Received: 31 January 2018 | Revised: 21 August 2019 | Accepted: 12 December 2019 | Published on: 5 March 2020

DOI: 10.1002/smj.3142

RESEARCH ARTICLE

STRATEGIC MANAGEMENT JOURNAL WILEY

Get rich or die trying... finding revenue model fit using machine learning and multiple cases

Ron Tidhar | **Kathleen M. Eisenhardt**

Department of Management Science and Engineering, Stanford University, Stanford, California, USA

Correspondence
Ron Tidhar, Department of Management Science and Engineering, Stanford University, Stanford, CA 94305.
Email: rtidhar@stanford.edu

Abstract

Research Summary: While revenue models are strategically important, research is incomplete. Thus, we ask: "What is the optimal choice of revenue model?" Using a novel theory-building method combining machine learning and multi-case theory building, we unpack optimal revenue model choice for a wide range of products on the App Store. Our primary theoretical contribution is a framework of high-performing *revenue model-activity system configurations*. Our core insight is the *fit* between value capture (revenue models) and value creation (activities) at the heart of successful business models. Contrastingly, low-performing products avoid complex value capture (i.e., freemium) and misunderstand value creation (e.g., overweight effort). Overall, we contribute a theoretically accurate and empirically grounded view of successful business models using a pioneering method for theory building using large, quantitative data sets.

Managerial Summary: Revenue models are critical for product performance. Yet, the high-performing choice is often unclear. We combine machine learning with multiple-case deep-dives to unpack optimal revenue model choice for a wide range of products on the App Store, a significant setting in the digital economy. Our primary insight is that high-performing products *fit* value capture (revenue models) and value creation

Glueck Best Paper Award, STR Division, AOM, 2019

(activity systems) to form coherent business models. Contrastingly, low-performing products avoid complex value capture (i.e., freemium) and misunderstand value creation (e.g., overweight effort and price). We also identify the importance of *user resources*, *marketing*, *offline brand*, and *product complexity* for specific revenue models. Overall, we contribute a framework for the optimal choice of revenue model and spotlight the revenue model-activity system configurations of successful business models.

KEYWORDS

business models, machine learning, mobile application products (apps), multi-case theory building, revenue models

1 | INTRODUCTION

In 2011, music streaming service, *Spotify*, launched its product to U.S. listeners. Although similar to other music streaming services, Spotify was notably different in its revenue model. While Pandora used an *advertising* revenue model and Apple Music chose *paid*, Spotify selected *freemium*. That is, it offered its basic service for *free*, and a *paid* premium version that provided more content, greater functionality, and a better user experience. Given these similar products, a puzzle emerges: Why do they have different revenue models? As other examples like Netflix and Hulu, Stitch Fix and Rent the Runway, and 23andMe and Ancestry.com suggest, rivals across many industries may choose surprisingly different revenue models. Yet, resolving this puzzle and the related optimal choice of revenue model remain elusive.

Consistent with others (Casadesus-Masanell & Zhu, 2010), we define a *revenue model* as the monetization approach by which a firm derives sales from its products. Thus, revenue models are the means by which firms capture value, and so earn the revenue essential for superior financial performance. Revenue models are part of the broader concept of business models. By *business model*, we mean the system of interconnected activities performed by a focal firm (and often by users and partners) to create value, with part of that value captured by the firm (i.e., revenue model; Massa et al., 2017; McDonald & Eisenhardt, 2019).

Research suggests that revenue models are a novel source of innovation (Casadesus-Masanell & Zhu, 2013; Snihur & Zott, 2019), and that revenue models (and broadly business models) are early and critical strategic choices (McDonald & Eisenhardt, 2019). Yet choosing a revenue model can be challenging. Falling communication costs and the internet have enabled new ways to create value (beyond the activities of producer-firms) such as with users and partners in settings like marketplaces and ecosystems (Gambardella & McGahan, 2010; Hannah & Eisenhardt, 2018; Ott & Eisenhardt, 2020; Zott, Amit, & Massa, 2011). These changes add flexibility in how to *create* value, but also complicate how to *capture* it since more actors (e.g., users, partners) are involved. In addition, many goods have low marginal costs or are experience goods that further complicate pricing (Massa et al., 2017; Shapiro & Varian, 1998).

Prior work recognizes multiple types of revenue models, but there is limited agreement on the core ones. Practitioner work often favors long atheoretical lists that may be too many to be meaningful (e.g., Gassmann, Frankenberger, & Csik, 2014 indicate 55). In contrast, academic work pares the list to a few: (a) *paid*, (b) *advertising*, and (c) *freemium* (e.g., Arora, Ter Hofstede, & Mahajan, 2017; Rietveld, 2018). Yet, this may be too few for capturing essential differences. Research also suggests several factors (e.g., product quality, consumer ad-aversion, and advertising rates) that shape the optimal choice of revenue model (e.g., Casadesus-Masanell & Zhu, 2010; Eckhardt, 2016). Yet while helpful, this work has inconsistent results such as for product quality. Its perspectives are sometimes incomplete. For example, although this work points to advertising rates, it neglects the advertiser's view such as why an advertiser would pay high (or low) rates. Finally, it under-theorizes the freemium model including whether it is even a unique revenue model. The freemium model is also puzzling because it is widely used, but its effectiveness is inconsistently supported by research.

More deeply, revenue model research is largely atheoretical. By neglecting the link to the overall business model, this research misses the theoretical relationship between value capture (i.e., revenue model) and value creation (i.e., activity system) at the heart of successful business models. Finally, it leaves open the empirical puzzle of why similar products (e.g., Spotify and Apple Music, Stitch Fix and ThredUp, Netflix and Hulu) have different revenue models. Overall, research lacks a theoretically coherent and empirically supported view of revenue models. Thus, we take a question-driven approach to ask, "What is the optimal choice of revenue model?"

We tackle our research question by studying products in Apple's App Store. This setting is appropriate for several reasons. First, the App Store (and the broad digital goods industry) is economically significant (Arora et al., 2017; Askalidis, 2015), generating over $120 billion in cumulative earnings for developers (Apple, 2019). In addition, many products, like ride-hailing service Uber and routing service Waze, sell only through mobile applications. Second, the App Store covers a wide range of products in multiple markets (Hallen, Davis, & Yin, 2018), yielding a rich view of revenue models across settings. Third, the App Store provides accurate measures of key constructs (Boudreau, Jeppesen, & Miric, 2019b; Yin, Davis, & Muzyrya, 2014) such as revenue models and performance.

We use a *novel combination of methods* to build theory: (a) Exploratory data analysis reveals trends and provides a roadmap for theoretical sampling of cases. (b) Multi-case theory-building identifies theoretical constructs and mechanisms, and seeds machine learning. (c) Machine learning provides large-scale analysis to corroborate and extend case studies with more precise effect sizes, equifinal paths, nonlinearities, and configurations.

We contribute to *strategy*. First, we identify unique and high-performing revenue model-activity system configurations. These configurations fit value creation (i.e., activity system) and value capture (i.e., revenue model) to form coherent business models. In contrast, past research often decouples these business model elements. Thus, a key theoretical insight is the *fit* between revenue models that capture value and activities that create this value. Finally, low-performing configurations avoid complex value capture (i.e., freemium) and misunderstand value creation (e.g. over-weight effort, price), thus forming ill-fitting and low-performing business models.

Second, we contribute by re-conceptualizing two of three core revenue models. We bifurcate the freemium revenue model into the *bundled freemium* and *fragmented freemium* revenue models. This neglected distinction may explain why prior research (e.g., Arora et al., 2017; Liu, Au, & Choi, 2012; Rietveld, 2018) has conflicting results for freemium. We re-conceptualize the advertising revenue model as the broader and more accurate *third-party* revenue model, highlighting that third-party payers are not limited to advertisers.

Third, we contribute constructs and novel theoretical relationships. We find that *user resources*, *offline brand*, *marketing*, and *product complexity* relate to specific value-creating activities. As such, they are associated with particular high-performing revenue model-activity system configurations in successful business models. These constructs differ from those in prior research like ad rates (e.g., Casadesus-Masanell & Zhu, 2010; Lin, Ke, & Whinston, 2012), ad-aversion (e.g., Gabszewicz, Laussel, & Sonnac, 2005), and product quality (e.g., Chen, Fan, & Li, 2016; Eckhardt, 2016).

As important, we contribute a novel combination of complementary methods: *Machine learning*, *multi-case theory building*, and *exploratory data analysis*. This combination enables theory building with large quantitative data sets. We also add a suite of three ML approaches that emphasize robustness and interpretability (not just prediction). We further add a new conception of a "*case*." While most studies use a few "thick," longitudinal cases about process, we use many "thin," mostly cross-sectional cases about content. Most importantly, we highlight the strong similarities between two powerful methods: Machine learning and multi-case theory building (Table 1).

TABLE 1 Comparison: Multi-case theory building and machine learning

	Multi-case theory building	**Machine learning**
Definition	Process for finding patterns in data using two or more cases	Algorithms for finding patterns in quantitative data
Objectives	Robust, accurate, and generalizable theory	Robust, accurate, and generalizable prediction
Guards against overfitting	Replication logic	Cross-validation
Guards against excess complexity	Construct abstraction	Regularization techniques
A priori assumptions	Few	Few
Sampling	Theoretical	Balanced random
Theoretical constructs	Identified from data	Selected from among many possible features in data
Final model selection	Researcher	Algorithm
Scale	Very small	Small to very large
Strengths	Empirically grounded Rich description Identification of major theoretical constructs, relationships, and mechanisms	Empirically grounded Large scale Selection of major constructs and identification of precise patterns (e.g., effect size and direction, nonlinearities, equifinality, configurations)
Weaknesses	Small scale Researcher dependent Imprecise identification of patterns (e.g., interactions, nonlinearities, effect sizes)	Cross-sectional Atheoretical, often "black box" No causal inference or significance tests as of yet

2 | BACKGROUND

Prior research examines revenue models primarily with formal analytic modeling and hypothesis-testing methods. It identifies three primary revenue models: (a) *advertising*—product is free to the user and revenue is earned from paying advertisers, (b) *paid*—product or service is sold at an up-front price, and (c) *freemium* revenue model—combines two products—a free version with a paid "premium" version. There are several research streams.

