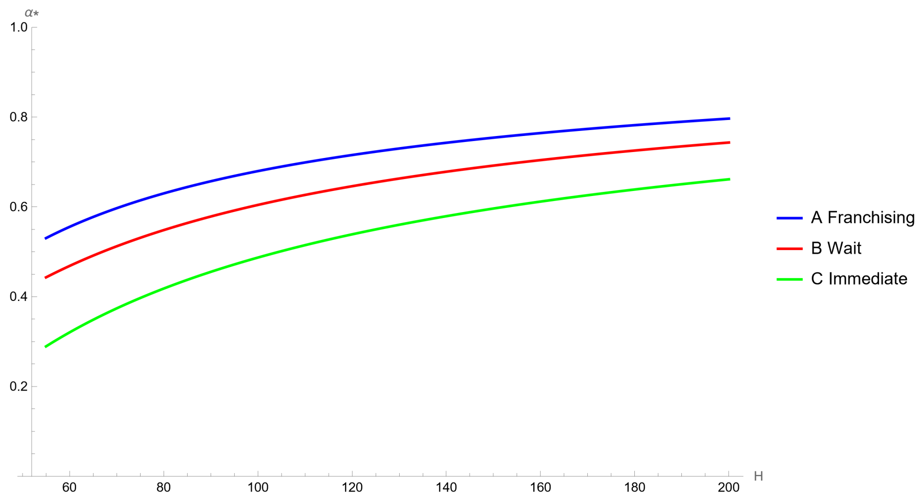

Figure A.4: Optimal royalty α^* as a function of H under $\lambda = 0.6$.

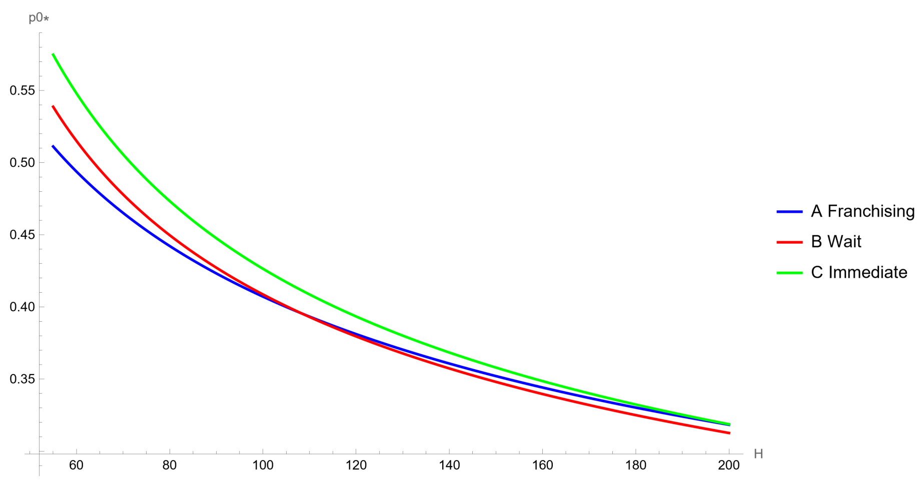

Figure A.5: Franchisee break-even threshold p_0 as a function of H under $\lambda = 0.6$.

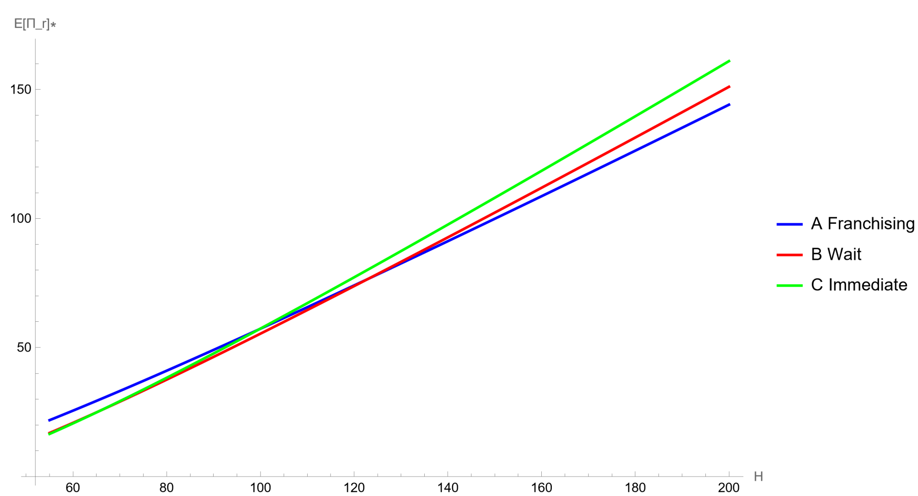

Figure A.6: Franchisor optimal expected profit $E[\Pi_r]^*$ as a function of H under $\lambda = 0.6$.

total market size and surplus, and thus enabling additional profit extraction opportunities. Immediate franchisor investment (Strategy C) allows the franchisor to capitalize directly and promptly on latent market potential, amplifying gains particularly at higher demand levels (H). Consequently, franchisee participation thresholds (p_0) rise more moderately compared to the pure cannibalization scenario. At sufficiently high market potential ($H > H^*$), Strategy C emerges as optimal, outperforming delayed investment (Strategy B) and the franchising-only approach (Strategy A).

This scenario, where $\lambda > 0.5$, more closely resembles our empirical setup. In practice, a franchisee's decision to open a store often reflects private knowledge about a market's future growth potential. Our empirical findings align closely with this perspective, showing that franchisee openings reliably predict subsequent house-price appreciation and other indicators of local economic expansion. In our model, high demand (H) not only represents current market attractiveness but also signals future growth potential, as indicated by the demand growth term $(pH+(1-p)L)(2\lambda-1)$, which increases with H. Franchisee-generated insights thus offer valuable private information about market prospects, enabling franchisors to strategically leverage and materialize this growth through timely direct investment. Viewed through this broader empirical lens, learning from franchisees and strategically expanding based on their insights could significantly enhance franchisor profitability relative to the short-run, fixed-demand benchmark initially presented in our model.

