

Online Appendices - Simulating competitive collocation strategies with deep reinforcement learning

Appendix I: Simulation approach

This section below here provides a more elaborate description (compared to the shortened version in the main paper) of the firm-world environment, the deep reinforcement learning, and the simulation logic:

Firm-world environment

We simulate a world comprising 50 countries (markets) under varying conditions and run the simulation for 50 time periods. A firm starts in a randomly allocated home country, and in each period it can expand to another country.

Initially, a single **1** in a string of **0**s represents a firm's home country. The corresponding vector $\mathbf{p}_{i,t}$ summarizes the countries in which a firm (i) is present at time t . At time t_0 , this could look like:

$$\mathbf{p}_{i,t_0} = (0,0,\dots,0,1,0,0,\dots,0).$$

If firm (i) decides to expand internationally to, for example, a neighboring country in t_1 , the string will change to:

$$\mathbf{p}_{i,t_1} = (0,0,\dots,0,1,1,0,\dots,0).$$

For each country, the simulation assumes a specific market size (in terms of the firm's product market). We allow different relationships across country markets to make the simulation more realistic (and hence more complex). In all cases, there is a vector of possible returns for all countries (j) in the simulation for every year t (i.e., the *market size*). The matrix representing these vectors is $\mathbf{\Pi}_t$. It will always have a randomly drawn component (following different multivariate distributions with mean α and covariance matrix $\mathbf{\Sigma}$), multiplied by a constant that represents the expected market potential (μ_j) of a specific country. The mean (α) for the multivariate distribution is randomly drawn from a uniform distribution ($0.8 \leq \alpha < 1.2$) allowing for differences between realized and potential market size. The market size

variation msv ($100 \leq msv_j < 2000$) determines the variation of the constant (μ_j) allowing for market size to vary across countries due to differences in GDP, country size, or demand. For some industries or markets, firms might face markets with similar potential across countries, while in others the market potential might be more directly coupled to country size. Consequently, the expected market potential ranges between 500 and 500 plus the market size variation. While the mean is a technical parameter of the distribution, the msv parameter is chosen to align with the range of entry barriers. In other words, we have calibrated the range of market size variation and entry barriers as a payback period – that is, the time needed to recover the upfront investments (entry barriers) when entering a new market. Although more sophisticated methods are available for comparing investment options, the payback period remains among the most commonly used (Hoang et al., 2024). Additionally, due to its simplicity, the payback period keeps the simulation tractable. The respective values for combinations of minimum and maximum market-size variation and entry barriers indicate payback periods of about 2 to 3 years, consistent with various surveys on these topics (e.g., Fotsch, 1984; Lefley, 1996).

There will also be a degree of autocorrelation (β) from one iteration to the next to represent the path dependency of economic growth over time. β is randomly drawn from a uniform distribution ($0.7 \leq \beta < 1$). This distribution largely reflects the set of autocorrelation values for the economies of the world from 1970-2023 (World Bank current GDP). In the international economy (and consequently in this simulation), the relationship among returns across countries is somewhat more complex. Economic developments in neighboring countries have a strong common factor due to economic integration (Lumsdaine & Prasad 2003), while the economies of more distant countries are less correlated. This means that there is a correlation between country returns that decreases with distance (the number of rows between a pair of countries in the vector Π_t). This is what we can largely observe in reality, where, for instance, economies of European countries are tightly coupled, resulting in a degree of comovement. The simulation factors this in via the covariance matrix Σ in the multivariate distribution above. The market sizes are thus:

$$\Pi_t = \Pi_{t-1} \cdot \beta + \mathcal{N}(\alpha, \Sigma)_t \cdot \mu \cdot (1 - \beta), \text{ with } \Pi_0 = \mathcal{N}(\alpha, \Sigma)_0 \cdot \mu.$$

The return a firm generates in a particular year (indicated by $\pi_{i,t}$) is the sum of the returns in all countries where a firm is present. Returns represent cash flows that already account for the cost of sales, which are homogeneous across firms. Hence, the model assumes that variation in performance and, consequently, internationalization cannot result from variations in technology or product offerings. The positive (or negative) cash flows a firm generates in one year are added to a firm's resource base, which represents cumulative cash flows. This approach keeps the model parsimonious and straightforward, as it facilitates the interpretation of overall performance consequences in terms of changes in the firm's cumulative cash flows. Additionally, it aligns with our approach to conceptualizing and calibrating market size variation and entry barriers later. We represent a firm's total returns by the scalar product of a firm's presence $p_{i,t}$ and the country returns Π_t at time t .

$$\pi_{i,t} = p_{i,t} \cdot \Pi_t$$

A canonical assumption in IB is that a foreign firm suffers from *LoF* (Hymer, 1976; Zaheer, 1995). A firm can partially overcome this LoF by adapting to and learning from the host environment, yet it will not fully overcome its liability (Zaheer & Mosakowski, 1997). The simulation factors this relationship into the return that a firm can appropriate in a country by deducting a loss that results from LoF. Four factors affect the extent of this loss. First, the simulation randomly draws the strength of the LoF for the different scenarios from a uniform distribution ($0.5 \leq \text{LoF strength} < 5$). Second, within each scenario (and thus firm-world), the extent of LoF is a positive function of the distance (be it psychic, institutional, cultural, or geographic) between a firm's home country and the respective foreign country (Eden & Miller, 2004). By definition, LoFs are zero in a firm's home country. They increase with the number of countries between the home country and a specific foreign country. The simulation considers this distance effect for LoFs equivalent to the common factor in economic development above (via the covariance matrix proxying for distance between countries). LoF (Λ_i) is the distance from the home to the foreign country to the power of $\frac{1}{\text{LoF strength}}$. This scales the LoF according to the distance between countries. The parameter for *LoF strength* is chosen based on indications in previous studies and to reflect a wide range of possible values. At low levels of LoF strength (0.5), a firm starting in country #1 would lose 0% in the next neighbouring country and 26% in the 25th most distant country, while at high levels of LoF strength

(5) this firm would lose 46% in the next neighbouring country and 87% in the 25th most distant country.

Appendix VI.D provides more details on how LoF, in combination with distance, affects return loss.

Third, the mechanisms from strategic interaction research suggest that firms follow a leader to a country due to reduced uncertainties (Head et al., 2002; Hansen & Hoenen, 2016). The simulation considers the potential benefit followers gain by discounting the LoF by a certain factor. This *LoF discount* (δ_i) is randomly drawn from a uniform distribution ($0 \leq \delta_i < 0.25$) and reduces the LoF (Λ_i) for each country, if the rival is already present in the country. In the lack of empirical guidance on the range, we calibrated this parameter via sensitivity analyses (an alternative range of 0-0.5 has no effect on the outcome). The vector $s_{i,t}$ represents the countries in which a given firm enters after the competition (1 for being a follower vs. 0 for being a first mover). The return of firm i in year t (i.e., $\pi_{i,t}$) corrected for the LoF (Λ_i) will consequently be ($\circ$ indicates a line-wise multiplication):

$$\pi_{i,t} = p_{i,t} \cdot \Pi_t - [p_{i,t} \circ \Pi_t \cdot (\Lambda_i \circ (1 - \delta_i \circ s_{i,t}))].$$

Fourth, the longer a firm has been present in a specific target country, the closer the LoF will approach zero (Zaheer & Mosakowski, 1997). The LoF will, however, approach zero only asymptotically; that is, there will always remain a small LoF, except for in a firm's home country (Butler, 2016). The simulation considers the number of years that a firm has been present in a foreign country to incorporate this reduction in the LoF. This ratio is expressed by $x_{i,t}$, which indicates the ratio of years that have passed since a firm (i) entered a country (j) over the currently observed year of the simulation (t). We incorporate this as the ratio of the sum of all presence vectors $p_{i,\tau}$ for the past years τ (with $\tau < t$) over t . If, for example, the simulation is in year 10 and a firm has entered a market j_0 in year 7, the firm's experience in this market will be 30% (i.e., $\frac{10-7}{10}$). In year 50 of the simulation, the experience will be 86% (i.e., $\frac{50-7}{50}$). In more general terms, the experience ratio is expressed as:

$$x_{i,t} = \frac{1}{t} \cdot \sum_{\tau=1}^{t-1} p_{i,\tau}.$$

As a result, LoFs will decrease. In the simulation, this means that experience ($x_{i,t}$) reduces the LoF (Λ_i).

Consequently, the return of a firm (i) in year (t) takes the form:

$$\pi_{i,t} = p_{i,t} \cdot \Pi_t - \left[p_{i,t} \circ \Pi_t \cdot \left(\Lambda_i \circ (1 - \delta_i \circ s_{i,t}) \circ (1 - x_{i,t}) \right) \right].$$

Given the simulation's focus on competition, there are two varying competitive factors that impact firm returns whenever the two firms collocate in the same country: *competitive intensity* (c) and persistence of *FMA* (a). Depending on the degree of competitive intensity, possible returns in a country decrease by the factor c for both firms ($0 \leq c < 1$). The range chosen represents the full range of all feasible values (sensitivity analyses show structurally equivalent results for alternative ranges). Competitive intensity maps onto margin erosion due to, for instance, price wars, increased marketing spending, demand-splitting, or the possibility to differentiate, which can vary by industry and product and could be interpreted as a decrease in price-cost margin of the firm.

The firm that has entered a country first, however, will benefit from a FMA (Isobe et al., 2000; Pan et al., 1999): losses to the competition will only gradually approach c , reaching the full effect after a years of direct rivalry (r). This reflects the possibility for firms to acquire key scarce assets, accounts, establish high switching costs, brand image, patents or network effects that will only fade slowly. For instance, if the persistence of the FMA (a) is 5 years and the focal firm is already present in market j_0 once a competitor enters, the focal firm will only lose 20% (i.e., $\frac{1}{5}$) of c in the first year of direct rivalry, 60% (i.e., $\frac{3}{5}$) of c in the third year, and 100% of c starting from the fifth year onwards. By contrast, the follower will see returns reduced by factor c starting from the first year of rivalry in country j_0 . The strength of the FMA in terms of years is randomly drawn from a discrete uniform distribution ($1 \leq a < 16$). Studies have shown that a FMA can be anywhere from relatively marginal for technologically dynamic products (Gomez et al., 2016) to 20 or up to 40 years for highly regulated or entrenched industries (Huff & Robinson, 1994; Robinson et al., 1994; VanderWerf & Mahon, 1997). Given the study's focus on market-seeking industries emphasizing branding and competitive positioning, the simulation parameter was chosen as a realistic middle ground between these two extremes while allowing for substantial variation to occur. The vector representing countries that firm i has entered first is represented by $f_{i,t}$ (again, $s_{i,t}$ represents the countries entered after the competition). Therefore, the return of firm i in year t corrected for the LoF (Λ_i) and competitive intensity (c) will be:

$$\pi_{i,t} = p_{i,t} \cdot \Pi_t - [p_{i,t} \circ \Pi_t \cdot (\Lambda_i \circ (1 - x_{i,t}) \circ (1 - \delta_i \circ s_{i,t}))] - [f_{i,t} \circ \Pi_t \cdot c \circ \frac{r_{i,t}}{a}]$$

$$- [s_{i,t} \circ \Pi_t \cdot c], \text{ with } \frac{r_{i,t}}{a} \leq 1 \text{ and } p_{i,t} = f_{i,t} + s_{i,t}$$

In the simulation, a firm can only expand internationally if its accumulated returns from the past are above a certain threshold. This threshold, which we term *entry barriers* (ϵ), is another varying contextual factor. If a firm has enough resources to expand internationally, it does so, and the model deducts the expansion costs from the firm's resource base. These entry barriers map onto the capital expenditure necessary to start operations in a market efficiently, varying from investments in an office in reality to tariffs, obtaining permits and investment in machinery. The capital required to enter a new country is randomly drawn from a discrete uniform distribution ($1,000 \leq \epsilon < 7,000$). The range of entry barriers or capital required is chosen based on potential payback periods linked to the market size parameter (see elaboration for market size variation).

The model assumes that a firm cannot leave a country once it has entered it. In this sense, this study aligns with previous work that assumed (for reasons of keeping model complexity manageable) that firms cannot exit a market once they have entered it (e.g., Alcácer et al., 2015). Naturally, this is a simplification of reality, as market exit is indeed a regular phenomenon for firms. Previous literature cites poor performance abroad, reduced managerial and coordination needs, and refocusing as among the core reasons for divesting subsidiaries (Kafourous et al., 2022). In our simulation, the expected average return in all countries in the simulation will always be positive, and thus there is no reason for a firm to exit a country ($\mu > 0$). In addition, the simulation model assumes rational decision-making and, thus, cognitive constraints or the need to refocus would not be reflected in the model. Still, we accept that ruling out exits is a simplification of reality.

After each iteration (t), the simulation captures a firm's positive (or negative) returns cumulatively in annual values, which represent the resource base of a firm in terms of cumulative cash flows.

Additionally, individual annual returns per firm are captured as vectors for analysis. The simulation also stores a firm's internationalization steps and possible country returns over time. The value of each environmental parameter is held constant over time (unless otherwise specified) but varies across 10,000 simulation runs.

A crucial question concerns the representativeness and calibration of the selected simulation parameters. We approach this issue from two sides. First, we select parameter ranges that allow for variation within a realistic range while also reflecting empirical realities and previous studies. Table 1 summarizes our approach in choosing the respective ranges (*Calibration approach*). Second, we conduct extensive sensitivity checks with alternative simulation parameters. This means we re-run the same 10,000 simulation runs as in the original analysis for each parameter and substitute the respective parameter with an alternative range.

The sensitivity analyses provide equivalent results for the mean of the normal distribution, the autocorrelation of market sizes, LoF strength, LoF discount, and competitive intensity. We test a narrower as well as a broader (the latter not tabulated in the appendix) range for the mean of the normal distribution, yielding identical results. We repeat the same procedure for the autocorrelation of market sizes, using narrower and broader ranges (the latter not tabulated in the appendix), and again obtain equivalent results in both cases. Testing a broader range of LoF strength and LoF discount does not alter the results either. In the main model, we take the broadest possible range for competitive intensity. However, narrowing the range by half on each end does not affect our simulation outcomes.

There are three factors that need some further attention. Extending the possible years of FMA to 40 years increases the level of avoidance as it becomes less attractive to follow a firm to a market. This means that, when paired with high competitive intensity, high FMA can also trigger complete avoidance. An FMA of 40 years, however, seems a rather unrealistic scenario for market-seeking industries. Greater variation in market size increases potential returns from a market and thus makes it more attractive to follow a rival into a market, thereby increasing collocation. Similarly, increasing entry barriers in the simulation makes market entry costly, reducing the overall level of internationalization and, consequently, mechanically increasing the overall level of avoidance. However, the subsequent analyses are equivalent for both variables. Because market size variation and entry barriers are linked, we also run a sensitivity analysis in which we vary both factors by the same amount.

Table I.1 Overview: Simulation variables

Simulation variables	Calibration approach
Market potential (μ)	
Market size variation (msv)	Calibrated as a payback period (Hoang et al., 2024) relative to entry barriers: the lower and upper boundaries of msv and entry barriers correspond to ~2 years (lowest possible msv and entry barriers) and ~3 years (highest possible values). This would be in line with previous studies (reporting common payback periods of 18 months to 4 years; e.g., Fotsch, 1984; Lefley, 1996). Naturally, this ratio can vary depending on the exact values.
Mean of normal distribution (α)	Calibrated via sensitivity analyses (no effect on outcome)
Auto-correlation of market sizes (β)	World Bank GDP (current) from 1970-2023 for all countries available: Autocorrelation (lag-1) (see Figure VI.C.0. in Appendix VI)
LoF: Liabilities of foreignness (Λ)	Figure VI.D.0 in Appendix VI illustrates how the strength of LoF work in conjunction with distance. Various studies provide different indications on the potential effect sizes of LoF (see Appendix VI.D)
LoF Discount (δ)	Calibrated via sensitivity analyses (no effect on outcome)
Competitive intensity (c)	Simulation considers range of all possible values
Persistence of FMA (a)	Previous studies in consumer markets report FMAs anywhere from 3 and up to 40 years, while suggesting they deteriorate sharply at least after 20 years (e.g., Huff & Robinson, 1994; Robinson et al., 1994; VanderWerf & Mahon, 1997)
Entry Barriers (ϵ)	Calibrated relative to potential market sizes (see point above for market size variation)

Deep reinforcement learning application

The two agent firms in each simulated firm-world internationalize and enter foreign markets. Opposed to modeling a specific rule according to which the two firms select markets, we use neural networks with DRL to optimize a firm's internationalization strategy for each simulated firm-world environment.

Although not yet part of the mainstream methodological repertoire of IB researchers, some studies have already utilized neural networks to examine internationalization decisions (Brouthers et al., 2009; Veiga et al., 2000). In this study, we leverage recent advances in DRL (e.g., Silver et al., 2016) to train the internationalization steps of two firms.

The two firms are trained in a process known as multi-agent reinforcement learning (Gronauer & Diepold, 2022). This means the two firms (i.e., the agents) interact in the same environment and learn in a 'self-play' set-up. Each simulation run (or seed) is trained in 61,000 episodes. Each episode involves 50 time steps, and in each time step, a firm can take one of the (technically) maximum of 50 possible actions. Technically, the action space consists of 51 actions as a function of the 50 countries the firm could enter plus the option of inaction. Naturally, a firm starts in one country (thus max 50 possible actions) and the

action space also becomes smaller over time as firms cannot enter a country where it is already present. These actions represent the policies the neural network trains to optimize for each time step. The firm observes four factors at each step, which can inform its strategic decision. For each of the 50 countries, the firm sees the market size, the distance to the country and two indicators for whether the firm or the competitor already have a presence in the country. The characteristics of these four factors are a function of the environmental characteristics and the endogenous location strategies of the two agents. The relative importance of each factor in driving the decision to enter a market and in determining a firm's position at a specific point might shift over time. In each episode, one firm's policies are fixed, and the other firm is trained. After every episode, the distribution of returns between the two firms is converted into advantage estimates to update the neural network, reinforcing policies that were successful in achieving higher relative profits and exploring alternatives for unsuccessful policies. After 1,000 episodes, the trained and fixed firm switches, and the rival firm is trained while keeping the policies of the other firm fixed from its most recent training episode. Each firm is trained in 30 rounds of 1,000 episodes per firm with an additional 1,000 episodes for the focal firm at the end (totaling 61,000 episodes). Through this iterative process, firms learn to make better decisions relative to competition over time, allowing their expansion strategies to gradually converge to an equilibrium in which neither firm's change in location strategy yields a higher payoff.

The DRL process works by optimizing a firm's cumulative returns relative to the rival; these relative rewards are used as payoff signals when updating the parameters and weights of the neural network. In doing so, the firm tries to improve its cumulative returns vis-à-vis the rival over the episode by following previous policies that have proven to be successful and exploring alternative actions. Alternative actions are encouraged via an entropy bonus, which decreases over time due to decay, shifting the balance from exploring alternatives to exploiting previous successful actions. These adjustments during the training episodes shift the importance of the observation factors (market size, distance, own and competitor presence) determining market entries over time. Every 100 episodes, the trained policies are evaluated and compared to a benchmark (fixed future-market-size) to evaluate whether the learned policies are 'good' strategies, outperforming fixed benchmarks, and to track convergence of the performance-optimizing strategies. The fixed future-market-size benchmark is the selection rule that identifies the most

attractive country each year based on the average past returns of countries (accounting for rival presence). This comparison and the convergence serve as inputs to fine-tune and calibrate the parameters of the DRL process and the neural network. The Appendix provides more information on the DRL- and neural network-specific parameters used. The DRL-specific parameter values were set using common heuristics and subsequently calibrated based on the learning curves.

This whole procedure is repeated in 10,000 simulation scenarios with randomly varied contextual factors. Appendix V provides further intuition for this notion and presents an example from our simulation data on internationalization steps. We run the DRL simulation using a Python 3.10 distribution on a SLURM-managed computer cluster. We then analyze the final performance-optimizing location strategies for each of the 10,000 simulation scenarios and generate subsequent plots in R.

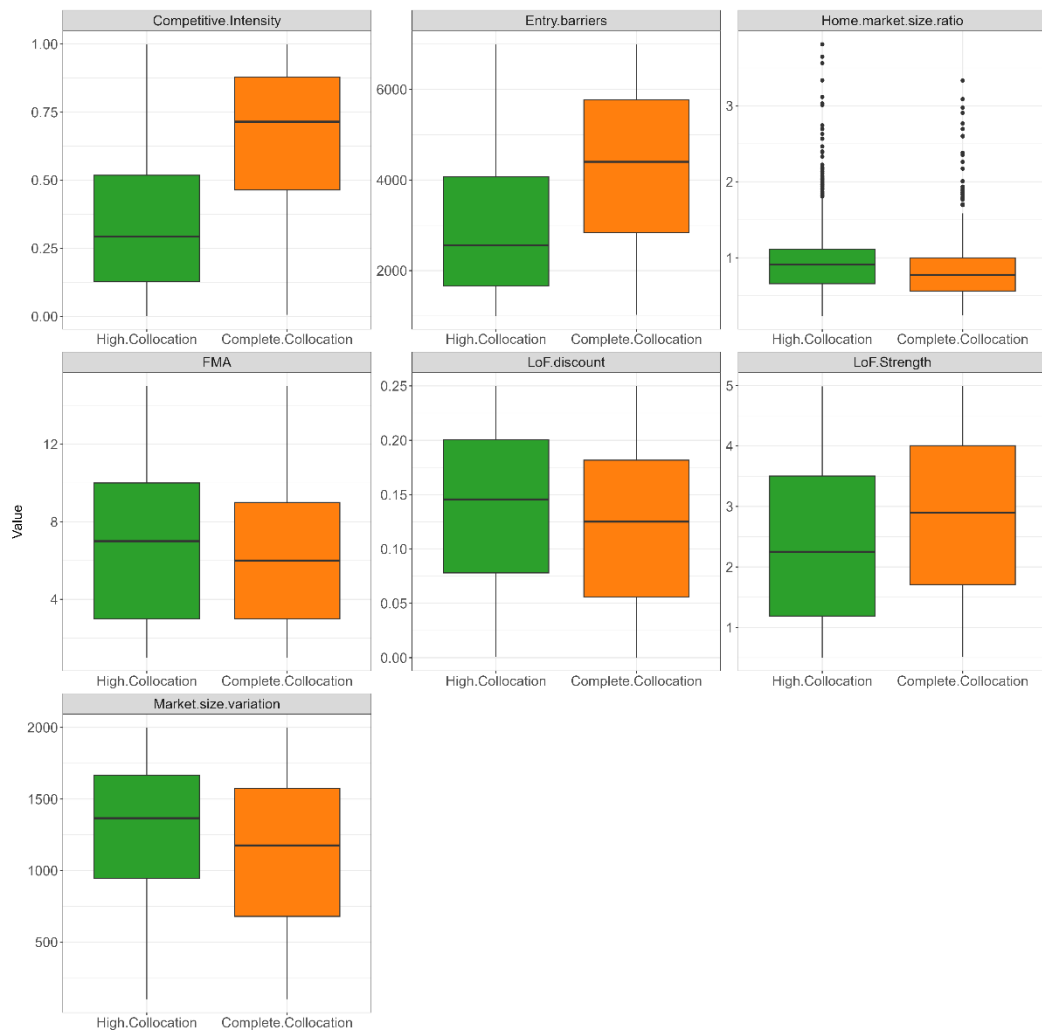
Appendix II: Data Generation Example

Table II.1 Data Generation Example

Company ID / Seed (i)	Home country ID (j)	Market potential – Constant (μ)	Market size variation (msv)	Mean of normal distribution (α)	Auto-correlation of market sizes (β)	LoF strength for liabilities of foreignness (Λ)	LoF discount (δ)	Competitive intensity (c)	Persistence of FMA (a)	Entry Barriers (ϵ)
491	26	500	1918.96	0.82	0.73	4.44	0.10	0.36	13	5952
3721	38	500	819.84	1.09	0.87	2.88	0.17	0.90	11	5350
5580	25	500	1494.13	0.87	1.00	1.33	0.03	0.72	8	5971
1608	10	500	1384.84	1.07	0.89	3.61	0.03	0.29	9	3049
1462	4	500	504.42	1.12	0.71	4.16	0.03	0.97	2	4883
7591	16	500	820.76	1.11	0.79	4.93	0.01	0.50	5	5868
9939	22	500	307.09	1.02	0.78	2.94	0.25	0.93	8	1404
4440	28	500	1174.23	0.83	0.87	3.52	0.08	0.85	10	6225
7439	30	500	266.53	1.04	0.99	4.48	0.16	0.88	6	6517
6287	43	500	1407.23	1.02	0.93	1.65	0.19	0.67	10	1627
7082	46	500	1623.20	1.02	0.86	1.37	0.07	0.19	11	1163
3402	48	500	165.89	0.95	0.92	4.76	0.07	0.08	10	5504
930	26	500	232.38	1.13	0.84	1.21	0.02	0.69	15	6930
9752	34	500	254.07	1.05	0.90	3.93	0.07	0.93	1	1199
5645	42	500	387.11	1.03	0.77	3.18	0.11	0.99	15	2337
7519	39	500	1435.00	1.16	0.73	2.67	0.11	0.66	4	2706
7560	14	500	829.18	0.88	0.91	4.82	0.00	0.05	4	5704
1646	40	500	399.82	1.03	0.90	4.56	0.17	0.12	4	1481
6919	50	500	841.22	1.05	0.88	2.90	0.14	0.86	8	1492
5787	23	500	1794.18	0.88	0.77	0.81	0.16	0.58	7	3190

Appendix III: High vs. Complete Collocation

Figure III.1 Sample Comparison: High collocation vs. complete collocation



Appendix IV: Performance Comparisons

The following Figures display the performance of the simulation scenarios for different environmental variables for the upper (>0.75) and lower quantiles (<0.25). The top panel in each Figure contains three local parametric fit lines; the green-dashed line (full sample) is the same as the black-solid line in the bottom panel of Figure 1; the solid (dotted) line shows the performance for the lower (higher) quantile of competitive intensity. The two bottom panels display the mean performance (fluctuating line) for the focal firm (black-solid) and the competitor (green-dashed), with a local parametric fit overlaid for the lower (left panel) and upper quantiles (right panel) of competitive intensity. In all cases, performance refers to the cumulative cash flows at the end of the simulation run (year 50).

Figure IV.1 Competitive Intensity

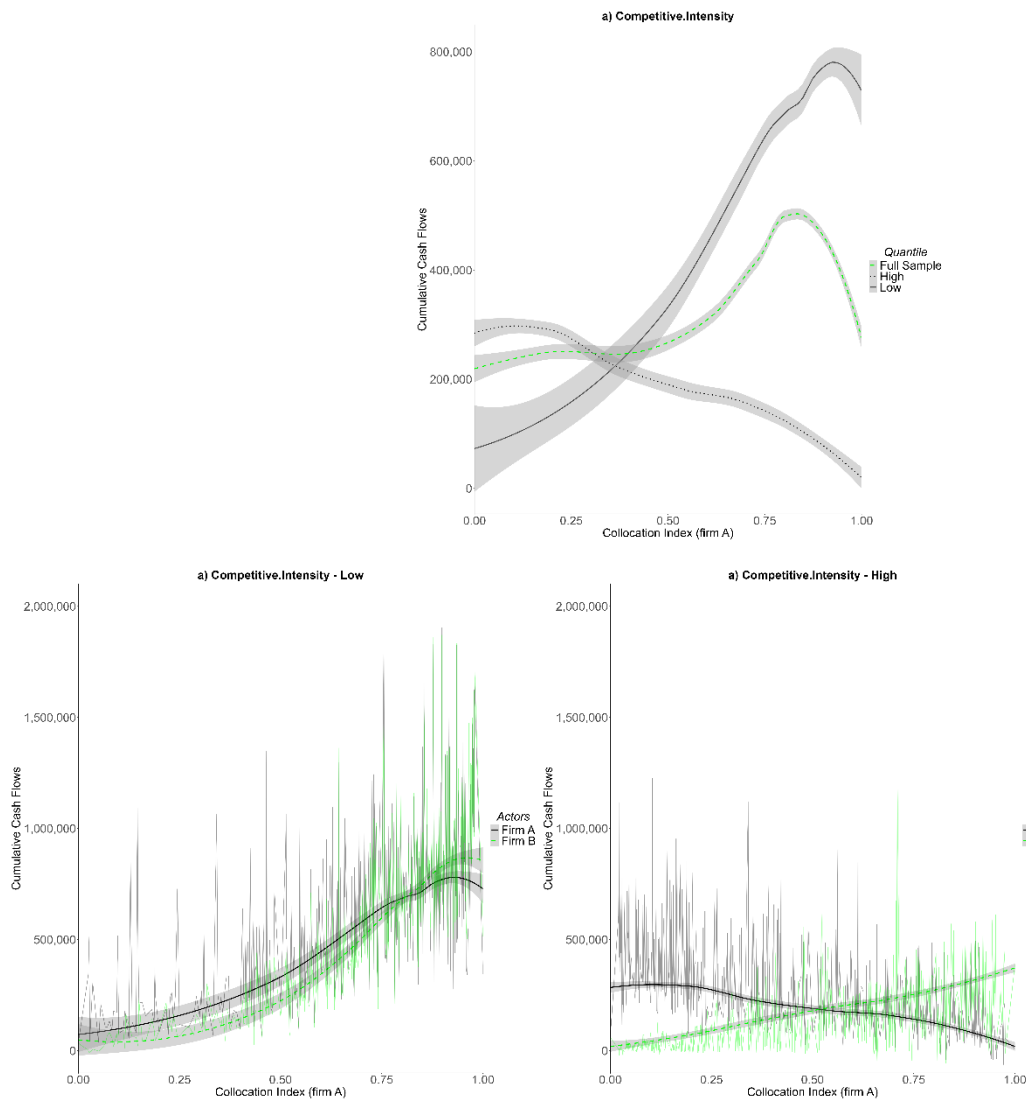


Figure IV.2 Entry Barriers

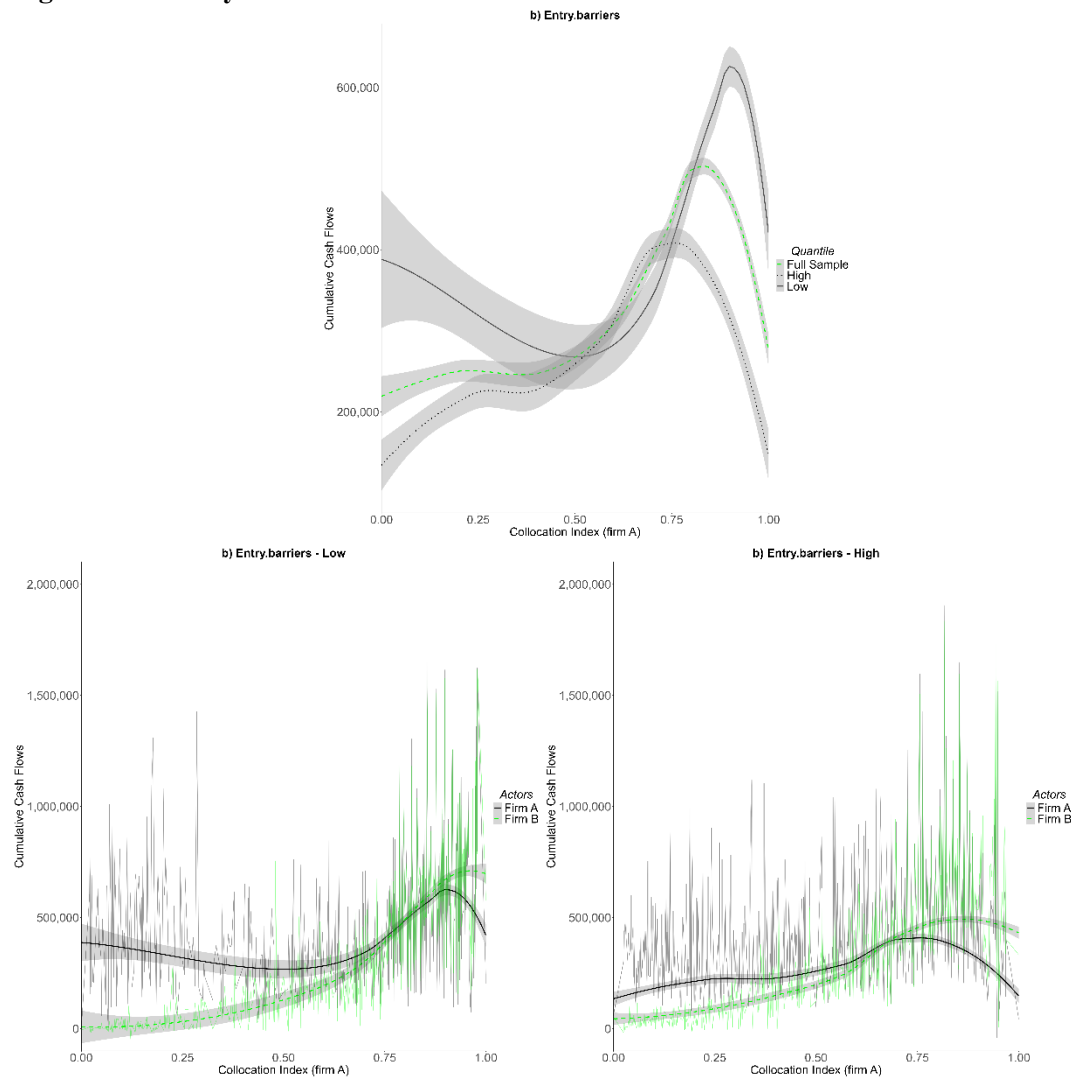


Figure IV.3 Market size variation

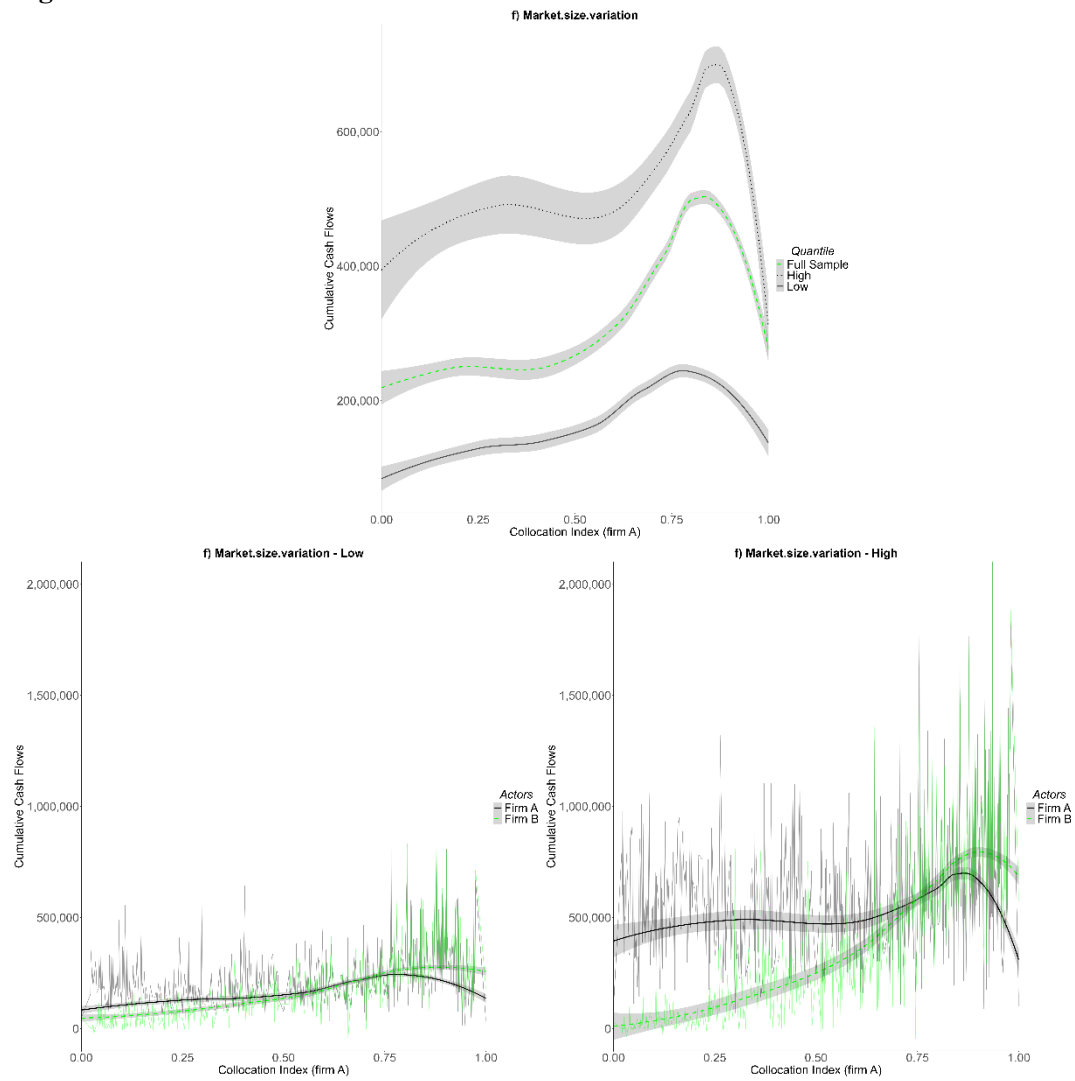
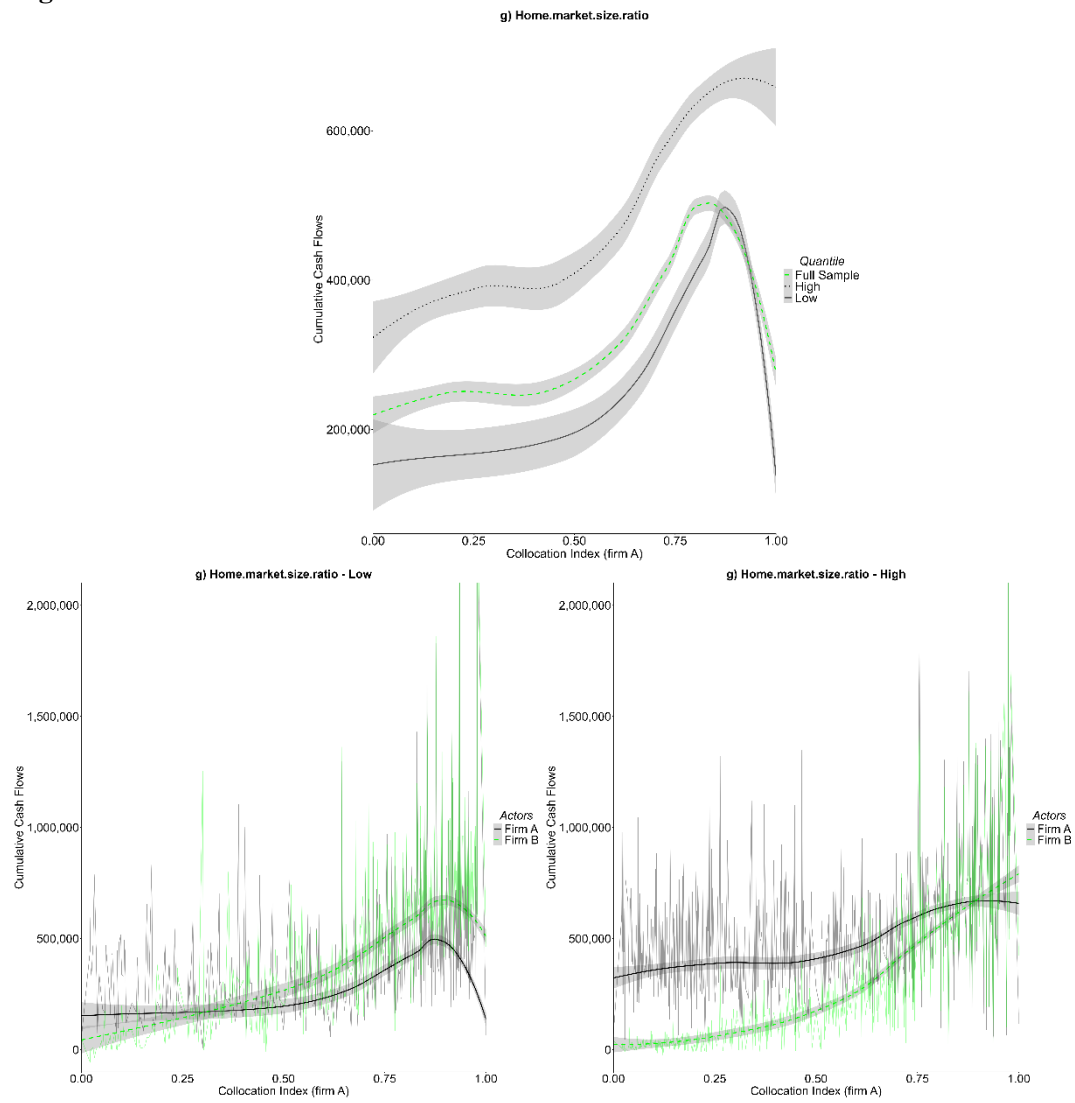