2.1 | Research on advertising revenue models

One research stream offers insight into the *advertising* revenue model. This model is effective with increasing the advertising rate (Casadesus-Masanell & Zhu, 2010; Lin et al., 2012), decreasing ad-aversion (Gabszewicz et al., 2005; Lin et al., 2012), and low product quality (Lin et al., 2012). For example, Casadesus-Masanell and Zhu (2010) use an analytic model to explore the performance of revenue models. The authors show that the incumbent should choose the advertising model if the advertising rate is high, since high revenue per ad maximizes overall revenue. Other work focuses on consumer preferences for advertising, especially ad-aversion. For example, using an analytic model of a monopolist, Gabszewicz et al. (2005) find that low ad-aversion (i.e., consumer dislike of ads) favors the advertising model. Still other work on Google Play (Lin et al., 2012) and ecommerce platforms (Chen et al., 2016) argues that the advertising revenue model is preferred when product quality is low since consumers are unlikely to pay.

2.2 | Research on paid revenue models

A second stream offers insight into the *paid* revenue model. Some research explores product quality and consumer awareness. For example, Duan, Gu, and Whinston (2008) study the effects of these factors on movie ticket sales. They find that consumer awareness (measured by number of reviews) has a positive and significant impact on later sales. Surprisingly, quality (measured by user ratings) does not. That is, consumer awareness, not quality, drives product sales (i.e., movies) with a paid revenue model. Relatedly, Rietveld (2018) examines the performance of paid v. freemium revenue models in gaming. Video games with a paid model receive more playtime and revenue than those with a freemium model. The argument is that anchoring and sunk costs lead users to value paid games more than freemium ones. Thus, perceptions and biases, not product quality, drive the success of paid revenue models. In contrast, Eckhardt (2016), who studies the Palm and smartphone ecosystems, finds that the paid revenue model is more effective when product quality is high. Analytic models find similar results, driven by assuming that product quality enables higher prices (e.g., Casadesus-Masanell & Zhu, 2010).

2.3 | Research on freemium revenue models

A recent stream parallels the rising importance of digital goods and platforms like the App Store, and offers insight into the *freemium* revenue model (Boudreau, Jeppesen, & Miric, 2019a). Some work highlights the advantages of freemium like customer segmentation and

allowing users to experience the product before paying. For example, Liu et al. (2012) study 1,597 Google Play products. They find that the freemium revenue model outperforms the paid one vis-a-vis downloads and revenue. Yet, others find freemium to be inferior to the paid model (Rietveld, 2018). For example, Casadesus-Masanell and Zhu (2010) use modeling to show that the freemium model is low performing relative to paid when there is rivalry. Similarly, Arora et al. (2017) analyze Google Play products, and find that freemium models slow adoption of the paid product, suggesting cannibalization. Yet, other work finds no results (Hamari, Hanner, & Koivisto, 2017).

2.4 | Summary

Overall, revenue model research offers insights into the types and choices of revenue models. An *advertising* revenue model is effective with increasing advertising rates (Casadesus-Masanell & Zhu, 2010; Lin et al., 2012) and decreasing ad-aversion (Gabszewicz et al., 2005; Lin et al., 2012), and when product quality is low because consumers are unlikely to pay (Casadesus-Masanell & Zhu, 2010; Chen et al., 2016; Lin et al., 2012). Yet, these explanations largely neglect the advertiser's perspective: They leave open when advertisers will be willing to pay high advertising rates, and why they would want to associate with a low-quality product.

The *paid* revenue model is effective when the ad rate is low (Casadesus-Masanell & Zhu, 2010; Gabszewicz et al., 2005; Lin et al., 2012). Yet, the advertiser's perspective is again missing. Other results are mixed. Some find that the paid model is effective when product quality is high (Chen et al., 2016; Eckhardt, 2016). Others find that consumer perception and biases (not product quality) drive effective paid models (Duan et al., 2008; Rietveld, 2018). In short, product quality has conflicting results, suggesting under-theorization.

Finally, some research highlights the advantages of the *freemium* revenue model (v. paid) like customer segmentation and ability to experience the product before paying (Liu et al., 2012). Yet others find the freemium model to be inferior to the paid model (Arora et al., 2017; Casadesus-Masanell & Zhu, 2010; Rietveld, 2018), especially under rivalry. Still other work finds no results (Hamari et al., 2017). These conflicting results suggest that the freemium revenue model is under-theorized. Finally, despite contrary research, freemium models are often successful in real firms like Spotify—a puzzling empirical fact.

In sum, prior research is often incomplete (e.g., missing advertiser perspective), conflicting (e.g., inconsistent results for product quality), and simplistic (e.g., likely under-theorizing freemium). More deeply, this work has a largely atheoretic focus on value capture (i.e., revenue models), but neglects value creation (i.e., activities). Finally, this research is silent on our empirical puzzle of why similar products (e.g., Spotify, Apple Music, and Pandora) have different revenue models. Overall, prior research lacks a theoretically coherent and empirically supported explanation of high-performing revenue model choice and broadly, business models. We address this gap with a novel theory-building approach.

3 | METHODS

We sampled from the entire population of mobile products available on Apple's App Store in November 2015. The App Store is a “store front” where consumers can browse and download products for their iOS devices. These products span 22 market categories like Travel, Finance,

and Games. Begun in 2008, the App Store has garnered much attention, investment, and success. There have been 2.2 million products developed for the App Store, with over $120 billion in cumulative earnings for developers (Apple, 2017, 2019).

The App Store is attractive for our research. First, the App Store (and other such stores) are economically significant and central to the digital economy (e.g., $143B in revenues and 12 million developers in 2016; Arora et al., 2017). Second, the App Store has a wide range of products across markets, from finance and medical products to games and travel, likely yielding rich understanding across a wide swath of products and categories. Importantly, the App Store is more than a sales channel. For example, ride-hailing services Uber and Lyft are exclusively mobile products, and offer no other way to access their services. Similarly, diverse products like stock trading platform Robinhood (valued at $1.3B in 2018), and mobile game Candy Crush (over $1B in revenue in 2017) are also "mobile first," such that their businesses are anchored in App Store products. In short, App Store products are both economically significant and diverse.

Third, the App Store is particularly useful for exploring both the under-theorized freemium revenue model and products for which revenue models may be challenging like goods with low marginal costs (e.g., software) and experience goods (Massa et al., 2017; Shapiro & Varian, 1998). Finally, the App Store provides accurate measures of constructs related to our research like revenue model, market category, and performance. Indeed, the growing number of studies across multiple topics using these data attest to their accuracy and value (e.g., Boudreau et al., 2019b; Bresnahan, Davis, & Yin, 2014; Hallen et al., 2018; Yin et al., 2014).

We sample the full population (13,195) of popular products (defined below), and a random sample of 53,457 unpopular products (from a population of over 1.5 million products). This yielded a sample of 66,652 products. We collected a second wave of data on these same products in November 2017 to capture any time trends and major changes. Since we saw almost no market category change and only a 6% change in revenue model among surviving products from 2015 to 2017, we focus on the 2015 sample.[1] We use a theory-building approach with a novel combination of methods: Exploratory data analysis, multiple-case theory-building, and machine learning. Each method provides unique insights and complements the others.

3.1 | Exploratory data analysis

We began with "exploratory data analysis" (EDA; Behrens, 1997). EDA involves examining data from nonparametric (often visual) lenses. The goal is a preliminary understanding of patterns within data. In contrast to "confirmatory data analysis" (e.g., hypothesis-testing), the emphasis is finding unexpected relationships. EDA begins with broad categories, and then systematically increases granularity by segmenting data into smaller subgroups along potentially relevant dimensions. Increasing granularity sharpens patterns that might otherwise be hidden.

Using data from App Store profiles for each product, we first measured variables likely to be important. We measure the performance of each product with a binary measure, *popularity*. Apple selects up to 240 products in each category to feature in a "popular" list and updates this

[1]While others (e.g., Andries, Debackere, & Van Looy, 2013; Berends, Smits, Reymen, & Podoynitsyna, 2016) find that business models benefit from learning and so change, our revenue models are stable. One reason may be that past work focuses on activities. Another may be that users are less tolerant of revenue model changes (e.g., free to paid). A third may be that firms iterate and learn before entering the App Store. Importantly, we find that a majority (60%) of popular products that change go from free to freemium with an increase in product complexity, consistent with our theoretical framework.

list several times daily.[2] Although the exact selection procedure is not revealed, Apple's popular ratings are seen by experienced observers as very closely linked to the number of downloads and commercial success of the product (Boudreau et al., 2019b; Bresnahan et al., 2014; Garg & Telang, 2013). We triangulated this performance measure with other measures from our cases (next section). Here, high (e.g., OpenTable, CandyCrush) and low (e.g., Medical Visual Books, LumberJack) performance are very clear and consistent over time.

We measure the revenue model of each product, *revenue model*. We assign each product to one of three revenue models consistent with prior research. We assess: (a) *paid* by whether the firm sets an up-front price (between $0.99 and $999.99) that the user must pay before downloading the product, (b) *free* by whether the user can obtain the product without payment (this includes the advertising model), and (c) *freemium* by whether the user can obtain the product for free, but can also purchase at least one up-sell product. Apple displays up to 10 types of up-sells termed "in-app purchases." We were able to categorize each product in our sample into one of these revenue models, indicating mutually exclusive and collectively exhaustive categories.