In summary, the numerical results generate interesting practical implications for franchisor strategy. Even though franchisees may possess valuable private information about local market potential, learning from this information to guide direct investment may not always be beneficial, especially when franchisor entry primarily cannibalizes existing franchisee revenue. Such investment duplicates operational costs, discourages franchisee participation, and limits the franchisor's flexibility to lower royalty rates due to the need for royalties to serve a screening function. In contrast, when a franchisor's investment effectively leverages franchisee signals to capitalize on genuine market growth opportunities, it can substantially enhance overall profitability and foster long-term growth.

A.5 Learning without Commitment

To simplify the analysis, we have assumed that in all three strategies, the franchisor would commit to her strategy through the contract so that the franchisee would face no ambiguity in his expected payoff. We now consider an alternative *Strategy D* in which the franchisor does not commit ex-ante to a specific investment timing strategy. In this scenario, franchisees would have to form initial beliefs about whether the franchisor will immediately invest (Strategy C) or wait until demand is realized (Strategy B), and calculate their expected payoff through backward induction. The timing and information structure are as follows:

- Period 0: The franchisor announces a royalty rate $\alpha \in (0, 1)$.
- Period 0: Franchisees of type p form beliefs and decide whether to accept or reject.
- Period 0.5: After acceptance but before demand realization, the franchisor chooses either immediate investment (Strategy C) or delayed investment (Strategy B).
- Period 1: Demand $D \in H, L$ is realized in period 1.
- Period 2: Decisions to operate a store or not are made based on the observed demand.

Franchisees evaluate their expected profitability based on their beliefs. If they anticipate delayed franchisor investment (Strategy B), their acceptance threshold is $p_0^B(\alpha)$. Conversely, if they anticipate immediate franchisor investment (Strategy C), their acceptance threshold is $p_0^C(\alpha) > p_0^B(\alpha)$. At stage $t = 1$, after observing acceptance, the franchisor chooses optimally between immediate and delayed investment. If franchisees accepted under the belief of delayed investment, those with types $p \in [p_0^B(\alpha), p_0^C(\alpha)]$ face negative profits if the franchisor chooses to invest immediately instead. Anticipating such encroachment, these marginal franchisees rationally reject the contract initially. Thus, equilibrium requires franchisees' beliefs and acceptance decisions to align precisely with the franchisor's optimal action.

Given the interdependence between franchisees' acceptance decisions and the franchisor's optimal investment timing, multiple equilibrium outcomes can emerge, depending on parameter values and announced royalty α:

1. Franchisees believe in delayed investment and adopt the threshold $p_0^B(\alpha)$. Under these conditions, the franchisor indeed prefers delayed investment, resulting in equilibrium profits $E[\Pi_r]_B^*$.

2. Franchisees believe in immediate investment and adopt the higher threshold $p_0^C(\alpha)$. Under these conditions, the franchisor indeed prefers immediate investment, resulting in equilibrium profits $E[\Pi_r]_C^*$.

Therefore, the equilibrium franchisor profits under this "learning without commitment" scenario should be $E[\Pi_r]_D^* = max(E[\Pi_r]_B^*,\ E[\Pi_r]_C^*)$.

From Observation 2, under complete cannibalization ($\lambda = \frac{1}{2}$), we have the following ordering:

$$E[\Pi_r]_A^* > E[\Pi_r]_B^*, \quad E[\Pi_r]_A^* > E[\Pi_r]_C^*.$$

Thus, under complete cannibalization, the franchising-only strategy (Strategy A) seems to dominates the learning without commitment strategy (Strategy D): $E[\Pi_r]_A^* > max(E[\Pi_r]_B^*, E[\Pi_r]_C^*) = E[\Pi_r]_D^*$. However, according to Observation 3, if cannibalization is partial ($\lambda > \frac{1}{2}$), immediate franchisor investment (Strategy C) can outperform franchising-only (Strategy A) for sufficiently high market potential H. Thus, under partial cannibalization and sufficiently high market potential, the "learning without commitment" strategy could also dominate franchising-only: $E[\Pi_r]_D^* = \max\{E[\Pi_r]_B^*, E[\Pi_r]_C^*\} \geq E[\Pi_r]_C^* > E[\Pi_r]_A^*$.

In sum, the possibility of franchisor learning without commitment does not alter our conclusion from the main analysis.

A.6 Supplementary Analytical Results

Lemma 1 (Cut-off ordering and monotonicity) *Fix parameters*

$$0.5 \;\leq\; \lambda < 1, \quad L \leq f, \quad \lambda H \geq c + f, \quad H > L > 0, \quad c > 0,\ f > 0,$$

then

$$0 < p_0^A(\alpha) \leq p_0^B(\alpha) \leq p_0^C(\alpha) \leq 1,$$

$$\textit{and} \quad \frac{dp_0^A}{d\alpha} < \frac{dp_0^B}{d\alpha} < \frac{dp_0^C}{d\alpha}$$

for any royalty $\alpha \in (0, 1)$ that ensure an interior solution of p_0 (i.e. $0 \leq p_0 \leq 1$).

Proof. The three break-even cut-offs are

$$p_0^A(\alpha) = \frac{c+f-L(1-\alpha)}{-f+(2H-L)(1-\alpha)},$$
$$p_0^B(\alpha) = \frac{c+f-L(1-\alpha)}{-f+(1-\alpha)\,(H-L+H\lambda)},$$
$$p_0^C(\alpha) = \frac{c+f-\lambda L(1-\alpha)}{-f+\lambda(2H-L)(1-\alpha)},$$

Cut-off orders. Define the two differences

$$\Delta_{(p)AB} := p_0^A - p_0^B, \qquad \Delta_{(p)BC} := p_0^B - p_0^C.$$

A single fraction gives

$$\Delta_{(p)AB}(\alpha) = -\frac{H\,(c+f-L(1-\alpha))\,(1-\alpha)\,(1-\lambda)}{\big[-f+(2H-L)(1-\alpha)\big]\,\big[-f+(1-\alpha)\,(H-L+H\lambda)\big]}, \tag{A.14}$$

$$\Delta_{(p)BC}(\alpha) = \frac{(1-\alpha)\,(1-\lambda)\,(-H(c+f)+L(c+2f)-\lambda HL(1-\alpha))}{\big[-f+\lambda(2H-L)(1-\alpha)\big]\,\big[-f+(1-\alpha)\,(H-L+H\lambda)\big]} \tag{A.15}$$

Given that $p_i \sim \text{Uniform}[0,1]$ and the numerators in $p_0^s(\alpha)$ $\;(s=(A,B,C))$ are all strictly positive, the denominators in $p_0^s(\alpha)$ are strictly positive unless α is large enough making some or all $p_0^s(\alpha)=1$. In that case, $0 < p_0^A(\alpha) \le p_0^B(\alpha) \le p_0^C(\alpha) \le 1$ holds. Therefore, we only consider the case when the denominators of (A.14) and (A.15) are strictly positive. The signs of these differences solely depend on the numerators.