Figure IV.4 Home market size ratio

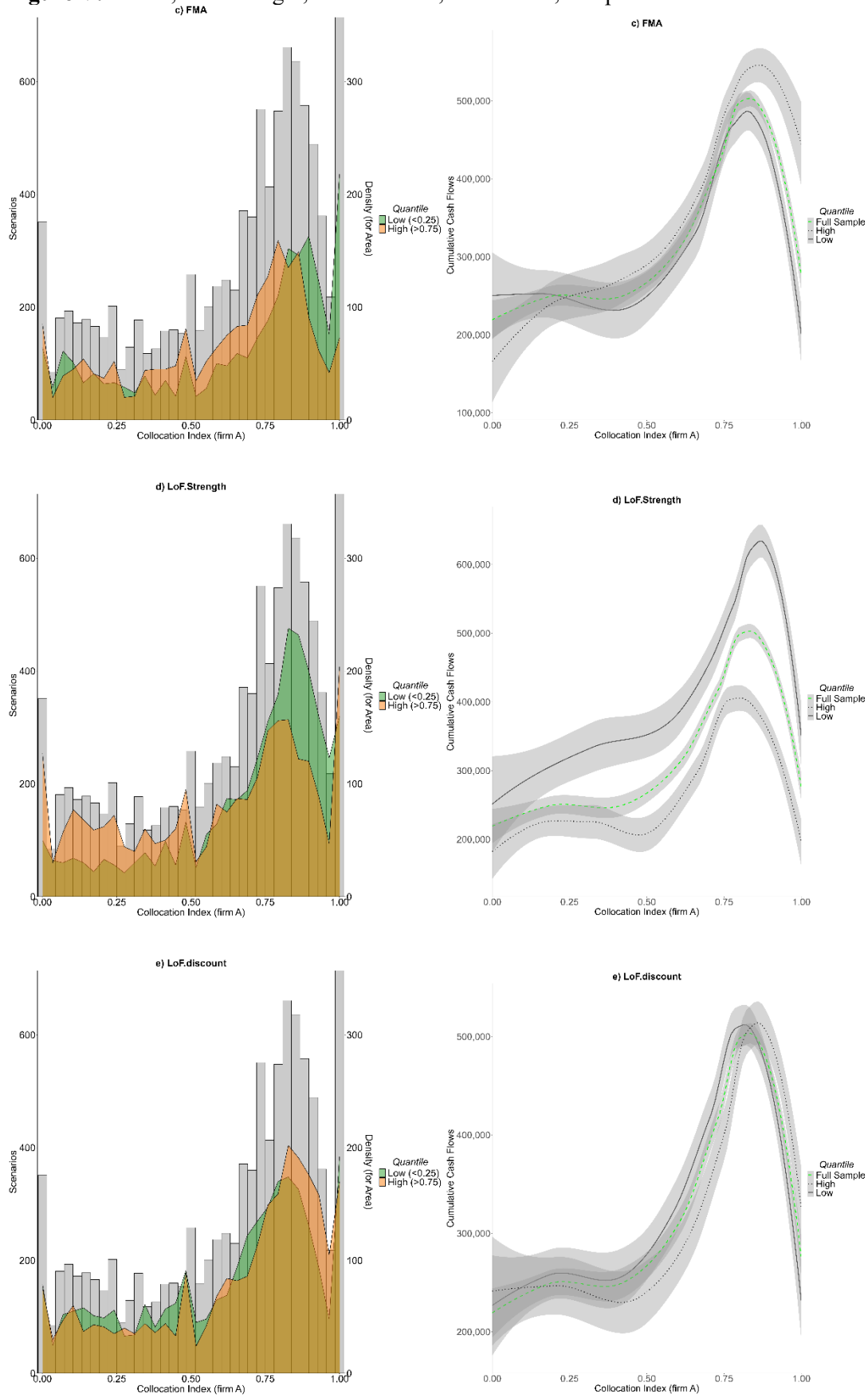


Appendix V: FMA, LoF, LoF Discount

Figure V.1 below shows the distribution of collocation scenarios and the associated performance outcomes for low and high values of *FMA*, *LoF strength*, and *LoF discount*. Despite FMA, LoF strength, or LoF discount being theoretically salient factors, differences in these factors do not materialize in structurally different degrees of optimal collocation. The persistence of FMA should, in theory, render a following-the-leader strategy more or less attractive. Although the left panel for FMA in Figure V.1 shows a spike at the end for complete collocation, subsamples with high and low values are structurally equivalent when compared to the full sample. Further sensitivity tests indicate that the FMA must be substantial (up to an unrealistic 40 years) to have a discernible effect on the degree of collocation by itself. In these extreme cases, FMA increases overall levels of avoidance. This suggests that the FMA (or entrenchment benefit) identified in prior models (Alcácer et al., 2015) must be sufficiently high to elicit different outcomes in firms' competitive location strategies.

LoF strength slightly evens out the distribution of simulation scenarios across degrees of collocation, as the density of scenarios at mid-to-high levels of collocation is less pronounced. This finding aligns with previous models, which conclude that a better fit to certain markets (due to lower or higher LoF) increases levels of avoidance (Alcácer et al., 2013). This finding would also align with studies that have found greater distance (and, consequently, greater LoF) decreases competitive actions (Yao & Li, 2016; Yu et al., 2009). However, our simulation outputs suggest that this effect is rather marginal compared to other factors, and strategic interaction with the rival predominates. Interestingly, LoF discount has absolutely no effect on the structural distribution of simulation outcomes across degrees of collocation. A discount on LoF represents a benefit for the follower firm, namely, reduced uncertainty about environmental conditions and, thus, fewer disadvantages associated with LoF in target markets. This mechanism, which reduces uncertainty about conditions abroad, is a key factor that should explain follow-the-leader behavior (Head et al., 2002; Yao et al., 2024). Still, this mechanism does not change the collocation behavior of our simulated firms nor the performance they achieve.

Figure V.1 FMA, LoF strength, LoF discount, collocation, and performance differences



Appendix VI: Sensitivity Checks for parameters

The sections below provide sensitivity checks for alternative parameter ranges of our simulation. The outputs for most alternative parameter ranges remain structurally equivalent. Each section replicates Figures 1,2,4,6,8,10,12 from the main paper and the respective Figures from Appendix IV and V with alternative parameter ranges. In each section, we briefly elaborate on choosing the main range, the alternative range, and (if necessary) the implications of the sensitivity analysis.

A. Market size variation (msv)

Market size variation (and the underlying market potential) is calibrated in terms of a payback period relative to entry barriers. Previous studies report common payback periods of 18 months to 4 years (e.g., Fotsch, 1984; Lefley, 1996). The values for msv and entry barriers are calibrated in such a way that the combination of both lower or higher values for both parameters aligns with this range. To put it differently, the lower possible values for market potential (500) and entry barriers (1,000) would amount to a payback period of ~2 years, and the highest possible values (market potential: 2,500 [msv: 2,000]; entry barriers: 7,000) would represent a payback period of ~3 years. These values can be higher or lower if the low/high msv and entry barriers are mixed differently.

For the sensitivity analysis of msv, we have increased the upper boundary to 10,000. This obviously changes the outputs, and it increases overall levels of collocation. If the variation in market size between the largest and smallest markets increases, firms focus on entering the largest markets, even if this means colluding with a rival, as the return is still higher than entering a small but unoccupied market. This is especially the case if the entry barriers remain unchanged (on a small level) and firms are unconstrained in their internationalization. However, on closer inspection, one can see the outputs are structurally equivalent to the main outputs in the main paper. For instance, the boxplots still show the same tendencies, and low competitive intensity still triggers high (but not full levels of collocation). In addition, we run an additional test (not tabulated here), where we increase entry barriers by roughly the same ratio, yielding equivalent outputs as in the main simulation models.

Overall, in terms of patterns, higher market size variation slightly increases collocation levels. However, the remaining distributional patterns remain equivalent and thus also the derived propositions

Figure VI.A.1 Collocation Index

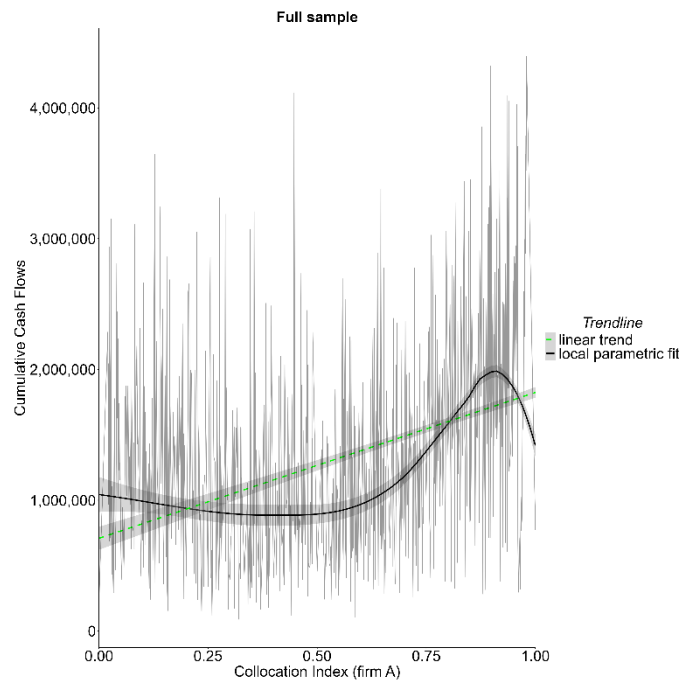
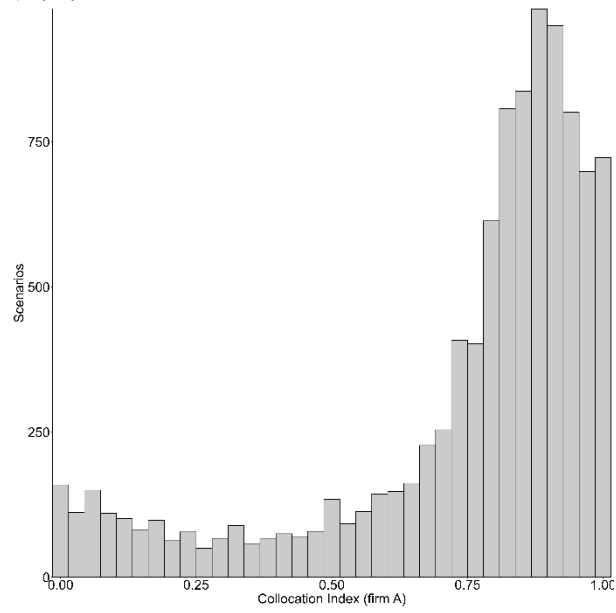


Figure VI.A.2 Sample comparison: All Scenarios vs. Complete Collocation

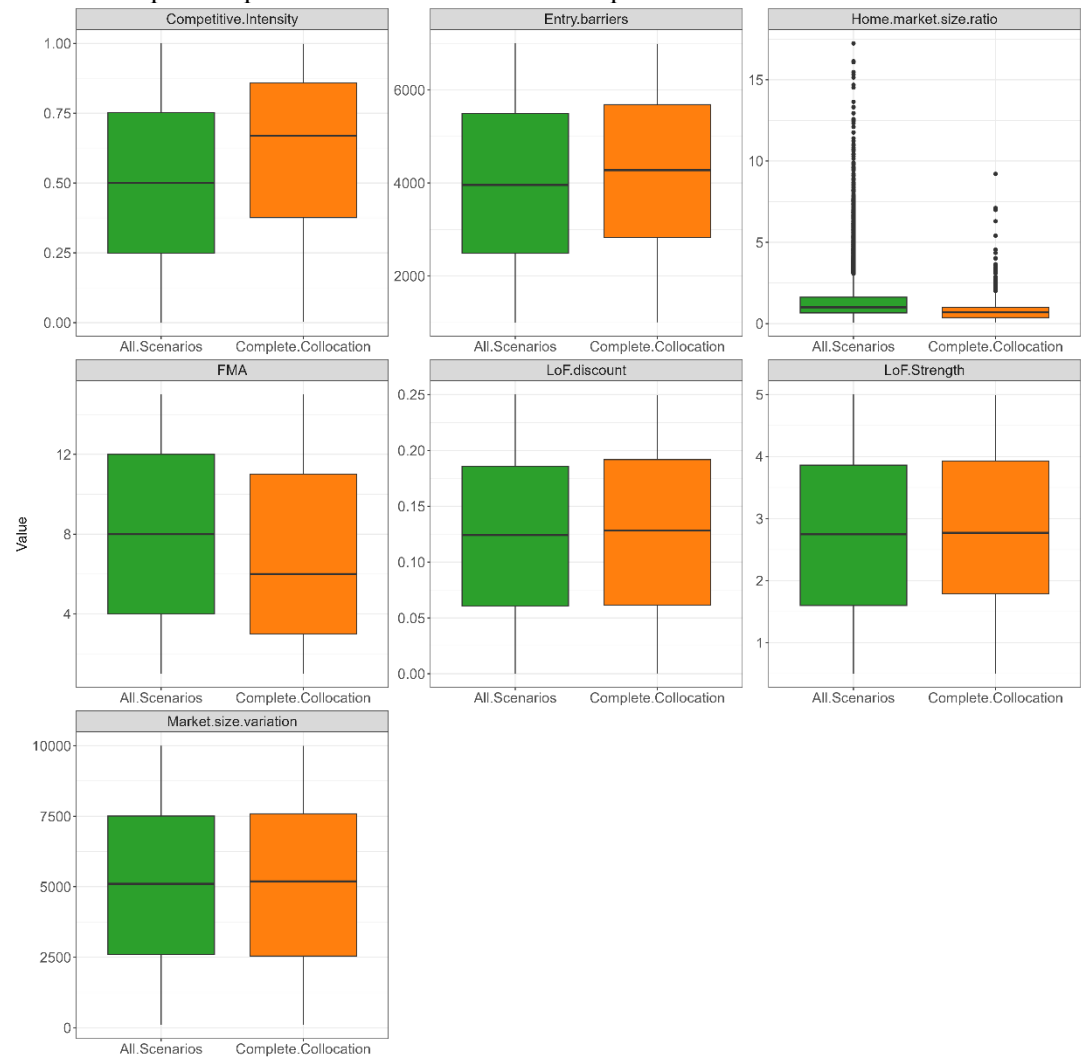


Figure VI.A.3 Sample Comparison: All Scenarios vs. Complete Avoidance

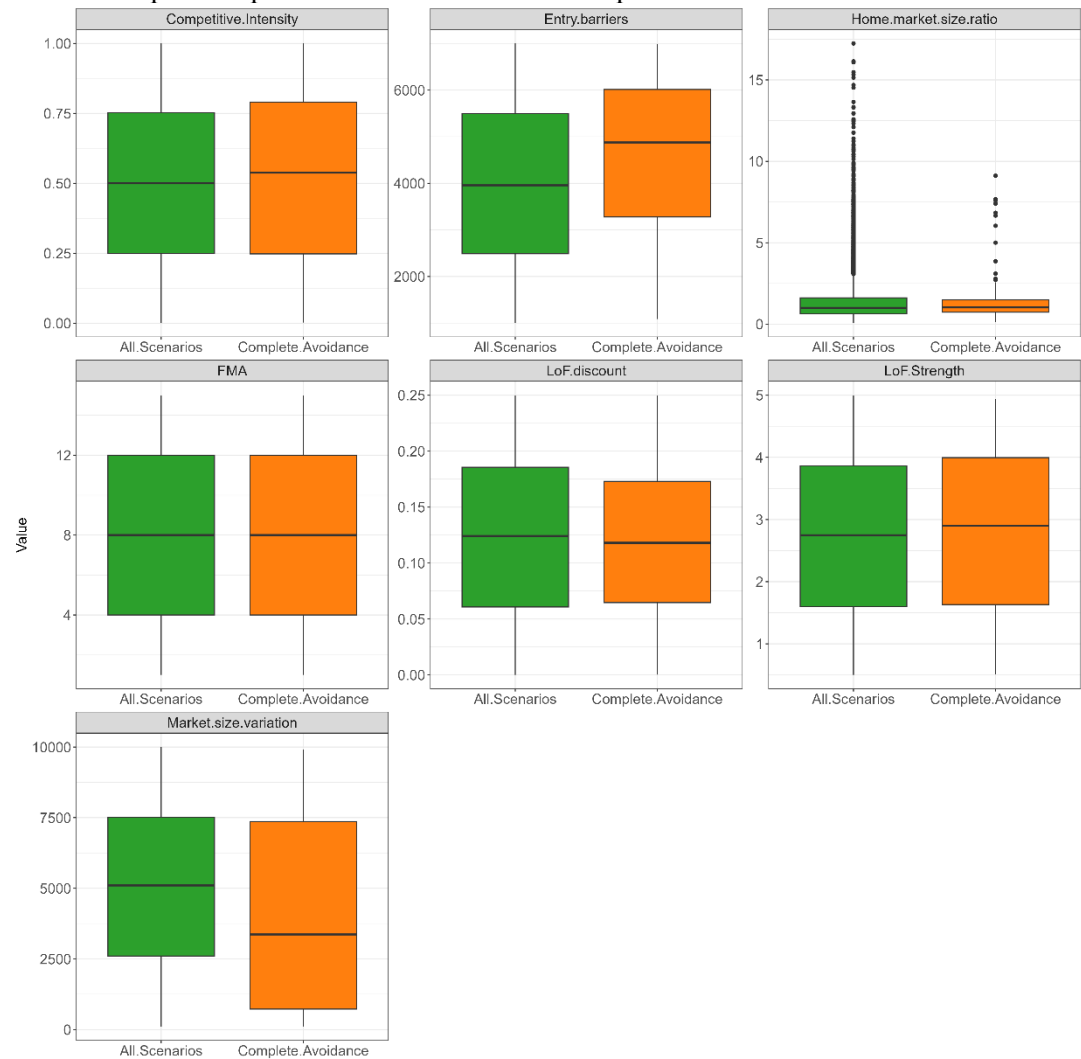


Figure VI.A.4 Competitive intensity, collocation, and performance differences

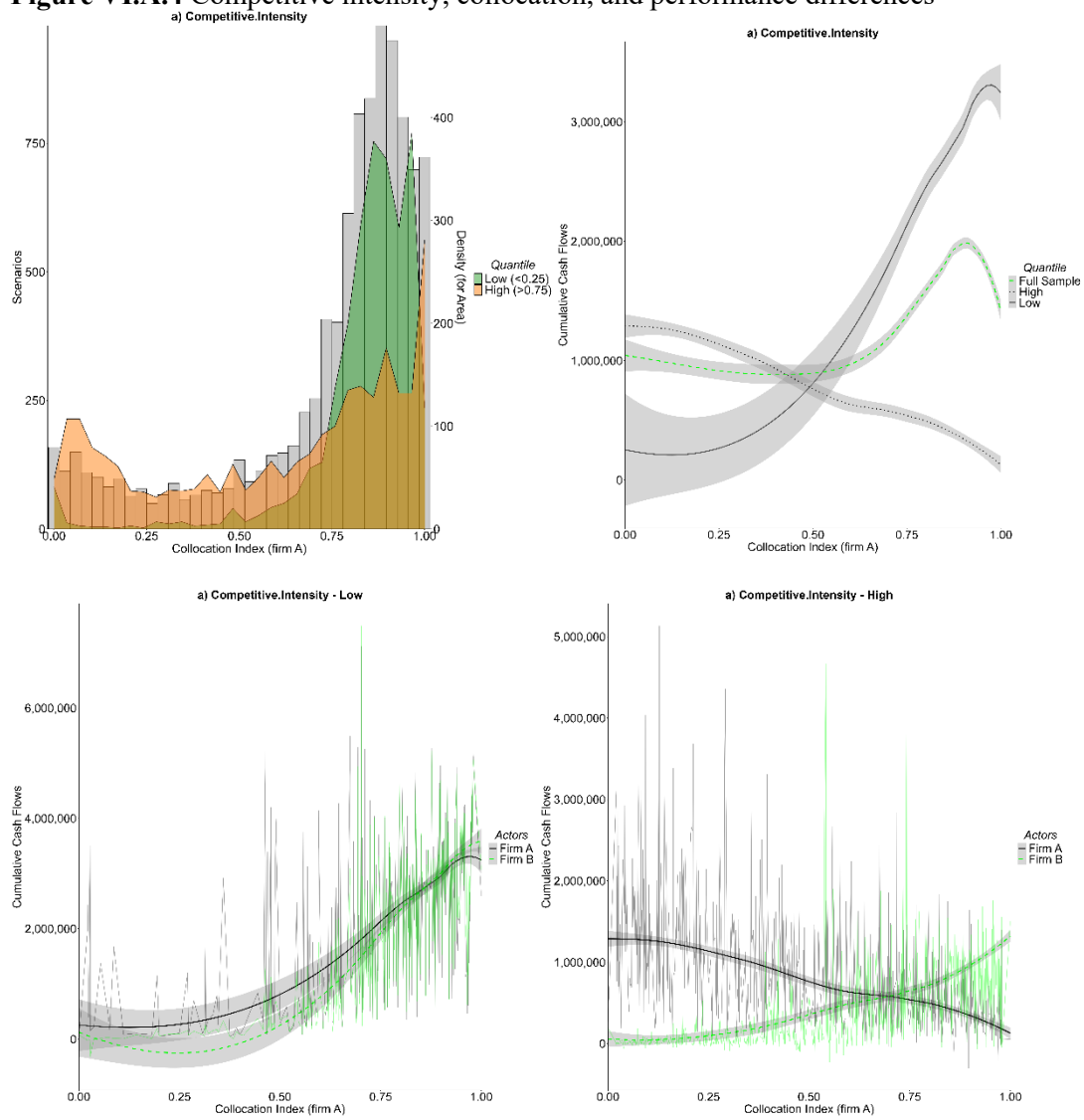


Figure VI.A.5 Entry barriers, collocation, and performance differences

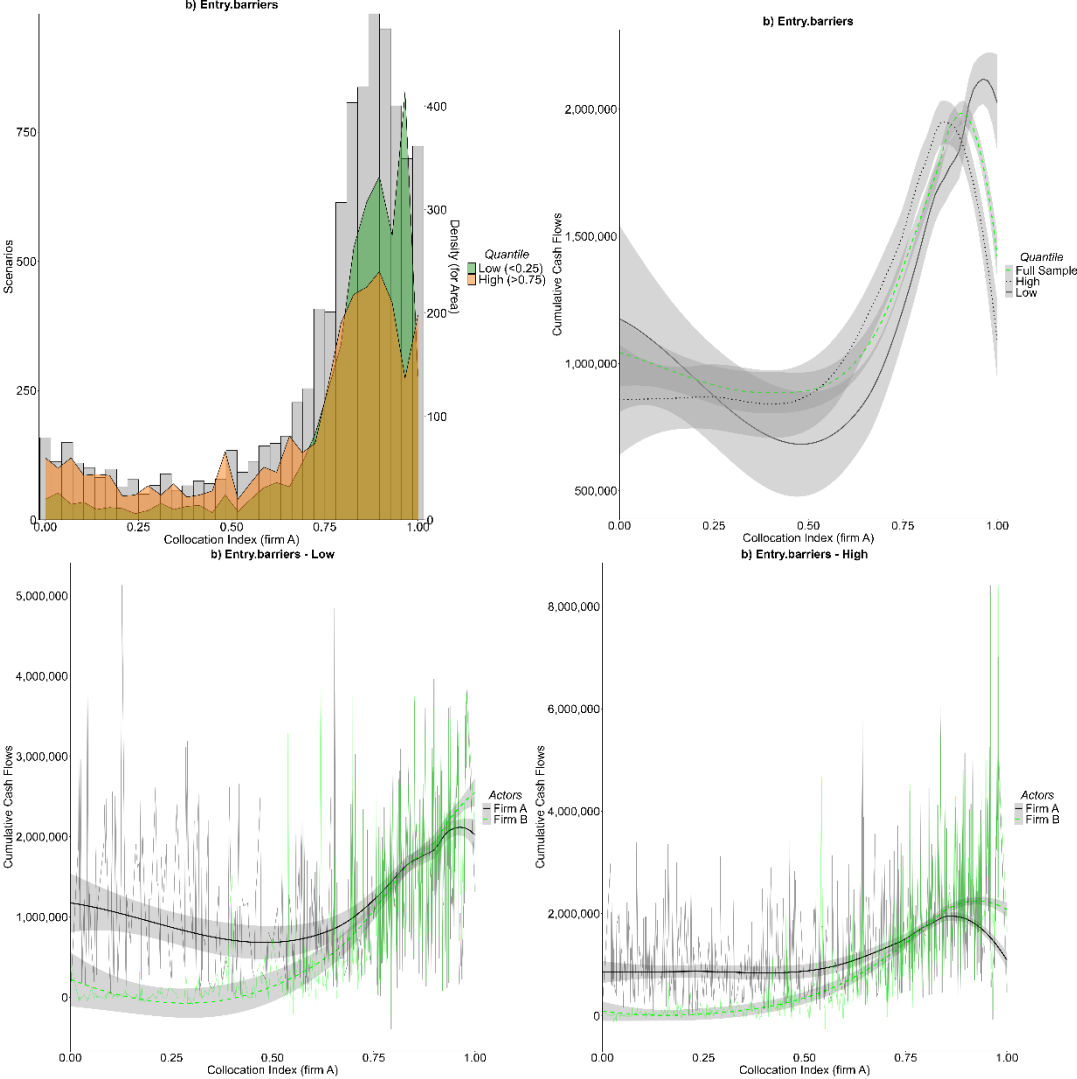


Figure VI.A.6 FMA, LoF strength, LoF discount, collocation, and performance differences

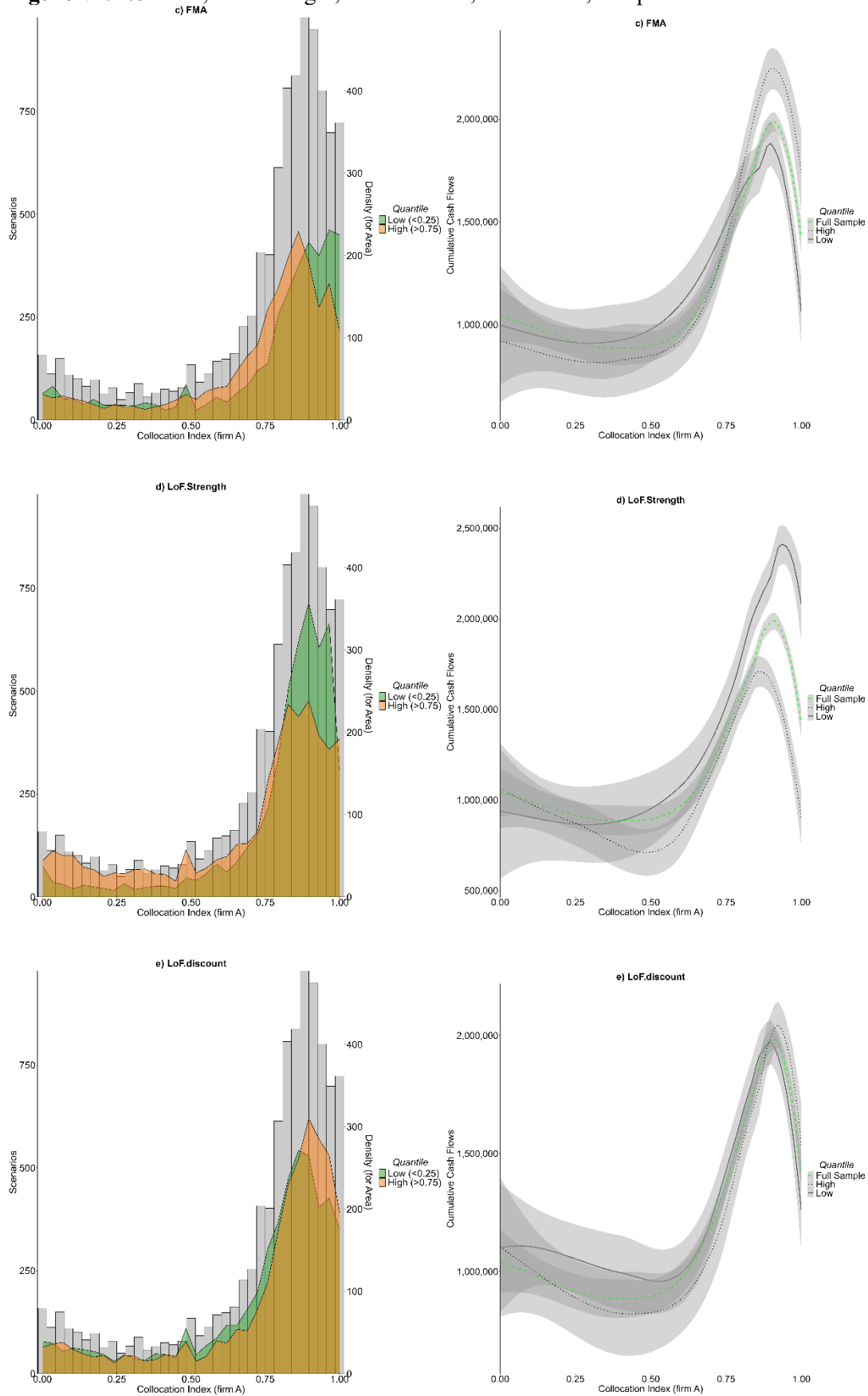


Figure VI.A.7 Market size variation, collocation, and performance differences

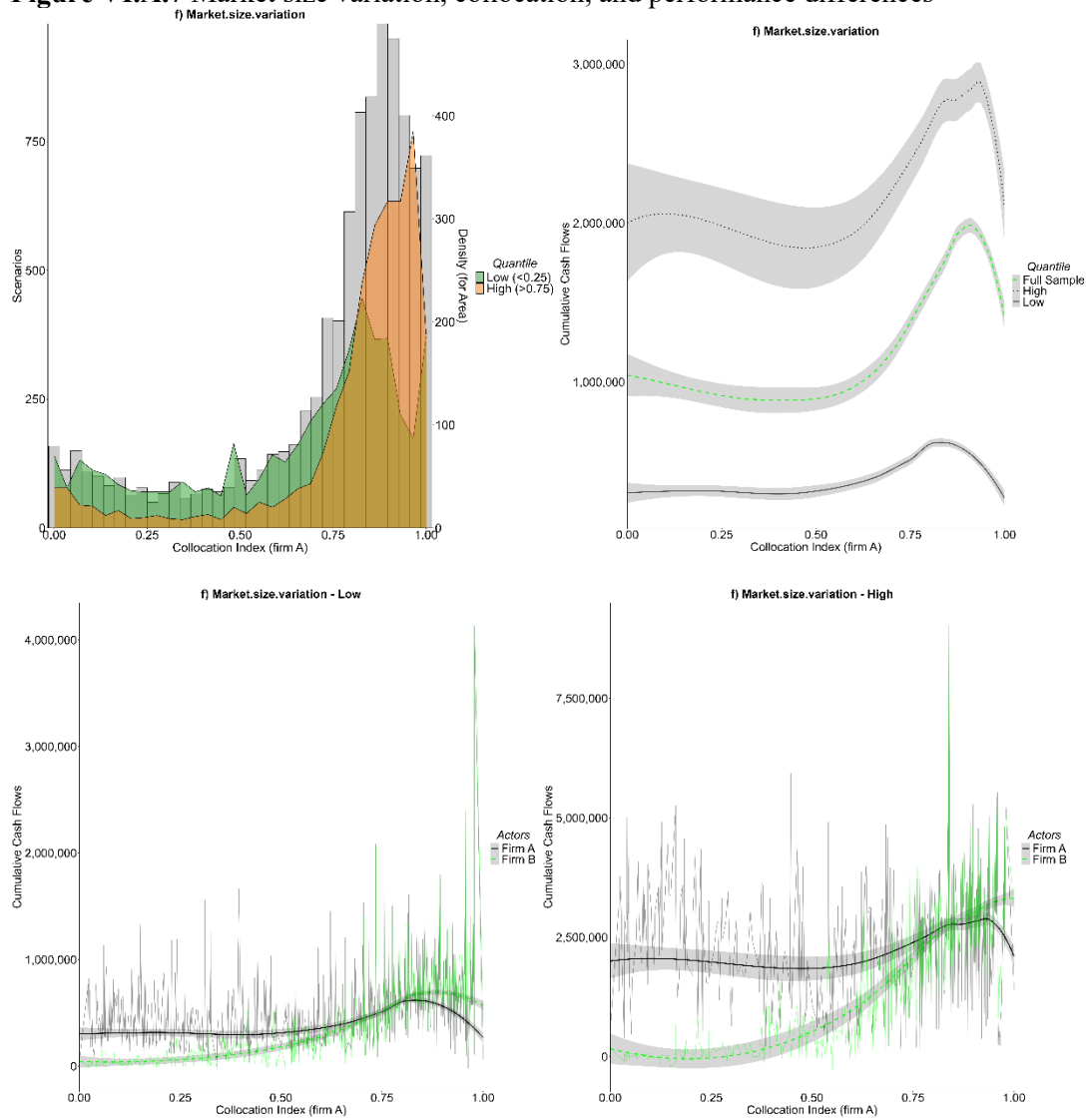
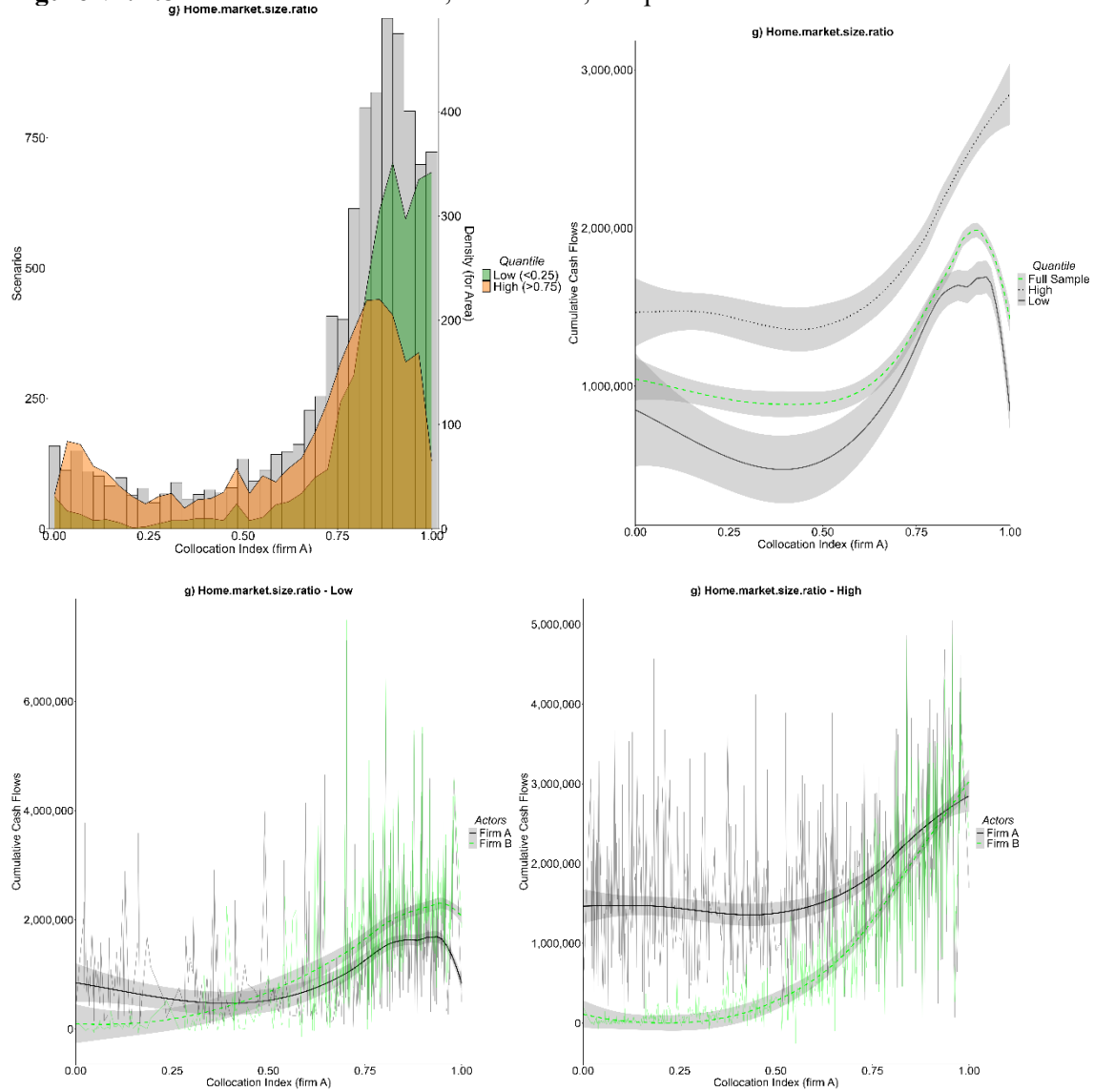


Figure VI.A.8 Home market size ratio, collocation, and performance differences



B. Mean of normal distribution (α)

In the absence of empirical, real-world guidance, we selected an appropriate parameter and verified its validity through sensitivity analyses. The alternative ranges of 0.9-<1.1 (tabulated below) and 0.7-<1.3 (not tabulated) provide entirely equivalent outputs to the main simulation models. Overall, the patterns show no deviation from the main results and thus also the propositions remain as stated.