We measure the market of each product, *category*. The App Store uses 22 market categories such as Lifestyle, Travel, and Sports with each product assigned to one category. The developer designates the category, and then Apple verifies this designation and rejects ones with incorrect categories. Since customers browse by category and Apple rejects products with incorrect categories, developers are highly motivated to make accurate selections.

Our next step was dividing the data into increasingly granular segments, and creating tables and visuals like frequency charts. First, we analyzed our data by *revenue model*, and then by performance measure, *popularity*. Increasing the depth, we divided our sample by *category* to observe rates of revenue models in each market and at different popularity levels. Next, we subdivided our data to explore particular categories and revenue models in more depth.

Overall, EDA revealed several patterns. First, popular v. unpopular products have very distinct profiles. Popular ones are likely to use freemium models (62%) while unpopular ones are likely to use paid (33%) or free (57%; Figure S1). Second, some market categories rely primarily on one revenue model while others use another or a mix, suggesting underlying theoretical distinctions (Figures S2–S3). For example, the free model dominates Finance (80%) whereas freemium dominates Games (75%). Third, by plotting the frequency of # of upsell products within the freemium model (Figure S4), we found an unanticipated bimodal distribution, suggesting two distinct freemium models that we term: *Fragmented* and *bundled*.

3.2 | Multi-case theory building

We next used EDA insights to guide the theoretical sampling of cases. Multi-case theory-building involves diving into a small number of rich cases to identify relevant constructs and theoretical relationships (Eisenhardt, Graebner, & Sonenshein, 2016). A key point is our unusual approach. Most multi-case theory-building studies (e.g., Graebner & Eisenhardt, 2004; Zuzul & Tripsas, 2019) focus on *process* with a few "thick," longitudinal cases. In contrast, we

[2]While the list is updated often, popularity is stable. For example, in a random sample of 200 popular and 200 unpopular products from our larger study, 75% of surviving popular products from 2015 are still popular in 2019. Only 3% of unpopular products from 2015 become popular. Importantly, most newly popular products either fit or changed such that they fit our theoretical framework. We describe several later.

focus on *content* with many "thin," mostly cross-sectional cases. This approach fits our sample where revenue models rarely change and our content (not process) research question. Thus, our cases are appropriately less rich and shorter (i.e., several pages), relative to the typical multi-case study. Nonetheless, we follow the basic steps of multi-case theory building such as theoretical sampling, within-case analysis, and cross-case analysis (Eisenhardt & Graebner, 2007).

We began by selecting cases using *theoretical sampling*. In contrast with random sampling, theoretical sampling involves selection of meaningful cases for building theory (Eisenhardt & Graebner, 2007). Specifically, we used patterns from EDA to select 8 market categories. We first selected categories that varied on the dominant (i.e., most common) revenue model. For example, the popular Games category is dominated by freemium (75%), while the popular Travel category is dominated by free (78%; Figure S2–S3). We balance among categories that tip toward one revenue model (e.g., Games, Travel) and those that do not (e.g., Music), and among categories where popular and unpopular product choices were more/less the same (Finance, Health & Fitness, Food & Drink). Specifically, we chose Games, Photo & Video, Social Networking, Health & Fitness, Music, Travel, Food & Drink, and Finance.

Within our eight categories, we then chose specific products for case studies. We randomly selected a popular and an unpopular product using the most common revenue model in each category. This theoretical sampling of *polar types* (i.e., extremes) typically clarifies patterns. We also randomly selected a counterfactual popular product (i.e., successful product with a revenue model counter to the most common one in the category). *Counterfactuals* help to eliminate alternative explanations. We also chose one well-known popular product in each category to ensure data availability. Thus, we purposefully chose (a) well-known cases (e.g., OpenTable, Candy Crush) to ensure data availability, and (b) randomly selected cases to improve robustness (e.g., iTranslate, Mortgage by Zillow). We repeated this sampling (i.e., two random polar types, one random popular counterfactual, one well-known popular) for each of the eight categories (i.e., four cases each, Table S2). We also randomly sampled eight unpopular products across all categories to improve robustness and balance. In sum, we selected 24 popular and 16 unpopular products, totaling 40 cases.

To develop our cases, we triangulated data from *multiple sources*. We collected online archival data (e.g., media articles, press releases, and online videos) for each case. We read over 150 unique pieces of data. Second, we downloaded and examined each product for its content, uses, and business model. We also read a random sample of about six user reviews from its App Store profile. Third, we interviewed 10 developers: Four popular product developers (one solo developer, three from large teams) and six unpopular product developers (two solo developers, four from large teams). We used a semi-structured interview that covered product history and choice of business model including revenue model and the rationale. We also included 25 online (e.g., YouTube) interviews. These were mostly with developers of well-known popular products like Candy Crush and Robinhood, and usually covered major product(s), product history including performance, and business model including revenue model.

Consistent with multi-case theory building methods (Eisenhardt & Graebner, 2007), we wrote each case independently, and then conducted *within-case analysis*. We focused on how and why the firm chose the product's revenue model. Next, we moved to *cross-case analysis* where we further developed tentative constructs and theoretical relationships from individual cases and compared them across cases (i.e., replication logic). For example, we iteratively compared products with the same revenue models, looking for similarities and differences. We also randomly compared products, across revenue models and categories. By cycling through

emerging theory and data, we strengthened our theoretical logic, constructs, and measures. Once we had reasonable correspondence (i.e., *theoretical saturation*), we ended analysis.

Multi-case theory-building revealed several insights. First, the advertising model is better conceptualized as the broader *third-party* model. For example, our *OpenTable* case indicates that its revenue model is not just advertising. It also charges restaurants for customer reservations. This and other cases indicate that advertisers are not the only third-party payers. Second, we identify novel constructs and theoretical relationships. For example, the third-party revenue model fits when users provide valuable *user resources* for which a third party will pay. *Product complexity* links to freemium. *Offline brand*, *marketing*, and *design* related to activities that create quality signals and fit the paid revenue model.

3.3 | Machine learning

We use machine learning to add scale and precision to our theory-building. Broadly, machine learning (ML) uses algorithms to systematically estimate and compare many alternative models, and then pick the best (i.e., most predictive; Athey, 2018). Thus, ML approaches find patterns in quantitative data, and yield robust models that predict out-of-sample (i.e., generalize). ML uses two broad techniques to improve prediction and generalizability by limiting overfitting and excess model complexity (Choudhury, Allen, & Endres, 2019; Varian, 2014). *Cross-validation* involves systematically (algorithmically) estimating models from training data and then testing on validation data to limit over-fitting. A common method is k-fold cross-validation that splits a data set into k equal parts. Each k part is then successively used as the validation data and the $k - 1$ parts as the training data, resulting in an averaged final estimate. *Regularization* "penalizes" excess model complexity such as by limiting coefficient size in regression-like techniques and leaves on decision trees. Regularization "tunes" the balance between over- and under-fitting.

ML is seeing growing use in strategy and organizations research (e.g., Choudhury, Wang, Carlson, & Khanna, 2019; Raj & Seamans, 2019). One use is showing that complex phenomena like high-impact patents can be predicted with higher accuracy by ML than other methods (e.g., Glaeser, Kim, & Luca, 2017; Lee, Menon, & Tabakovic, 2018), improving causal insight (Rathje & Katila, 2020). Choudhury, Allen, and Endres (2019) show that ML can uncover paths in data that are difficult to observe with other methods. Another use is natural language processing and recently facial recognition to assess hard-to-measure constructs like emotions (Menon, Nave, & Bhatia, 2019) and personality (Gow, Kaplan, Larcker, & Zakolyukina, 2016). These measures are then used to test hypotheses using traditional econometrics.

We use ML for a new application, *theory-building*. First, ML is a pattern recognition technique, thus fitting theory-building. It relies on many iterative comparisons to find the most predictive (i.e., generalizable) model very similar to multi-case theory building (but on a much larger scale with algorithms). Second, ML complements multi-case theory building (Table 1). Case studies (a) reveal constructs that narrow the search space (making ML more meaningful and efficient; Lettau & Pelger, 2018) and (b) provide theoretical mechanisms (complementing MLs atheoretical pattern recognition). In complement, ML adds (a) potential for large-scale corroboration and elaboration of multi-case insights and (b) precise identification of size effects, nonlinearities, equifinal paths, and configurations (all challenging for case studies). Finally, the current state-of-the-art of ML fits well with research questions such as ours that are classification problems using primarily cross-sectional data (Athey, 2018; Varian, 2014).

We began by choosing a random sample of 400 popular products and 400 unpopular ones from our sample of 66,652 products. Although ML is often associated with "big data," smaller samples of even 0.1% can produce accurate ML results (Varian, 2014). To ensure balanced data for training and cross-validation, we randomly chose products within each revenue model (i.e., we sampled 100 products for each revenue model; Choudhury, Allen, & Endres, 2019).[3] We further collected and set aside a separate random sample of 100 popular and 100 unpopular products to be the "out-of-sample" test data.[4] In sum, our full ML sample is 1,000 products.

Next, we hand-collected construct measures for each product as per our multi-case analysis and prior research. We used the App Store and online sources. Specifically, we added measures for *user resources*, *product quality*, *offline brand*, *marketing*, *design*, *product complexity*, and *fragmented* and *bundled* freemium models (Measures in Appendix).