Sign of $\Delta_{(p)AB}$. All factors in the numerator of (A.14) are positive except the leading "−", so $\Delta_{AB} < 0$ and therefore $p_0^A(\alpha) < p_0^B(\alpha)$.

Sign of $\Delta_{(p)BC}$. We evaluate $B_{\max} = -H(c+f)+L(c+2f)-\lambda HL(1-\alpha)$ in (A.15) by inserting the minimal admissible $H=(c+f)/\lambda$ (it maximises $B_{\max}$ by assumption $\lambda H \ge c+f$):

$$B_{\max} \;\le\; -\frac{(c+f)^2}{\lambda} + L\big(c+2f\big) - L\big(c+f\big)\big(1-\alpha\big) \;<\; 0 \qquad \big[\,0.5 \le \lambda < 1,\; L \le f\big].$$

Hence, the numerator of (A.15) is negative, and $p_0^B(\alpha) < p_0^C(\alpha)$.

Combining the two gaps and the corner solution case discussed above gives

$$0 < p_0^A(\alpha) \le p_0^B(\alpha) \le p_0^C(\alpha) \le 1.$$

Derivative orders. The α-derivatives of the three break-even cut-offs are

$$p_0^{A\prime}(\alpha) = \frac{(2H-L)(c+f)-Lf}{\big[-f+(2H-L)(1-\alpha)\big]^2},$$
$$p_0^{B\prime}(\alpha) = \frac{((1+\lambda)H-L)(c+f)-Lf}{\big[-f+(1-\alpha)\,(H-L+H\lambda)\big]^2},$$
$$p_0^{C\prime}(\alpha) = \frac{\lambda\big[(2H-L)(c+f)-Lf\big]}{\big[-f+\lambda(2H-L)(1-\alpha)\big]^2}.$$

Note that $p_0^{C\prime}(\alpha) = \frac{\lambda\left[(2H-L)(c+f)-Lf\right]}{\left[-f+\lambda(2H-L)(1-\alpha)\right]^2} > \frac{\lambda(2H-L)(c+f)-Lf}{\left[-f+\lambda(2H-L)(1-\alpha)\right]^2} = p_0^{\tilde{C}\prime}(\alpha)$. In order to prove $p_0^{A\prime}(\alpha) < p_0^{B\prime}(\alpha) < p_0^{C\prime}(\alpha)$, we can instead prove $p_0^{A\prime}(\alpha) < p_0^{B\prime}(\alpha) < p_0^{\tilde{C}\prime}(\alpha)$.

$$\text{Let }\ \Phi(k) = p_0^{s\prime}(\alpha)\ \ (s=(A,B,\tilde{C})) = \frac{\kappa_s(c+f)-Lf}{\big[\kappa_s(1-\alpha)-f\big]^2},$$

$\kappa_A = 2H-L > \kappa_B = (1+\lambda)H-L > \kappa_{\tilde{C}} = \lambda(2H-L)$.

Differentiate with respect to k:

$$\Phi'(\kappa) = -\,\frac{(c+f)\big[\kappa(1-\alpha)\big] + (c+f)f - 2Lf(1-\alpha)}{\big[\kappa(1-\alpha)-f\big]^3}.$$

To ensure we get an interior solution of $p_0(\alpha)$, $\kappa(1-\alpha)-f>0$,

$$\kappa(1-\alpha) > f \quad\Longrightarrow\quad (c+f)\kappa(1-\alpha) > (c+f)f,$$

so $(c+f)\big[\kappa(1-\alpha)\big] + (c+f)f - 2Lf(1-\alpha) > 2(c+f)f - 2Lf(1-\alpha) = 2f\big[(c+f)-L(1-\alpha)\big] > 0$, with the last inequality using $L(1-\alpha) < L \le f < c+f$. Hence $\Phi'(\kappa) < 0$ for every admissible κ.

Therefore $\Phi(\kappa)$ is strictly decreasing, thus

$$\kappa_A > \kappa_B > \kappa_{\tilde{C}} \implies p_0^{A\prime}(\alpha) < p_0^{B\prime}(\alpha) < p_0^{\tilde{C}\prime}(\alpha) < p_0^{C\prime}(\alpha).$$

■

Lemma 2 (Strict concavity of $E[\Pi_r]_s$ in α) *Let $\lambda \in [\,\frac{1}{2}, 1)$ and assume the baseline restrictions*

$$L \;\le\; f, \qquad \lambda H \;\ge\; c+f.$$

For every strategy $s \in \{A,B,C\}$ the franchisor's expected-profit function

$$E_s(\alpha) := \mathrm{E}\big[\Pi_r\big]_s(\alpha), \qquad \alpha \in (0,1),$$

is twice continuously differentiable and strictly concave*:*

$$\frac{d^2E_s}{d\alpha^2}(\alpha) \;<\; 0 \quad \textit{for all}\ \ \alpha\in(0,1).$$

Proof.

First and second derivatives of p_0^s. Write each acceptance threshold as

$$p_0^s(\alpha) \;=\; \frac{N_s(\alpha)}{D_s(\alpha)} \quad \big(s = A, B, C\big), \tag{A.16}$$

with

$$\begin{aligned} N_A &= c + f - L(1-\alpha), & D_A &= (2H - L)(1-\alpha) - f,\\ N_B &= c + f - L(1-\alpha), & D_B &= \big[(1+\lambda)H - L\big](1-\alpha) - f,\\ N_C &= c + f - \lambda L(1-\alpha), & D_C &= \lambda(2H - L)(1-\alpha) - f. \end{aligned}$$

Because $1-\alpha > 0$, $L \le f$, and $\lambda H \ge c + f$, all $N_s(\alpha)$ are positive. To ensure an interior solution of $p_0^s(\alpha)$, $D_s(\alpha) > 0$ for each s.