Figure VI.B.1 Collocation Index

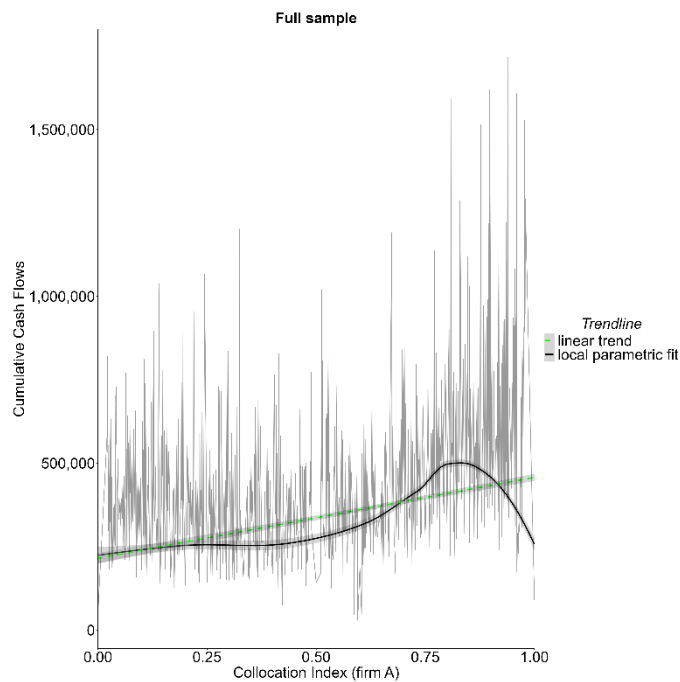
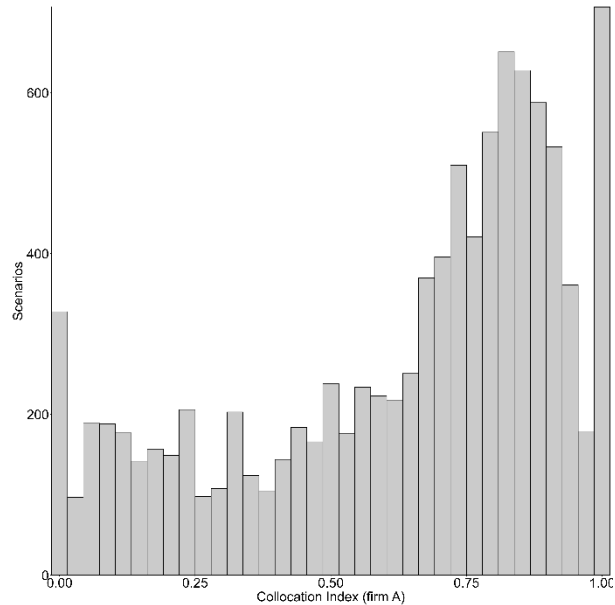


Figure VI.B.2 Sample comparison: All Scenarios vs. Complete Collocation

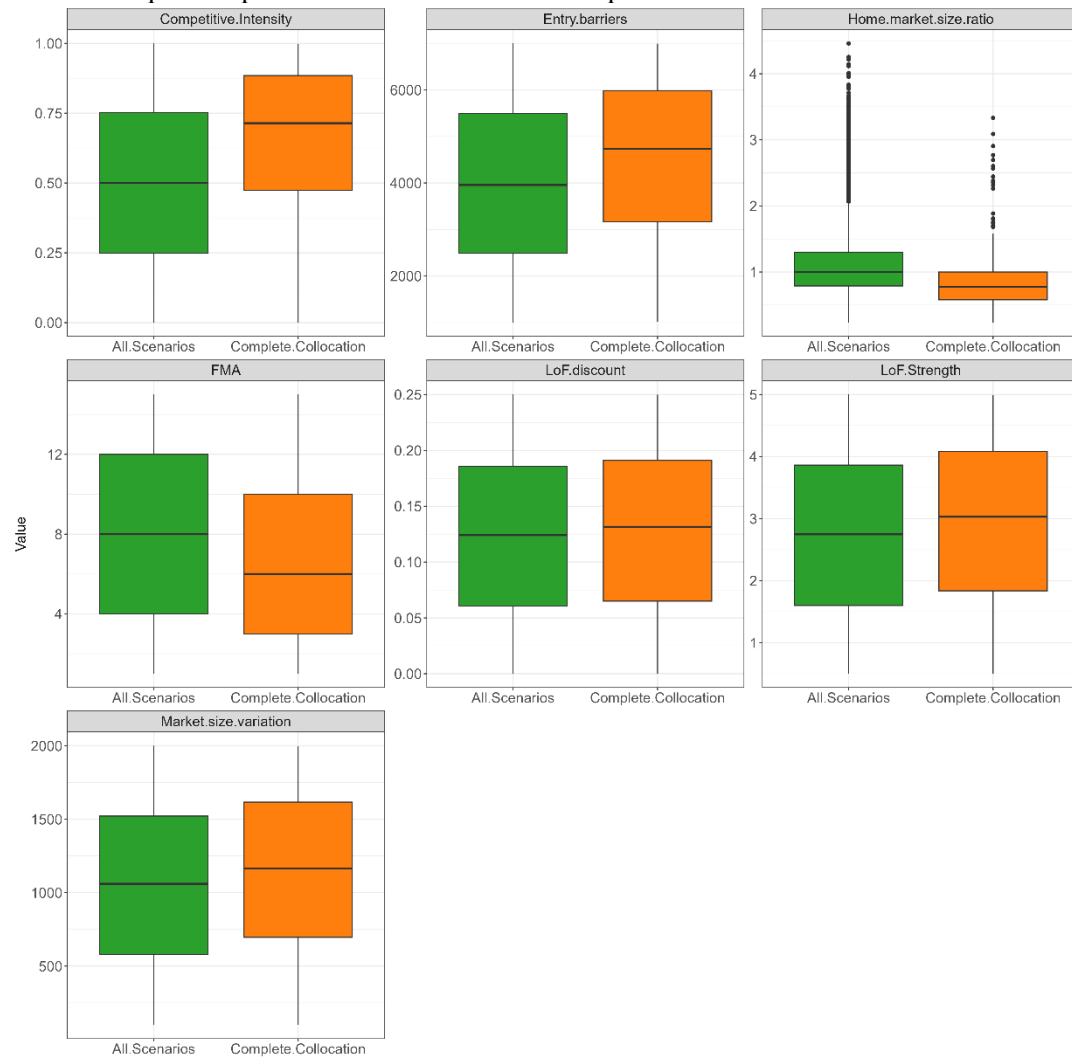


Figure VI.B.3 Sample Comparison: All Scenarios vs. Complete Avoidance

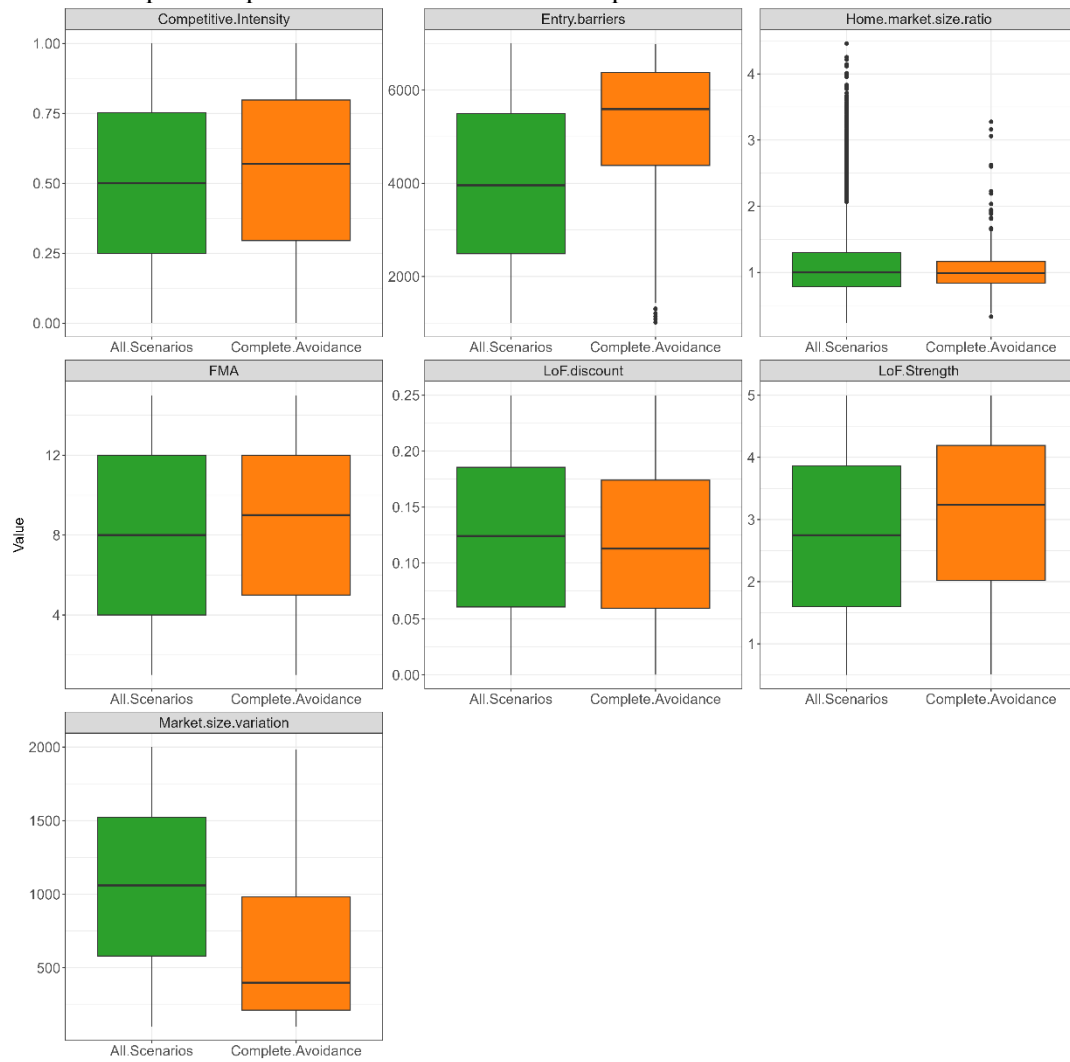


Figure VI.B.4 Competitive intensity, collocation, and performance differences

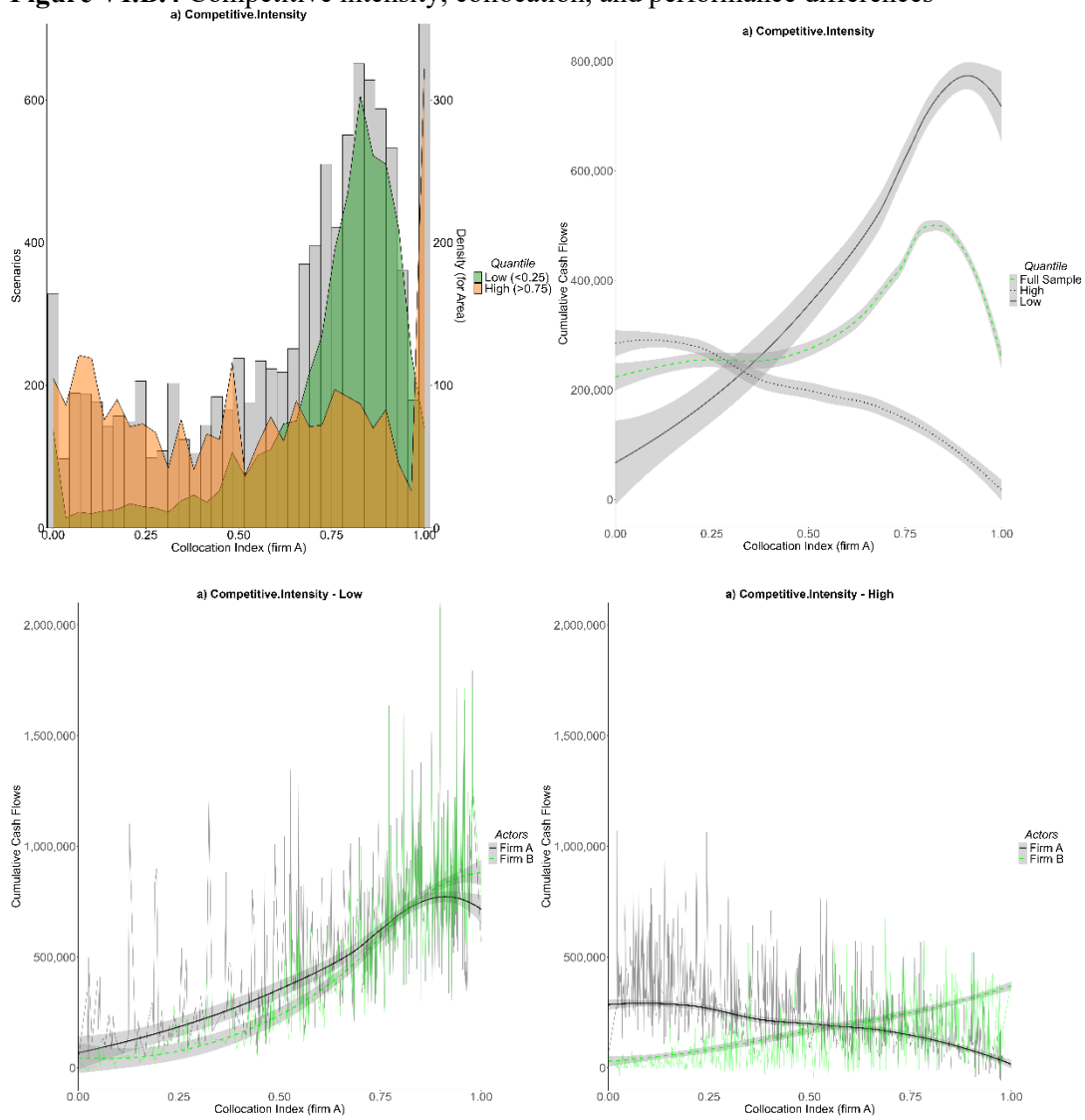


Figure VI.B.5 Entry barriers, collocation, and performance differences

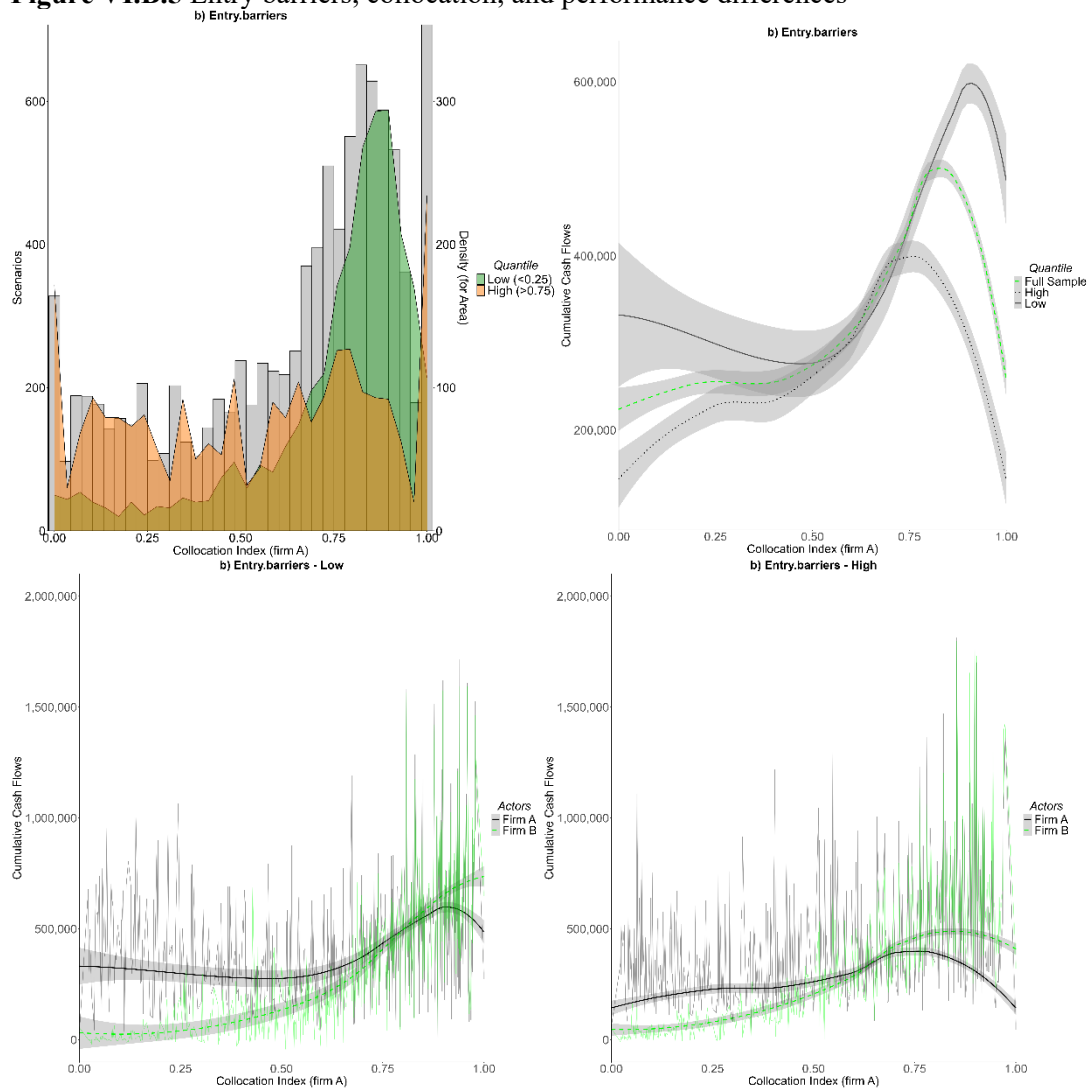


Figure VI.B.6 FMA, LoF strength, LoF discount, collocation, and performance differences

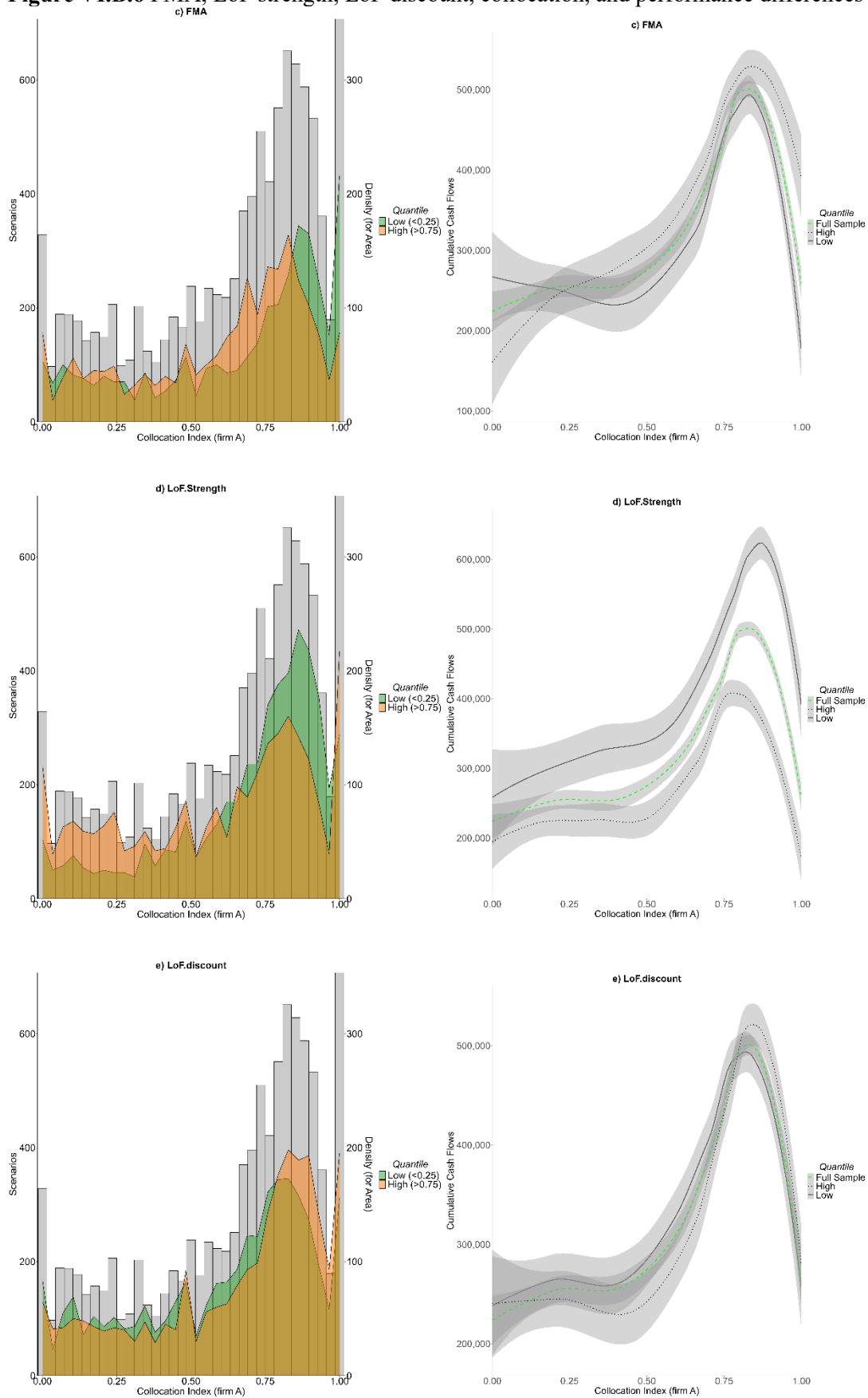


Figure VI.B.7 Market size variation, collocation, and performance differences

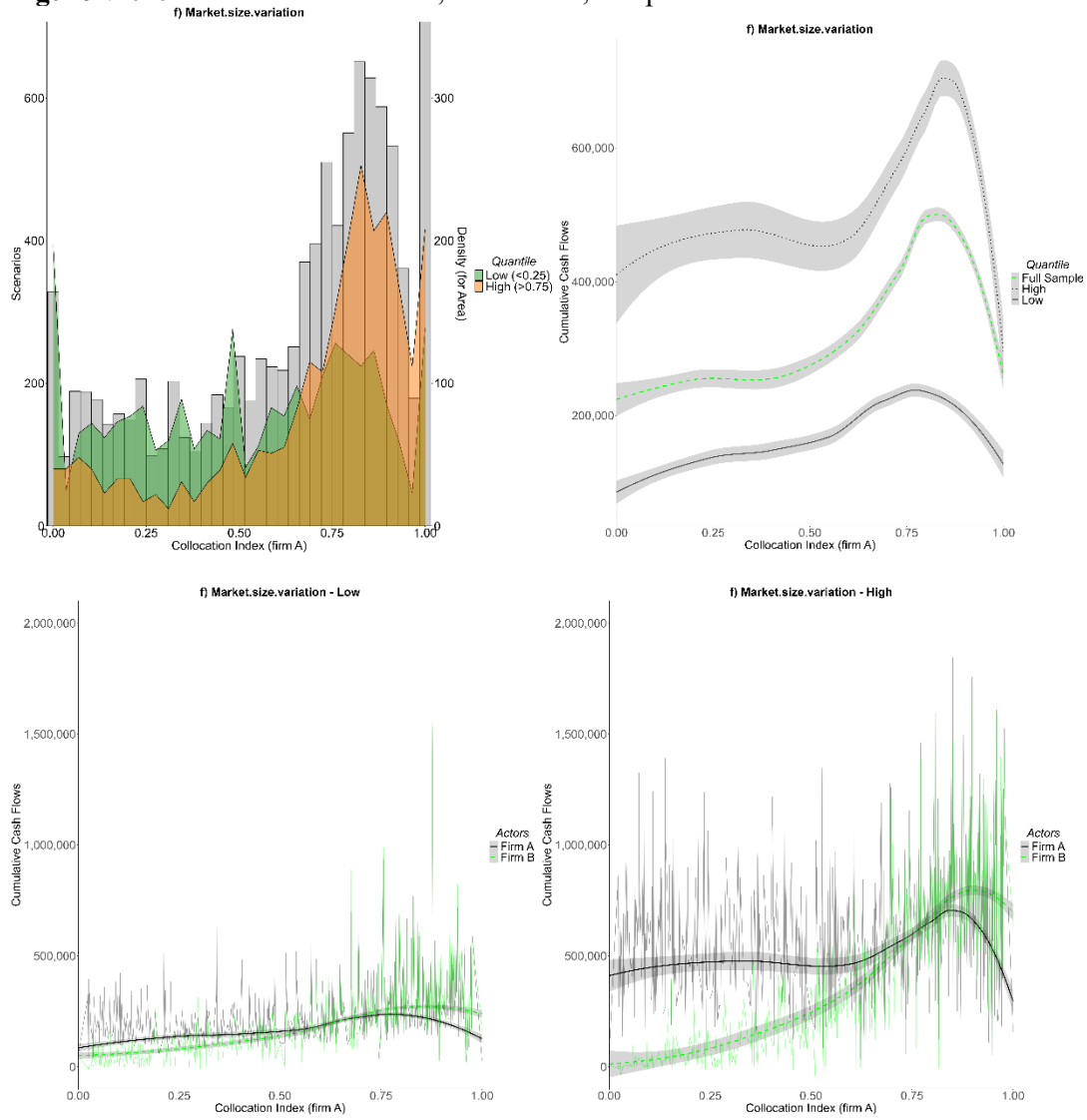
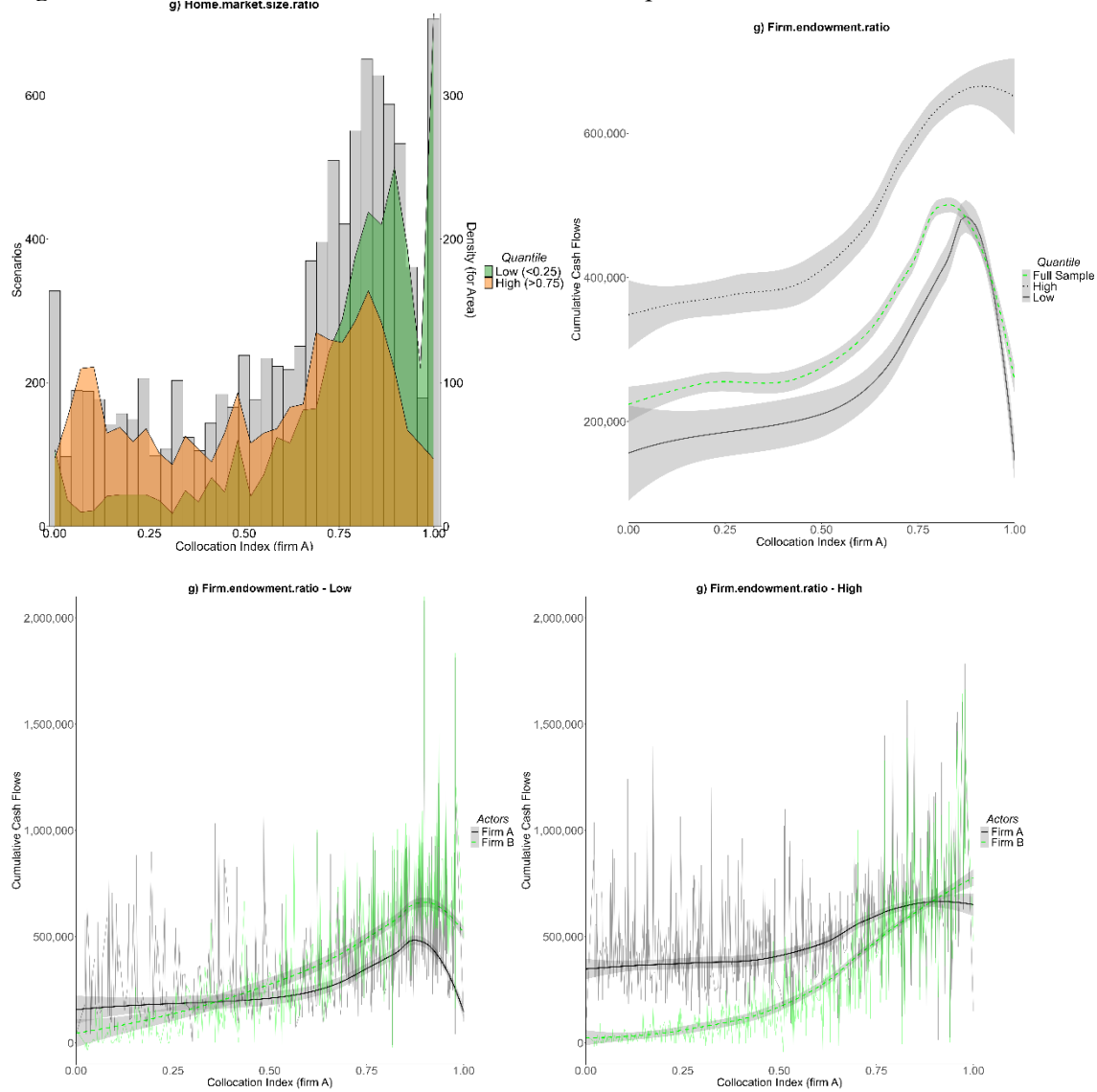


Figure VI.B.8 Home market size ratio, collocation, and performance differences



C. Auto-correlation of market sizes (β)

Figure VI.C.0 shows the autocorrelation values of the world economies based on current GDP (sourced from the World Bank) from 1970 to 2023. Each point represents one country (4 countries were excluded due to missing data, and one country yielded a strange outlier below zero). The main simulation run features a range of 0.7 to <1 . The dark, horizontal line indicates the lower boundary of 0.7 in Figure VI.C.0. Hence, we include essentially all possible values in the main simulation run. In the sensitivity check below, we narrow the range of values to the range within the red, dashed lines (0.8- <0.95). We also run an additional test (not tabulated here) with a broader range (0.6- <1). Both tests yield entirely equivalent outputs. Overall, the patterns show no deviation from the main results and thus also the propositions remain as stated.

Figure VI.C.0 Autocorrelation of world economies (current GDP; 1970-2023)

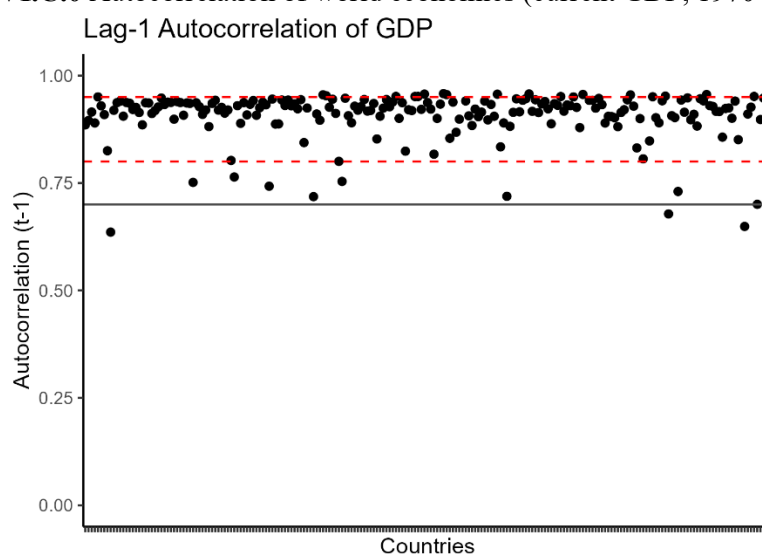


Figure VI.C.1 Collocation Index

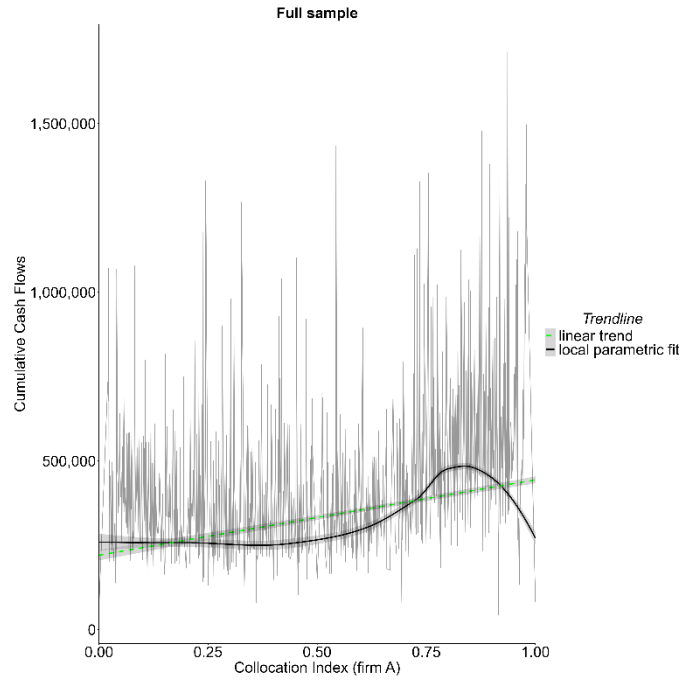
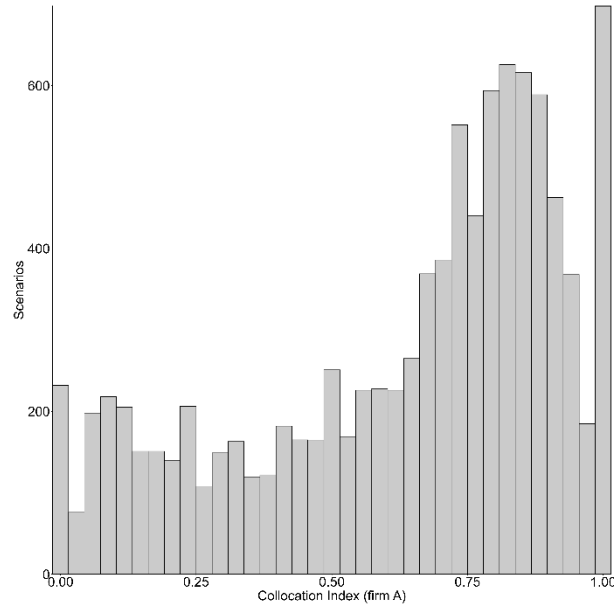