We chose three complementary ML approaches (see Choudhury, Allen, & Endres, 2019 for a review and Supporting Information) for several reasons. First, these approaches are robust and relatively interpretable, not just predictive, which fits theory-building aim (see also Jung, Concannon, Shroff, Goel, & Goldstein, 2017; Puranam, Shrestha, He, & von Krogh, 2018). In contrast, ML approaches like deep learning focus on prediction, often sacrificing interpretability. Second, they have complementary advantages. *Penalized multinomial logistic regression* highlights size effects and direction of predictors. This technique differs from traditional multinomial logit: (a) relies on a k-fold cross-validation to develop robust coefficients, (b) imposes a complexity penalty (regularization) on excessively large coefficients, and (c) highlights the most important variables (not all variables and controls) and reports correct classifications (not significance levels)—all consistent with our theory-building aim.

Decision trees classify the revenue model of each product with a series of true/false decision nodes. Decision trees are especially effective when nonlinearities like configurations and equifinal paths exist (Varian, 2014). Decision trees can reveal features in the data that are not apparent with other methods. Thus, decision trees complement penalized multinomial logit regressions by highlighting nonlinearities, and identifying configurations and equifinal paths.

Random forest ensembles generate many decision trees using random subsamples of data and predictors, and average these decision trees ($n = 100$ in our study). Random forests provide accurate estimates of the most important predictors because they are based on many uncorrelated decision trees. While less interpretable than decision trees, they are also less prone to overfitting. Random forests are particularly effective for highly nonlinear data such as with discontinuities, and for indicating the most important variables that contribute to prediction accuracy.

Collectively, these complementary ML approaches capture: (a) effect sizes and direction (penalized multinomial logit), (b) configurations and equifinality (decision trees), and (c) robust measures of feature (i.e., variable) importance (random forests). Importantly, they collectively yield robust and relatively interpretable results, rather than complicated outputs that achieve the best prediction at the cost of interpretability (cf. Lee et al., 2018). This fits our theory-building aim. We run these ML approaches in separate analyses of popular and less popular products. The result of our EDA, multi-case analysis and ML is an emergent theoretical framework (Table 1).

[3]Unbalanced data sets can lead to misleading measures of classification performance. Consider building a model to test for a rare disease (e.g., appears 1% of the time). A highly accurate (but poor) test would predict "no disease" 100% of the time. This would achieve an accuracy of 99%, but would miss all positive (i.e., "disease") samples.

[4]The test set is randomly selected from the full population (for popular and unpopular), and is disjoint from our training and validation data (i.e., of 400 popular and 400 unpopular products). This fully random sampling ensures model test accuracies reflect true out-of-sample performance.

4 | EMERGENT THEORETICAL FRAMEWORK

Our research question asks, "What is the optimal choice of revenue model?" Broadly, we find that popular v. unpopular products have distinct profiles (Figure S1). *Popular* products have coherent business models that fit value capture (i.e., revenue model) with value creation (i.e., activities). By activities, we mean the processes, skills, and resources by which firms (users and partners) create value (McDonald & Eisenhardt, 2019; Zott et al., 2011). Indeed, our primary theoretical insight is the *fit* of value creation and capture in high-performing business models. Consistent with fit, ML correctly predicts revenue model at a high level—that is, 80 v. naïve estimate of 33%. In contrast, *unpopular* products often use simple revenue models (i.e., paid or free) that lack fit with activities. Consistent with weak fit, ML correctly predicts only 47% of revenue models. Finally, consistent with performance, 75% of popular products survive during our study v. only 45% of unpopular products.

4.1 | Free revenue model: Third-party

As defined earlier, *free* is the revenue model in which the user does not pay, and there is no upsell product. Although we identify two free revenue models,[5] we focus on the interesting one, which we term *third-party*. This model relies on actors (other than users) to pay. Although past research studies the advertising model, our data show that advertising is too narrow. Rather, third-party is more accurate. Finally, a third (35%) of popular products use this revenue model. In contrast, over half (57%) of unpopular products use it, suggesting these developers often miss the central role of value creation for payers.

Our multi-case analysis indicates that the third-party model is effective when the user provides valuable *user resources* while using the product. The firm can then sell these resources to a third party that is willing to pay (e.g., advertiser, investor, or bank). Thus, the third-party model fits when users contribute to value creation, and the firm can monetize that value. User resources contrasts with prior research (e.g., Casadesus-Masanell & Zhu, 2010; Gabszewicz et al., 2005; Lin et al., 2012) that emphasizes ad rates and ad-aversion, but neglects the advertiser's perspective and misses the more accurate third-party lens.

A well-known example is *WebMD* in the Health & Fitness category. The free WebMD product provides physician-reviewed health information and a "symptom checker" that allows users to search diseases by symptoms. Since this product tracks a user's disease interests, the resulting data are a valuable (and difficult to obtain otherwise) user resource for pharmaceutical firms, making them willing to pay to advertise to these well-targeted users. Since the ads relate to health concerns, they are likely tolerated and even appreciated by users. As a user reviewed, "I'm in love with the 'news & lifestyle' section [of WebMD]. I appreciate the daily updates with the latest news [typically ad-related] around drugs, treatments or interesting tips to improve health." Revenue estimates for this popular product exceed $700 million (Reuters, 2017).

Another well-known example is *OpenTable* in Food & Drink. OpenTable is a popular restaurant directory and booking service that enables users to browse and make reservations. A key

[5]The free revenue model that we term *transaction* occurs when the focal product is an efficient and/or convenient sales channel to buy nonapp products. Examples include apps for Domino's Pizza and Target. Revenue comes from the end-product purchase (e.g., pizza, clothing), not the app. We eliminated products with a transaction revenue model from our study because they are a sales channel, not an actual product.

insight is that OpenTable gathers valuable information (i.e., user resources) regarding users' preferences. With these data, OpenTable is attractive for restaurants to advertise. Ads can be targeted to people who like to eat out (or prefer specific cuisines or prices), providing favorable matches for advertising restaurants. OpenTable also has a second third-party revenue model that captures more value from its user resources: Restaurants can pay for a booking service in which users reserve tables. While reservations are free for users, OpenTable "sells" users' reservations to restaurants for $1 per seat. These third-party revenue models netted OpenTable about $200 million in revenue, leading to acquisition by *Priceline Group* for $2.6 billion (Lachapelle, 2016).

A third example is *Robinhood* in Finance. Robinhood is a popular stock market investment platform, on which users can buy and sell stocks for free. The data collected on users' stock trades are not unusually valuable. But, since users give Robinhood access to their invested capital, the firm can hold investors' funds between trades and earn interest. Robinhood "sells" these user resources (i.e., funds) to banks, for which it receives revenue (i.e., interest). Its valuation is $7.6 billion (Verhage, 2019), consistent with the success of its third-party model.

The ML analysis—decision tree, random forest, and penalized multinomial logit—underscores the importance of user resources. *User Resources* is the most important variable for distinguishing the third-party model from the paid and freemium models among popular products. It is the first (thus most important) branch of the decision tree and directly predicts the third-party revenue model (Figure 1). User resources has the largest (and positive) coefficient in the penalized multinomial logit, and is the most important random forest "feature" (Table 2).

There are several reasons *why* valuable user resources *fit* with the third-party revenue model. First, since users do not pay, products must ultimately have other actors (i.e., third-party) who do pay. These actors are likely to pay when the products have user resources that

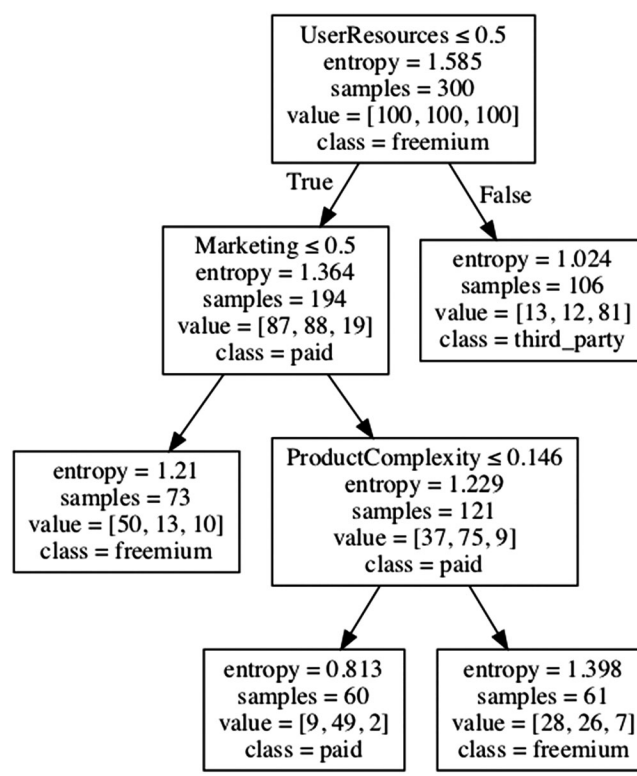

FIGURE 1 Decision tree (popular products): Classifies revenue models. Test accuracy (i.e., percent correctly classified) = 85%. Optimal hyperparameters: Minimum samples at leaf = 60, minimum samples at a split = 5, maximum depth = 3. Training accuracy = 69%. Cross-validation accuracy = 69% (standard deviation of 11%)

TABLE 2 Popular products: Penalized multinomial logit and random forest analyses

Penalized multinomial logistic regression						
Revenue model	**User resources**	**Marketing**	**Design**	**Offline brand**	**Product complexity**	**Product quality**
Third-party (free)	2.04	0.04	−0.08	−0.53	0.06	−0.20
Paid	−1.07	0.95	0.78	0.58	−0.13	0.19
Freemium	−0.97	−0.99	−0.70	−0.06	0.08	0.02

Random forest	
Feature	**Importance (mean impurity decrease)**
User resources	0.55
Marketing	0.18
Product complexity	0.15
Design	0.06
Offline brand	0.03
Product quality	0.03

Notes: *Penalized multinomial logistic regression*: Test accuracy (i.e., percent correctly classified) = 80%. Coefficients (log-odds) fit on popular products ($N = 300$). Optimal hyperparameter (inverse of regularization strength), $C = 1$. Training accuracy = 69%. Cross-validation accuracy = 68% (standard deviation of 12%). *Random forest*: Test accuracy (i.e., percent correctly classified) = 80%. Predictor importance ($n_{\text{trees}} = 100$) analysis of popular products ($N = 300$). Optimal hyperparameters: Minimum samples at leaf = 5, minimum samples at a split = 40, maximum depth = 3. Training accuracy = 71%. Cross-validation accuracy = 68% (standard deviation of 11%).

create sufficient value for them. In the case of advertisers as the third party, user resources (e.g., information) enable better matches with consumers (e.g., better-targeted ads) and so increase ad rates and advertisers' willingness to pay. Similarly, better matches increase user utility (e.g., less ad-aversion) because of more interest in the ad content. This reasoning fits with the facts that the third-party revenue model is common in Lifestyle, Food & Drink and Travel (where users often reveal personal data including purchase preferences and intentions), but rare in Games and Photo & Video (where users often provide few or no valuable resources like invested capital and personal information for well-targeted advertising).