Differentiating (A.16):

$$(N_s', D_s') = (X_s, -Y_s), \quad X_s \in \{L, L, \lambda L\},\ Y_s \in \{2H - L, (1+\lambda)H - L, \lambda(2H - L)\},$$

all of which are strictly positive. Hence

$$p_0^{s\,\prime}(\alpha) = \frac{N_s' D_s - N_s D_s'}{D_s^2} = \frac{X_s D_s + Y_s N_s}{D_s^2} > 0,$$

and a further derivative gives

$$p_0^{s\,\prime\prime}(\alpha) = \frac{2Y_s X_s D_s + 2Y_s^2 N_s}{D_s^3} > 0.$$

Thus every acceptance threshold is *increasing and convex* in α.

Integrand $\Pi_{r,s}(p,\alpha)$ and its α-slope. For each strategy the franchisor's single-market profit can be written as an affine function in α:

$$\Pi_{r,s}(p,\alpha) = a_s(p) + b_s(p)\,\alpha,$$

with

$$\begin{aligned} a_A(p) &= 0, & b_A(p) &= 2pH + (1-p)L;\\ a_B(p) &= p\big[\lambda H - f - c\big], & b_B(p) &= p\big[H + \lambda H\big] + (1-p)L;\\ a_C(p) &= p\big(2\lambda H - 2f\big) + (1-p)\big(\lambda L - f\big) - c, & b_C(p) &= \lambda\,[2pH + (1-p)L]. \end{aligned}$$

Noter that

$$\partial_\alpha \Pi_{r,A} = b_A(p), \quad \partial_\alpha \Pi_{r,B} = b_B(p), \quad \partial_\alpha \Pi_{r,C} = b_C(p), \qquad \text{all strictly positive.}$$

First derivative of E_s. With the uniform density 1 on $[0,1]$,

$$E_s(\alpha) = \int_{p_0^s(\alpha)}^{1} \Pi_{r,s}(p,\alpha)\,dp.$$

Leibniz' rule yields

$$E_s'(\alpha) = \int_{p_0^s}^{1} \partial_\alpha \Pi_{r,s}(p,\alpha)\, dp - \; \Pi_{r,s}\big(p_0^s(\alpha),\alpha\big)\; p_0^{s\,\prime}(\alpha). \tag{A.17}$$

Second derivative of E_s. Differentiating (A.17) yields:

$$E_s''(\alpha) = -\, 2\partial_\alpha \Pi_{r,s}\big(p_0^s,\alpha\big)\, p_0^{s\,\prime} \;-\; \Pi_{r,s}\big(p_0^s,\alpha\big)\, p_0^{s\,\prime\prime} \;-\partial_p \Pi_{r,s}\big(p_0^s,\alpha\big)\big(p_0^{s\,\prime}\big)^2. \tag{A.18}$$

Based on previous derivation, $\partial_\alpha \Pi_{r,s}(p_0^s,\alpha)$, $p_0^{s\,\prime}$ and $p_0^{s\,\prime\prime}$ are all strictly positive. $\partial_p \Pi_{r,s}\big(p_0^s,\alpha\big)$ is strictly positive for all three strategies: the higher p is, the more revenue one will generate. Finally, $\Pi_{r,s}(p_0^s,\alpha) \geq 0$: for all three strategies, at franchisee's break-even point, franchisor earns the entire total surplus.

Therefore each summand in (A.18) is strictly negative, and so

$$E_s''(\alpha) \;<\; 0 \qquad \forall\, \alpha \in (0,1).$$

■

Lemma 3 (Profit function derivative ordering) *For every $\alpha \in (0,1)$ and before p_0 reaches 1*

$$E_A'(\alpha) \;>\; E_B'(\alpha) \;>\; E_C'(\alpha).$$

Proof. Using the representation $E_s'(\alpha) = I_s(\alpha) - B_s(\alpha)$ just derived in Lemma 2,

$$I_s(\alpha) = \int_{p_0^s}^{1} \partial_\alpha g_s\, dp, \quad B_s(\alpha) = g_s(p_0^s,\alpha)\, p_0^{s\,\prime}.$$

(i) Integral part. From the explicit formulas $\partial_\alpha g_A = 2pH + (1-p)L$, $\partial_\alpha g_B = (1+\lambda)pH + (1-p)L$, $\partial_\alpha g_C = \lambda[2pH + (1-p)L]$ and $2 > 1+\lambda \geq \frac{3}{2}$, we have strict ordering $\partial_\alpha g_A > \partial_\alpha g_B > \partial_\alpha g_C$. Because the domain of integration also shrinks as s moves from A to C (Lemma 1),

$$I_A(\alpha) \;>\; I_B(\alpha) \;>\; I_C(\alpha) \;>\; 0.$$

(ii) Boundary part. Lemma 1 gave $p_0^A \leq p_0^B \leq p_0^C$ and $p_0^{A\,\prime} < p_0^{B\,\prime} < p_0^{C\,\prime}$. Moreover, because at the franchisee's break-even point, the franchisor earns all of the total surplus which increases with p_0^S, since $p_0^A \leq p_0^B \leq p_0^C$, we get $g_A(p_0^A,\alpha) \leq g_B(p_0^B,\alpha) \leq g_C(p_0^C,\alpha)$. Hence $B_A < B_B < B_C$.

Putting (i)–(ii) together gives the profit function's derivative ordering. ■

Proposition 1 (Optimal royalty ordering)

$$\alpha_A^* \;>\; \alpha_B^* \;>\; \alpha_C^*.$$

Proof. By Lemma 2 each E_s is strictly concave, so its maximiser is unique and characterised by $E_s'(\alpha_s^*) = 0$.

At $\alpha = \alpha_A^*$, $E_A' = 0$ and Lemma 3 implies $E_B' < 0$; concavity of E_B then leads to $\alpha_B^* < \alpha_A^*$.

At $\alpha = \alpha_B^*$, $E_B' = 0$ and Lemma 3 implies $E_C' < 0$, concavity of E_C then leads to $\alpha_C^* < \alpha_B^*$.