Figure VI.C.2 Sample comparison: All Scenarios vs. Complete Collocation

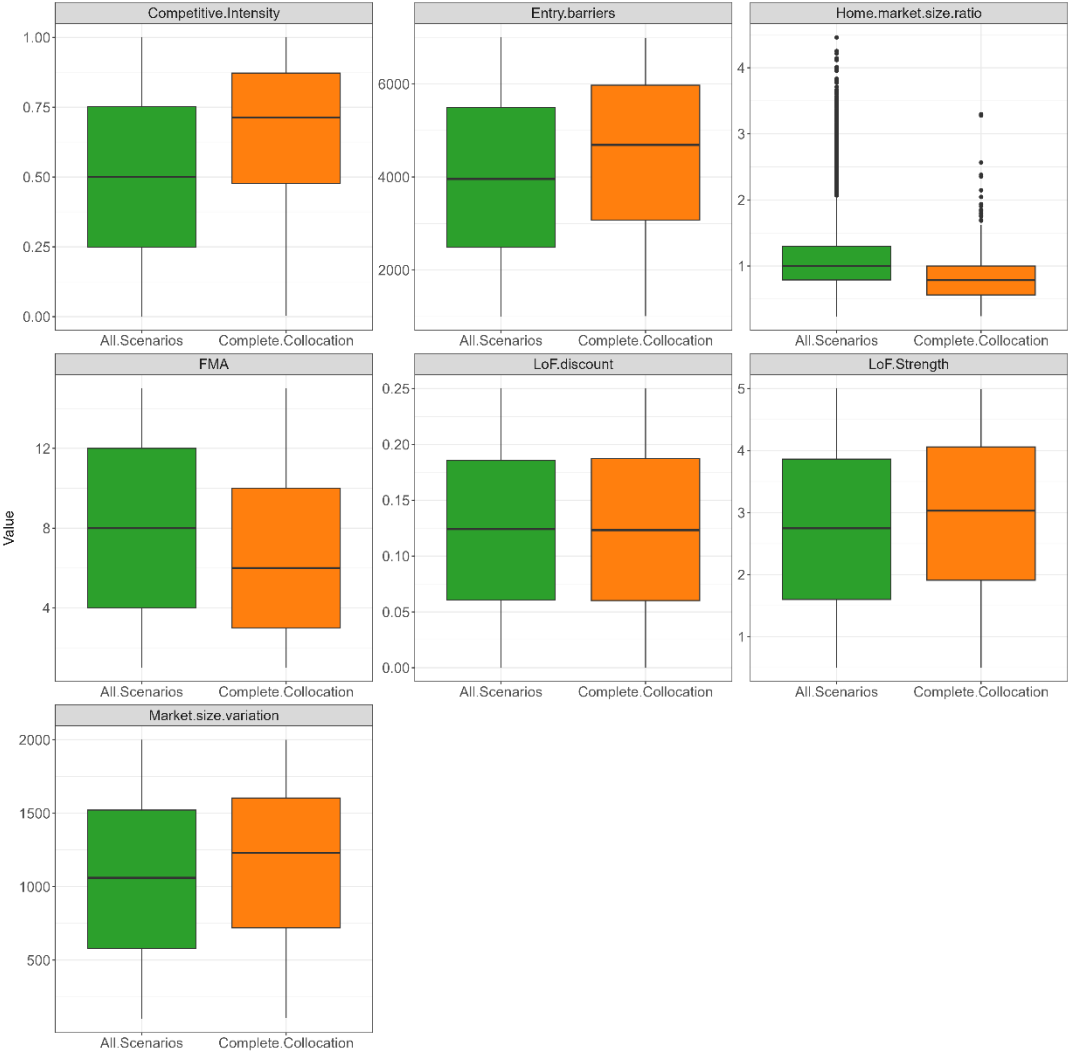


Figure VI.C.3 Sample Comparison: All Scenarios vs. Complete Avoidance

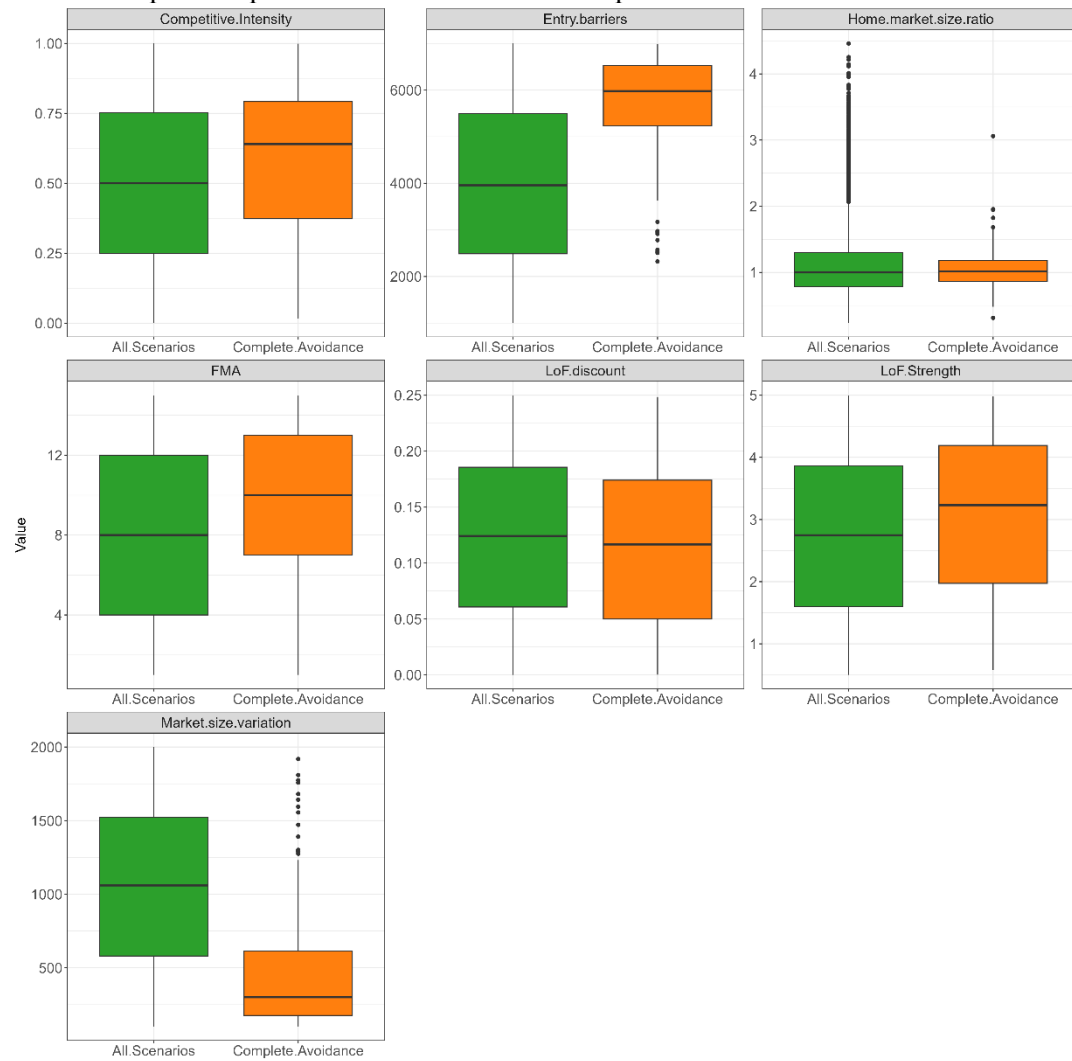


Figure VI.C.4 Competitive intensity, collocation, and performance differences

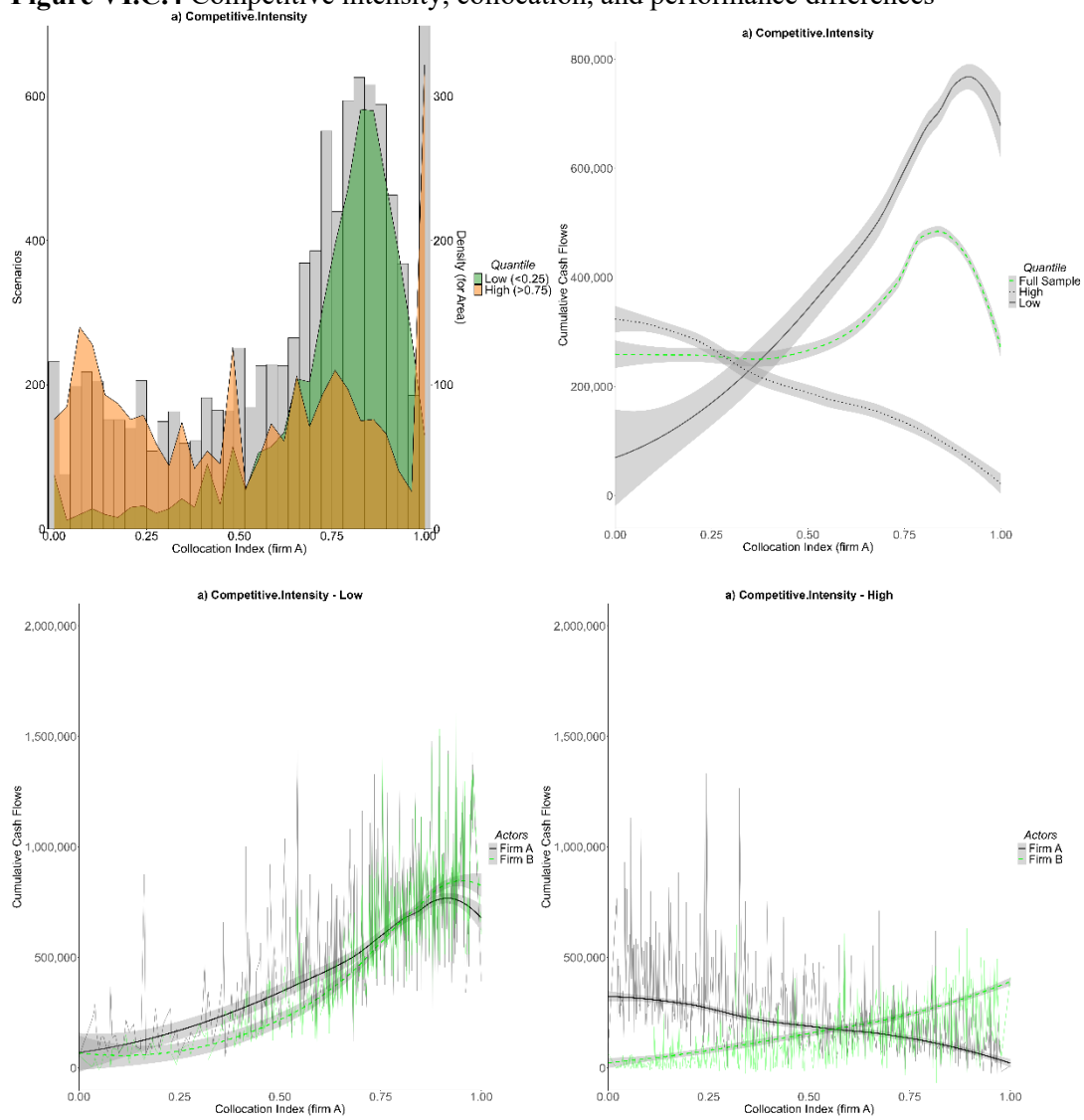


Figure VI.C.5 Entry barriers, collocation, and performance differences

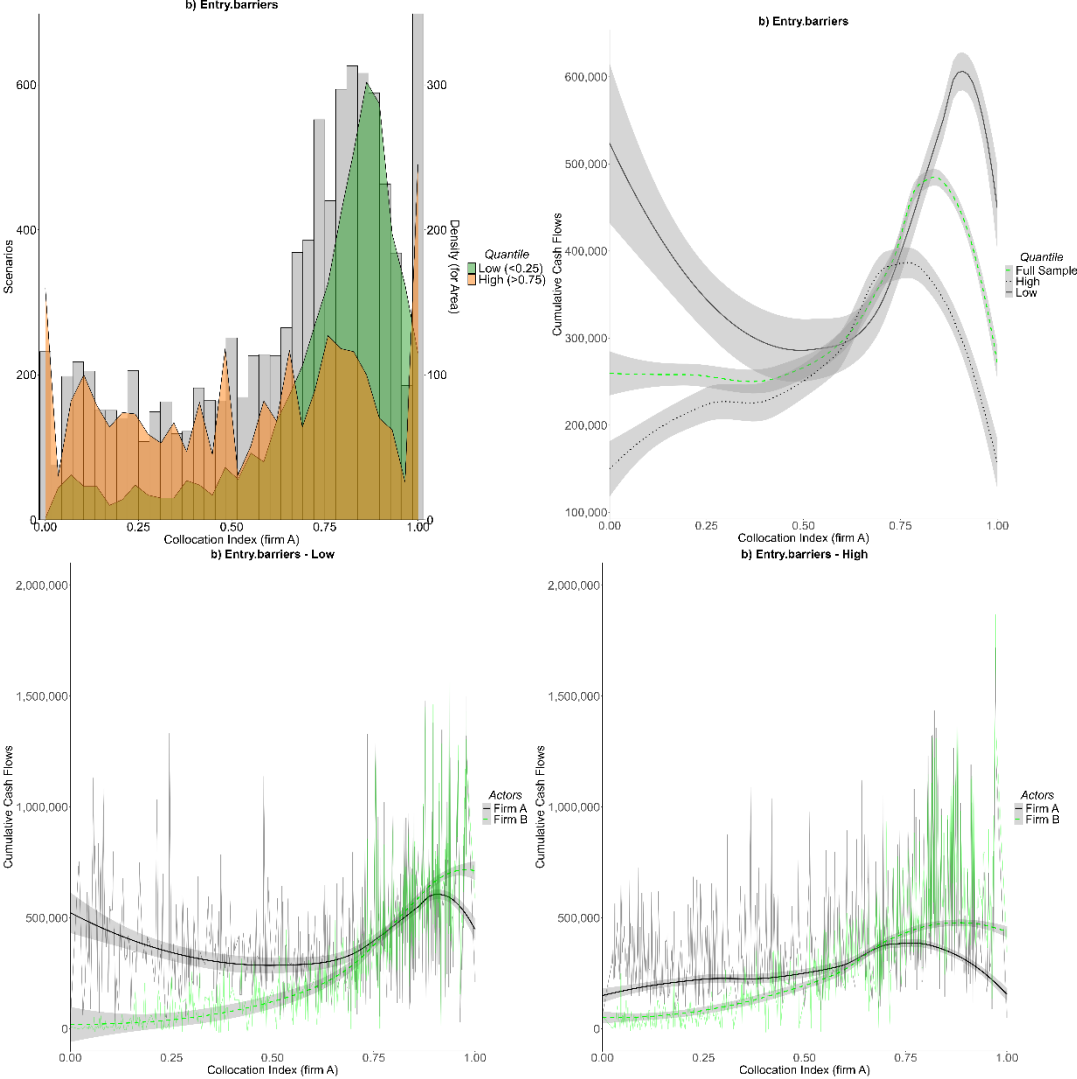


Figure VI.C.6 FMA, LoF strength, LoF discount, collocation, and performance differences

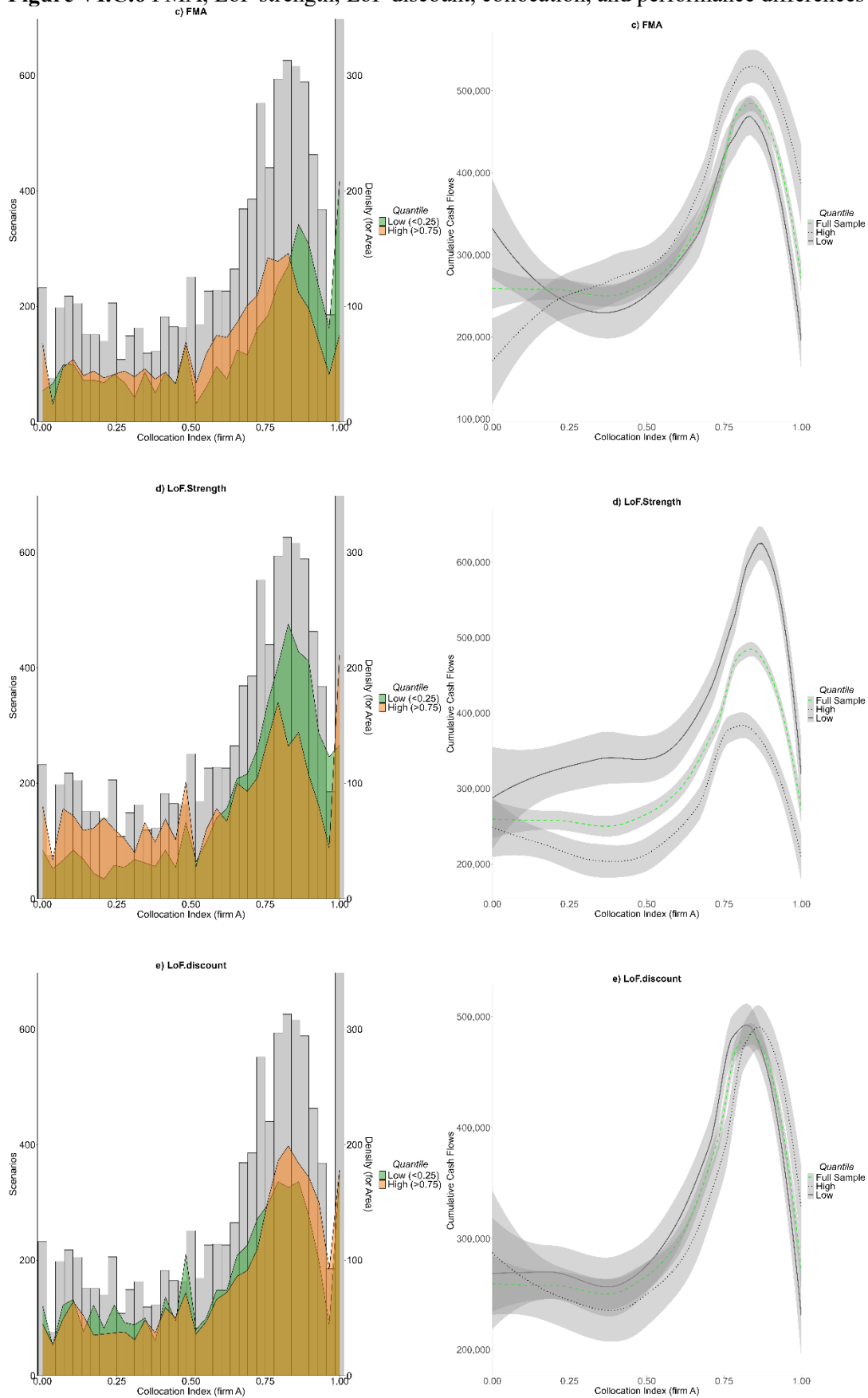


Figure VI.C.7 Market size variation, collocation, and performance differences

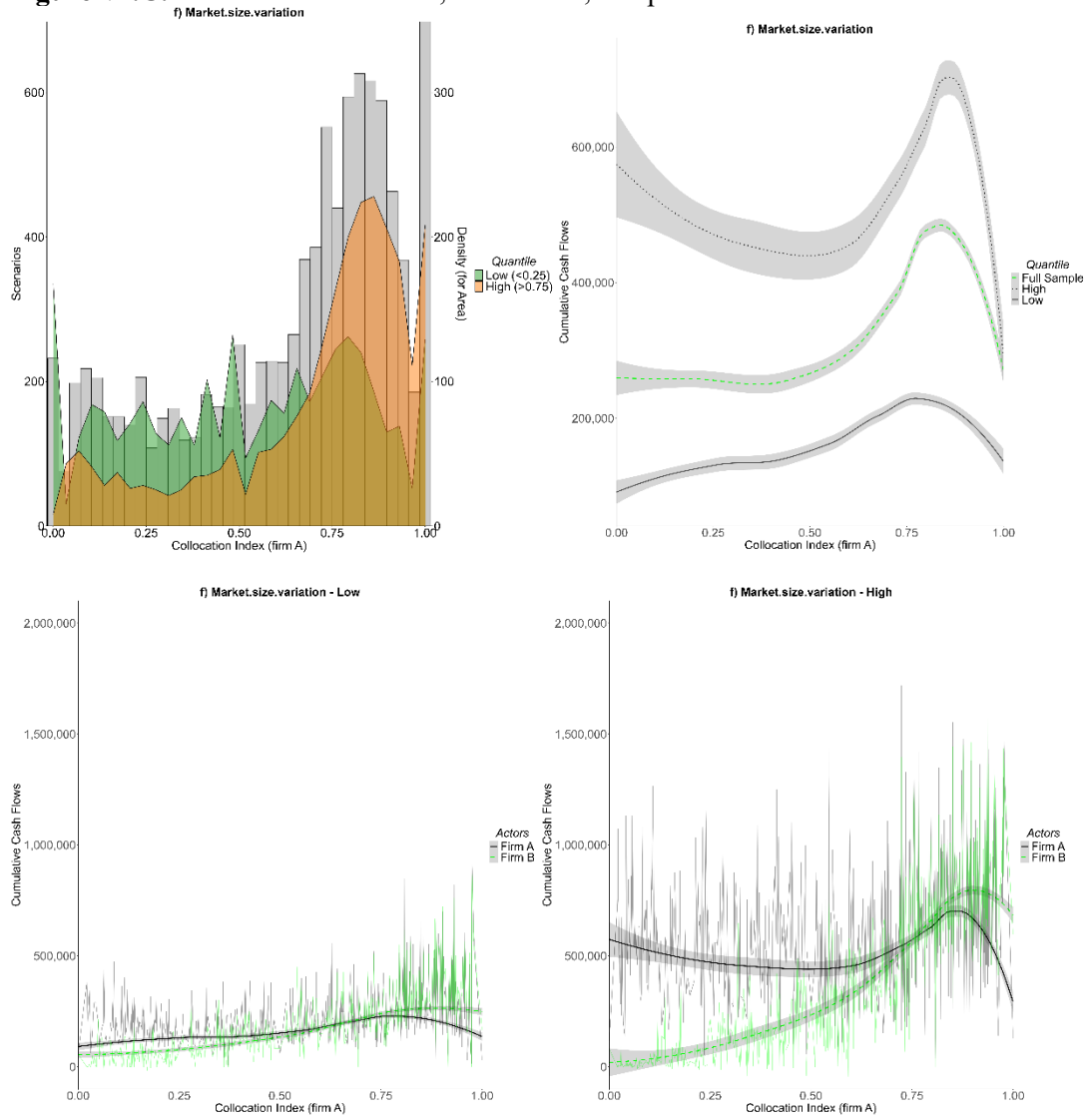
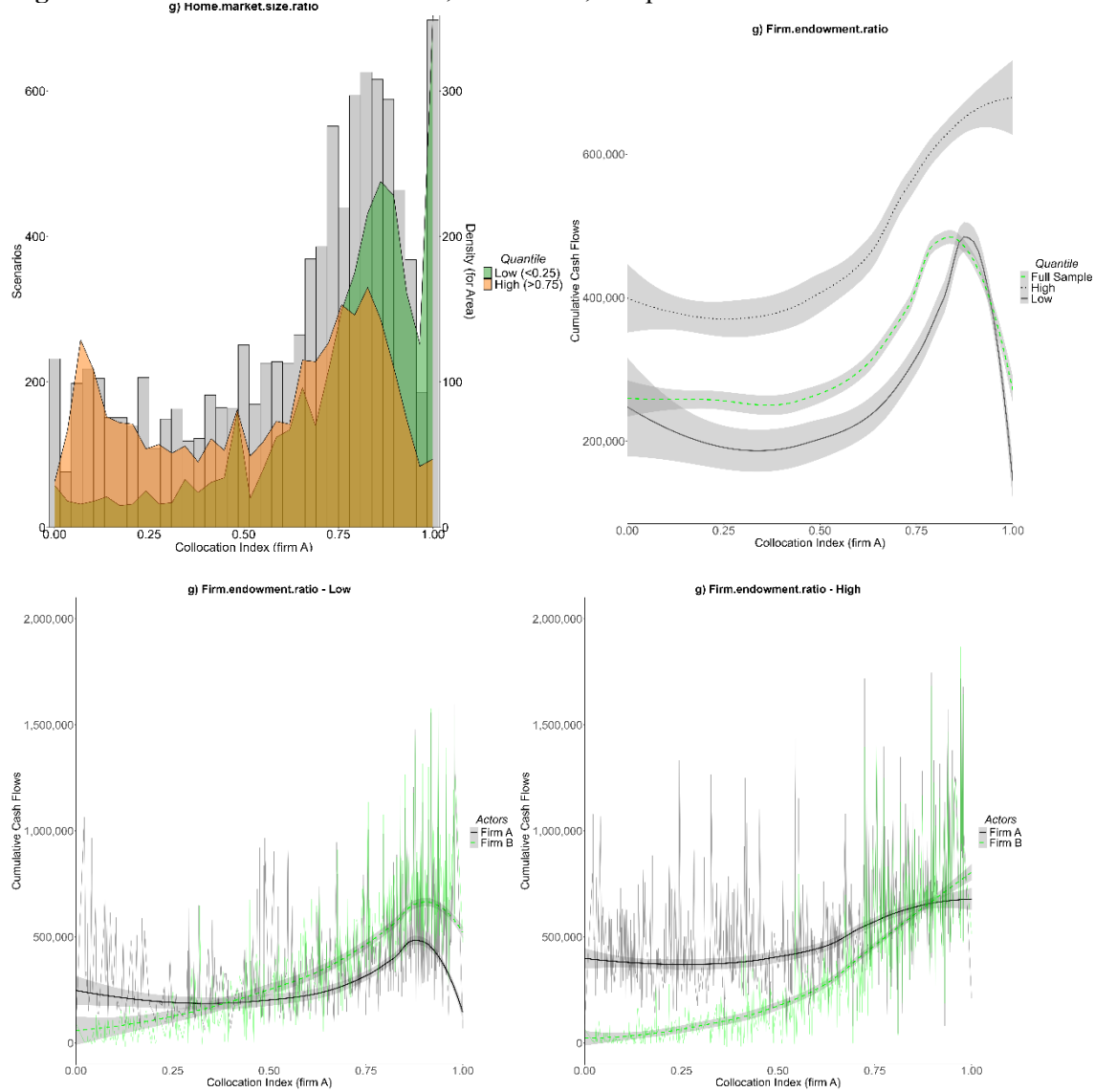


Figure VI.C.8 Home market size ratio, collocation, and performance differences



D. LoF: Liabilities of foreignness (Λ)

Calibrating the value for LoF presents a challenge; however, the sensitivity analysis below indicates that the strength of LoF does not significantly impact the findings of this study.

It is notoriously difficult to precisely quantify the exact loss due to LoF when operating abroad. Studies exist that provide insights into the potential strength of LoF. Buchner et al. (2018) find that foreign VCs underperform domestic ones by 16-27%, Wu and Salomon (2017) find that foreign banks are 27% more likely to receive regulatory sanctioning, or Kim (2019) finds that for every dollar spent on lobbying, foreign firms increase the contract amount by 10% (compared to 1% for domestic firms). These studies provided a rough anchor point for our operationalization of LoF as LoF cannot only differ in terms of strength in our simulation but also naturally vary with distance. Figure VI.D.0 illustrates the effect of LoF on firms in our simulation. The x-axis represents the distance to the home country (NB: in the simulation, we operationalize distance via the correlation matrix. To facilitate the interpretation of the Figure, we have re-labeled the x-axis to align with the notion of distance). The y-axis represents the loss in cash flows due to LoF. A firm that starts in its home country (distance 0) would lose 0% in the next neighbouring country and 26% in the 25th most distant country under low LoF (0.5), while at high levels of LoF strength (5), this firm would lose 46% in the next neighbouring country and 87% in the 25th most distant country. The full range of the x-axis would represent the most extreme case, where a firm would start in country #1 or #50 (the most remote options). In many cases, firms would not need to go beyond the middle of the x-axis.

The solid lines are the cases we use in the main simulation. We conduct a sensitivity analysis by expanding the range to consider LoF strengths of 0.2 and 10 (dashed lines), which, on the higher end, represent already highly extreme cases. Still, the outputs of our simulation remain equivalent to the main outputs. Overall, the patterns show no deviation from the main results and thus also the propositions remain as stated.

Figure VI.D.0 Illustration – Effect of LoF

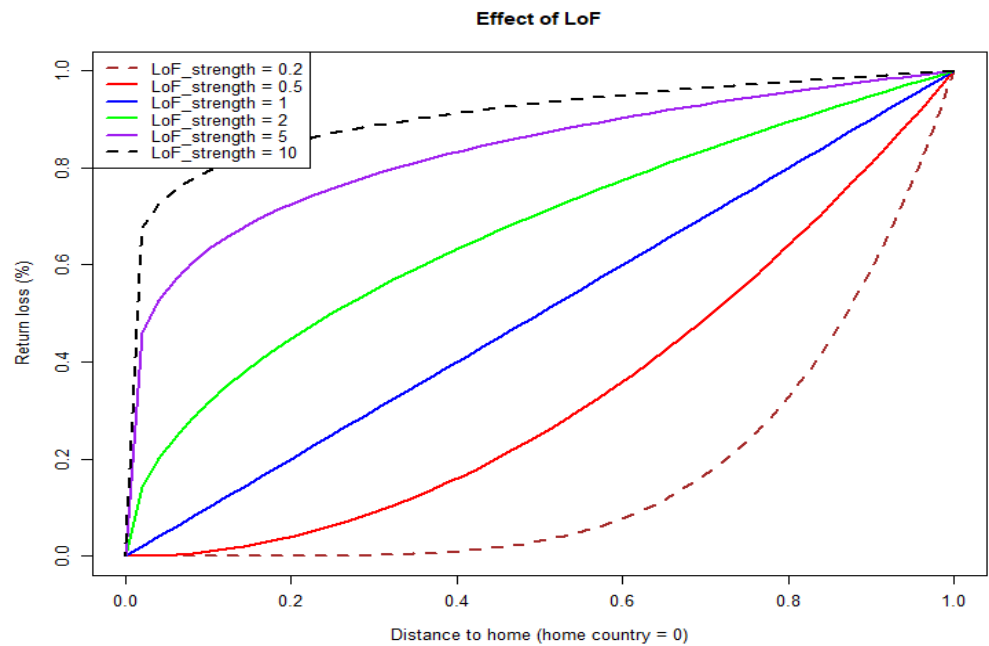
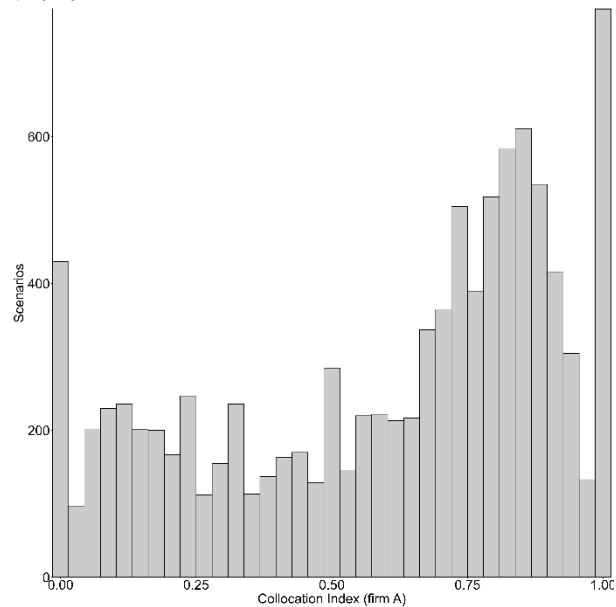


Figure VI.D.1 Collocation Index



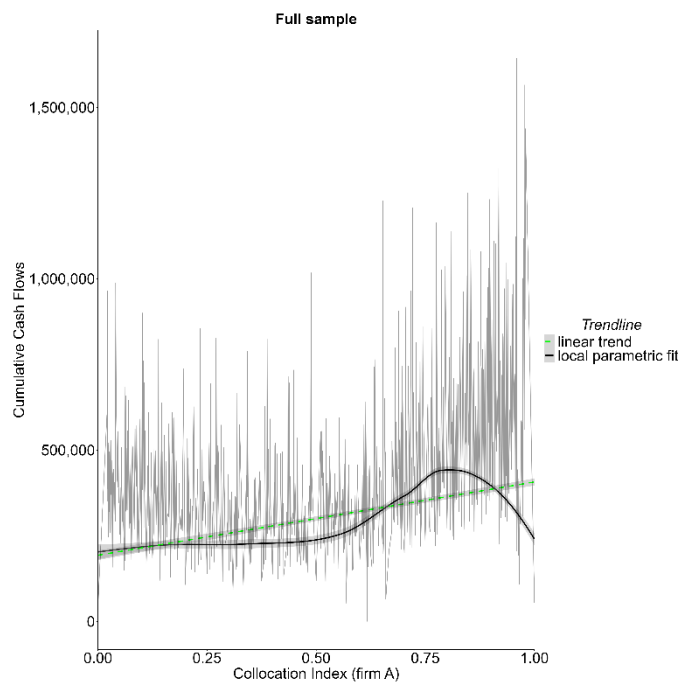


Figure VI.D.2 Sample comparison: All Scenarios vs. Complete Collocation

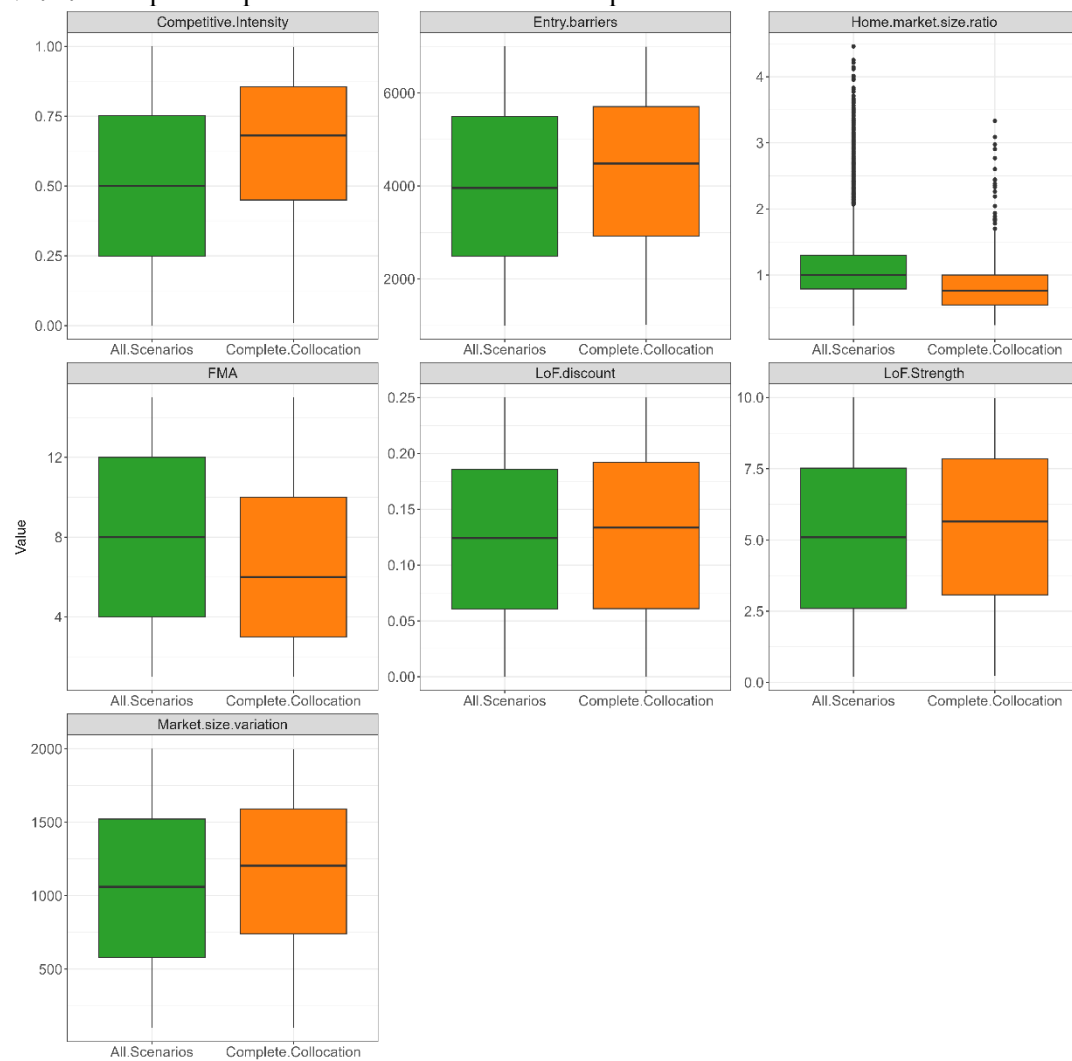


Figure VI.D.3 Sample Comparison: All Scenarios vs. Complete Avoidance

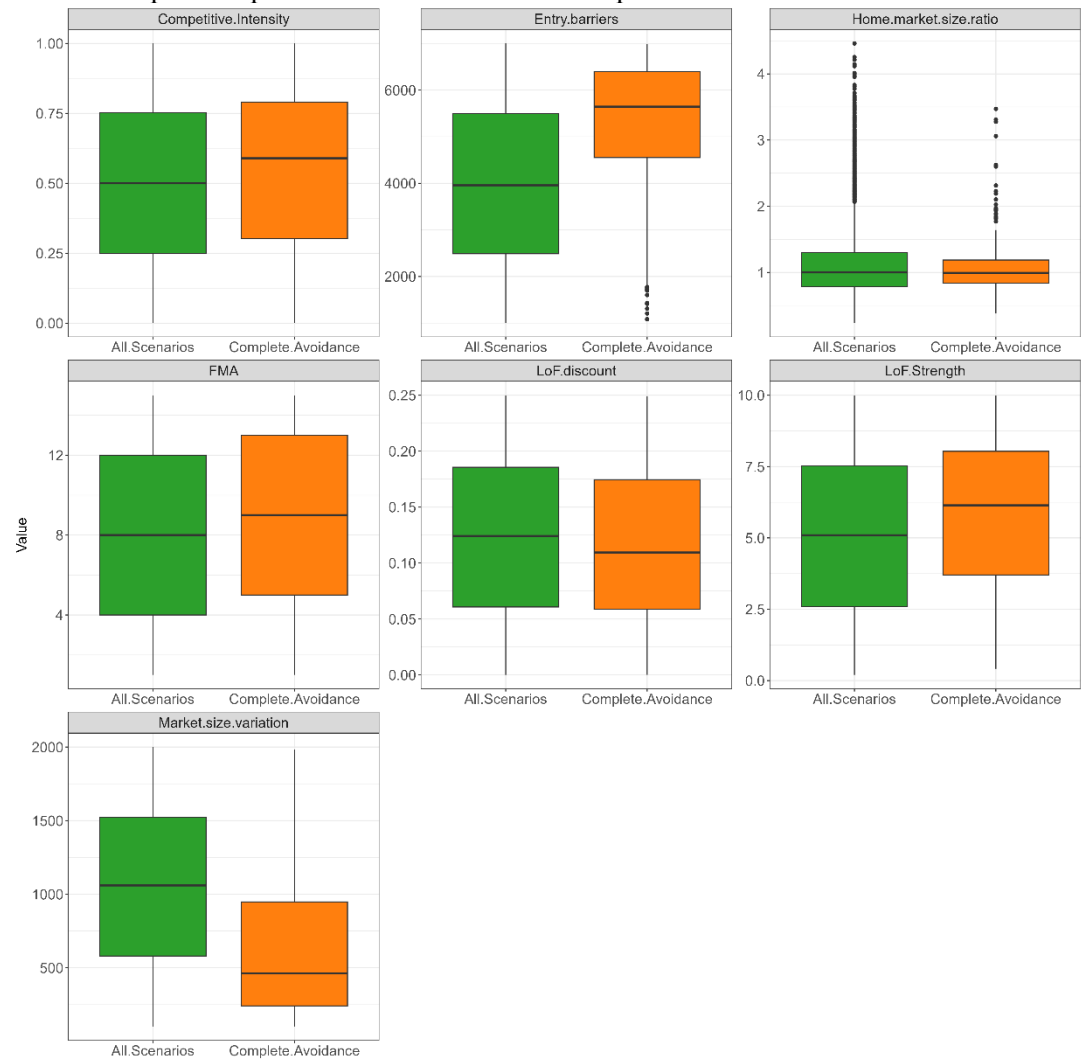


Figure VI.D.4 Competitive intensity, collocation, and performance differences

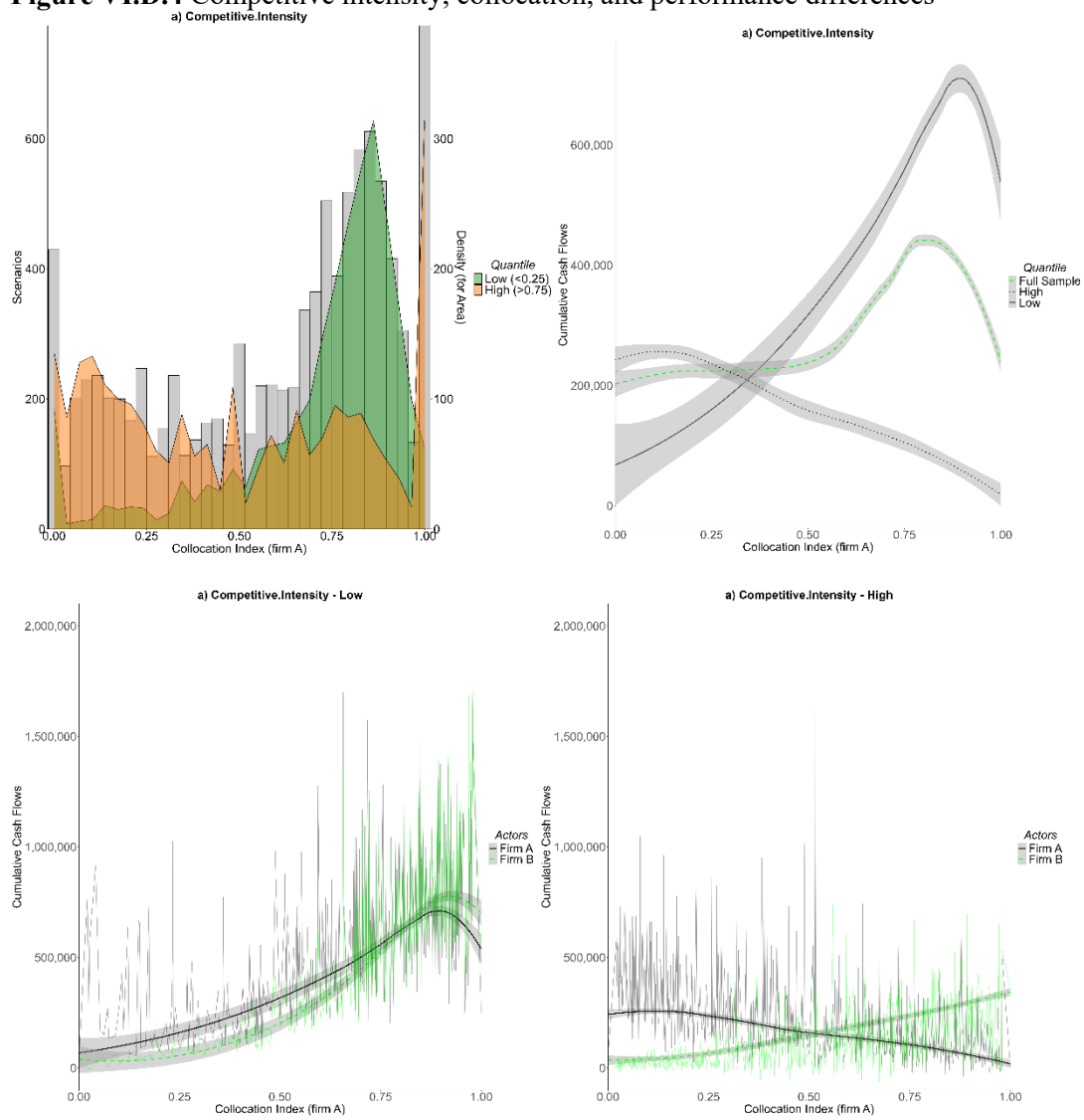


Figure VI.D.5 Entry barriers, collocation, and performance differences

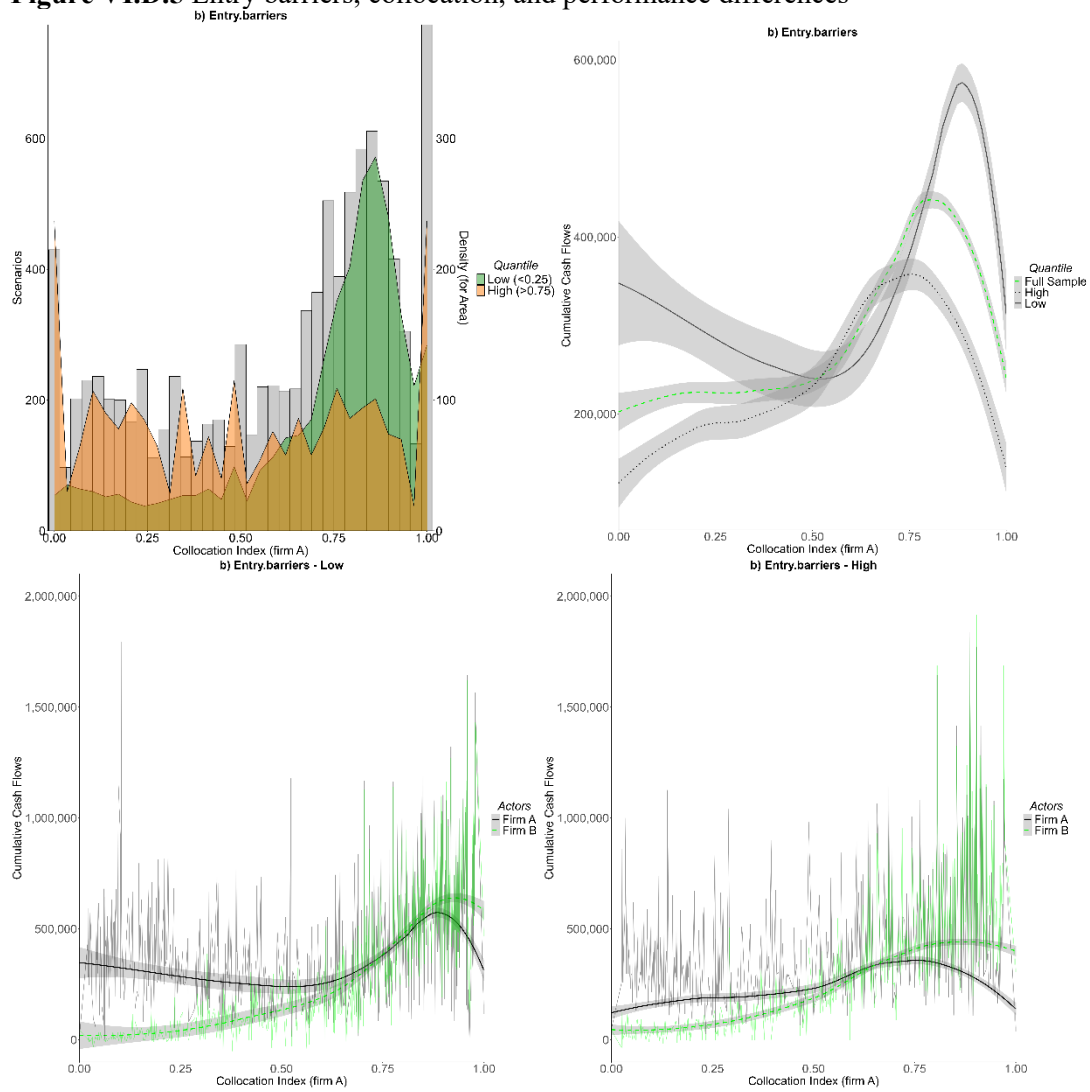


Figure VI.D.6 FMA, LoF strength, LoF discount, collocation, and performance differences