In contrast, user resources are not associated with the third-party revenue model for unpopular products (Table 3). While over half (57%) of unpopular products use the third-party revenue model, *only* 6% provide valuable user resources. Indeed, user resources is not a decision node on the decision tree (Figure A1). An example is *Coub*, an unpopular product in Social Networking. Its product—broadcast service for short videos (10 s or less)—entertains, but users reveal little information because the videos are so short and viewing choices lack personal information. Coub has struggled to find advertisers, unlike WebMD that gathers and sells valuable user information to third-parties. Our cases suggest that these developers often mistakenly think it is easy to attract advertisers, and neglect the necessity of creating value for advertisers (or other third parties). While they may attract users with free products, they ultimately fail without payers. As a typical developer of an unpopular free product told us: "We thought advertising would work best. But we had a lot of trouble finding advertisers.... We just couldn't find advertisers that could show good ads and would pay a lot for them. We eventually gave up."

Overall, the core challenge of the third-party revenue model is recognizing valuable user resources (if any) and finding a third-party to pay for them. For example, *LinkedIn* in Social Networking quickly added users who wanted to keep up with professional friends. While investors were initially pleased, they ultimately demanded revenue. By carefully segmenting its users, LinkedIn stumbled into realizing that the third-party willing to pay was employers seeking new hires (a tiny percentage of LinkedIn users), and the valuable user resource was information about job skills that virtually all LinkedIn users provide (Piskorski, 2007).

4.2 | Paid revenue model

As defined earlier, *paid* is the revenue model for which the user pays an upfront price. The paid revenue model is infrequent among popular products (only 3%). In contrast, a much higher percentage of unpopular products (33%) uses the paid model, suggesting that many unpopular product developers are too optimistic about their products.

Our multi-case analysis indicates that the paid revenue model is effective when the product is supported by particular activities that signal quality.[6] This contrasts with research that emphasizes low ad rates (e.g., Casadesus-Masanell & Zhu, 2010; Gabszewicz et al., 2005) and high product quality (e.g., Chen et al., 2016). For example, firms can signal quality by engaging in *marketing* activities like publicity campaigns and connecting with journalists that increase media coverage. This media coverage can signal product quality to potential customers before they buy. Marketing is likely especially effective for creating media coverage when the firm also has superior product development activities that lead to better or more unique products.

A popular example in Games is *Monument Valley* (counterfactual case—i.e., paid revenue model in a category dominated by freemium). Monument Valley is a puzzle adventure game. The developers used unique Escher-inspired graphics that distinguish it from other games. Indeed, the firm spent a significant amount—over $800,000—for product development and marketing activities (Etherington, 2015). Its marketing activities are particularly striking. The team worked with Apple to release Monument Valley when there were few rival games, and so more available media coverage. With ties to influential journalists like at TechCrunch for prelaunch articles plus creation of three polished prelaunch trailers on YouTube (over 3 million views), the marketing team secured early media coverage. In fact, Monument Valley was highly anticipated in the media before its launch (Lomas, 2014). Upon release, Monument Valley won awards, including Apple's *Worldwide Developers Conference Design*, triggering more marketing and media coverage. Its developer, Ustwo, saw Monument Valley as a unique experience like a blockbuster movie or roller coaster ride, and successfully chose a paid revenue model (supported by extensive marketing activities to generate media coverage) for this popular product.

Firms can also signal quality with *design* activities such as superior product development skills that support creating multiple popular products. These prior products can signal quality for the focal product. For example, the developer *Fitness 22* has products in Health & Fitness. The firm explicitly eschews marketing, and instead concentrates on skillfully creating simple, reliable products at a range of prices (Bort, 2015). In fact, five Fitness 22 products have become

[6]In the case of *subscription*, quality signals are relevant for initial purchase while updated content (e.g., Netflix) or ongoing costs (e.g., Dropbox cloud storage) are needed for later subscription. Given space limits and few cases (<1% of sample), these results are available from the authors.

TABLE 3 Unpopular products: Penalized multinomial logit and random forest analyses

Penalized multinomial logistic regression						
Revenue model	**User resources**	**Marketing**	**Design**	**Offline brand**	**Product complexity**	**Product quality**
Third-party (free)	1.34	0.46	1.00	0.39	0.42	0.54
Paid	0.55	2.55	1.00	3.46	1.20	1.23
Freemium	1.22	0.63	1.00	0.52	1.19	1.37

Random forest	
Feature	**Importance (mean impurity decrease)**
Product complexity	0.30
Offline brand	0.26
Product quality	0.21
Marketing	0.19
User resources	0.04
Design	0.00

Notes: *Penalized multinomial logistic regression*: Test accuracy (i.e., percent correctly classified) = 47%. Coefficients (log-odds) fit on unpopular products ($N = 300$). Optimal hyperparameter (inverse of regularization strength), $C = 1$. Training accuracy = 52%. Cross-validation accuracy = 50% (standard deviation of 19%). *Random forest*: Test accuracy (i.e., percent correctly classified) = 52%. Predictor importance ($n_{\text{trees}} = 100$) analysis on unpopular products ($N = 300$). Optimal hyperparameters: Minimum samples at leaf = 8, minimum samples at a split = 5, maximum depth = 1. Training accuracy = 52%. Cross-validation accuracy = 52% (standard deviation of 11%).

popular. Supported by its design skills and reputation for excellent products, Fitness 22 released its next product, *5 K Runner*, with a paid model and found similar success (Bort, 2015).

A third path by which firms can signal quality is an *offline brand*, typically supported by offline activities in venues like movies and print. Examples include video games ported from consoles, and products related to well-known print products. For example, National Geographic released its popular *World Atlas* product in the Reference category as a complement to its well-regarded offline magazines. The offline brand can signal quality for the online product, and its offline activities can also improve online product quality. Consistent with these arguments, National Geographic successfully used a paid revenue model for *World Atlas*.

The ML analysis—decision tree, random forest, and penalized multinomial logit—reinforces the importance of quality signals and their supporting activities. *Marketing* (i.e., early media coverage) is particularly relevant for the paid revenue model (Table 2, Figure 1). It is the most important feature distinguishing paid v. freemium revenue models in the decision tree. The penalized multinomial logit indicates that *Marketing* (0.95) has the highest positive coefficient for identifying the paid revenue model. *Design* (0.78) and *OfflineBrand* (0.58) also have large, positive coefficients. The random forest finds all three—*Marketing*, *Design*, and *OfflineBrand*—important. In contrast, *Product Quality* per se has little relationship to the paid model, or to *any* revenue model among popular products. It is the least important feature in the random forest and has a small coefficient in the penalized multinomial logit. It is absent in the decision tree, suggesting it is not predictive of any revenue model. Thus, users seem attuned to quality signals, not quality itself, and firms with the requisite activities can create these quality signals.

There are several reasons *why* quality signals and their supporting activities *fit* with a paid revenue model. First, users are reluctant to buy without assessing quality first. Yet since users often have difficulty assessing quality, they use quality signals. Research across settings supports this argument, especially if attention is limited and quality signals are strong (e.g., Boudreau et al., 2019b; Rindova, Williamson, Petkova, & Sever, 2005). Second is salience. Quality is very perceptual, making quality signals (and their supporting activities) more salient to users than quality. A popular developer emphasized, "Value is more easily perceived on the surface, from a customer perspective."

In contrast, while many unpopular products use the paid model (33%), they often lack the requisite quality signals and supporting activities to create those signals. For example, Health & Fitness product, *How to Surf Like a Pro*, is not part of a portfolio of popular online products by the same firm (e.g., suggesting unproven design skills). It has no media coverage (e.g., consistent with few or no marketing activities), and no related products in other venues (i.e., no offline brand and supporting activities). Similar unpopular paid products such as *Yes!! I Know Plumbin* (Reference), *Ballet Dancing* (Music), and *Stem Cell Therapy* (Medical) also lack quality signals.

The ML analysis reinforces these observations. *Marketing*, *OfflineBrand*, and *Design* are not decision tree nodes (Figure A1). In the penalized multinomial logit (Table 3), *Marketing* and *OfflineBrand* predict the paid revenue model, but the prediction rate is quite low. *Design* is not predictive at all. It seems that firms mistakenly believe that their products are valuable such that users should or will pay for them. For example, a developer explained his reasoning for choosing paid (i.e., effort), "I put a lot of effort into this and I thought people should pay for it." Another described his thinking (i.e., low price), noting "Five dollars [product's price] isn't that much more than a cup of coffee." Yet they neglected quality signals. In contrast, popular products (3%) rarely use the paid model. As a popular-product developer recognized "Even $2 is a barrier." He explicitly avoided the paid model in favor of freemium.