Chaining the two inequalities gives the claimed ordering. ■

Proposition 2 (Optimal profit ordering under $\lambda = \frac{1}{2}$) *Fix parameters (L, c, f, H) that satisfy*

$$L \le f \quad \text{and} \quad \tfrac{1}{2}H \;>\; c + f.$$

Let $\lambda = \frac{1}{2}$ (pure cannibalisation). Denote by $E[\Pi_r]^*_s = \max_{\alpha\in(0,1)} E[\Pi_r]_s(\alpha)$, $s \in \{A, B, C\}$, the franchisor's optimal expected profits under Strategies A (franchising only), B (wait–to–invest), and C (immediate investment). Then

$$E[\Pi_r]^*_A \;>\; E[\Pi_r]^*_B; >\; E[\Pi_r]^*_C.$$

Proof. Write $R(p) = 2pH + (1-p)L$ for the two-period gross revenue when a *single* store operates in market type $p \in [0, 1]$. The proofs below compare profits *market-by-market* after matching the marginal franchisee.

Strategy A dominates Strategy C. Fix any royalty $\alpha_C \in (0, 1)$ in Strategy C and let p_0^C be the franchisee cut-off that solves $\Pi_{e,C}(p_0^C, \alpha_C) = 0$. Choose

$$\hat{\alpha}_A = \frac{1+\alpha_C}{2} \in (0, 1).$$

A franchisee of type p_0^C is indifferent under Strategy A if offered $\hat{\alpha}_A$, hence every $p \ge p_0^C$ accepts it. For such p the franchisor's per-market profit gap is

$$\begin{aligned}\Delta_{AC}(p) &= \Pi_{r,A}(p, \hat{\alpha}_A) - \Pi_{r,C}(p, \alpha_C)\\ &= \tfrac{1+\alpha_C}{2}R(p) - \Big[\tfrac{\alpha_C}{2}R(p) + \tfrac{1}{2}R(p) - c - (1+p)f\Big]\\ &= c + (1+p)f \;>\; 0.\end{aligned}$$

(The inequality is strict except at $p = p_0^C$.) Hence $E[\Pi_r]^*_A \ge E[\Pi_r]_A(\hat{\alpha}_A) > E[\Pi_r]_C(\alpha_C)$; because α_C was arbitrary, $E[\Pi_r]^*_A > E[\Pi_r]^*_C$.

Strategy B dominates Strategy C. Fix α_C and the same cut-off p_0^C. Set

$$k = \frac{p_0^C H + \frac{1}{2}(1-p_0^C)L}{1.5\, p_0^C H + (1-p_0^C)L}, \qquad \hat{\alpha}_B = 1 - k\big(1-\alpha_C\big) \in (\alpha_C, 1).$$

The franchisee of type p_0^C is indifferent under Strategy B if offered $\hat{\alpha}_B$, so every $p \ge p_0^C$ signs. For those markets

$$\Delta_{BC}(p) = \Pi_{r,B}(p, \hat{\alpha}_B) - \Pi_{r,C}(p, \alpha_C) \;=\; f + (1-p)c \;>\; 0.$$

(The inequality is strict except at $p = p_0^C$.) Thus $E[\Pi_r]^*_B > E[\Pi_r]^*_C$.

Strategy A dominates Strategy B. Retain the royalty $\hat{\alpha}_B = 1 - k\big(1-\alpha_C\big)$ just defined and the associated cut-off $p_0^B = p_0^C$. Choose

$$\hat{\alpha}_A = \frac{1+\alpha_C}{2}.$$

The franchisee of type p_0^B is indifferent under Strategy A if offered $\hat{\alpha}_A$, so every $p \geq p_0^B$ accepts it. For any such p

$$\Delta_{AB}(p) = \Pi_{r,A}(p, \hat{\alpha}_A) - \Pi_{r,B}(p, \hat{\alpha}_B) \;\; = \;\; p(c + f) \; > \; 0.$$

(The inequality is strict except at $p = p_0^B$.) Consequently $E[\Pi_r]_A^* > E[\Pi_r]_B^*$.

Combining the steps above yields $E[\Pi_r]_A^* > E[\Pi_r]_B^* > E[\Pi_r]_C^*$, establishing the proposition. ∎

B Appendix: Additional Empirical Results

B.1 Geographic distribution of McDonald's outlets

Table B.1 compares state-level statistics on McDonald's establishments from our dataset with official McDonald's data disclosed in 2016. The correlation in the number of franchisee-owned (franchisor-owned) establishments between our data and the official data is 0.996 (0.861). Extant literature has discussed the quality issues of the NETS data, such as imputation of sales and employment data and the imprecise timing of firm entries and exits (Crane and Decker, 2019). In our context, if the entry or exit timing is recorded inaccurately, this measurement error would introduce noise into our franchisee openings variable, likely leading to attenuation bias in the estimated relationship between franchisee investment and future house price growth.

B.2 Franchised establishments by industry

We classify franchise establishments into 16 industries following the definitions provided by *Entrepreneur*, the source of our franchise data. The top three industries—food (e.g., McDonald's, Pizza Hut), personal service (e.g., Great Clips, The UPS Store), and retail (e.g., Ace Hardware, 7-Eleven)—together account for 95.4% of all franchise establishments in our sample. Notably, these three sectors exhibit statistically significant positive associations between the number of franchisee-owned new establishments and subsequent local house price growth (see Table B.2). Specifically, the estimated coefficients are 0.139 (t = 1.99) for food, 0.113 (t = 2.29) for personal services, and 0.149 (t = 1.66) for retail. These industries predominantly cater to local consumer bases and are highly sensitive to location-specific conditions. Thus, these results underscore the importance of franchisee decisions in location-sensitive industries as meaningful predictors of local economic vitality, reflected by house price growth.

B.3 Alternative house price measure

Table B.3 shows that our baseline estimates from Model (1) remain robust when using Zillow's Home Value Index (ZHVI) as an alternative proxy for local house price levels. This confirms that our main findings are not driven by the choice of house price measure.[B.2]

B.4 County level house price predictability

We define ZIP codes as our primary unit of analysis to capture the fine-grained, localized nature of franchisee investment decisions. Franchisees often rely on hyper-local knowledge, such as foot traffic patterns, neighborhood reputation, and local economic momentum, that can vary meaningfully within counties. By conducting our analysis at the ZIP code level, we are better able to detect the informational advantage held by franchisees in choosing specific local markets. To assess whether our main findings persist under broader geographic definitions, we conduct a robustness check using county-year level data. The results are reported in Table B.4, where the dependent variable is the change in county-level house prices over one-, two-, and three-year horizons. These regressions include

[B.2] According to Zillow, the Zillow Home Value Index (ZHVI) is "a smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35th to 65th percentile range." Source: https://www.zillow.com/research/data/.

the same county- and state-level controls as our ZIP code-level model (Table 2, columns (4)–(6)), with fixed effects now specified at the county and year levels.