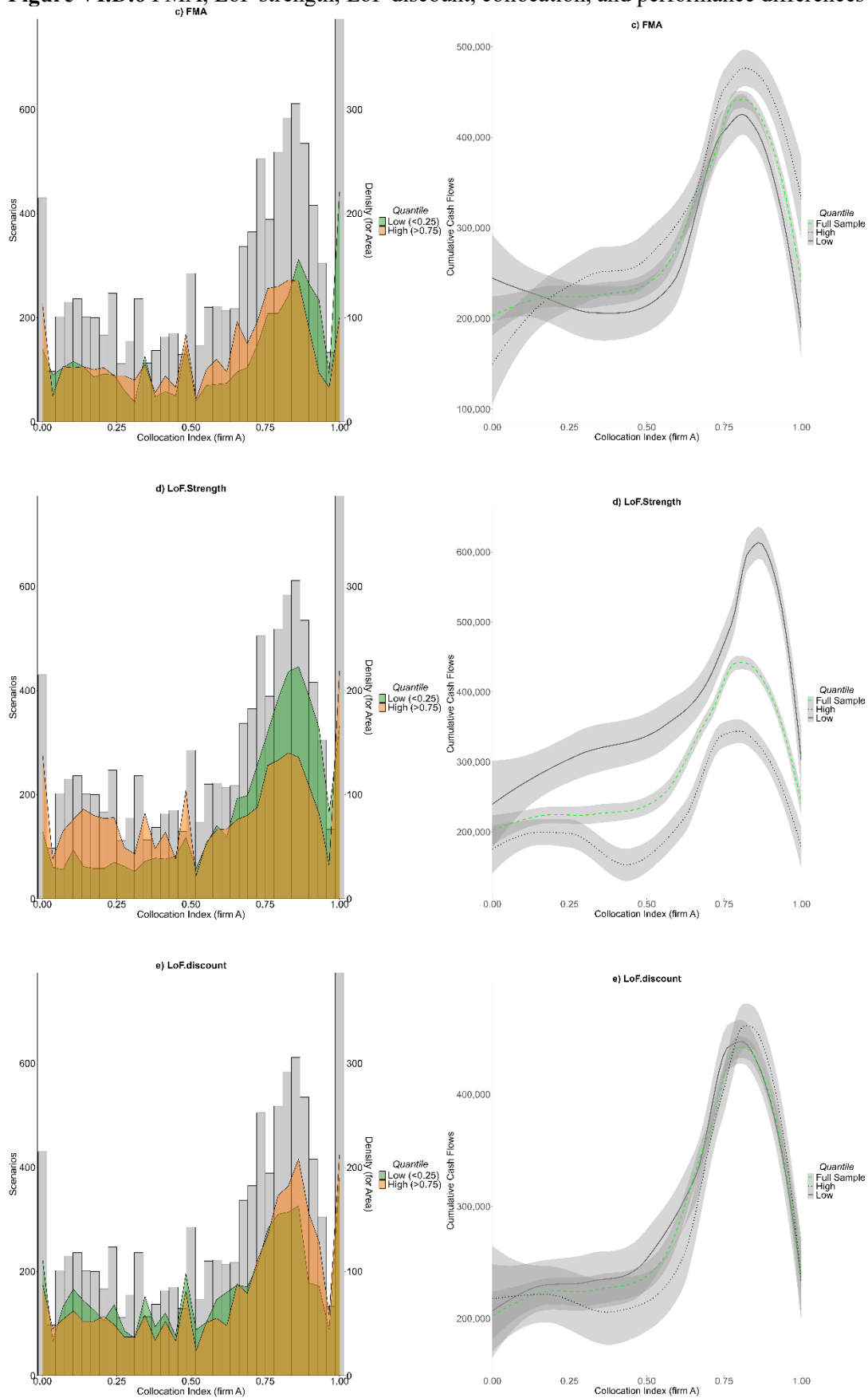


Figure VI.D.7 Market size variation, collocation, and performance differences

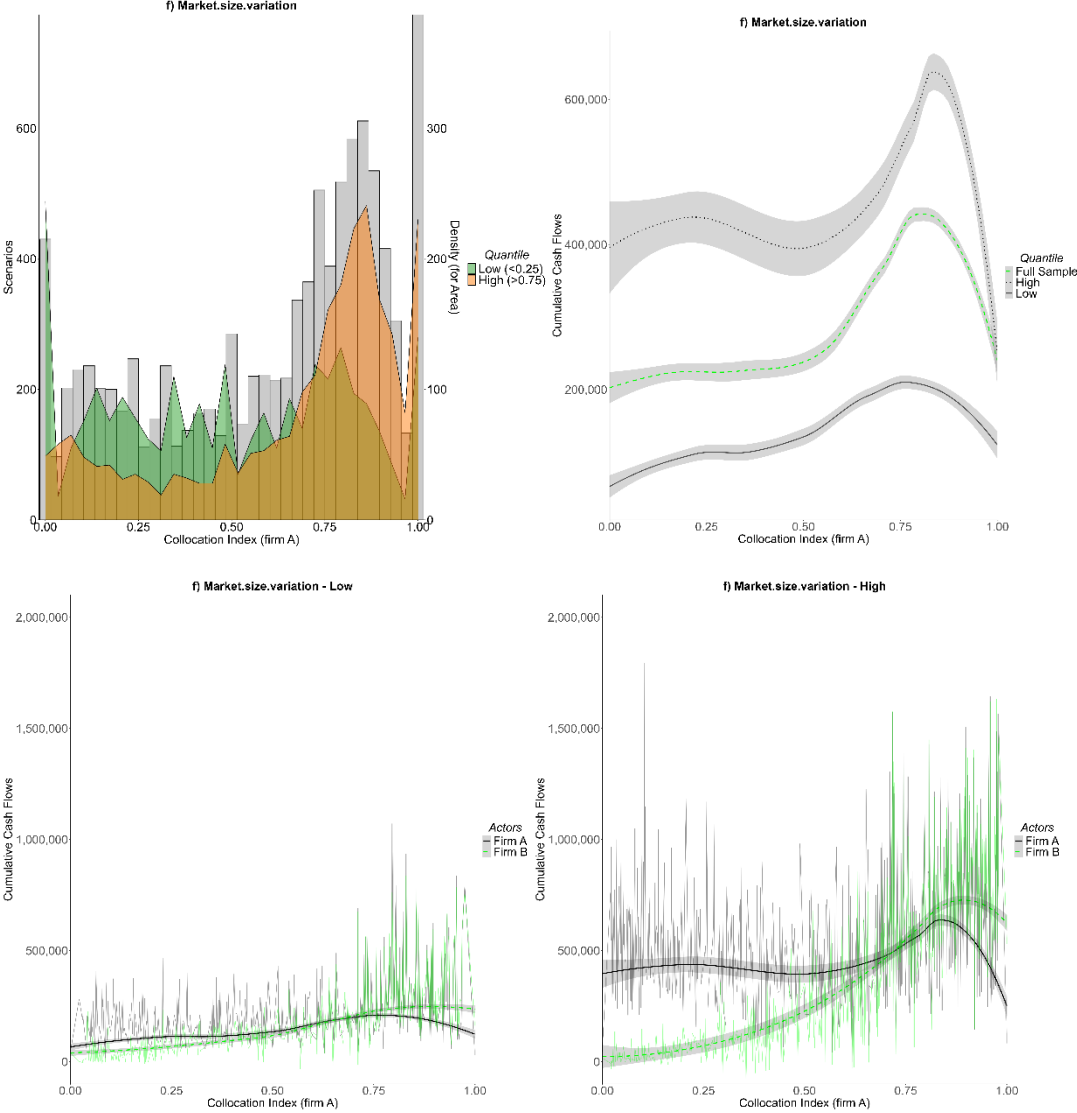
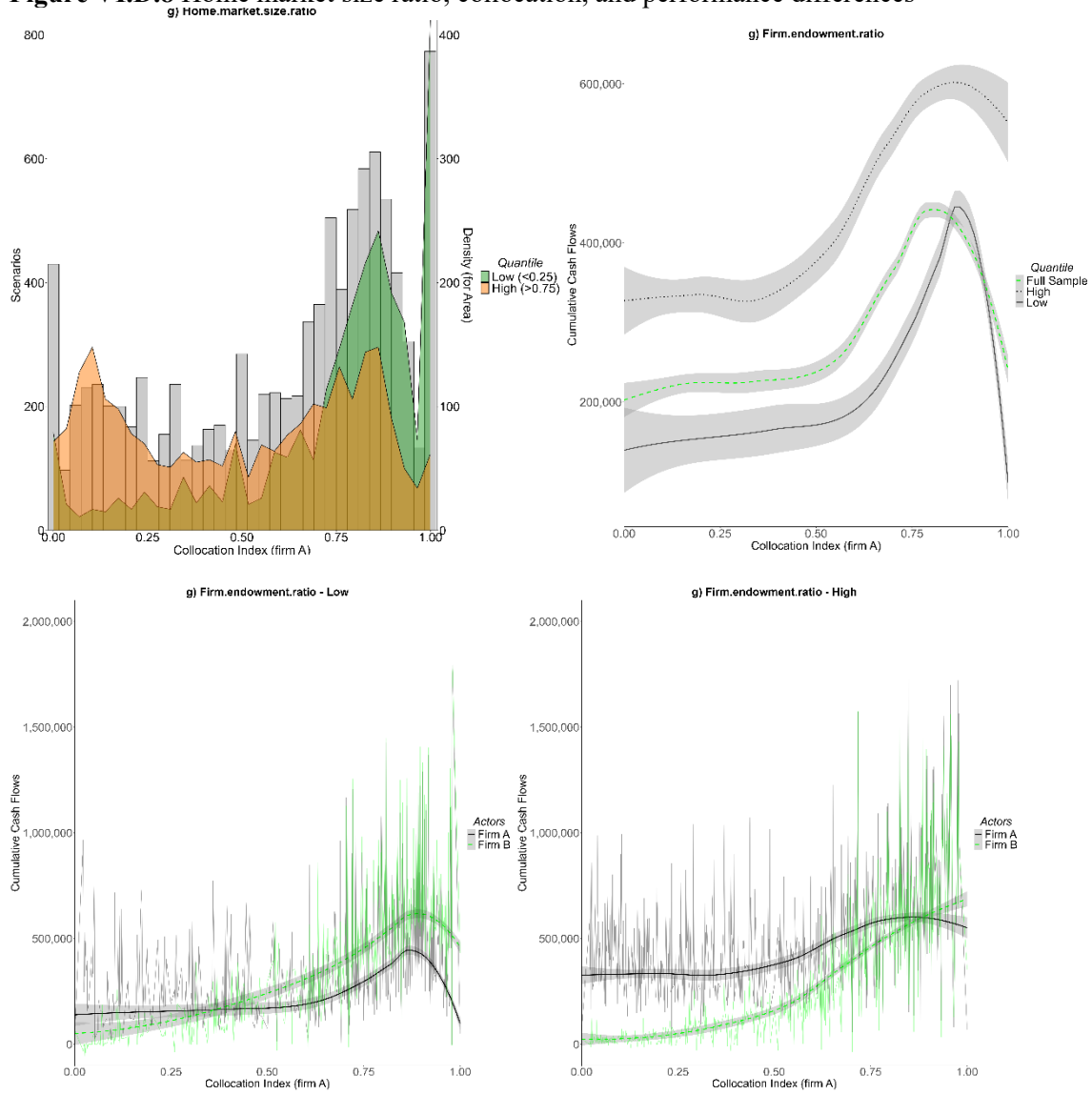


Figure VI.D.8 Home market size ratio, collocation, and performance differences



E. LoF discount (δ)

In the absence of empirical, real-world guidance, we chose a parameter that intuitively makes sense and verified its validity through sensitivity analyses. The alternative range of $0 < \delta < 0.5$ provides entirely equivalent outputs to the main simulation models. Overall, the patterns show no deviation from the main results and thus also the propositions remain as stated.

Figure VI.E.1 Collocation Index

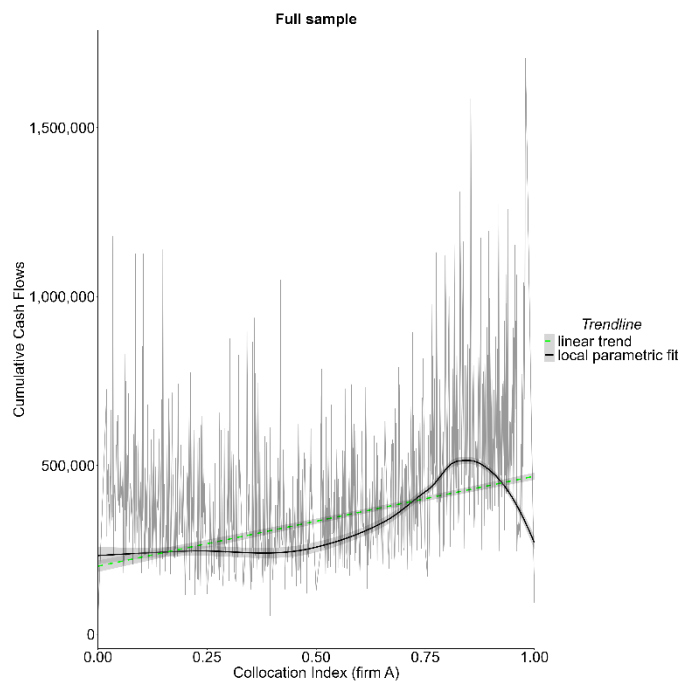
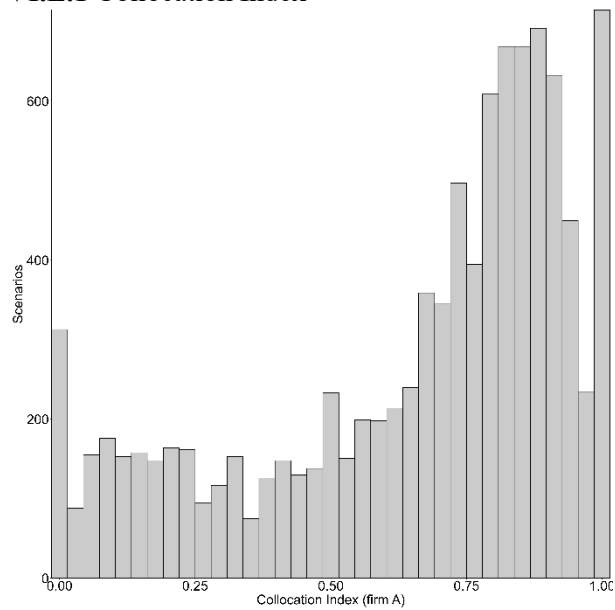


Figure VI.E.2 Sample comparison: All Scenarios vs. Complete Collocation

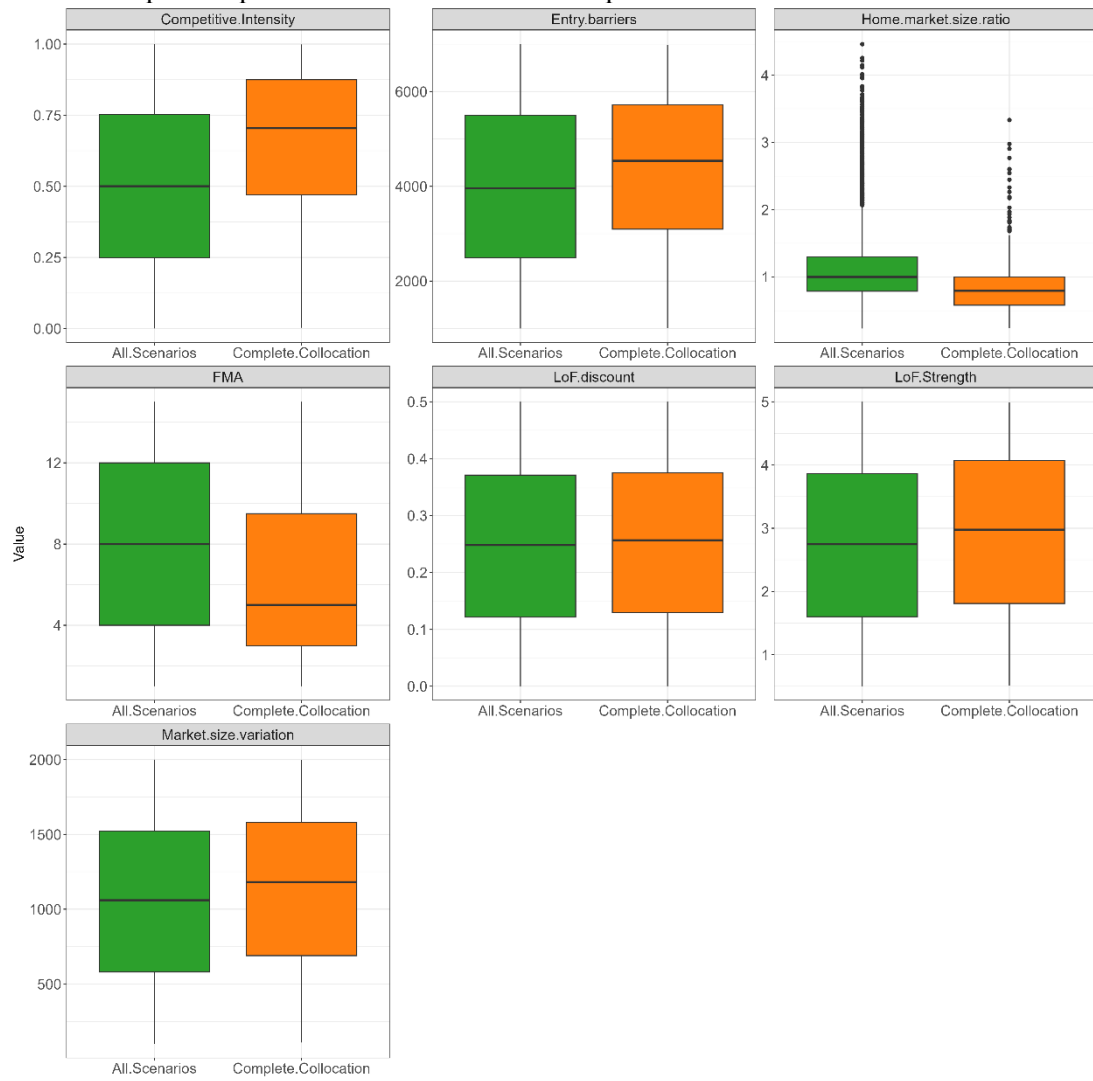


Figure VI.E.3 Sample Comparison: All Scenarios vs. Complete Avoidance

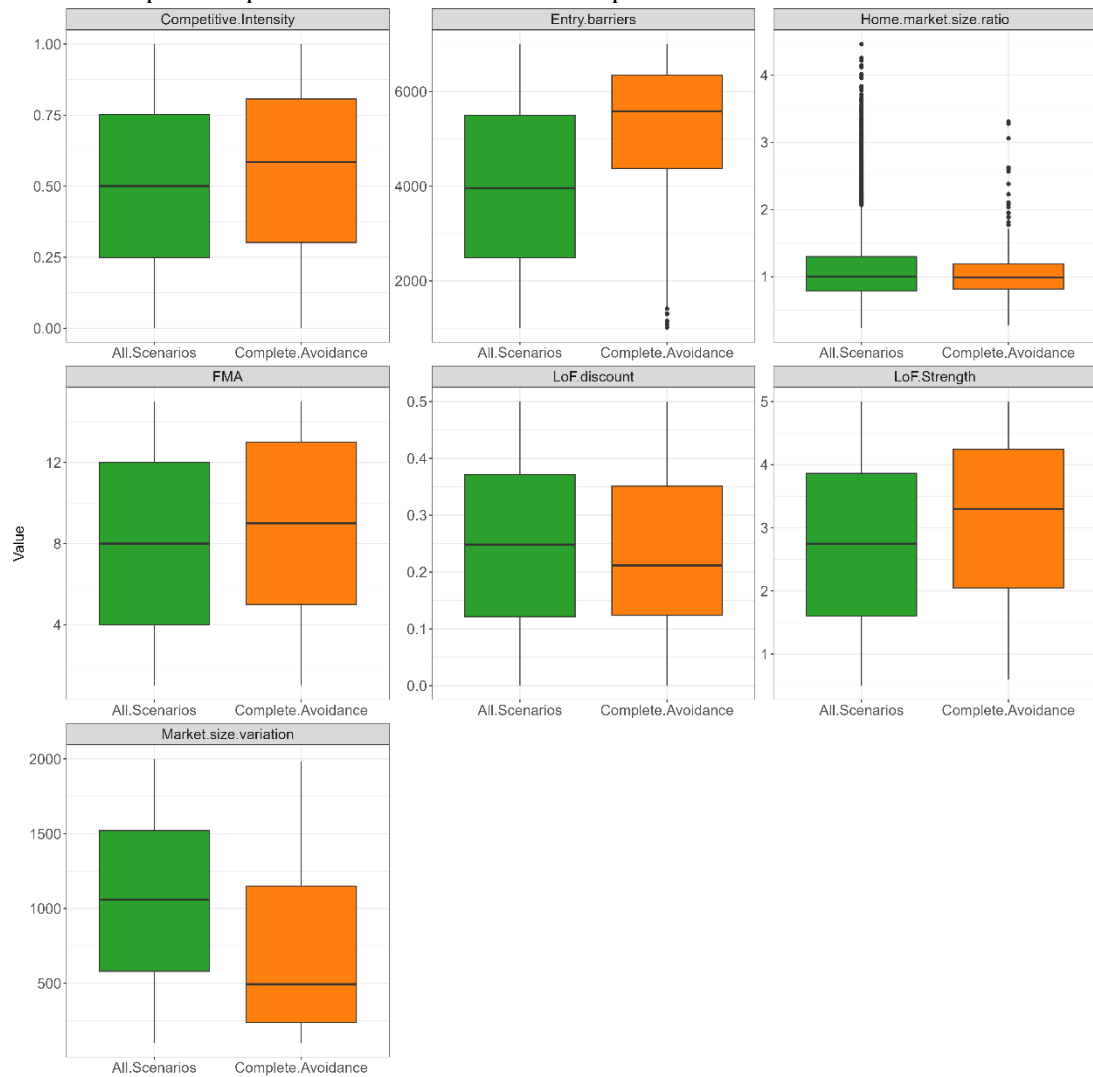


Figure VI.E.4 Competitive intensity, collocation, and performance differences

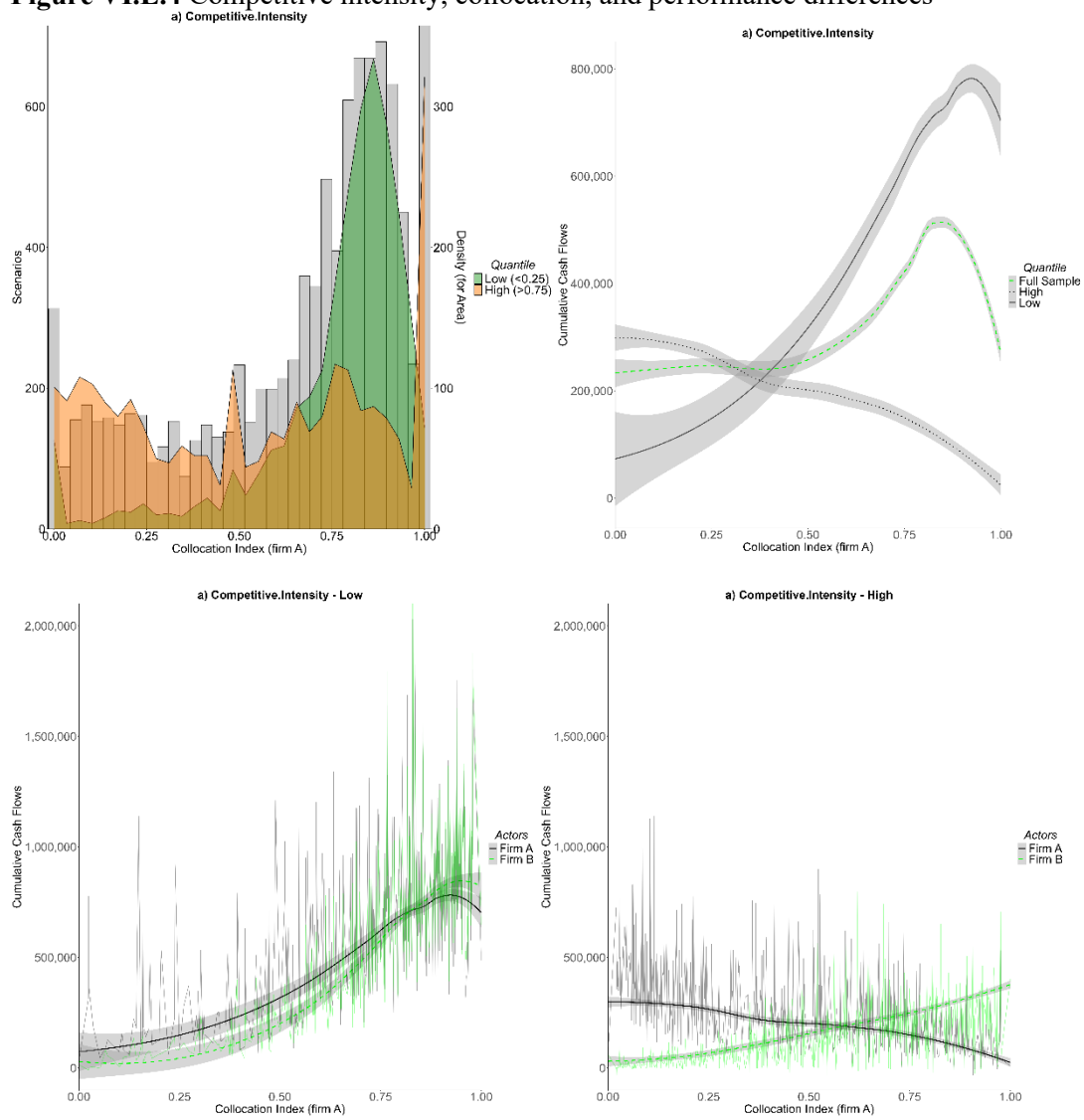


Figure VI.E.5 Entry barriers, collocation, and performance differences

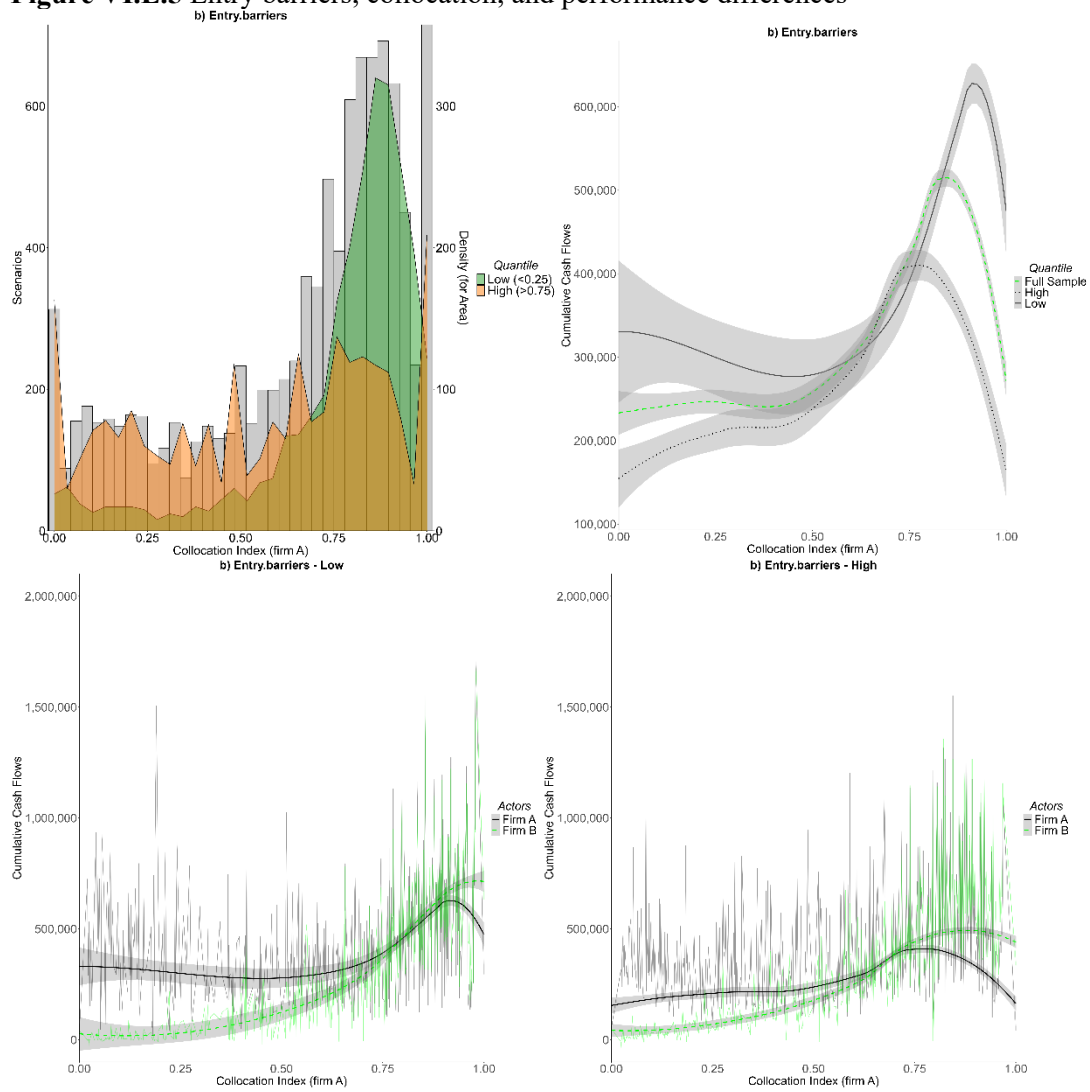


Figure VI.E.6 FMA, LoF strength, LoF discount, collocation, and performance differences

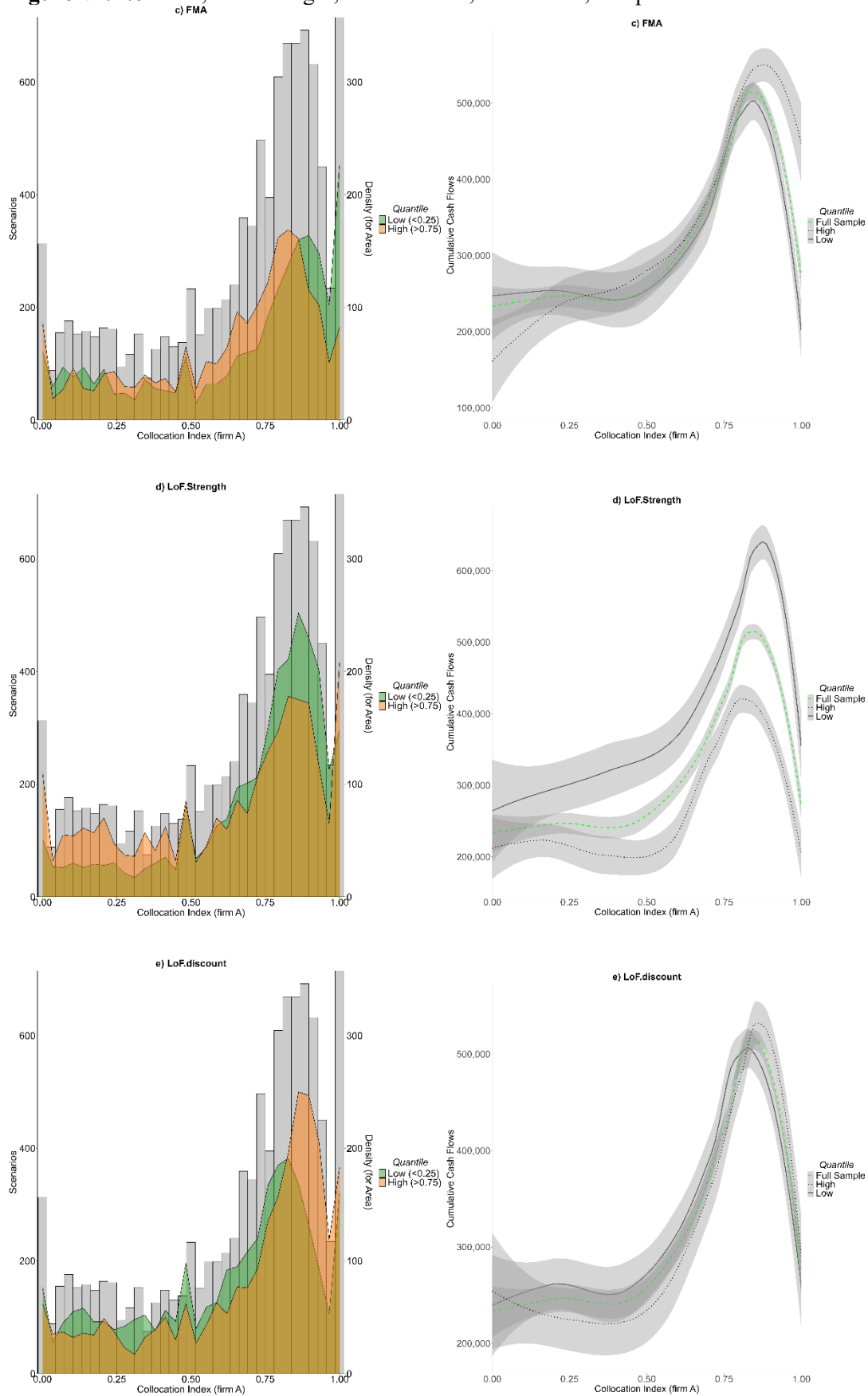


Figure VI.E.7 Market size variation, collocation, and performance differences

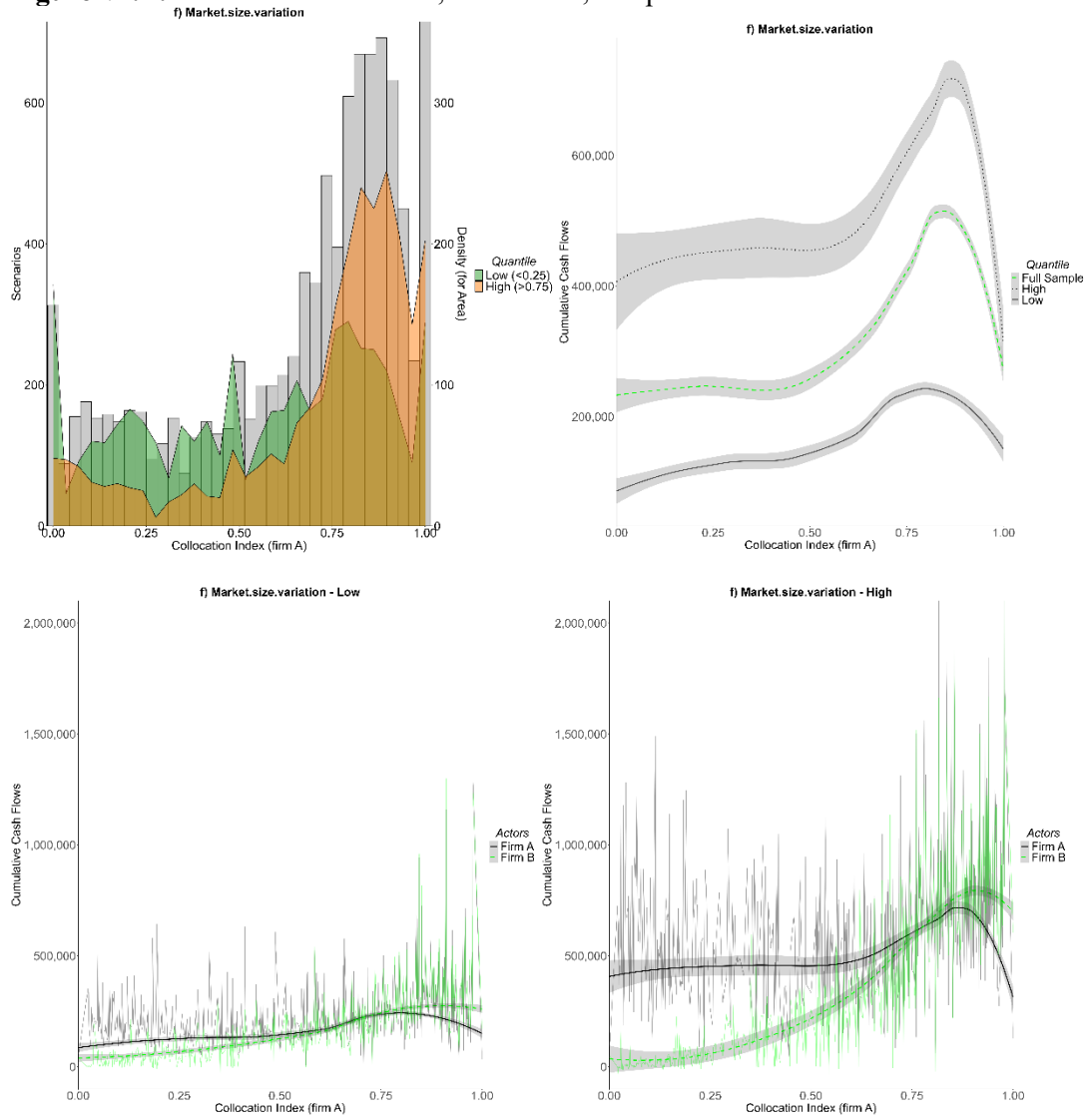
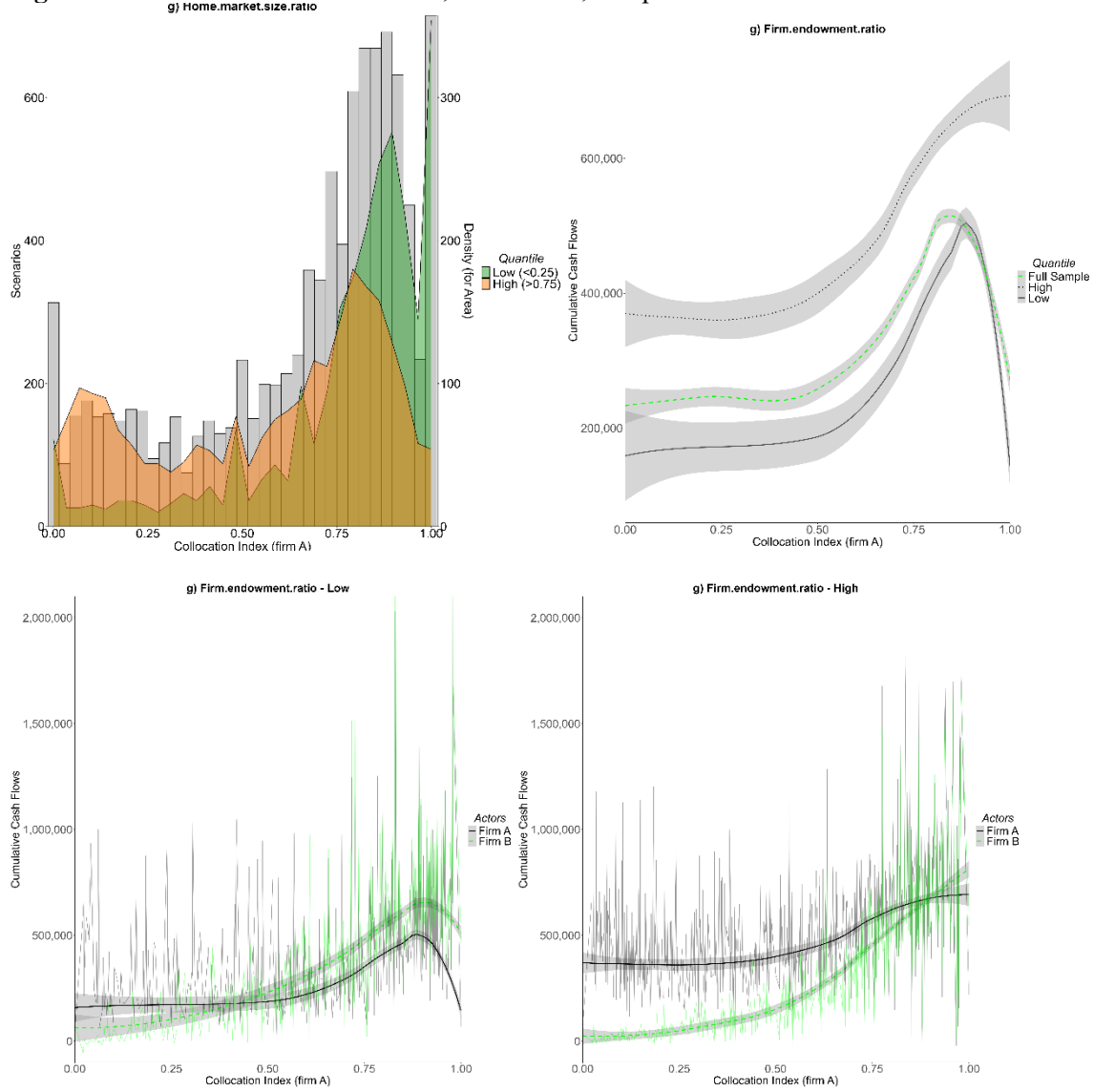


Figure VI.E.8 Home market size ratio, collocation, and performance differences



F. Competitive intensity (c)

In the main simulation model, we allow competitive intensity to assume any feasible value. In the sensitivity analysis, we constrain the range to $0.25 < 0.75$, which provides equivalent outputs. Overall, the patterns show no deviation from the main results and thus also the propositions remain as stated.