Overall, the core challenge is building activities (e.g., marketing processes, product design skills) to generate sufficient quality signals. Monument Valley illustrates. These developers consciously chose to create a high-quality product with a paid revenue model. The lead developer noted, "A premium price made sense because we were developing a traditional premium product...more similar to a film. Pay once, see amazing things." Yet per above, they also invested extensively in marketing to build media coverage to signal that quality.

4.3 | Freemium revenue model: Bundled and fragmented

Freemium is a revenue model in which users can use the product without charge, but also pay for premium products (i.e., up-sells) that unlock added content. Freemium is the most common revenue model for popular products (62%), and fits well when firms have the design and marketing skills to create valuable *product complexity*. In contrast, the freemium model is the least common among unpopular products (11%), suggesting that firms with less popular products avoid the more complex freemium model and its demanding activities in favor of the simpler revenue models, paid and free. As one unpopular developer regretted, "I do wish we could have pulled freemium off. It's complex, it's hard to do.... We couldn't pull it off."

We uncovered two freemium models (Figure S4). In one, which we term *bundled*, the firm creates value with a free product, and adds substantially more value with one (or a few) premium up-sells from which it captures value. The bundled model fits well with complex upsells

that have many interrelated and reinforcing features (i.e., super-modularity) that users often combine. Here, it is difficult to separate the premium features and sell them separately.

A well-known popular example is *Spotify* in Music. Users can listen to shuffled music free of charge, but they can also purchase an up-sell with features like on-demand streaming (i.e., user can play a specific song) and offline mode (i.e., users can download songs). These features are highly interrelated and more useful together than separate. For example, downloading a single song is useless without the ability to play that individual track. These features also create more overall product value in the long run because the service learns which songs users play, and so provides increasingly better recommendations. The firm has surpassed 100 million subscribers and $1.9 billion in quarterly revenues (Sawers, 2019).

A random example is *iTranslate*, a popular product in Productivity. iTranslate lets users translate phrases in 90 languages, and includes voice output and bookmarking. Users can obtain a free "slim" product, but can also purchase an up-sell with voice input, offline translation, and verb conjugation features that are mutually reinforcing. For example, the value of voice input increases if it can be used offline while the value of offline translation increases if it is easy to use—for example, voice input. Freemium works well for this top grossing product (iTranslate, 2018).

In contrast, less popular products often offer many up-sell products even when their features are highly interrelated. Yet, fragmenting inter-related premium products places too many features behind separate paywalls. This makes it difficult for users to understand which features they might prefer. It is also frustrating to face extra steps and payments. As a result, users frequently avoid the up-sell product. For example, *Medical Visual Books* in the Medical category provides human anatomy graphics. It offers many up-sells, including *3D Body Anatomy*, *2D Skeleton System*, and *Lymphatic System*. Since most features are split into many small up-sells, it is difficult for users to experience the value of these interrelated human systems in the complete product. Few users upgrade, and this unpopular product has languished.

A second freemium revenue model is what we term *fragmented freemium*. Here, the firm creates value with a free product, and adds more value with many "fragmented" upsell products from which value is captured. Fragmented freemium fits when (while the free product is often complex), multiple and diverse product features can be sold separately (i.e., modular). These features are often consumables—that is, purchases that deplete after use. This model works well, in part, because users can self-select up-sells based on their interests and willingness to pay.

A well-known example is *Candy Crush Saga* in Games. This popular puzzle arcade game lets players move pieces to create three-of-a-kind matches, earn points, and beat levels. When players lose, they must wait 24 hr before playing again. Candy Crush offers many upsells such as extra lives, moves, and power-ups with which users can avoid the timeout. The variety and modularity of up-sells mean that users can choose the ones they prefer. The core game is complex (i.e., many interrelated features), but its extra features are modular. With this freemium model, *Candy Crush Saga* has generated over a billion dollars in revenues (Statista, 2018).

Another popular example is *Udemy* in Education. It offers a wide range of online courses. Using the free version, consumers can browse, read reviews, and preview lectures. Yet Udemy up-sells individual courses—that is, from $9.99 to $199.99 or more. The segmentation of courses into modules and their consumable character makes fragmented freemium an effective choice. Udemy has attracted over 10 million users taking courses across 190 countries (Udemy, 2016).

The ML analysis adds insight. The decision tree for popular products identifies *two equifinal paths* to the freemium model when *User Resources* are low (Figure 2). In the first path, popular

products with high media coverage (*Marketing* > 0.5) and *Product Complexity* can choose paid or freemium. In the second path, popular products that lack quality signals like high media coverage (paid) and user resources (third party) are forced into freemium—that is, freemium is the last resort. The penalized multinomial logit is consistent (Table 2). While *Product Complexity* is the largest positive coefficient and predicts freemium, the coefficients for *User Resources*, *Design*, *OfflineBrand*, and *Marketing* are negative, indicating that freemium is the default.

There are several reasons *why* product complexity *fits* with the freemium revenue model. First, the free version needs enough complexity to attract users and provide positive experiences that encourage purchase of the upsell(s). Second, the upsell product(s) needs added complexity to provide sufficient value beyond the free version for users to pay. Such products rely on product development activities to build a complex product and effectively segment it, and on marketing activities to identify desirable features. If the free product is too simple with too few features, users will not experience enough value to pay for the up-sell. Conversely, if the free product is too inclusive, they will not need to upgrade.

Finally, few unpopular products use the freemium model (10% vs. 62% of popular products). Our ML analysis supports this: *Product Complexity* is more predictive of paid than freemium, suggesting these developers avoid freemium. For those using freemium, some products poorly separate free and premium (e.g., Medical Visual Books) while others offer a moderate upsell number, straddling the freemium models and giving a confused user experience.

Overall, the core challenge is organizing product complexity to strike a balance between the free and premium versions. Yet creating this balance is not easy. A well-known example is *Snapchat* in Photo & Video. While the firm has a successful free version, it has struggled to find added value-creating features (for which users will pay) in its fragmented freemium revenue model. The firm tried animated photo filters for $0.99, and photo replays for $4.99. Yet even at these low prices, they were unsuccessful in earning sufficient revenue (Valinsky, 2016).[7]

5 | DISCUSSION

Intrigued by the puzzle of differing revenue models for Spotify, Pandora, and Apple Music, we asked, "What is the optimal choice of revenue model?" Past research focuses on factors like ad rates, ad-aversion, and product quality. Yet, this work has inconsistent results (e.g., product quality), lacks essential perspectives (e.g., advertiser), and proposes under-theorized constructs (e.g., freemium). More deeply, this work misses the central theoretical relationship between value creation and value capture at the heart of successful business models.

Our primary theoretical contribution is an emergent framework of several high-performing revenue model-activity system configurations (Table 4). While research often decouples these elements, our insight is that high-performing products *fit* value capture (i.e., revenue model) with value creation (i.e., activities of firm, partners, and users) to form coherent business models. In contrast, low-performing products avoid complex value capture (i.e., freemium) and misconstrue value creation (e.g., overweight effort and price), thus forming ill-fitting and

[7]We further analyzed bundled versus fragmented freemium revenue models. The penalized multinomial logit highlights that popular bundled freemium products are likely to have *Interrelated* premium (up-sell) features while fragmented freemium products have many *Consumable* up-sell features. The decision tree and random forest are consistent and highly predictive. Results available from authors.

TABLE 4 Emergent theoretical framework: Revenue model-activity system configurations

Revenue model (value capture)	Activities (value creation)	Theoretical logic	Core challenge	Unpopular products
Third-party	User activities create valuable resources	Since users do not pay, there must be a third-party payer Third-parties are likely to pay when users create value for them Firms capture some of that value	Recognizing valuable user resources (if any) Finding a third-party willing to pay for them	Over-estimate the value of their user resources Under-estimate the challenge of finding a third-party payer
Paid	Marketing activities create media coverage Design activities create a series of successful products Offline brand creates positive reputation	Users are reluctant to buy w/o assessing quality Since quality is hard to assess, users use observable and relevant quality signals (esp. when attention is limited or quality signals are strong) When activities provide sufficient quality signals of value creation, users buy Firms capture some of that value	Building activities (e.g., marketing, design, brand) that provide sufficient quality signals to persuade potential users to buy	Over-estimate importance of effort, product quality, product complexity, and low price Under-estimate importance of quality signals and their supporting activities
Freemium	Design and marketing activities create valuable product complexity Equifinal paths: #1 Products with quality signals and product complexity can choose freemium #2 Firms without valuable user resources and quality signals must choose freemium	Free version needs sufficient complexity to attract users and provide positive experiences that encourage user to buy upsell Paid (upsell) version needs added complexity to create sufficient added value for user to buy Firms capture value from paid version	Organize product complexity to strike a balance (i.e., segment) between value creation by free and paid versions	Avoid freemium revenue model even w/complex products Mis-estimate balance of free and paid (e.g., free creates too much value, paid creates too little value)

unsuccessful business models.[8] Our primary methods contribution is a novel combination of complementary methods (i.e., machine learning, multi-case theory building, EDA) to build theory using large quantitative data sets and for question-driven research.