The results show a notably different pattern. At the county level, franchisor-owned outlet openings are positively and significantly associated with future house price growth across all horizons, while franchisee-owned openings show no predictive power and, in the short run, a negative association. This contrast highlights a key trade-off in the level of geographic aggregation. Broader geographic units, such as counties, may better capture regionally scaled investment decisions made by the franchisors, such as a franchisor's strategic choice to enter a large metropolitan market (e.g., "the Atlanta area"), but obscure the more granular signals that franchisees rely on to identify specific high-growth neighborhoods or ZIP codes within those regions. This pattern reinforces our decision to define markets at the ZIP code level as the appropriate unit for measuring franchisees' localized information advantage.

B.5 Alternative explanation: home equity as collateral for business ventures

A potential concern is that franchisees might indirectly influence local house prices by using home equity as collateral for financing their business ventures. Conceptually, however, this mechanism is unlikely to drive our results. Pledging a home as collateral does not mechanically increase its appraised market value, as appraisals are determined by banks based on existing comparable sales in the local market. Moreover, the typical number of franchisee openings per ZIP code per year is small (mean = 0.486, SD = 0.918). This limited scale makes it implausible that collateral-based borrowing would materially affect aggregate housing market dynamics at the ZIP code level.

To empirically address this collateral-based explanation, we further examine whether increased mortgage refinancing activity, which could potentially reflect home-equity financing for business investment, precedes franchisee investments. Our baseline regression results remain robust to extensive controls for local mortgage market conditions. Additionally, we conduct a placebo analysis explicitly examining trends in mortgage refinancing activity in the years prior to franchisee openings. As reported in Table B.5 in the Internet Appendix, there is no significant increase in refinancing activity leading up to franchisee investments. This lack of elevated refinancing activity prior to franchisee entry alleviates concerns that mortgage market dynamics drive franchisee investment decisions.

B.6 Characteristics of franchise versus non-franchise business groups

Given that franchising provides franchisors with access to local franchisees' information, an important follow-up question is how they differ from companies that do not choose franchising as their organizational form. We address this question by comparing the operational characteristics of franchise and non-franchise business groups within the same industry. Our sample consists of a cross-sectional snapshot of 2,880 public business groups from 2016, including 84 franchise groups, each matched with one or more non-franchise groups from the same 4-digit SIC industries, totaling 2,796 non-franchise groups. We estimate the following regression at the business-group level:

$$\text{Characteristic}_g = \beta \text{Franchise Group}_g + \gamma \# \text{ Est.}_g + \phi_s + \delta_i + \varepsilon_g, \tag{B.1}$$

where *Franchise Group* is a binary indicator equal to one if the business group operates as a franchise. The dependent variables include geographic scope (number of states and counties), the logarithm of average distance from establishments to headquarters, and average local market characteristics (population and income). The regression controls for the total number of establishments ($\#\text{ Est.}_g$), headquarters state fixed effects (ϕ_s), and industry fixed effects (δ_i).

The results reported in Table B.6 indicate significant differences between franchise and non-franchise business groups. Franchise groups operate across significantly broader geographic scopes (12 additional states and 102 additional counties, both significant at the 1% level) and have establishments located further from their headquarters ($e^{54\%} - 1 = 72\%$ greater average distance, significant at the 1% level). However, we find no statistically significant differences in the average population or income levels of markets entered by franchise versus non-franchise groups, suggesting that franchise organizations do not simply target systematically different types of local markets.

These findings raise the question of why some firms choose not to franchise despite the apparent benefits of geographic expansion and local market insights provided by franchisees. One plausible explanation is that non-franchise firms prioritize tighter managerial control and direct oversight to maintain consistent brand quality, customer experience, and corporate culture. These firms may also strategically choose to operate within geographically concentrated areas where direct management is more feasible and cost-effective. Additionally, some firms might possess proprietary business practices or operational secrets that they prefer to protect through direct ownership rather than risk exposure to third-party franchisees. Consequently, the choice not to franchise may reflect a deliberate strategy focused on factors other than broad geographic expansion.

Table B.1: State distribution of McDonald's establishment in 2016

This table presents the state-level number of McDonald's establishments in 2016 based on our data and the official data from franchise disclosure documents.

	(1) Franchisor-owned	(2) Franchisor-owned (Official data)	(3) Franchisee-owned	(4) Franchisee-owned (Official data)
AL	3	0	224	254
AK	5	0	26	31
AZ	14	11	257	280
AR	0	0	166	175
CA	94	100	1176	1216
CO	4	14	200	200
CT	8	0	121	142
DC	0	2	26	29
DE	0	3	31	35
FL	67	107	789	793
GA	28	31	382	443
HI	16	23	51	51
ID	0	0	58	63
IL	137	93	546	579
IN	34	63	303	292
IA	20	5	128	149
KS	10	0	140	248
KY	18	30	212	227
LA	3	0	214	244
ME	2	0	49	62
MD	58	62	236	233
MA	34	0	195	240
MI	66	95	471	458
MN	24	33	199	194
MS	0	0	129	148
MO	24	0	264	316
MT	0	0	45	48
NE	12	21	65	59
NV	7	16	121	128
NH	0	0	56	58
NJ	16	5	227	260
NM	0	0	89	105
NY	39	17	600	622
NC	29	41	415	454
ND	0	0	23	25
OH	64	79	568	545
OK	13	24	183	181
OR	3	0	153	166
PA	37	12	440	500
RI	0	0	28	31
SC	5	9	190	223
SD	0	0	28	30
TN	9	21	290	321
TX	32	78	1078	1131
UT	0	0	95	118
VT	0	0	25	28
VA	31	41	362	371
WA	27	43	240	226
WV	2	2	89	102
WI	20	25	261	275
WY	0	0	30	29
Correlation	0.861		0.996	