Figure VI.F.1 Collocation Index

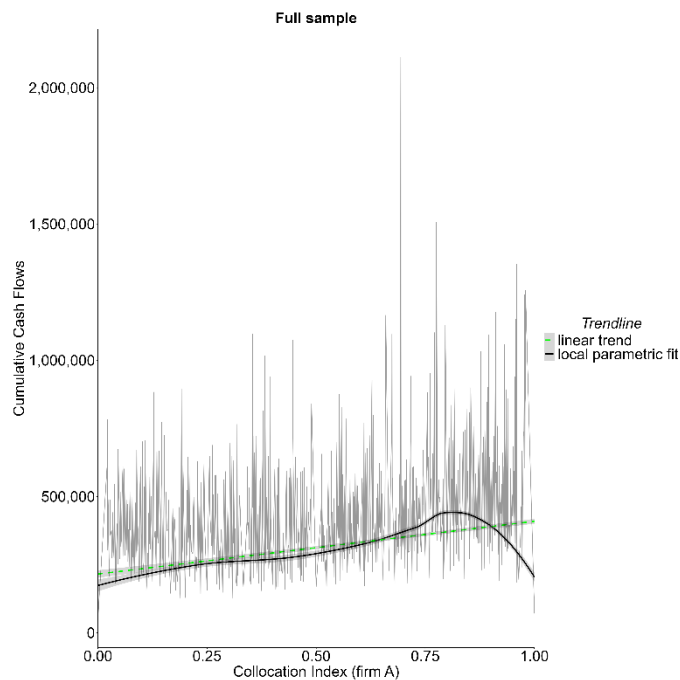
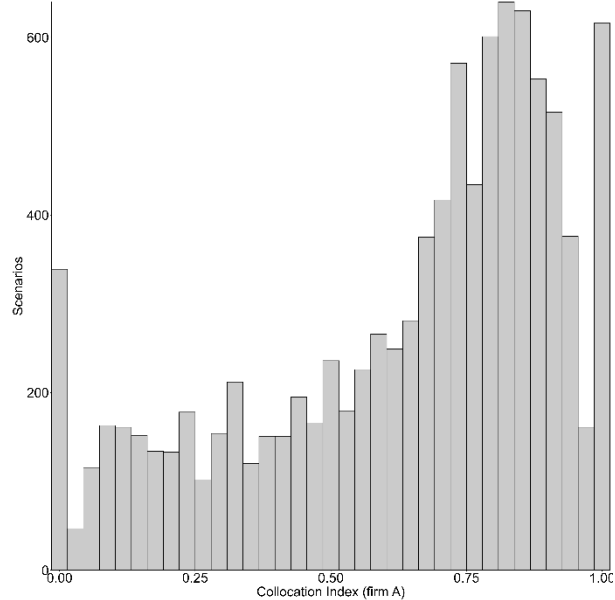


Figure VI.F.2 Sample comparison: All Scenarios vs. Complete Collocation

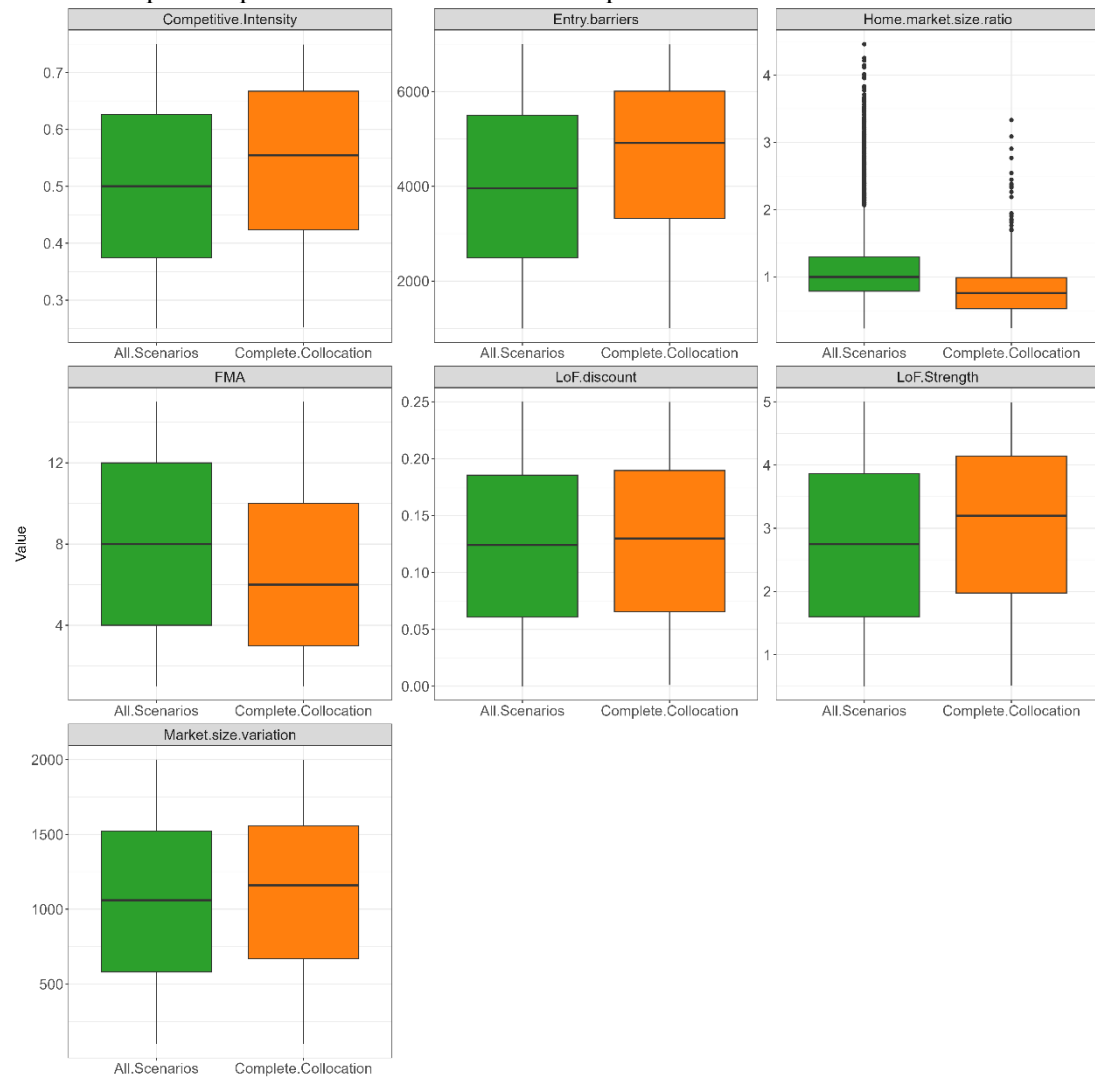


Figure VI.F.3 Sample Comparison: All Scenarios vs. Complete Avoidance

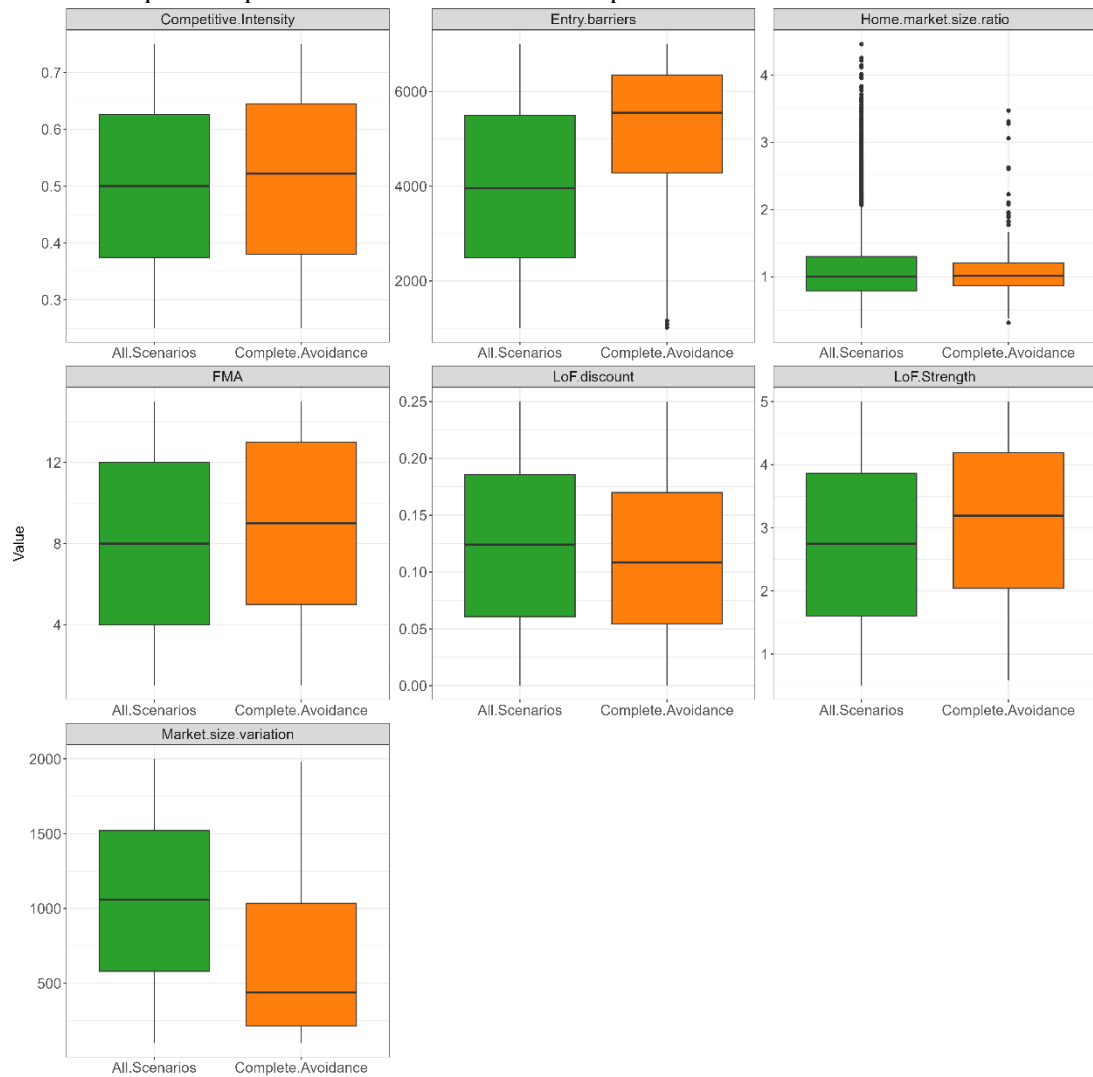


Figure VI.F.4 Competitive intensity, collocation, and performance differences

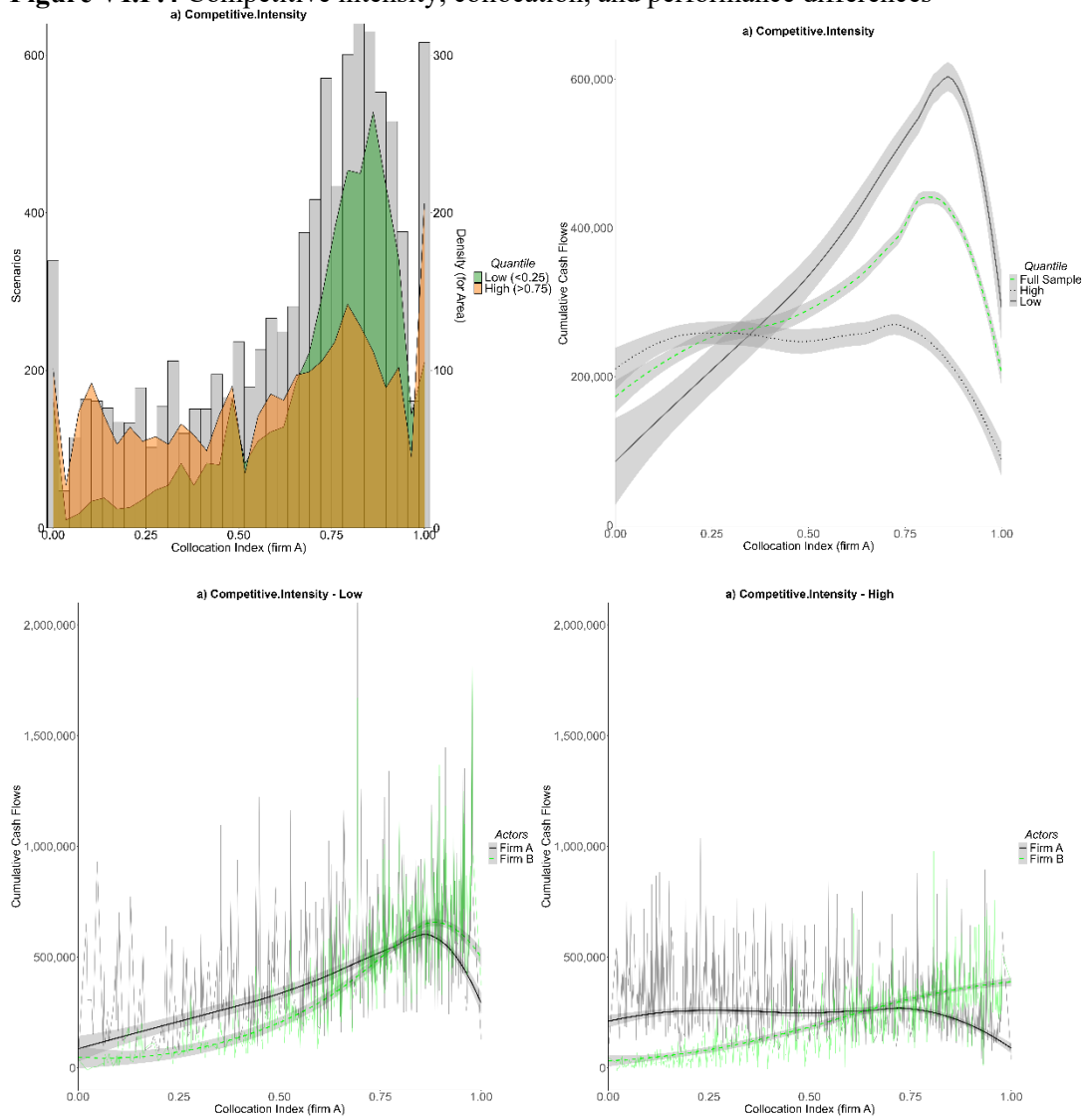


Figure VI.F.5 Entry barriers, collocation, and performance differences

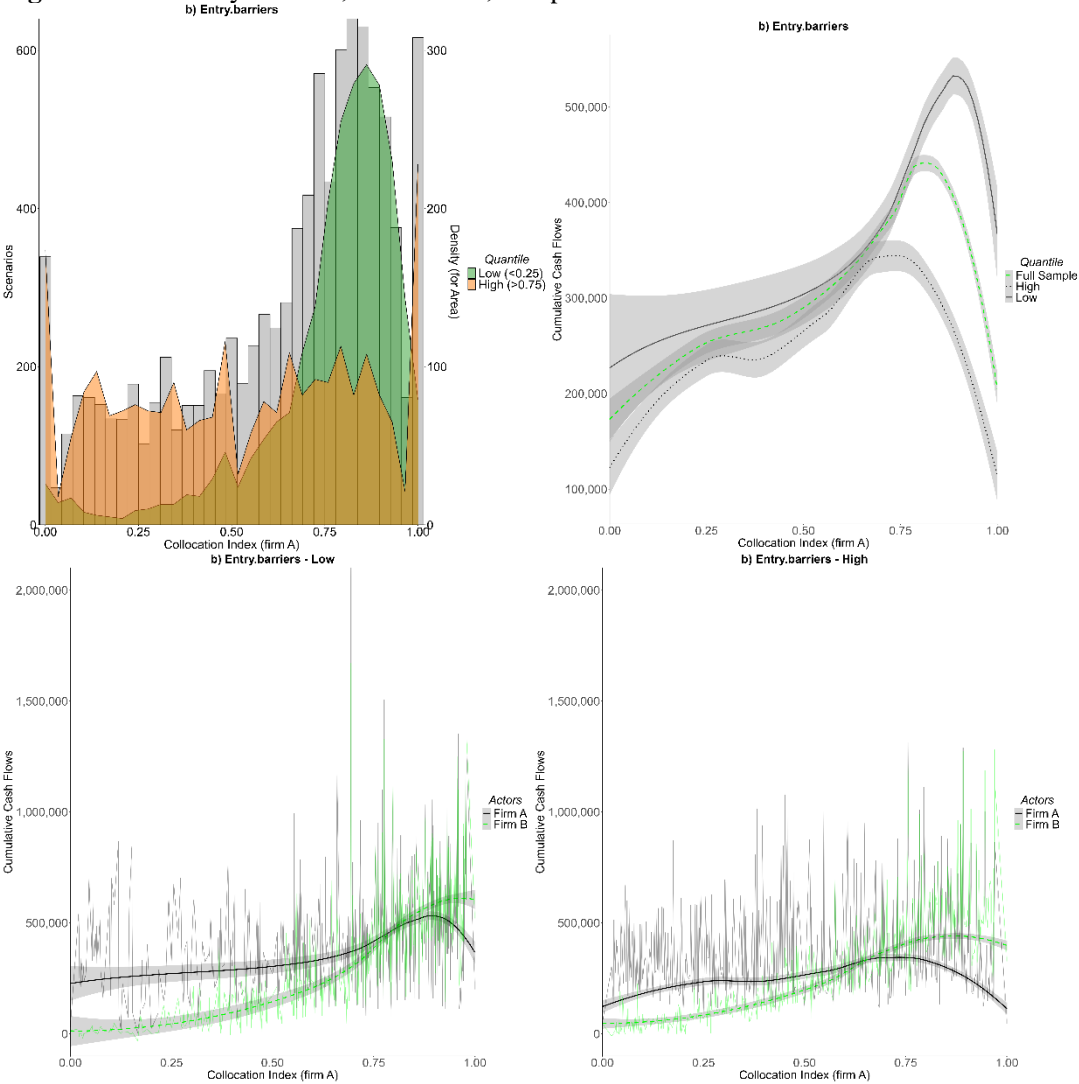


Figure VI.F.6 FMA, LoF strength, LoF discount, collocation, and performance differences

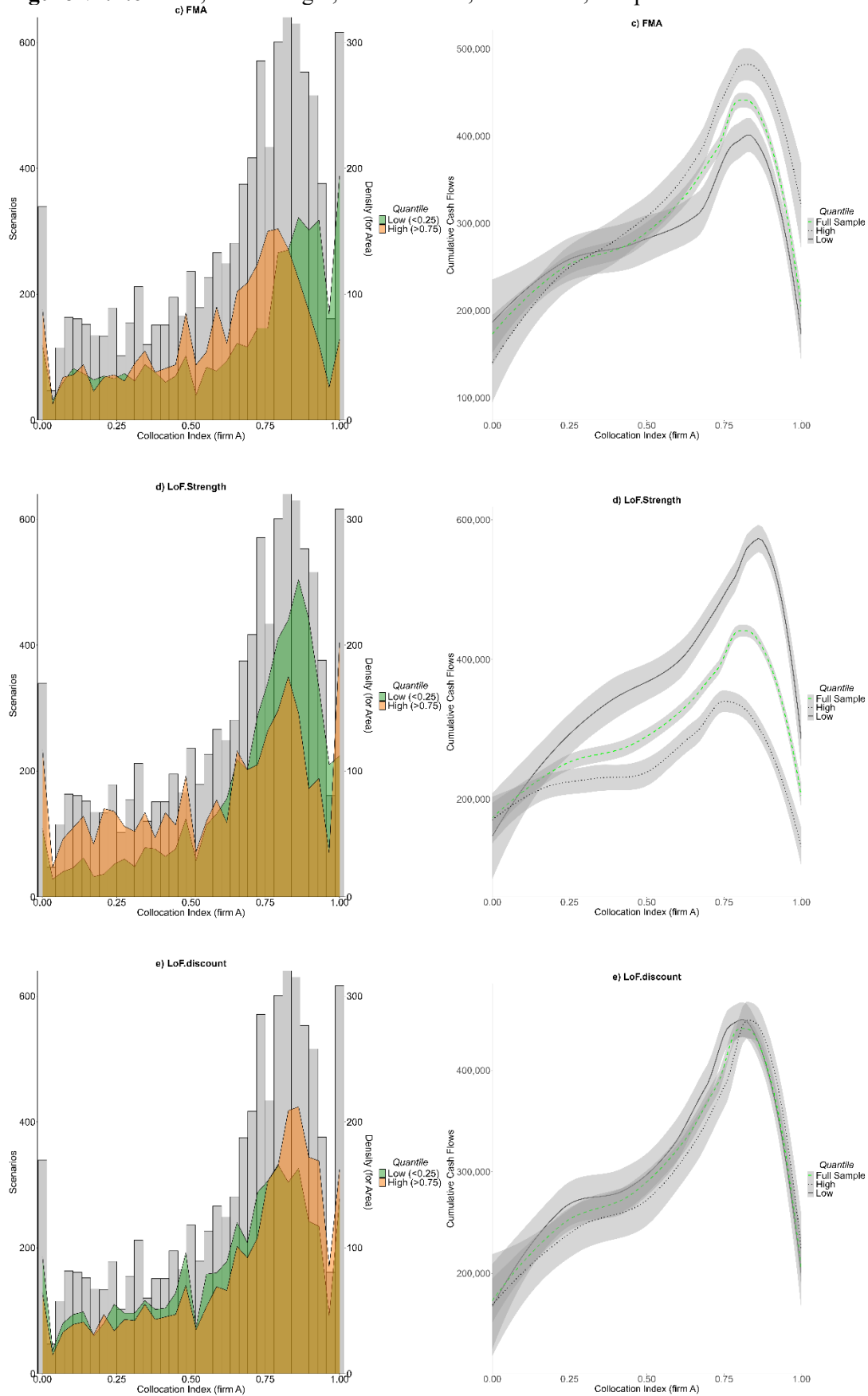


Figure VI.F.7 Market size variation, collocation, and performance differences

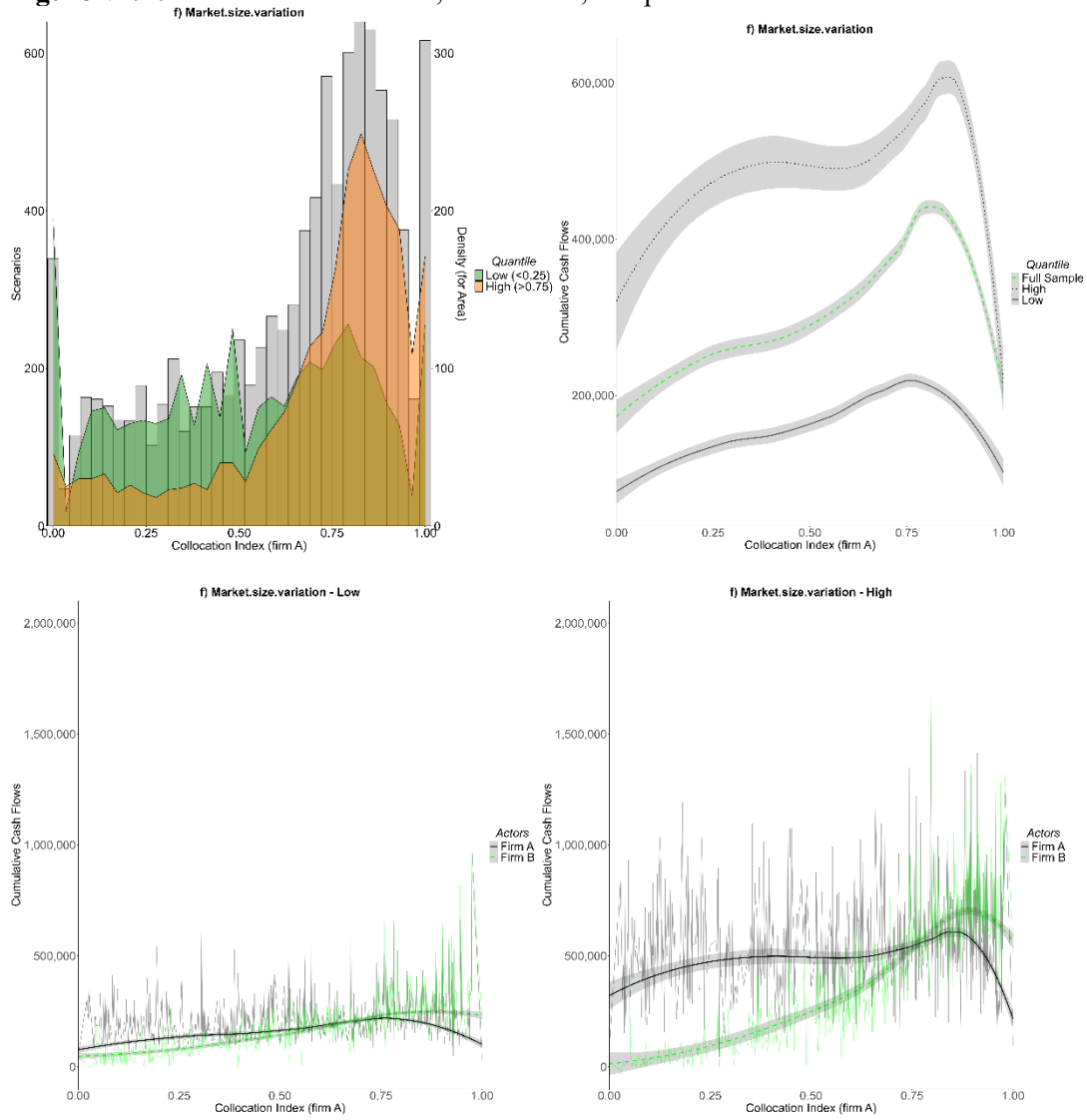
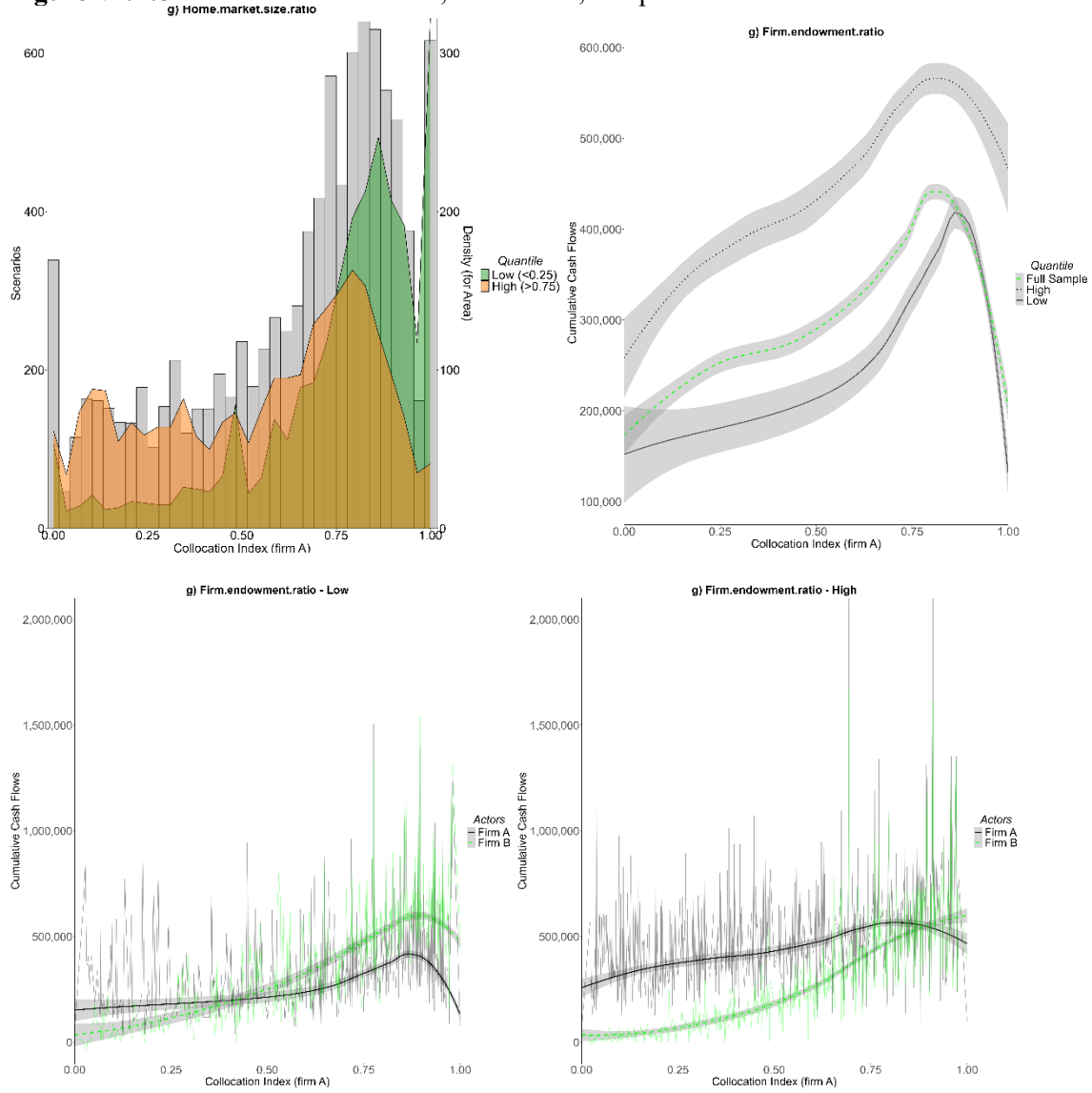


Figure VI.F.8 Home market size ratio, collocation, and performance differences



G. Persistence of FMA (a)

Previous studies have shown that the persistence of a FMA can be anywhere from relatively marginal for technologically dynamic products (Gomez et al., 2016), a couple of years, and up to 40 (or even 50 years) for highly regulated or entrenched industries (Huff & Robinson, 1994; Robinson et al. 1994; VanderWerf & Mahon, 1997). Given the study's focus on market-seeking industries, which emphasize branding and competitive positioning, the simulation parameter was chosen as a realistic middle ground between these two extremes, allowing for substantial variation to occur.

In the sensitivity analysis, we set <40 as the upper boundary for the persistence of FMA. In contrast to the main model, we can observe an effect of FMA on the simulation outcome. Particularly, a higher FMA increases avoidance. Without surprise, the higher the persistence of a FMA, the less attractive it is for a firm to follow a firm to a market. This also has an ensuing effect on competitive intensity, as higher competitive intensity paired with high FMA entices firms to avoid each other. Hence, complete avoidance is not only triggered by low market size variation and high entry barriers but also by high FMA and competitive intensity in this test. For the remainder of the outputs, we can observe structural equivalence. While increasing the possible persistence of FMA yields additional outputs, we would consider 40 years a rather unrealistic duration of FMA for market-seeking industries like consumer goods.

Overall, we can observe that FMA increases the significance of FMA and levels of avoidance in general. However, structurally, the main results from the paper hold and thus also the propositions stated. At the same time the results show that in industries with particular high FMA, we might see different outcomes, although we would see the level of FMA necessary to be unrealistically high (unless it is a protected or entrenched industry).