5.1 | Emergent theoretical framework of revenue model-activity system configurations

Our framework proposes three unique and high-performing revenue model-activity system configurations that link value capture (i.e., revenue model) and value creation (i.e., activities of focal firms, users and partners) into coherent business models. First, the *third-party revenue model* fits with user resources that create value for third-party payers, with some of that value captured by the focal firm. When user resources (e.g., personal data, revealed preferences, financial assets) provide value for which a third party (e.g., advertiser, employer, investor, marketplace seller) will pay, then a third-party revenue model fits. The core challenge is figuring out what user resources (if any) are valuable and for whom.

The third-party revenue model extends prior research in two ways. First, it contributes by recognizing that advertisers are simply one type of third-party actor who may pay for user resources. Thus, it subsumes the advertising model. Second, it contributes *user resources*. While prior work argues that high ad rates (e.g., Casadesus-Masanell & Zhu, 2010; Lin et al., 2012), low ad-aversion (e.g., Gabszewicz et al., 2005), and poor product quality (e.g., Chen et al., 2016; Lin et al., 2012) matter, the fundamental source of value creation is user resources. User resources are, in fact, the *most* predictive variable for distinguishing among revenue models.

Second, the *paid revenue model* fits with focal firm activities that create enough signals of the product's value (i.e., quality) to entice users to pay before they use. We identify three signals, all tied to activities. Marketing activities (e.g., connections to influential journalists, skills in creating marketing materials) can generate media coverage about the product's value. Design activities (e.g., product development process, skilled developers) can create successive popular products that build the firm's reputation for high-quality products. Third, a well-established offline brand and its supporting activities can create a positive view of the focal product. The core challenge is building activities (i.e., processes, resources, skills) that can create sufficiently strong quality signals to persuade potential users to pay before they use.

We add to prior research. First, we contribute by identifying the relevance of *quality signals and their supporting activities* to the paid revenue model. While past work emphasizes low ad rates (e.g., Lin et al., 2012), we instead find that quality signals and their related activities fit with the paid model while ad rates actually tie to user resources and the third party model. Second, we contribute reasons for why *product quality* is less important. One is that users have difficulty assessing actual quality, and so resort to quality signals. In our study, for example, product quality measures like user ratings are often similar across products, providing little information. Indeed, research across products suggests that users often resort to quality signals, particularly when their attention is limited and quality signals are strong (Rindova et al., 2005).

[8]Our theoretical framework also solves the initial puzzle: Apple has a *paid* revenue model for its Music product fitting its design skills in past products and offline brand, both of which signal quality. Spotify initially lacked quality signals and so focused on developing a complex product with free and premium versions in the *freemium* revenue model. Pandora pioneered the science of music taste and sold its users' taste preferences with sophisticated marketing (e.g., client ad teams, tying music taste to selling concert tickets and music merchandise) in the *third-party* model.

Another reason is salience—that is, users buy based on perceptions of quality, not actual quality. Consistent with this reasoning, popular-product developers recognize the relevance of quality signals for the paid model (even for very high-quality products like Monument Valley) while unpopular-product developers focus mostly on product features and their own effort.

The *freemium revenue model* fits with a complex product supported by product development and marketing activities to design a free version that attracts users while creating enough added value in an upsell version to capture value. The core challenge is organizing product complexity to strike a balance between the free and paid versions. We extend prior research in two ways. First, we identify two unexpected *equifinal paths* to the freemium model. The first path is for well-endowed firms—that is, those with the requisite activities to create both quality signals and product complexity can choose paid or freemium. The second path is for firms with no choice—that is, they lack both valuable user resources and quality signals, and so are forced into freemium. Second, we contribute by *bifurcating freemium* into two models with distinct ways of organizing value-creating product complexity: *Bundled* with one (or a few) interrelated upsells and *fragmented* with many modular upsells. Overall, we offer a better-theorized freemium model that may clarify why there have been inconsistent results (e.g., Arora et al., 2017; Rietveld, 2018).

A key question is whether freemium is a combination of revenue models or theoretically distinct. We think that freemium is distinct. First, freemium uniquely uses product complexity (and related activities) to create value. In contrast, the third-party and paid revenue models are not associated with product complexity, and rely on different constructs and activities. Second, freemium is effective independent of user resources (needed for third-party) and quality signals (needed for paid). The ML analysis supports this independence: Freemium is directly predicted (v. other revenue models) by *Product Complexity*. Third, freemium has two distinct modes of organizing product complexity, revealing a more nuanced model than simply "free plus paid." In short, the freemium revenue model appears theoretically (and empirically) distinct.

Finally, we contribute insights into the revenue models of *less popular products*. First, their developers avoid the difficult freemium model and choose (90%) simple revenue models (i.e., third-party and paid). Second, they misalign their revenue model with activities. For example, although 57% of unpopular products use the third-party model, only 6% have the requisite user resources. Although 33% use the paid model, many lack quality signals and activities like marketing skills that can create them. Overall, their revenue models are not predictable much above chance (47% vs. 80% popular products).

What is occurring? First, developers of unpopular products are often unrealistic. Some assume that third-parties like advertisers will pay without considering what they will pay for. As a developer told us, "We thought ads were easy, but they're not." Others focus on adding users while hoping to figure out revenue. While this worked for LinkedIn (earlier), it has not for most others like Coub. Some are overly optimistic (e.g., price 30% more on average than popular products despite lower average product quality, and few if any quality signals). They justify the paid model by effort (e.g., "I put a lot into this") or price (e.g., "less than a cup of coffee"), but miss the relevance of quality signals for busy potential users to buy.

Second, these developers pay attention to the wrong variables. For example, the ML analysis finds that *Product Quality* and *Product Complexity* are the most predictive for unpopular products. This contrasts with popular products where *User Resources* matter most, followed by *Marketing* and other quality signals, and *Product Quality* is not predictive. While developers of popular products fit *Product Complexity* with freemium, developers of unpopular ones fit it with *paid*. Thus, the revenue models of unpopular products are often (a) too simple, (b) chosen for

unrealistic reasons, (c) predicted by the wrong variables (i.e., lack ties to theoretically relevant activities), and (d) less predictable (47% vs. 80%).

Finally, we examined several of the very few unpopular products that became popular (about 3%) during our study. Some like Vent (in Social Networking) began as small free products. They became popular after (a) adding product complexity (e.g., Vent added 6× more code and multiple consumable upsells), and (b) shifting to the freemium model. Others like Siddur (in Books) later tied to an offline brand (e.g., Chabad, a prominent Jewish group) and continued with the paid model. Others like Stop (in Games) began with a viable revenue model-activity system (e.g., freemium with a complex product), but were slow to take off. In sum, these turnaround products either shifted to or already had a high-performing revenue model-activity system configuration. The overall implication is that high-performing business model configurations are a necessary condition for a successful product.

5.2 | Toward a novel theory-building method

Our primary methods contribution is a novel methods combination. *Exploratory data analysis* is a systematic (and under-utilized) approach for revealing preliminary patterns in quantitative data. It can also improve multi-case theory building by identifying relevant theoretical sampling of cases. *Multi-case theory-building* recognizes theoretical patterns by systematically analyzing a small number of cases. Since this method can reveal empirically grounded theoretical patterns, it can improve ML with theoretical guidance that both narrows the search space (and so upgrades model performance; Lettau & Pelger, 2018) and enhances interpretability with meaningful theoretical constructs and mechanisms. *Machine learning* is also a pattern recognition technique. It adds precision to multi-case theory building with (a) large-scale corroboration and elaboration of patterns, and (b) precise identification of size effects, equifinal paths, nonlinearities, and configurations (all challenging with cases). In sum, we contribute a novel methods combination, useful for theory building with large quantitative data sets.

We also contribute by adding to the applications of *machine learning* in strategy and organizations (Raj & Seamans, 2019). Following Varian (2014), we note that ML is not restricted to "big data," but rather can be useful with less data such as we have. More important, we extend ML to theory-building, an application that complements theory-testing uses like better measures (e.g., Choudhury, Wang, et al., 2019), predictions (e.g., Lee et al., 2018), and causal paths (e.g., Choudhury, Allen, & Endres, 2019). Our suite of complementary, robust and relatively interpretable (not just predictive) approaches fits a theory-building aim. *Decision trees* can reveal nonlinearities like configurations and equifinal paths, *random forests* show variable importance even in highly nonlinear data, and *penalized multinomial logit* indicates size and direction effects.

We also add to *multi-case theory building* by expanding the conceptualization of a case. Prior work typically relies on a few "thick" cases with longitudinal data and a process focus (e.g., Graebner & Eisenhardt, 2004; Ott & Eisenhardt, 2020; Zuzul & Tripsas, 2019). In contrast, we use (a) many "thin" cases with largely cross-sectional data and a content focus, and (b) embed these cases within a larger quantitative analysis. Yet while our cases look "different," we use well-known theory-building steps like theoretical sampling and within- and-cross-case analysis (Eisenhardt & Graebner, 2007).

Finally, we contribute by highlighting the important complementarities and similarities of multi-case theory building and machine learning. Although distinct, both are pattern

recognition methods that use iterative comparison while guarding against overfitting and excess complexity. That is, each uses techniques to reduce overfitting—for example, *replication logic* for multi-case and *cross-validation* for ML. These techniques repeatedly revisit data slices to generate possible theory, and then test that theory on the remaining data. This reduces "noise" and finds the "best" (i.e., most predictive) model. Similarly, each reduces excess complexity—for example, *construct abstraction* for multi-case and *regularization* for ML (e.g., pruning leaves on decision trees). Yet these two powerful methods are also highly complementary: Multi-case theory-building provides theoretical constructs and mechanisms while ML provides large-scale validation and precision.

5.3 | Boundary conditions and elaborations

Like all research, ours may have boundary conditions. A potential one is digital goods, a significant, growing segment of the global economy. They are attractive for exploring the under-theorized freemium model and products for which revenue models are challenging (e.g., low marginal costs like software, experience goods). Yet as the rare use of the paid revenue model by popular products implies, digital, and physical goods (paid revenue model is common) differ.