Table B.2: Openings of franchised establishments and future house price growth: By industry

This table reports the coefficient estimates of *ln(Franchisee New$_{t-1}$)* from Model (1), with the variable capturing only new openings within the specified industry. The dependent variable is the growth rate of HPI from year t to $t+2$. Industry classifications follow the definitions provided by *Entrepreneur*, the source of our franchise data. Industries are sorted in descending order by the t-statistic of their respective coefficient estimates. The *Cumulative Percent* column reports the cumulative share of all franchisee-owned establishments represented by each industry, summing from the top down. All regressions include ZIP code and county × year fixed effects, and control variables identical to those in Table 2, column (5). Standard errors are clustered at the county level, and the corresponding t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	$\%\Delta HPI_{t,t+2}$	
Industry	Coefficient	Cumulative Percent (%)
Personal Service	0.113**	37.2
	(2.29)	
Food	0.139**	87.5
	(1.99)	
Retail	0.149*	95.4
	(1.66)	
Business Service	0.171	96.2
	(1.59)	
Home Improvement	0.129	97.2
	(1.03)	
Automotive	0.280	97.8
	(1.00)	
Financial Service	0.104	97.8
	(0.88)	
Pet Business	0.270	97.9
	(0.70)	
Personal Care	0.059	99.0
	(0.57)	
Maintenance	0.017	99.1
	(0.13)	
Recreation	-0.006	99.1
	(-0.05)	
Children's Service	-0.068	99.2
	(-0.25)	
Tech Business	-0.053	99.2
	(-0.31)	
Education	-0.816	99.2
	(-1.06)	
Healthcare	-0.263	99.2
	(-1.18)	
Hotel	-0.561	100.0
	(-1.24)	

Table B.3: Openings of franchised establishments and future house price growth: Zillow Home Value Index

This table presents the estimates of Model (1) using Zillow Home Value Index (*ZHVI*) an alternative measure of house prices. The dependent variables are the growth rate of *ZHVI* from year t to $t+1$, $t+2$, and $t+3$. The key independent variable is ln(*Franchisee New*$_{t-1}$). We include the same control variables as those in Table 2, columns (4)–(6). All regressions include ZIP code and year fixed effects. Standard errors are clustered at the county level, and the corresponding t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	%ΔZHVI$_{t,t+1}$	%ΔZHVI$_{t,t+2}$	%ΔZHVI$_{t,t+3}$
	(1)	(2)	(3)
ln(Franchisee New$_{t-1}$)	0.2170***	0.5505***	0.5504***
	(3.14)	(4.52)	(3.98)
ln(Franchisor New$_{t-1}$)	-0.0587	-0.1628	-0.0766
	(-0.58)	(-1.04)	(-0.36)
Controls	Y	Y	Y
ZIP code FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	149,468	149,468	149,468
Adjusted R-squared	0.633	0.694	0.733

Table B.4: Openings of franchised establishments and future house price growth: County level evidence

This table presents the estimates of Model (1) using county-year level data. The dependent variables are county-level *HPI* growth rates from year t to $t+1$, $t+2$, and $t+3$. We include the same county- and state-level control variables as those in Table 2, columns (4)–(6). All columns include county and year fixed effects. Standard errors are clustered at the county level, and the corresponding t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	%ΔHPI$_{t,t+1}$	%ΔHPI$_{t,t+2}$	%ΔHPI$_{t,t+3}$
	(1)	(2)	(3)
ln(Franchisee New$_{t-1}$)	-0.1654***	-0.0453	-0.0554
	(-3.00)	(-0.53)	(-0.53)
ln(Franchisor New$_{t-1}$)	0.2455***	0.6470***	0.7600***
	(2.96)	(5.00)	(4.78)
Controls	Y	Y	Y
County FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	28,986	28,986	28,986
Adjusted R-squared	0.386	0.590	0.670

Table B.5: Refinancing and openings of franchised establishments

This table presents the estimates of Model (1) using lagged mortgage growth as the dependent variable. Specifically, the dependent variables are the growth rate of the total number and the dollar amount of refinancing loans (all mortgage loans) approved from year $t-4$ to $t-1$ in columns (1) and (2) (columns (3) and (4)). The key independent variable is ln(*Franchisee New*$_{t-1}$). We include the same control variables (except the mortgage variables) as those in Table 2, columns (4)–(6). All regressions include ZIP code and year fixed effects. Standard errors are clustered at the county level, and the corresponding t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	%Δ#Refinancing$_{t-4,t-1}$	%Δ\$Refinancing$_{t-4,t-1}$	%Δ#Mortgage$_{t-4,t-1}$	%Δ\$Mortgage$_{t-4,t-1}$
	(1)	(2)	(3)	(4)
ln(Franchisee New$_{t-1}$)	3.3752	4.1004	2.9380	2.4770
	(1.34)	(1.54)	(1.05)	(0.84)
ln(Franchisor New$_{t-1}$)	22.6122***	22.5365***	26.5645***	27.8890***
	(4.36)	(4.06)	(4.67)	(4.66)
Controls	Y	Y	Y	Y
ZIP code FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	132,890	132,890	132,964	132,964
Adjusted R-squared	0.279	0.270	0.285	0.280

Table B.6: Business group characteristics: Franchise vs. non-franchise

This table presents estimates from Model (B.1) using business-group level data. The dependent variables include the number of states and counties where the business group operates (*# States* and *# Counties* in columns (1) and (2)); the natural logarithm of the average distance between the establishments and the group's headquarters (*ln(avg.Dist to HQ)* in column (3)); the natural logarithm of the average population of counties where the group operates (*ln(avg.Population)* in column (4)); and the natural logarithm of the average income of ZIP codes where the group operates (*ln(avg.Income)* in column (5)). The key independent variable, *Franchise Group*, is a binary indicator equal to one if the business group operates as a franchise. The model also controls for the total number of establishments (# Est.) All specifications include fixed effects for the headquarters state and the four-digit SIC industry. Standard errors are clustered at the four-digit SIC level, and the corresponding t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	# States	# Counties	ln(avg.Dist to HQ)	ln(avg.Population)	ln(avg.Income)
	(1)	(2)	(3)	(4)	(5)
Franchise Group	12.0820***	101.8385***	0.5421***	-0.0056	-0.0065
	(6.93)	(3.34)	(6.30)	(-0.10)	(-0.37)
# Est.	0.0052***	0.1442***	0.0002***	-0.0000	-0.0000***
	(2.66)	(5.43)	(2.77)	(-0.93)	(-4.53)
HQ state FE	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Observations	2,880	2,880	2,868	2,880	2,875
Adjusted R-squared	0.322	0.733	0.058	0.313	0.238

C Appendix: Detailed Data Cleaning Process

There are two parts in this data cleaning process, (1) linking establishments to franchise brands and (2) identifying ownership of the establishments.