Figure VI.G.1 Collocation Index

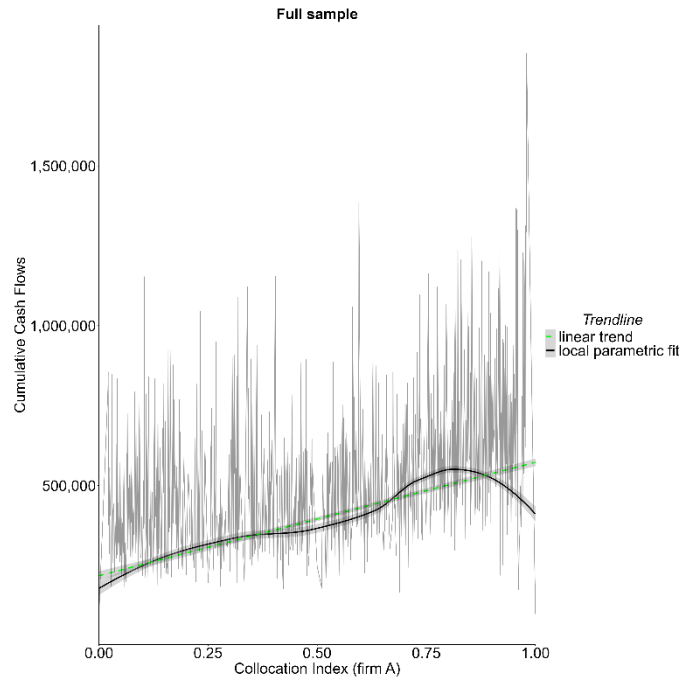
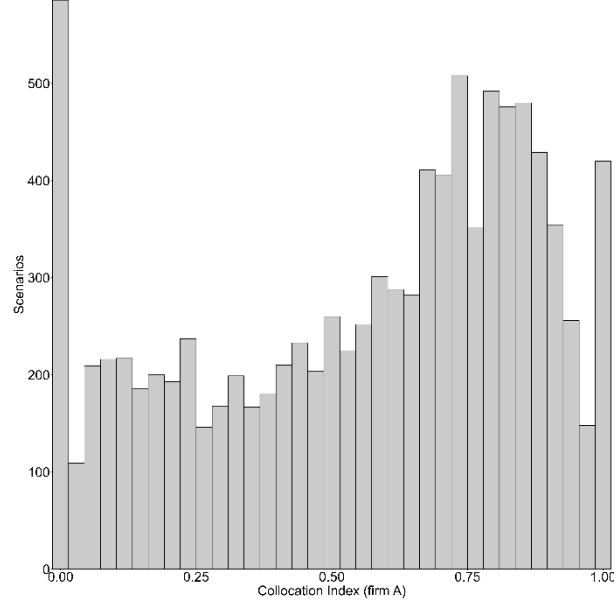


Figure VI.G.2 Sample comparison: All Scenarios vs. Complete Collocation

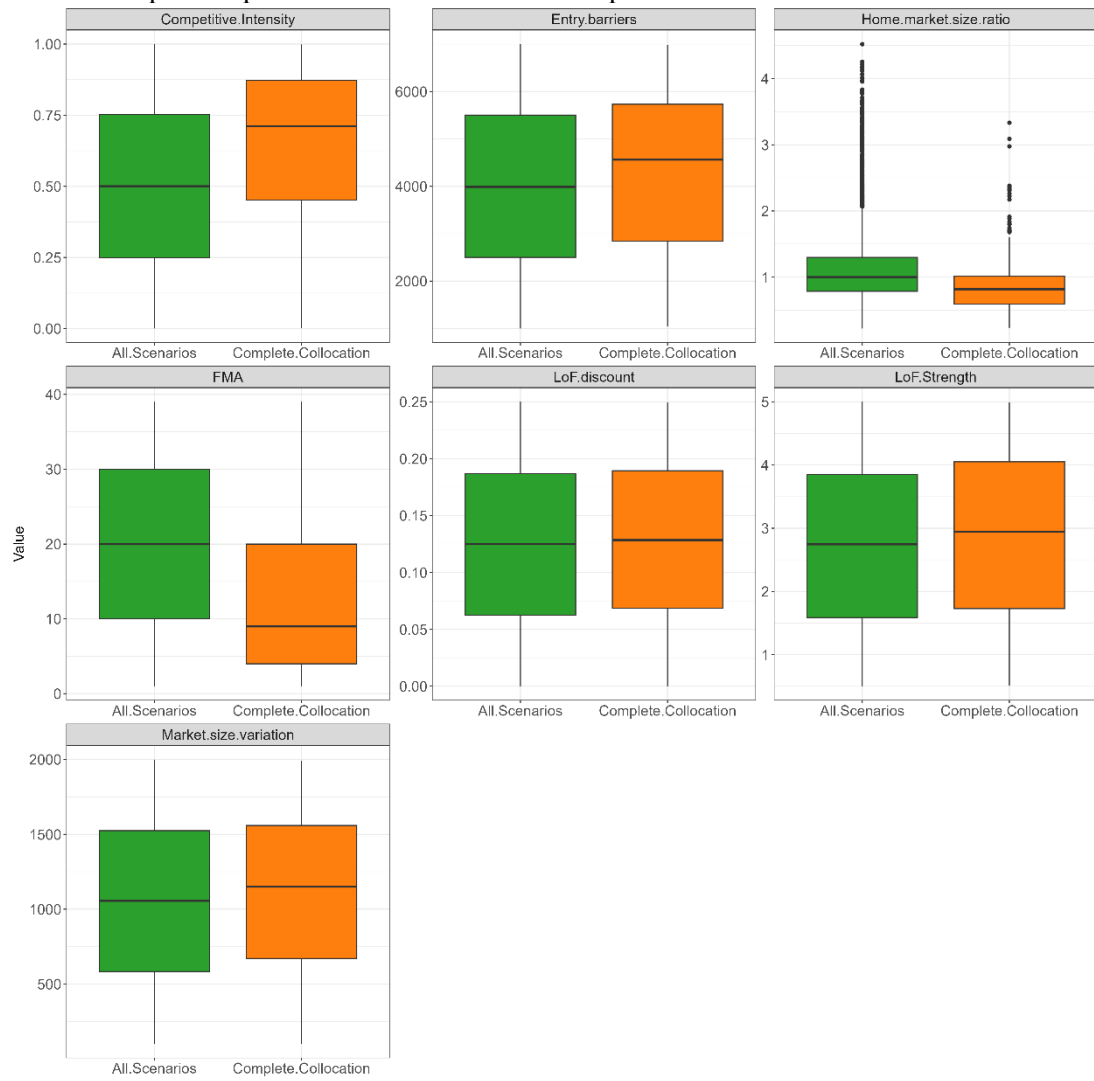


Figure VI.G.3 Sample Comparison: All Scenarios vs. Complete Avoidance

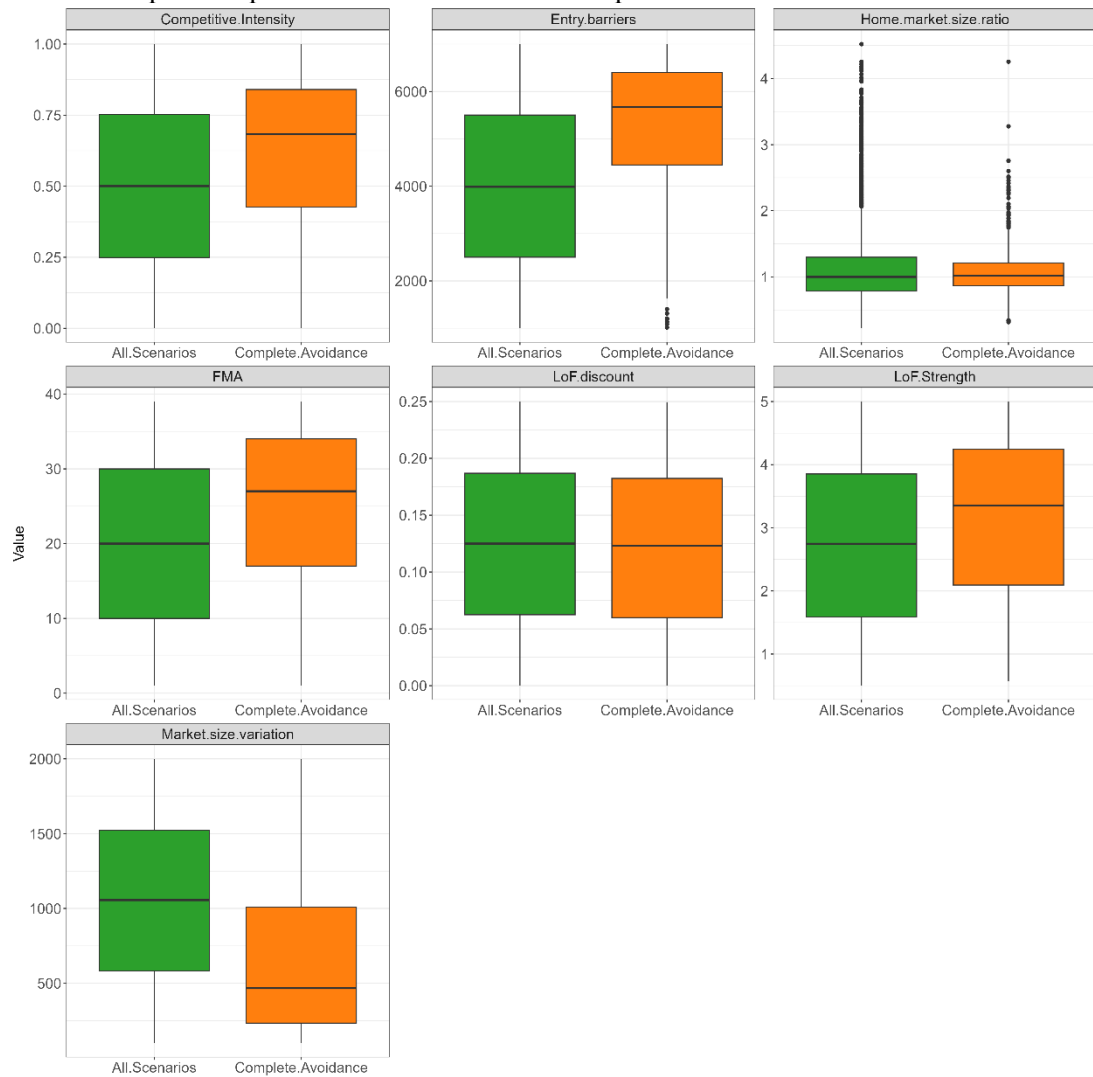


Figure VI.G.4 Competitive intensity, collocation, and performance differences

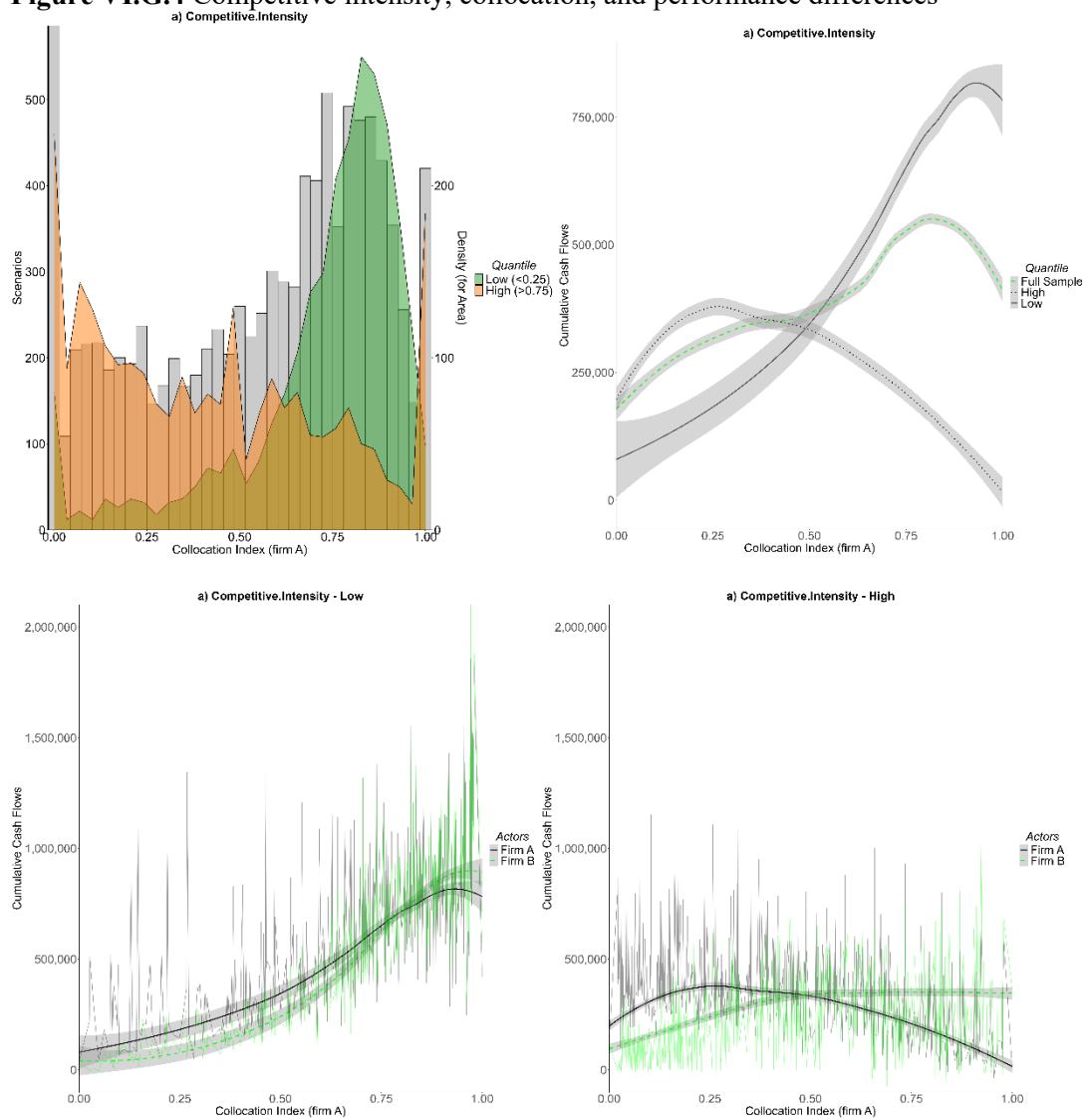


Figure VI.G.5 Entry barriers, collocation, and performance differences

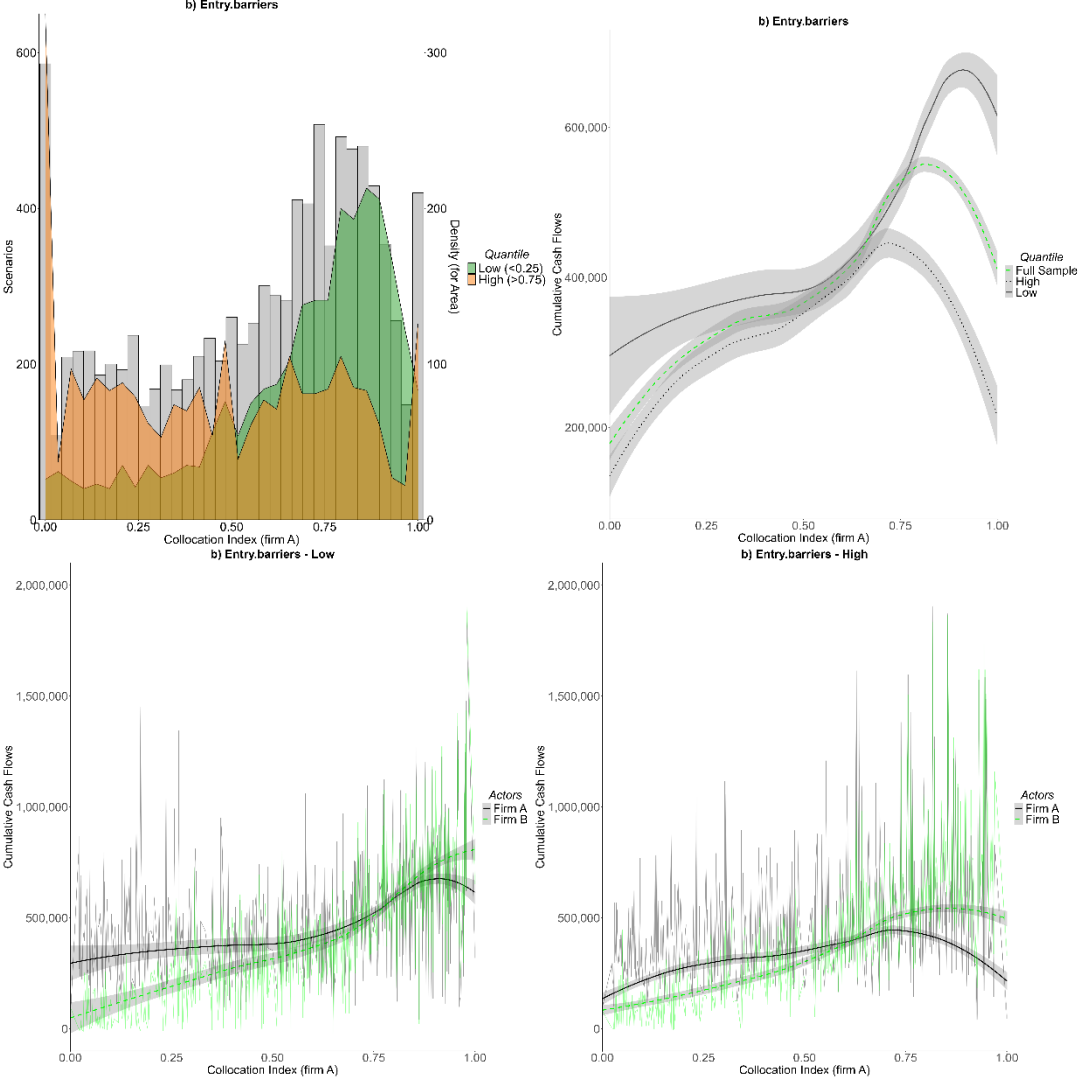


Figure VI.G.6 FMA, LoF strength, LoF discount, collocation, and performance differences

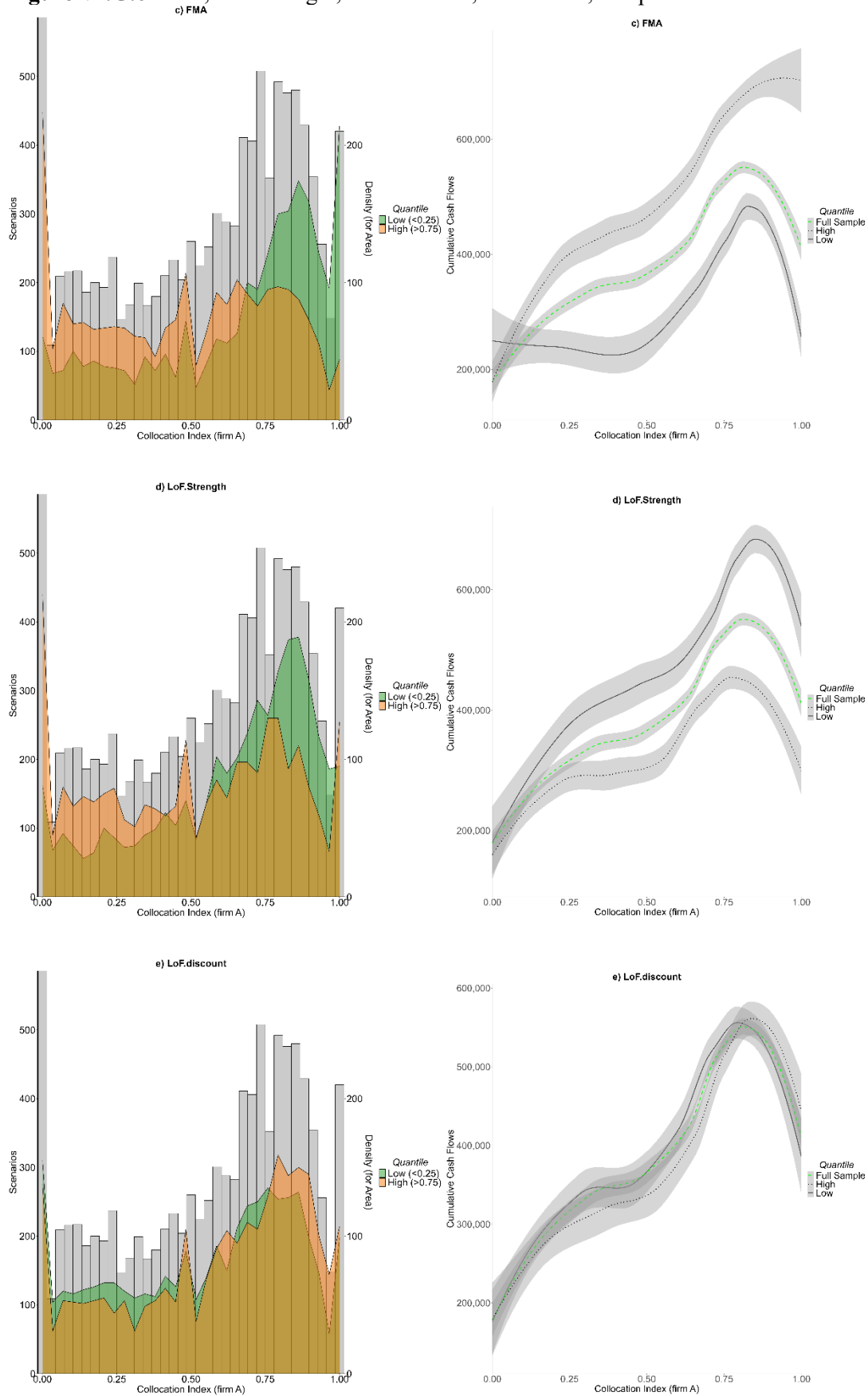


Figure VI.G.7 Market size variation, collocation, and performance differences

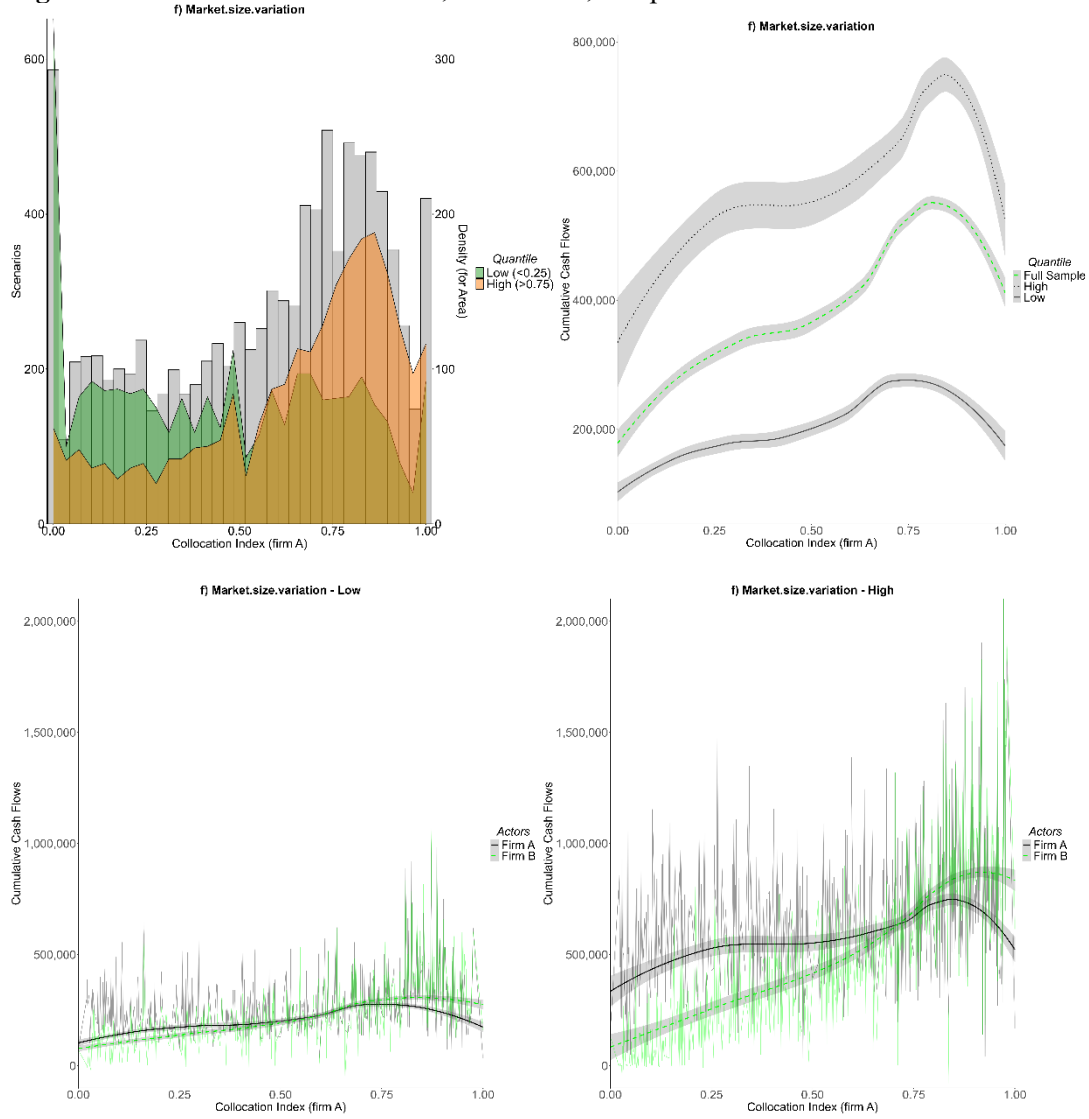
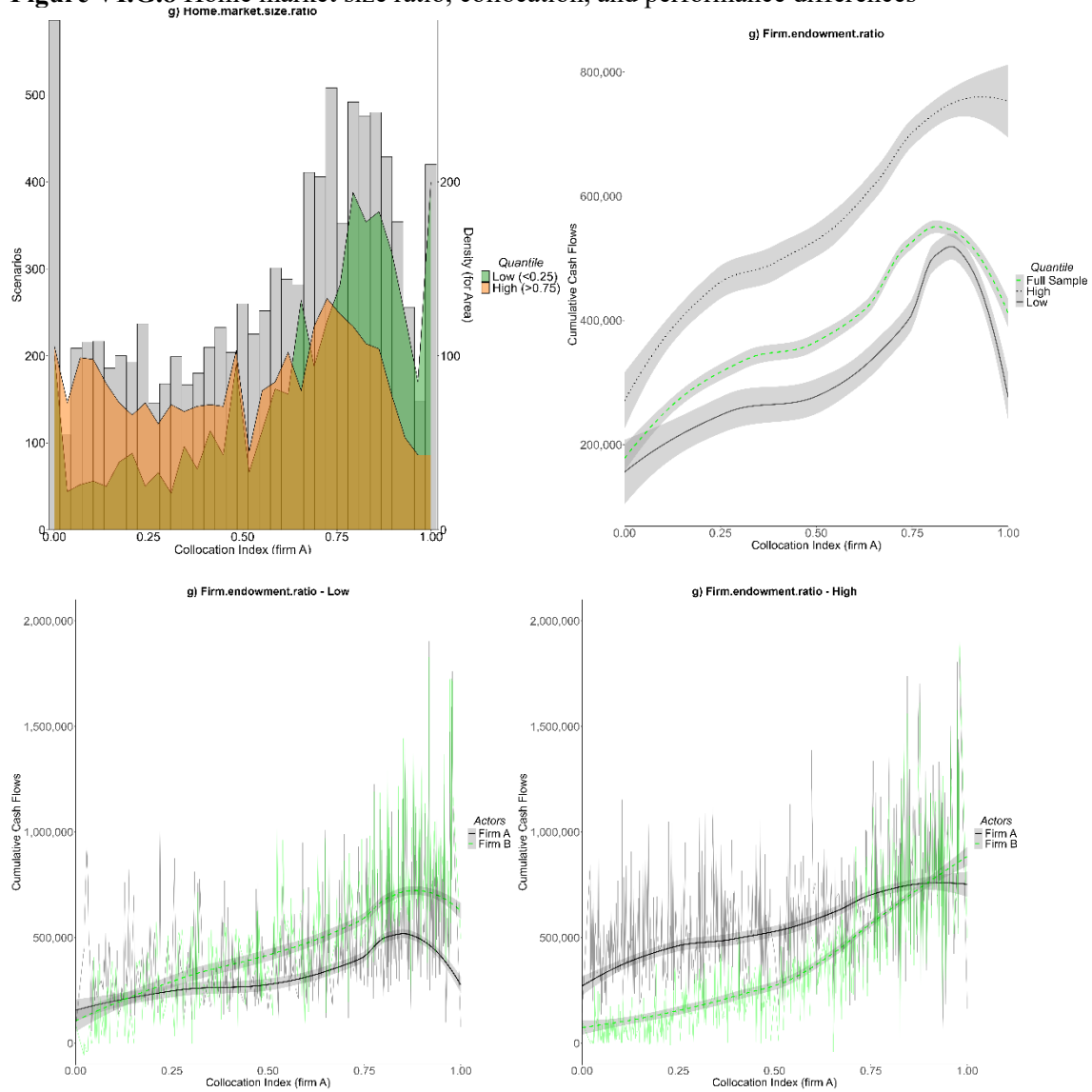


Figure VI.G.8 Home market size ratio, collocation, and performance differences



H. Entry Barriers (€)

The entry barriers are calibrated relative to market size variation via the payback period. For more details, refer to point A in this appendix (msv).

The effect of increasing the possible amount of entry barriers represents the contrast to msv. Higher entry barriers increase the likelihood of complete avoidance, as firms do not generate sufficient returns to enter many markets due to the high entry barriers. However, the remaining outputs are largely structurally equivalent to the main simulation outputs and thus also the propositions stated.

Figure VI.H.1 Collocation Index

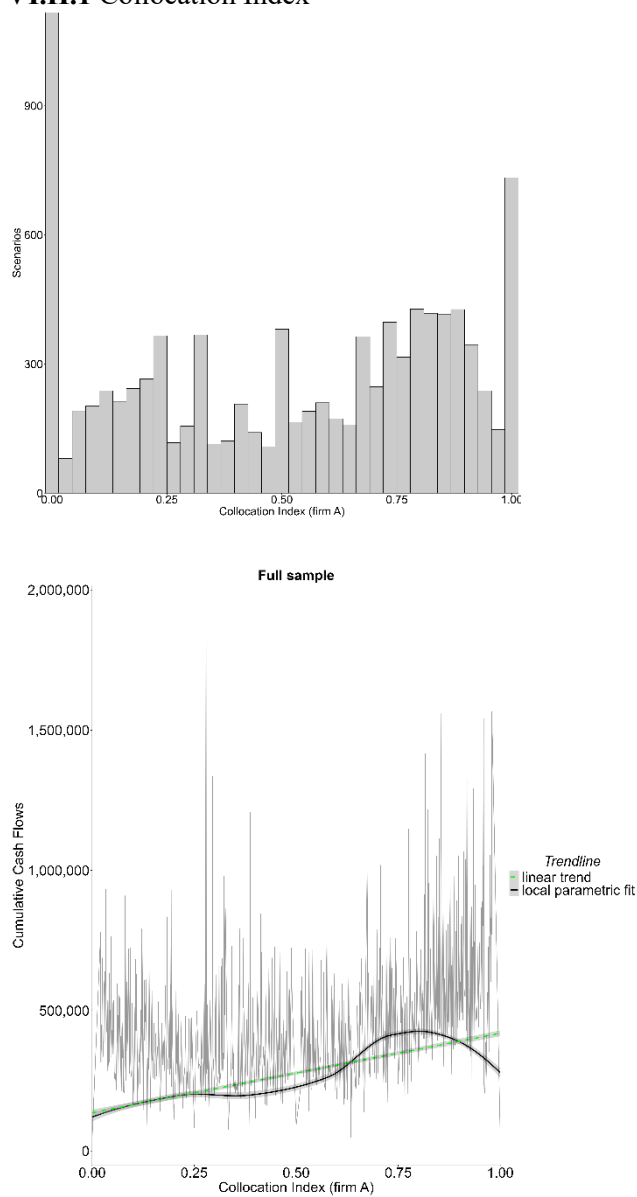


Figure VI.H.2 Sample comparison: All Scenarios vs. Complete Collocation

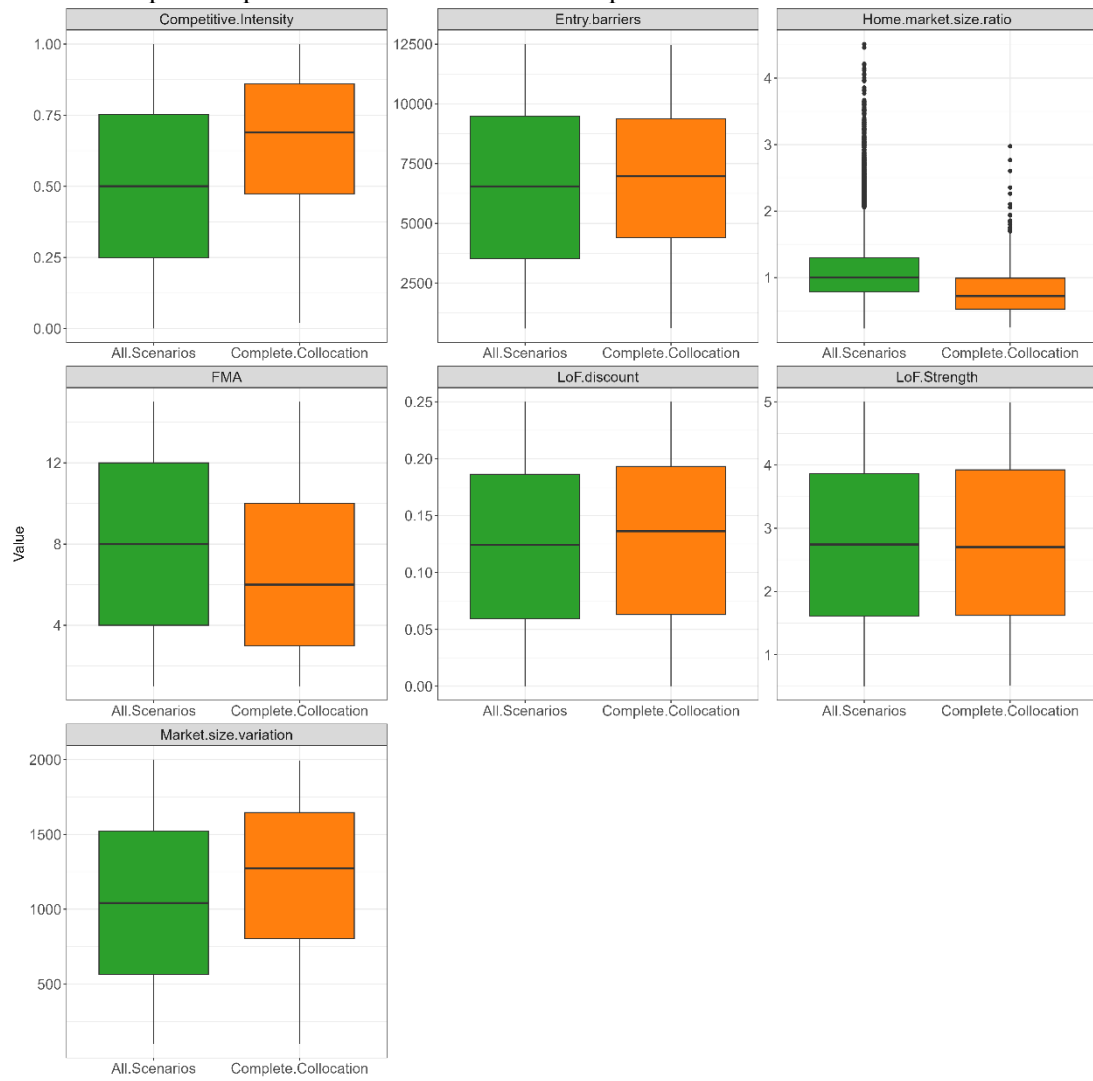


Figure VI.H.3 Sample Comparison: All Scenarios vs. Complete Avoidance

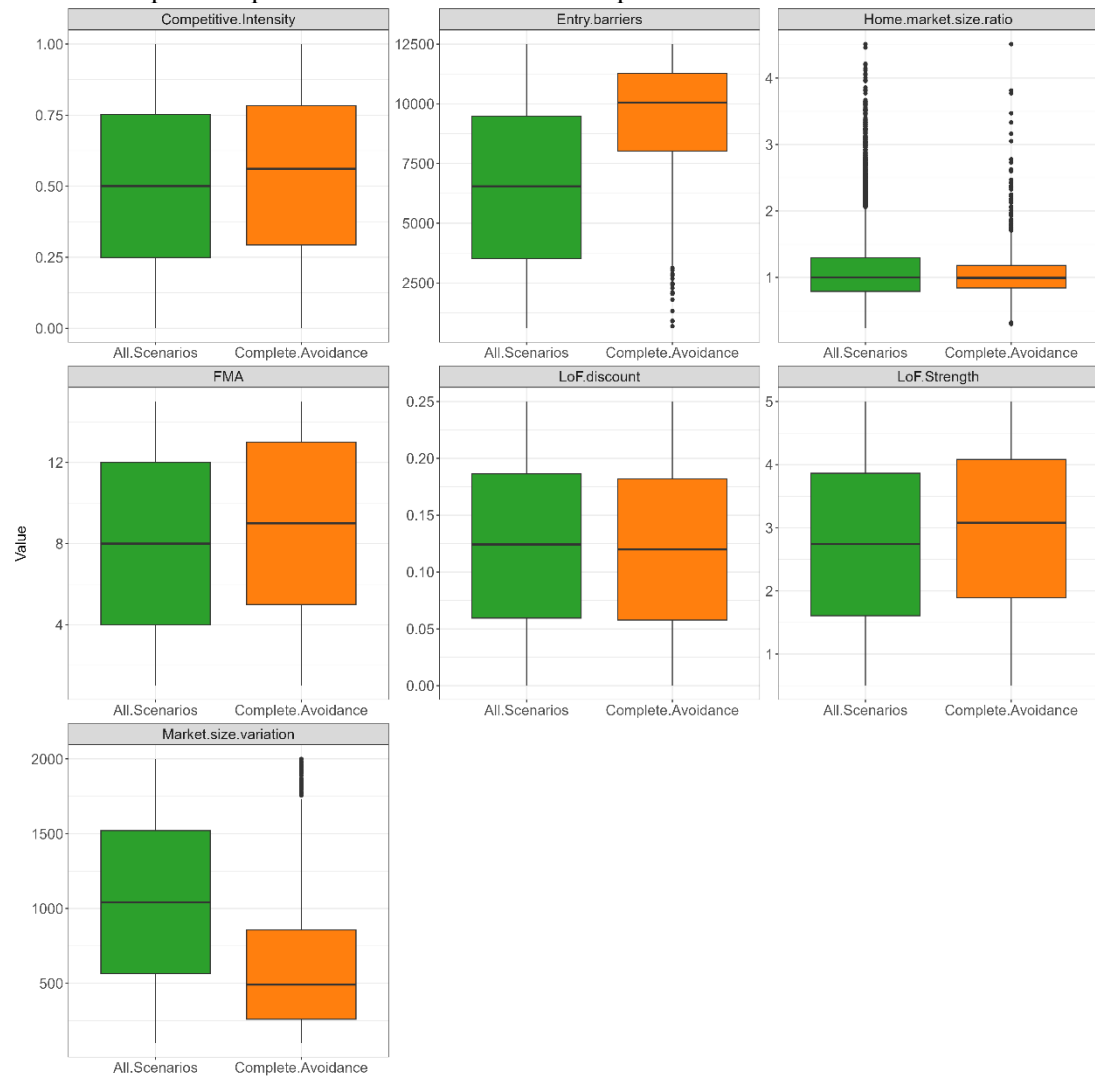


Figure VI.H.4 Competitive intensity, collocation, and performance differences

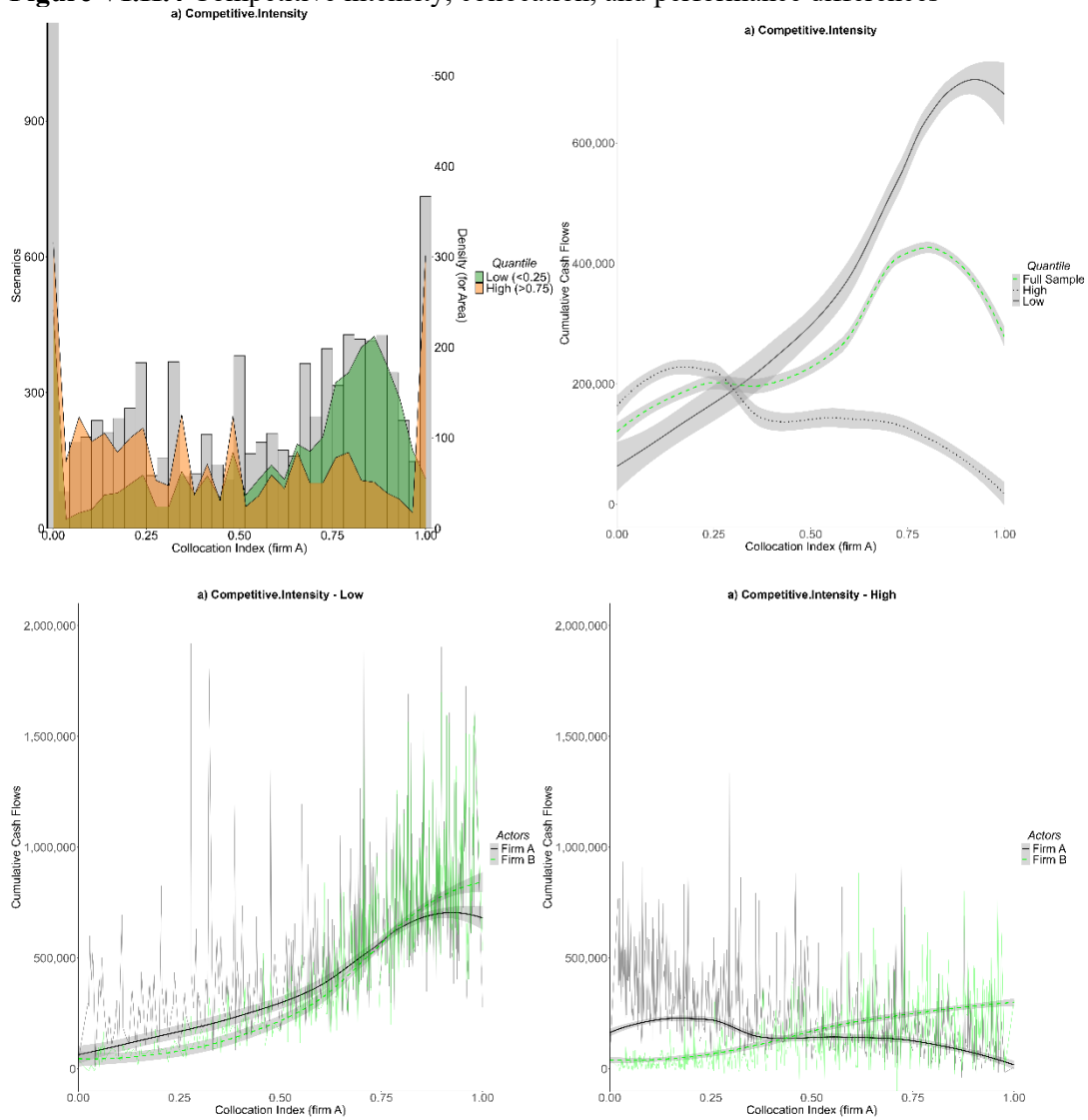


Figure VI.H.5 Entry barriers, collocation, and performance differences

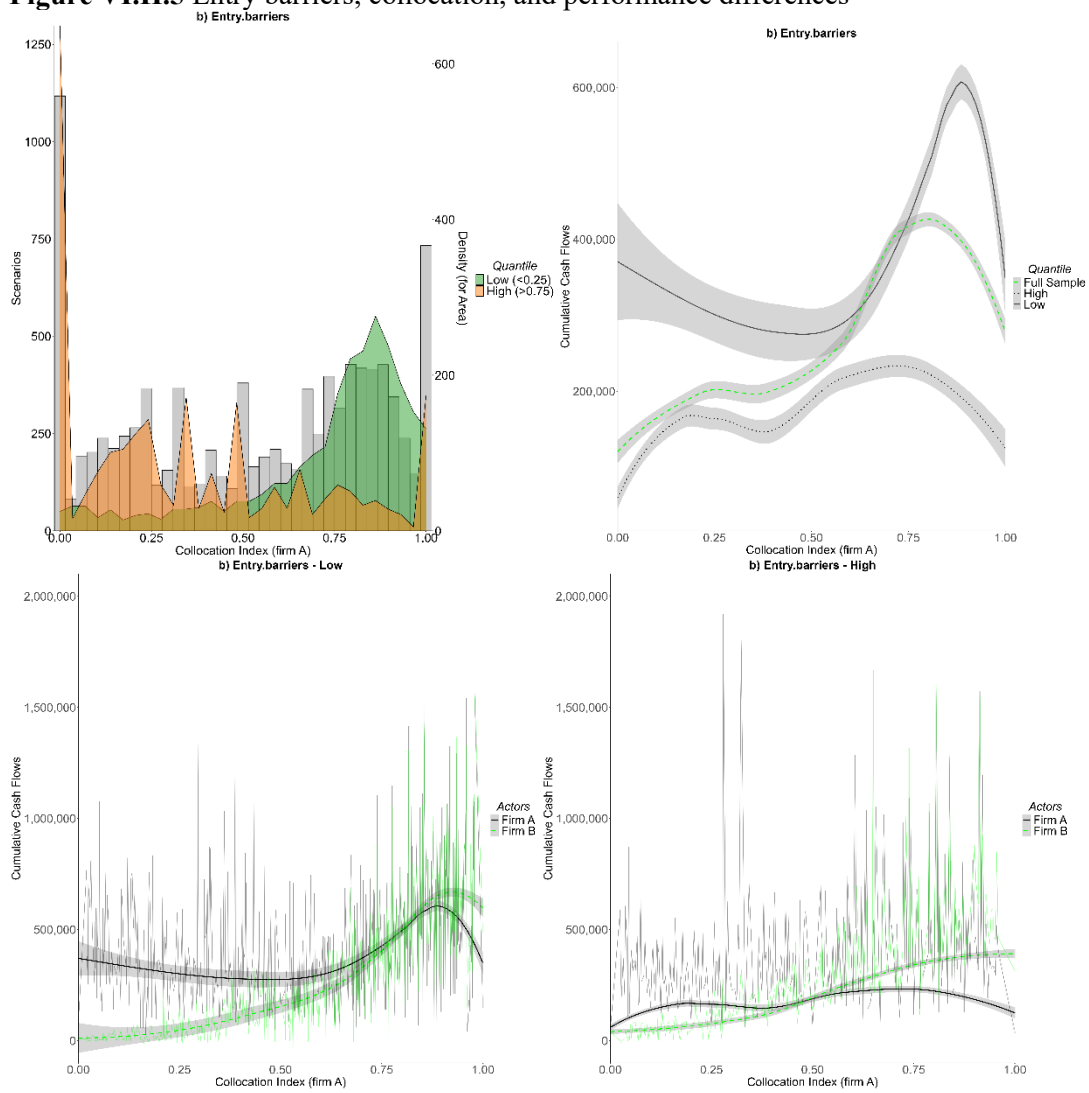


Figure VI.H.6 FMA, LoF strength, LoF discount, collocation, and performance differences