C.1 Linking establishments to franchise brands

1. We start with a comprehensive list of franchise brand names compiled by *Entrepreneur Magazine* in 2019. *Entrepreneur* is an American magazine that focuses on news stories regarding entrepreneurship, small business management establishments. There are 2,084 franchise brands, each with a unique key (FKey). This list contains franchise brand information such as headquarters address and CEO name. Each FKey identifies a unique franchise brand, but different franchise brands could be associated with the same business group. For example, KFC, Pizza Hut, and Taco Bell are owned by Yum! Brands, Inc., but they have their own unique FKey's. We also obtain a list of establishments from NETS database. There are 64,493,329 establishments in total with 46,447,581 unique names. Each of these establishments is assigned a unique ID (Duns number) and is linked to a parent firm that also has a Duns number (HQDuns). The information on parent firms from NETS allows us to identify establishments under the same business group.

2. We extract the keywords from the list of franchise brands from the *Entrepreneur Magazine* and search for these keywords in the names of establishments from NETS. In some cases, the company names of the establishments from NETS are missing. In these cases, we use the "trade name" instead. There are 499,347 establishments that contain one of the keywords in the names. Among these results, 191,036 are exact matches (i.e., the establishments' names contain only the exact keywords).

3. From the non-exact matches of the 499,347 observations, we manually look for cases containing the exact keyword of a franchise brand plus the location's name. For example, the "McDonald's of Richardson" should be matched to the McDonald's. After the manual check, we find another 5,609 matches and now have 196,645 matches total.

4. We then extract the identifier (HQDuns) of the parent firms of these matched establishments. This procedure gives us 150,659 unique links between the franchise brand identifier (FKey) and the NETS parent firms (HQDuns). Using these FKey-HQDuns links, we can find other parent franchise company subsidiaries from the 499,347 observations that contain a keyword but are not precisely matched. We identify an additional 9,717 establishments from this approach.

5. After the keyword search, we perform another round of name matching between the name list from *Entrepreneur Magazine* and the name list from NETS. First, we standardize the NETS company names by removing stop words such as "the", ".com", "inc", "ltd", "llc", "division", "corporation", "corp", "company", and "co". We then perform a fuzzy name match in SAS using the SPEDIS function and allow the estimated name distance to be up to 20. Doing so allows us to find imperfect matches due to misspellings. Furthermore, we repeat Step 4 to find other establishments affiliated with the same parent firms as the matched establishment from the fuzzy name matching. We find another 7,808 matches from this step, adding up to 272,438 matches.

6. After removing franchise brands for which we only find one matched establishment, the final sample consists of 272,196 observations for 1,220 franchisors. This conservative approach re-

quires minimal manual checks and relies primarily on exact keyword matches. Thus, while our sample could miss some valid observations from the data (i.e., false negatives), the matched observations in our sample are of high reliability (i.e., low false positives).

C.2 Identifying ownership type

We identify the ownership type of an establishment by determining whether its parent firm establishment (i.e. HQDuns from NETS) can be identified as the franchise company. In what follows, we categorize establishments as franchisor-owned if we determine their parent establishment as one of the franchise companies. We then organize the rest of the establishments as franchisee-owned.

1. We identify 185,577 unique FKey-HQDuns pairs from the 272,196 observations. Since we have information on the headquarters address and CEO from the *Entrepreneur Magazine* and NETS, we look for FKey-HQDuns pairs that share the same ZIP code and either the CEO's first or last name. Then we manually verify the matched pairs and find 349 parent firms matched to 313 franchise companies.[C.3] There are 8,644 establishments affiliated with these 349 HQDuns that we tentatively categorize as franchisor-owned establishments.

2. To keep the data checking work manageable, we restrict the list of HQDuns that we consider as potential franchisors based on a rule of thumb. Specifically, we first count the number of subsidiary establishments for each HQDuns. For each franchise brand, there exists one HQDuns that has the largest number of subsidiary establishments. We then only consider HQDuns affiliated with the same franchise brand that has at least one-third of the number of establishments as the largest one as potential franchisors. 4,647 FKey-HQDuns links meet this criterion, and 180 franchise brands have only one HQDuns meeting this criterion. In other words, for these 180 franchise brands, the second largest HQDuns is more than 66% smaller than the largest one in the number of subsidiary establishments. The remaining 4,467 HQDuns have at least one peer HQDuns that can be considered as a potential franchisor.

3. We assign a score for the 4,647 FKey-HQDuns links that we keep from Step 2. The score equals one if the FKey and HQDuns can match on either headquarters ZIP code or CEO last name, two if they match only on headquarters city, three if they match only on headquarters state, and five if they do not match based on any of the above information.[C.4]

4. For FKey's that have been matched to an HQDuns with a score of 1, we keep those FKey-HQDuns as franchisors and remove other matched HQDuns from consideration. There are 143 such FKey-HQDuns matches. For FKey's that do not have any matched HQDuns with scores equal to 1, we keep those with scores equal to 2 or 3. There are 122 such FKey-HQDuns matches. We have so far found 614 FKey-HQDuns matches from Steps 1-4 that we consider potential franchisors (349+143+122=614).

5. We further confirm the status of these FKey-HQDuns based on the establishment category status from NETS. Specifically, we keep only FKey-HQDuns for which the establishment category is "headquarters". Ultimately, we identify 418 FKey-HQDuns as franchisors and categorize the rest of FKey-HQDuns as franchisees.

6. Finally, we merge all the franchisor and franchisee FKey-HQDuns links with NETS establishments identified from Section C.1. The final sample consists of 208,598 establishments from 389 franchise brands.

[C.3] There are a few franchise companies with more than one HQDuns identified as the franchise company in this step.
[C.4] We manually check these name matches to ensure there is no error.