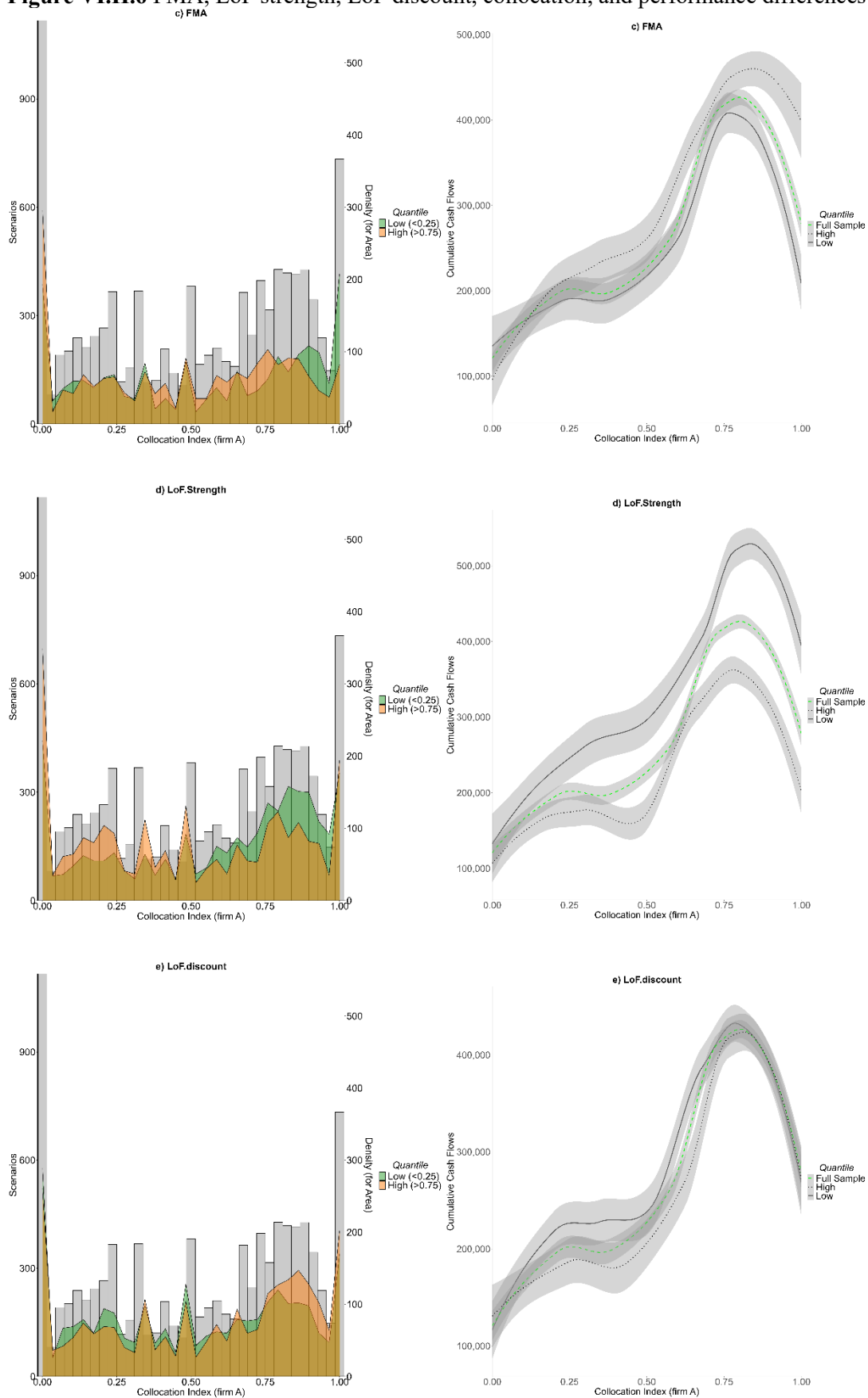


Figure VI.H.7 Market size variation, collocation, and performance differences

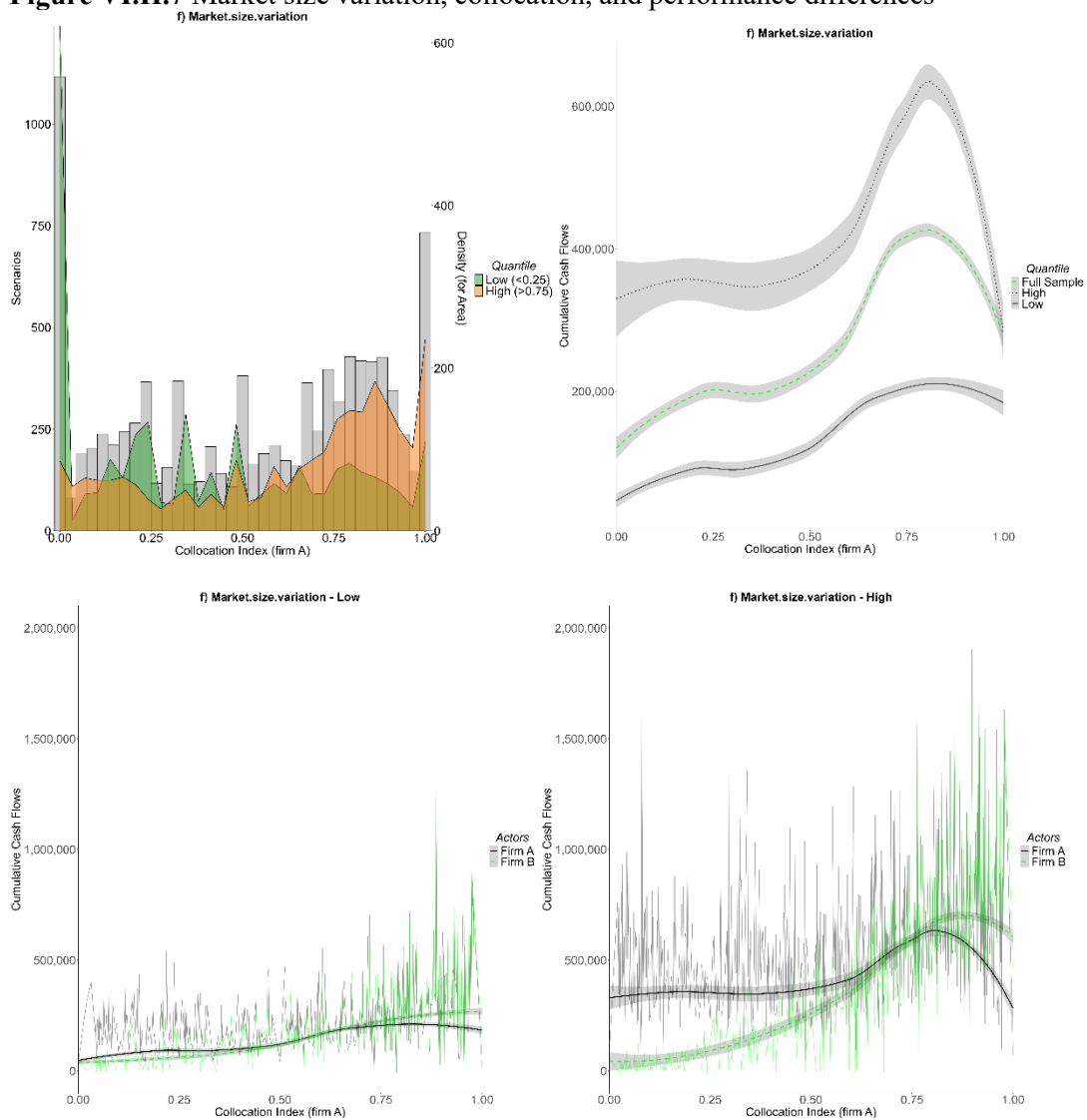
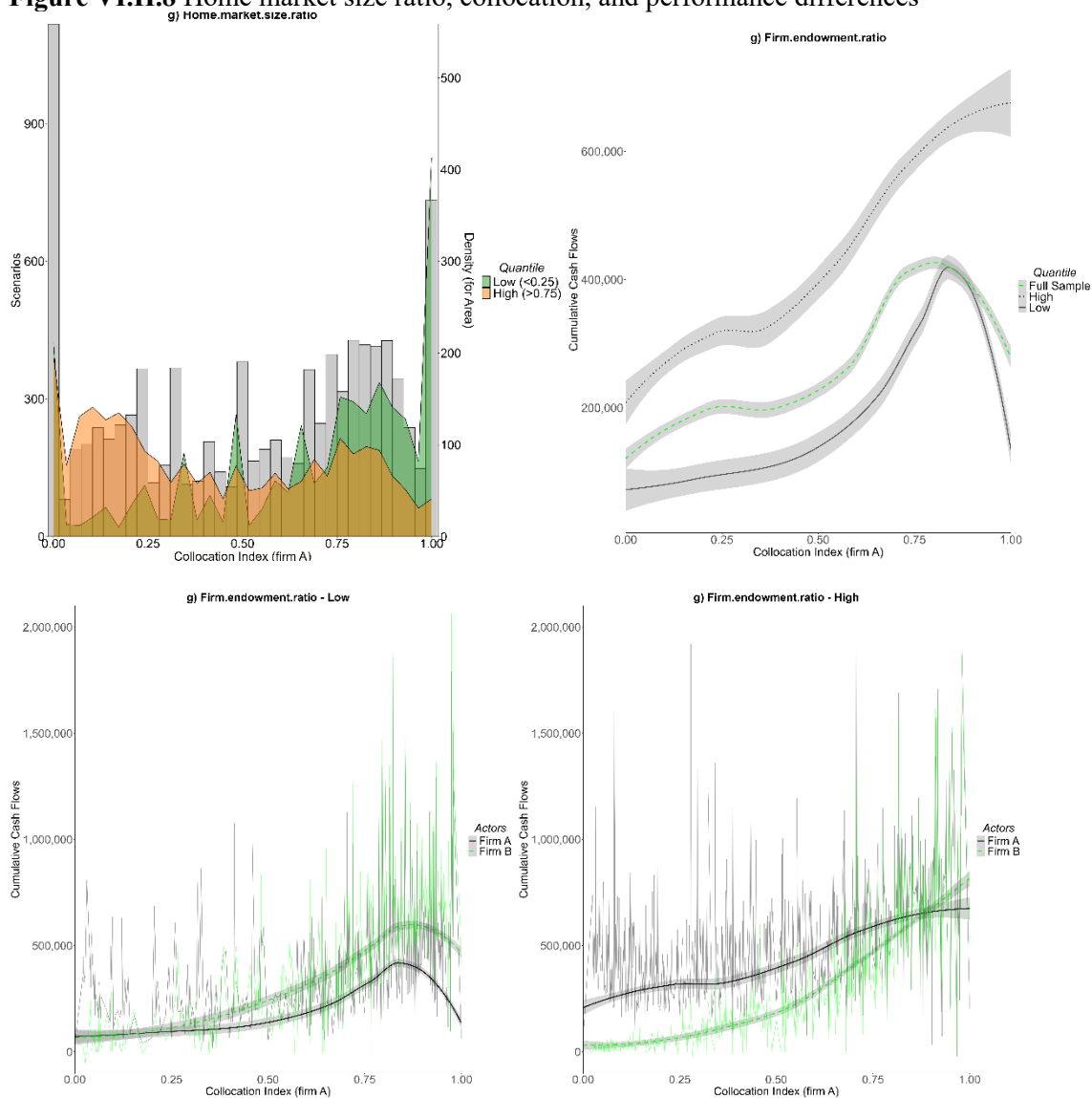


Figure VI.H.8 Home market size ratio, collocation, and performance differences



Appendix VII: Summary and comparison of findings

Table VII.1 Overview of findings and propositions

	<i>Empirical studies</i>	<i>Formal models</i>	<i>Simulation using DRL</i>
Overall levels of collocation	Various studies provide evidence for the occurrence of collocation in various industries and contexts (e.g., Flowers, 1976; Knickerbocker, 1973; Ito & Rose, 2002; Rose & Ito, 2008; Yang et al., 2023; Yu & Ito, 1988) Studies mostly investigate single outcomes (Yao et al., 2024): level of market concentration (e.g. Flowers, 1976; Knickerbocker, 1973), probability of investing into or following a firm to a country (e.g., Gimeno et al., 2005; Li & Guisinger, 1992; Rose & Ito, 2008; Yang et al., 2023; Yu & Ito, 1988)	Overall avoidance most common (Alcácer et al., 2013, 2015)	Simulation suggests mid-high collocation levels, complete collocation, and avoidance as most frequent outcomes Higher levels of collocation best for performance Proposition 1a <i>Without the consideration of environmental contingencies, high degrees of collocation are the performance-optimal outcome on average.</i> Proposition 1b <i>Without the consideration of environmental contingencies, low levels of collocation (but not avoidance) are rarely the performance-optimal outcome on average.</i> Simulation suggests a different regime in terms of contextual factors that underpins the distinction between high levels of collocation and complete collocation Proposition 1c <i>Complete collocation represents a distinct theoretical outcome from mid-to-high levels of collocation.</i>
Complete Collocation	Instances of complete collocation rarely studied; various studies found evidence of tendencies for collocation in different contexts (e.g., Ito & Rose, 2002; Knickerbocker, 1973; Rose & Ito, 2008; Yang et al., 2023; Yu & Ito, 1988)	Complete collocation as a consequence of high market size variation (Alcácer et al. 2013, 2015)	Complete collocation (follow-the-leader mechanism) as a consequence of high competitive intensity (limit rival's growth possibilities), originating in a smaller home market (leader initiates market entry moves), and low FMA (follower can follow and take up leader's market share) Simulation suggests market entry decisions are overly driven by market size before blending in competitive considerations Proposition 2a <i>Complete collocation emerges under conditions of high competitive intensity, lower home-market size, and a low first-mover advantage.</i>

Complete Avoidance	Instances of complete avoidance rarely studied; studies found evidence of tendencies for mutual avoidance or less collocation (e.g., Dike & Rose, 2019; Rose & Ito, 2008; Yao & Li, 2016; Yu & Ito, 1988)	Complete avoidance as a consequence of low market size variation (Alcácer et al., 2013, 2015)	Complete avoidance as a consequence of low market size variation (firms can avoid direct competition) AND high entry barriers (limiting the number of overall market entries), together with LoF (enter adjacent markets) Simulation suggests market entry decisions are driven by market attractiveness across the whole internationalization process Proposition 2b <i>Complete avoidance emerges under high entry barriers, little variance in market size, and high liabilities of foreignness.</i>
Competitive intensity	Mostly continuous effect of competitive intensity on collocation. High (low) competition or low (high) concentration reduces (increases) collocation behavior (e.g., Gimeno et al., 2005; Yu & Ito, 1988). Few studies note declining collocation behavior above certain industry concentration (Flowers, 1976; Knickerbocker, 1973; Rose & Ito, 2008) Low competitive intensity due to mutual forbearance (collocating with the same rival across different countries) increases collocation and performance (e.g., Ghemawat & Thomas, 2008; Kim et al., 2024; Yu & Cannella, 2013)	Implicit assumption of a continuous effect of competitive intensity on collocation Firms avoid each other due to high competitive intensity and thus lower profits (Alcácer et al., 2015)	Simulation suggests non-linear effect of competitive intensity on collocation: Low competitive intensity leads to high (but not complete) collocation. High competitive intensity causes many cases of complete collocation Proposition 3a <i>Low competitive intensity leads to high degrees of collocation but not complete collocation.</i>
Entry barriers	High entry barriers increase avoidance: Firms in the African telecommunications industry avoid head-on competition (Dike & Rose, 2019)	-	Low entry barriers reduce possibility of avoidance strategy and increase collocation. → Inability of avoidance strategy Low entry barriers suggest a mix of market attractiveness and competitive positioning; high entry barriers favor market size factors in decision-making Proposition 3b <i>Low entry barriers reduce the probability that firms can follow an avoidance strategy, despite being a relatively attractive option.</i>
Market size variation	Studies underscore the importance of competitive and market characteristics in market entry decisions (e.g., Ito & Rose, 2002; Li & Guisinger, 1992; Yu & Ito, 1988)	Low (high) market size variation leads to more avoidance (collocation) (Alcácer et al. 2013, 2015)	Low market size variation reduces levels of collocation but does not result in avoidance per se. → Market-seeking (size) effect

**Home market
size ratio**

Resource dissimilarity reduces competitive moves (Baum & Korn, 1999; Young et al., 2000), smaller firms exhibit less collocation behavior (e.g., Gimeno et al., 2005), and larger firms tend to follow smaller firms (Rose & Ito, 2008)

Different results depending on differences in firm size (Alcácer et al., 2013, 2015) or home country size (Graham, 1998)

Proposition 3c *Low variance in market size leads to low degrees of collocation on average.*

Simulation suggests that firms originating in smaller home markets follow leaders and collocates in a strategic attempt to limit leaders' growth potential. → Follow-the-leader effect

Low market size variation favors market attractiveness factors in initial phases before considering competitors' positioning later; high market size variation blends competitive and market attractiveness factors equally throughout internationalization

Proposition 3d *Firms from smaller home markets benefit more from collocating with the rival compared to firms from bigger home markets.*

Appendix VIII: Dictionary & Terms

(Convolutional) Neural Network

A neural network consists of extensively interconnected units (neurons) that create a vast network for pattern recognition. A neural network typically consists of input neurons, output neurons, and one or more layers of hidden neurons. In our case, the input neurons are country and firm characteristics, which are part of our firm world. The hidden layers of neurons connecting input with output neurons are not ‘observed’, but compress input features into learned traits (e.g., easy to enter). A neural network typically consists of two hidden layers of neurons.

For our study, we use a Convolutional Neural Network (CNN) with an actor-critic architecture. A CNN is a type of neural network that learns patterns by moving small filters across the data. It is thus very efficient and strong in recognizing patterns in different places. The first layer usually learns simple patterns, while the latter layer converts them into more complex signals. An actor-critic architecture means that the CNN provides two sorts of output layers: the policy for an actor in a specific state, which is essentially a probability distribution for actions (for 50 possible countries plus the option of inaction); and, a critic value that evaluates the policy by estimating the expected cumulative returns from a specific state onwards.

Training the CNN then implies adjusting the parameters within the CNN so that the firm makes better choices. The reference for training and thus optimizing the CNN comes from the value loss, which is a function of the critic value and the comparative rewards based on the agent’s policy (i.e., actions). To train the CNN, we need a reinforcement learning algorithm and a numerical optimizer. We use a proximal policy optimization-based (PPO) algorithm (Schulman, Wolski, Dhariwal, Radford & Klimov, 2017) and an ADAM optimizer (Kingma & Ba, 2014), which have become a standard repertoire for reinforcement learning applications.

PPO

PPO stands for Proximal Policy Optimization, which is an algorithm used in reinforcement learning. To implement reinforcement learning (RL) with a neural network, you need an algorithm that

updates the policies (market entry decisions) of the neural network. In other words, you want to increase the probabilities of actions that resulted in beneficial outcomes and reduce the probabilities of bad actions. PPO belongs to the family of policy gradient methods, which directly learn a policy function (instead of a value function), and has become the standard for RL due to its implementation simplicity (computationally), its stability (in terms of parameter tuning), and its sample complexity (requiring less data to train policies) (Schulman et al., 2017). The appeal of PPO lies in the clipped surrogate objective that prevents large, destructive policy updates (common for simple policy gradient methods), while still allowing for the exploration of new policies (via an entropy bonus) (Schulman et al., 2017). PPO is also suitable in situations where the neural network architecture shares parameters between the policy and value function, as in our case with the actor-critic style CNN. Ultimately, PPO provides the loss function to minimize, indicating what to optimize.

ADAM

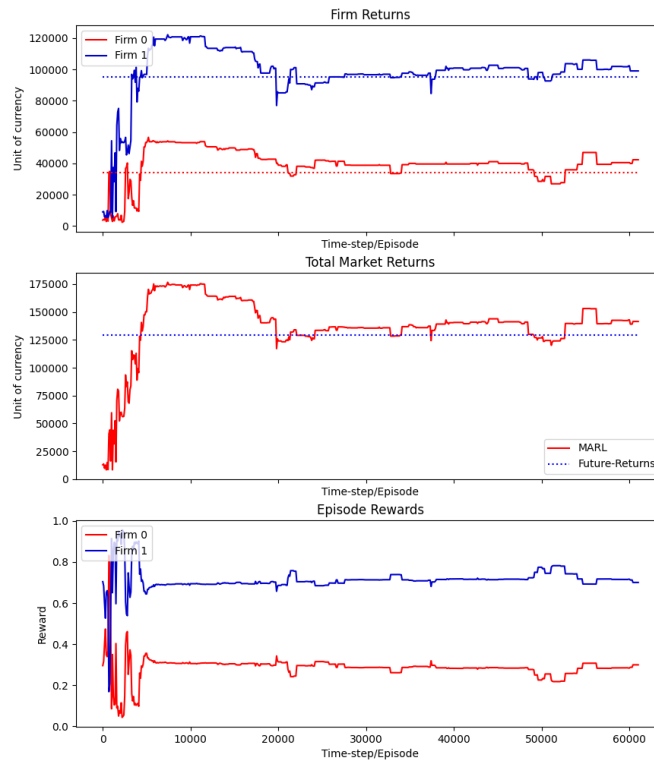
ADAM stands for ‘Adaptive Moment Estimation’ and is a numerical optimizer that facilitates adjustments to the weights and biases in neural networks based on the PPO loss function. In simple words, ADAM takes the signal provided by the PPO and updates the parameters of the CNN to optimize it. The gradients provided by backprop with respect to the loss function can be quite noisy. ADAM facilitates the convergence to an optimal performance outcome by making adaptive per-parameter steps. It dynamically adapts the learning rate, making more extensive updates for infrequent and vital parameters and minor updates for frequent and less important parameters. This adaptive adjustment of learning rates helps achieve faster and more effective training of the firms. In simple terms, ADAM determines how much the parameters of the CNN need to be updated to improve and minimize the loss function. The clip function in the PPO prevents ADAM from making overly large updates that deviate from the current behavior policy.

Appendix IX: Illustrative example – Reinforcement Learning

Process

Figure IX.1 illustrates the high-level reinforcement learning process for a single seed. The x-axis on all three panels of Figure IX.1 depicts the 61,000 episodes of full simulation training runs used for the DRL of each firm. The first panel of Figure IX.1 shows the cumulative cash flows per firm at the end of each episode, on the y-axis, in red and blue. The second panel shows the total cash flow in the market generated by the two firms. The third panel's y-axis shows a measure of relative reward (the input for the advantage estimates) – that is, red's cumulated cash flows at the end of each episode divided by total market cash flows and vice versa. Figure IX.1 clearly shows that the firms try varying combinations of country footprints with varying effects on (relative) performance. However, the firms gradually converge towards an equilibrium where neither player's competitive move can tilt their (relative) returns in their respective favor.

Figure IX.1 Illustrative reinforcement learning process



The dotted line for future returns is the fixed future-market-size benchmark used in the evaluation step.

Appendix X: Simulation – Replication package description

The simulation was implemented in Python. The paper comes with a replication package to implement the simulation. The key components and (key scripts) are the firm-world environment (`firm_world.py`) described in the main body of the paper, a neural network (`ppo_cnn_agent`), a DRL-execution (`clean_ppo_selfplay.py`) and a master script (`train.py`) that executes the simulation. In addition, the replication package contains a few helper scripts that contain utility functions (`selfplay_utils.py`), compute the comparison-baseline (`fixed_profit_max.py`), and merge the results of all seeds into one file (`merge_csv.py`)

The first step in the master script (`train.py`) is developing the folder structure: the script will set up a folder for overall ‘results’ and the ‘learning_curves’. Structurally, the script will store the result for each seed in a separate folder containing the seed number (with a subfolder for the models and the results). The master script then loops through the list of seeds (defined in a list at the top), with each seed reproducing the same result. This process can be made faster by parallelizing the computation for each seed.

The first step in the loop (after the subfolder structure) is setting up the respective firm-world environments. This essentially means creating a Python object (*FirmWorldEnv*) that draws the simulation parameters from the specified distribution and ranges (as outlined in Table 1 of the main paper). Three equivalent firm-world environments are created; the first one computes a fixed-baseline comparison (without DRL); the second environment is the training environment, where agents are trained, and policies are updated; the third evaluation environment is used to measure and assess performance periodically during training periods. In our case, these two environments are the same, but one could also keep the evaluation environment fixed (in terms of parameters). Subsequently, we define the number of actions each firm can take and the observation space the firm has. In our case, each firm can take 51 actions (enter each of the 50 countries plus the option of inaction). However, practically the maximum is 50, as each firm starts in one country. The observation space relates to the first layer of the CNN. These two values are derived from the training environment object and serve as input for the CNN later.

We then set up the CNN architecture for each agent. We set a manual seed for the PyTorch models (neural network library). Then we create a policy and value network for the first agent and copy it

for the second firm. This is again a Python object (*CNNAgent*), where we specify the respective CNN parameters besides the number of countries and policies (actions). We use a basic CNN architecture based on the common pytorch library (Paszke et al., 2019): an input layer, two hidden layers, and an output layer. The next step is to define a numerical optimizer for each firm. In our case, we use an ADAM optimizer, which has become a standard approach. The ADAM optimizer is again available from the pytorch library. We assign an optimizer for each agent and apply a rather conservative learning rate (*lr*). We then assign an entropy coefficient to each agent (which determines how explorative the agents are during the training process) and define the extent of the training episodes. The extent of the training episodes determines how long the agents are trained and is a function of the number of rounds, the training periods, and the number of firms. Eventually, we set up a PPO algorithm (*CleanPPOSelfPlay*), where we input the training/evaluation environments, the neural networks, and the optimizer for each agent.

Training – Multi-agent reinforcement learning (MARL)

The next step in the master script is to then train the CNN for the two agents using the PPO and ADAM optimizers based on the created firm-world environments. The PPO algorithm has a few hyperparameters that were carefully tuned to enhance the training process – they are defined at the beginning of the master script. The DRL-specific parameter values were set using common heuristics and subsequently calibrated based on the learning curves. In most cases, we set the values to common or default values or opted for a more conservative approach. The training then works as follows. The two firms are trained for 61,000 episodes in total (1,000 training periods x 30 rounds x 2 firms + 1,000 periods for the focal firm at the end).

The two firms are trained in a ‘self-play’ format. This means in each round, one firm is trained, while the other firm is kept fixed. After each round, the policy and value network for the trained firm are optimized in three epochs. The number of epochs defines how many times the policy and value network are updated using the same experience data to improve learning. After 1,000 periods, this is switched, and the other firm is trained. This procedure is repeated in 30 rounds for each firm. Every 100 training periods, the policy quality is evaluated against a fixed-future benchmark, which selects the most attractive market based on the past average returns of countries (considering rival presence). We use this

comparison and the convergence of the learning curves to fine-tune the hyperparameters of the DRL. The internationalization moves of the final policy are then stored and saved together with the environmental and training data for further analysis. The script training the neural network also saves the values of the neural network for the last 1000 episodes to allow for the computation of the SHAP values later (in the analysis we only use the last 50 episodes).

Once the 61,000 training episodes are done, the master script closes the environments. After all seeds are computed, the script merges the single seed results into one CSV file and does some pre-processing for the analysis.

References

- Alcácer, J., Dezső, C. L., & Zhao, M. (2013). Firm rivalry, knowledge accumulation, and MNE location choices. *Journal of International Business Studies*, 44(5), 504–520.
- Alcácer, J., Dezső, C. L., & Zhao, M. (2015). Location choices under strategic interactions. *Strategic Management Journal*, 36(2), 197–215.
- Baum, J. A. C., & Korn, H. J. (1999). Dynamics of dyadic competitive interaction. *Strategic Management Journal*, 20(3), 251–278.
- Brouthers, L. E., Mukhopadhyay, S., Wilkinson, T. J., & Brouthers, K. D. (2009). International market selection and subsidiary performance: A neural network approach. *Journal of World Business*, 44(3), 262–273.
- Buchner, A., Espenlaub, S., Khurshed, A., & Mohamed, A. (2018). Cross-border venture capital investments: The impact of foreignness on returns. *Journal of International Business Studies*, 49(5), 575–604.
- Butler, K. C. (2016). *Multinational finance: Evaluating the opportunities, costs, and risks of multinational operations*. John Wiley & Sons.
- Dike, M. C., & Rose, E. L. (2019). Cross-border expansion and competitive interactions of indigenous mobile network operators in sub-Saharan Africa. *Thunderbird International Business Review*, 61(1), 29–42.
- Eden, L., & Miller, S. R. (2004). Distance matters: Liability of foreignness, institutional distance and ownership strategy. In M. A. Hitt & J. L. C. Cheng (Eds), *Theories of the multinational enterprise: Diversity, complexity and relevance* (pp. 187–221). Emerald Group Publishing Limited.
- Flowers, E. B. (1976). Oligopolistic Reactions in European and Canadian Direct Investment in the United States. *Journal of International Business Studies*, 7(2), 43–55.
- Fotsch, R. J. (1984). Machine tool justification policies: Their effect on productivity and profitability. *Journal of Manufacturing Systems*, 3(2), 169–195.
- Ghemawat, P., & Thomas, C. (2008). Strategic Interaction Across Countries and Multinational Agglomeration: An Application to the Cement Industry. *Management Science*, 54(12), 1980–1996.
- Gimeno, J., Hoskisson, R. E., Beal, B. D., & Wan, W. P. (2005). Explaining the clustering of international expansion moves: A critical test in the U.S. Telecommunications Industry. *Academy of Management Journal*, 48(2), 297–319.
- Gomez, J., Lanzolla, G., & Maicas, J. P. (2016). The role of industry dynamics in the persistence of first mover advantages. *Long Range Planning*, 49(2), 265–281.
- Graham, E. (1998). Market Structure and the Multinational Enterprise: A Game-theoretic Approach. *Journal of International Business Studies*, 29(1), 67–83.
- Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2), 895–943.
- Hansen, M. W., & Hoenen, A. (2016). Global oligopolistic competition and foreign direct investment: Revisiting and extending the literature. *Critical Perspectives on International Business*, 12(4), 369–387.
- Head, K., Mayer, T., & Ries, J. (2002). Revisiting oligopolistic reaction: Are decisions on foreign direct investment strategic complements? *Journal of Economics & Management Strategy*, 11(3), 453–472.
- Hoang, D., Gatzert, S., & Ruckes, M. (2024). The Economics of Capital Allocation in Firms: Evidence from Internal Capital Markets. *Management Science*, 71(8), 6392–6425.

- Huff, L. C., & Robinson, W. T. (1994). Note: The Impact of Leadtime and Years of Competitive Rivalry on Pioneer Market Share Advantages. *Management Science*, 40(10), 1370–1377.
- Hymer, S. H. (1976). *The international operations of national firms*. MIT Press.
- Isobe, T., Makino, S., & Montgomery, D. B. (2000). Resource Commitment, Entry Timing, and Market Performance of Foreign Direct Investments in Emerging Economies: The Case of Japanese International Joint Ventures in China. *Academy of Management Journal*, 43(3), 468–484.
- Ito, K., & Rose, E. L. (2002). Foreign Direct Investment Location Strategies in the Tire Industry. *Journal of International Business Studies*, 33(3), 593–602.
- Kafourous, M., Cavusgil, S. T., Devinney, T. M., Ganotakis, P., & Fainshmidt, S. (2022). Cycles of de-internationalization and re-internationalization: Towards an integrative framework. *Journal of World Business*, 57(1), 101257.
- Kim, J. H. (2019). Is your playing field unleveled? U.S. defense contracts and foreign firm lobbying. *Strategic Management Journal*, 40(12), 1911–1937.
- Kingma, D. P., & Ba, J. 2014. *Adam: A Method for Stochastic Optimization*.
<https://doi.org/10.48550/arXiv.1412.6980>
- Knickerbocker, F. T. (1973). *Oligopolistic reaction and multinational enterprise*. Harvard University Press.
- Lefley, F. (1996). The payback method of investment appraisal: A review and synthesis. *International Journal of Production Economics*, 44(3), 207–224.
- Li, J., & Guisinger, S. E. (1992). The Globalization of Service Multinationals in the "Triad" Regions: Japan, Western Europe and North America. *Journal of International Business Studies*, 23(4), 675–696.
- Lumsdaine, R. L., & Prasad, E. S. (2003). Identifying the Common Component of International Economic Fluctuations: A new Approach. *The Economic Journal*, 113(484), 101–127.
- Pan, Y., Li, S., & Tse, D. K. (1999). The Impact of Order and Mode of Market Entry on Profitability and Market Share. *Journal of International Business Studies*, 30(1), 81–103.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, Sasank, Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: an imperative style, high-performance deep learning library, *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 32, 8026-8037.
- Robinson, W. T., Kalyanaram, G., & Urban, G. L. (1994). First-Mover Advantages from Pioneering New Markets: A Survey of Empirical Evidence. *Review of Industrial Organization*, 9(1), 1–23.
- Rose, E. L., & Ito, K. (2008). Competitive interactions: The international investment patterns of Japanese automobile manufacturers. *Journal of International Business Studies*, 39(5), 864–879.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. 2017. *Proximal Policy Optimization Algorithms*. <https://doi.org/10.48550/arXiv.1707.06347>
- VanderWerf, P. A., & Mahon, J. F. (1997). Meta-Analysis of the Impact of Research Methods on Findings of First-Mover Advantage. *Management Science*, 43(11), 1510–1519.
- Veiga, J. F., Lubatkin, M., Calori, R., Very, P., & Tung, Y. A. (2000). Using neural network analysis to uncover the trace effects of national culture. *Journal of International Business Studies*, 31(2), 223–238.

- Wu, Z., & Salomon, R. (2017). Deconstructing the liability of foreignness: Regulatory enforcement actions against foreign banks. *Journal of International Business Studies*, 48(7), 837–861.
- Yang, J. Y., Kim, M., Li, J., & Lu, J. W. (2023). Information voids and cross-border bandwagons of foreign direct investment into an emerging economy. *Strategic Management Journal*, 44(11), 2751–2782.
- Yao, F. K., Chen, M.-J., Li, J., Combs, D., & Li, Q. (2024). A review of 50 years of research since Knickerbocker (1973): competitive dynamics in international business. *Journal of International Business Studies*, 55(5), 522–550.
- Yao, F. K., & Li, J. (2016). Multi-market contact and foreign entry location decisions in China. *Management International Review*, 56(1), 95–122.
- Young, G., Smith, K. G., Grimm, C. M., & Simon, D. (2000). Multimarket contact and resource dissimilarity: A competitive dynamics perspective. *Journal of Management*, 26(6), 1217–1236.
- Yu, T., & Cannella, A. A. (2013). A Comprehensive Review of Multimarket Competition Research. *Journal of Management*, 39(1), 76–109.
- Yu, T., Subramaniam, M., & Cannella, A. A. (2009). Rivalry Deterrence in International Markets: Contingencies Governing the Mutual Forbearance Hypothesis. *Academy of Management Journal*, 52(1), 127–147.
- Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal*, 38(2), 341–363.
- Zaheer, S., & Mosakowski, E. (1997). The Dynamics of the Liability of Foreignness: A Global Study of Survival in Financial Services. *Strategic Management Journal*, 18(6), 439–464.