Nonetheless, our theoretical framework likely generalizes beyond digital goods because of its *configurational logic*. That is, although the frequency of specific configurations varies across contexts, the theoretical logic within each configuration generalizes. For example, quality signals may be easier to achieve for physical goods, thus increasing use of the paid model and its configuration. Similarly, time-periods, geography, and "fashion waves" may imprint specific configurations (see Eisenhardt, 1988 for an example of founding date predicting compensation configurations). To illustrate, freemium may be "fashionable" now, but high-performing configurations using freemium likely include activities for creating complex products. In sum, while various factors (e.g., physical goods, fashion, geography) may affect the *frequency* of specific configurations, the theoretical logic within these configurations still likely holds.

Another potential boundary condition is product simplicity. For example, industries like finance and real estate may have more complex revenue models than we study. The residential solar industry provides a possible counter-example. When the industry was disrupted by tax law changes, some firms developed a novel revenue model—that is, free electricity to homeowners, but sell user resources (i.e., rooftops with solar panels) to third-party investors who could benefit from tax credits (Hannah & Eisenhardt, 2018). While the details of this finance revenue model are novel (and even complex), it is still a third-party model. Similarly, real estate brokers combine two core revenue models: Paid for services like listing (sellers) and third-party ("selling" interested buyers to sellers; buyers). Broadly, our revenue model-activity system configurations may well fit complex products (e.g., residential solar) and/or are building blocks for complicated models (e.g., real estate). This is an avenue for future research.

Strategic context is another potential boundary condition. One context is firms with product portfolios that underprice some products like the well-known "razor and blades." Yet at its core, this is a fragmented freemium model—that is, "razor" has a very low price and revenue comes mostly from selling many (consumable) "blades." Another context is firms that launch a free product to grow users, and then hope to find a third-party payer (e.g., LinkedIn) or transition to freemium (e.g., Vent). But these are *paths* to the high-performing business models that we identify, *not* different configurations. Future research could explore other strategic contexts.

Finally, we address causality. Is the revenue model chosen first? Since our theoretical framework is configurational, it is about *fit*. It is inherently not causal. Nonetheless, our cases offer insight into how firms design fit. For some products like Monument Valley, firms choose the revenue model and elaborate the activities. For others like Pandora, firms choose activities and then figure out the revenue model. Consistent with research on how entrepreneurs design business models (McDonald & Eisenhardt, 2019) and broadly, strategies (Ott & Eisenhardt, 2020), some like OpenTable iteratively design fit by experimentation and trial-and-error. Others like many Game products follow templates of successful predecessors. Overall, multiple paths (not a specific causal direction) exist by which firms can reach a successful revenue model-activity system configuration.

We began with an empirical puzzle: Why do similar products have differing revenue models. Our primary contributions are (a) an emergent theoretical framework of unique and high-performing business model configurations of *value capture* (i.e., revenue models) and *value creation* (i.e., activity systems) and (b) a novel theory-building method that combines machine learning, multi-case analysis, and exploratory data analysis. The next steps are to test our framework and apply our novel method in emerging theory-building applications.

ACKNOWLEDGMENTS

We thank Yorgos Askalidis, Kevin Boudreau, Prithwiraj Choudhury, Gary Dushnitsky, Milan Miric, Jason Rathje, Joost Rietveld, Robert Seamans, Mary Tripsas, and Eric Volmar for helpful comments. We appreciate feedback from participants at the Wharton Technology and Innovation Conference, Strategy Science Conference, Academy of Management meetings, and INSEAD. Kira Kalkus provided excellent research assistance. We are grateful to our editors and anonymous reviewers for their guidance, and the Stanford Technology Ventures Program and Stanford Graduate Fellowship for financial support. A prior version won the 2019 Glueck Best Paper Award.

ORCID

Ron Tidhar https://orcid.org/0000-0001-8122-9840

REFERENCES

Andries, P., Debackere, K., & Van Looy, B. (2013). Simultaneous experimentation as a learning strategy: Business model development under uncertainty. *Strategic Entrepreneurship Journal*, *7*(4), 288–310.

Apple. 2017. App Store shatters records on New Year's Day. Retrieved from https://www.apple.com/newsroom/2017/01/app-store-shatters-records-on-new-years-day/.

Apple. 2019. Developer earnings from the App Store top $70 billion. Retrieved from https://www.apple.com/newsroom/2019/01/apple-entrepreneur-camp-kicks-off-as-app-developer-earnings-hit-new-record/.

Arora, S., Ter Hofstede, F., & Mahajan, V. (2017). The implications of offering free versions for the performance of paid mobile apps. *Journal of Marketing*, *81*(6), 62–78.

Askalidis, G. (2015). The impact of large scale promotions on the sales and ratings of Mobile apps: Evidence from Apple's app store. *arXiv 1506.06857*. https://arxiv.org/pdf/1506.06857.pdf.

Athey, S. (2018). The impact of machine learning on economics. In A. K. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence—An Agenda*. Chicago, IL: University of Chicago Press.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131–160.

Berends, H., Smits, A., Reymen, I., & Podoynitsyna, K. (2016). Learning while (re) configuring: Business model innovation processes in established firms. *Strategic Organization*, *14*(3), 181–219.

Bort J. 2015. How Fitness22 is making 'millions of dollars a year' selling iPhone apps with no marketing. *Business Insider*. Retrieved from http://www.businessinsider.com/how-fitness22-is-making-millions-on-apps-2015-8.

Boudreau, K, Jeppesen, LB, Miric, M. 2019a. Competing on free(mium): Digital competition with network effects. Available at SSRN 2984546.

Boudreau, K. J., Jeppesen, L. B., & Miric, M. (2019b). *Sustaining competitive advantage in turbulent digital marketplaces [Working paper]*. Boston, MA: Northeastern University.

Bresnahan, T. F., Davis, J. P., & Yin, P. L. (2014). Economic value creation in mobile applications. In A. B. Jaffe & B. F. Jones (Eds.), *The changing frontier: Rethinking science and innovation policy* (pp. 233–286). Chicago, IL: University of Chicago Press.

Casadesus-Masanell, R., & Zhu, F. (2010). Strategies to fight ad-sponsored rivals. *Management Science*, *56*(9), 1484–1499.

Casadesus-Masanell, R., & Zhu, F. (2013). Business model innovation and competitive imitation: The case of sponsor-based business models. *Strategic Management Journal*, *34*(4), 464–482.

Chen, J., Fan, M., & Li, M. (2016). Advertising Versus Brokerage Model for Online Trading Platforms. *MIS Quarterly*, *40*(3), 575–596.

Choudhury, P., Allen, R., & Endres, M. G. (2019). *Machine learning methods framework for management research: Application to exploratory pattern detection [Working paper]*. Harvard Business School: Boston, MA.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, *40*(11), 1705–1732.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016.

Eckhardt, J. T. (2016). Welcome contributor or no price competitor? The competitive interaction of free and priced technologies. *Strategic Management Journal*, *37*(4), 742–762.

Eisenhardt, K. M. (1988). Agency- and institutional-theory explanations: The case of retail sales compensation. *Academy of Management Journal*, *31*(3), 488–511.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.

Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal*, *59*(4), 1113–1123.

Etherington, D. 2015. Monument Valley team reveals the cost and reward of making a hit iOS game. *TechCrunch*. Retrieved from https://techcrunch.com/2015/01/15/monument-valley-team-reveals-the-cost-and-reward-of-making-a-hit-ios-game/.

Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2005). Does advertising lower the price of newspapers to consumers? A theoretical appraisal. *Economics Letters*, *87*(1), 127–134.

Gambardella, A., & McGahan, A. M. (2010). Business-model innovation: General purpose technologies and their implications for industry structure. *Long Range Planning*, *43*(2), 262–271.

Garg, R., & Telang, R. (2013). Inferring app demand from publicly available data. *MIS Quarterly*, *37*(4), 1253–1264.

Gassmann, O., Frankenberger, K., & Csik, M. (2014). *The business model navigator: 55 models that will revolutionise your business*. London: Pearson.

Glaeser, E., Kim, H., & Luca, M. (2017). *Nowcasting the local economy: Using yelp data to measure economic activity [Working paper 18-022]*. Boston, MA: Harvard Business School.

Gow, I. D., Kaplan, S. N., Larcker, D. F., & Zakolyukina, A. A. (2016). *CEO personality and firm policies [Working paper 22435]*. Cambridge, MA: National Bureau of Economic Research.

Graebner, M. E., & Eisenhardt, K. M. (2004). The seller's side of the story: Acquisition as courtship and governance as syndicate in entrepreneurial firms. *Administrative Science Quarterly*, *49*(3), 366–403.

Hallen, B. L., Davis, J. P., & Yin, P. L. (2018). *Network isolates: Entrepreneurial bootstrapping and the social disconnection of new organizations in the Mobile app ecosystem [Working paper]*. Washington, DC: University of Washington.

Hamari, J., Hanner, N., & Koivisto, J. (2017). Service quality explains why people use freemium services but not if they go premium- an empirical study in free-to-play games. *International Journal of Information Management*, *37*(1), 1449–1459.

Hannah, D. P., & Eisenhardt, K. M. (2018). How firms navigate cooperation and competition in nascent ecosystems. *Strategic Management Journal*, *39*(12), 3163–3192.

iTranslate. 2018. Your Passport to the World!. *iTranslate*. Retrieved from https://www.itranslate.com/.

Jung, J, Concannon, C, Shroff, R, Goel, S, Goldstein, DG. 2017. Simple rules for complex decisions. SSRN 2919024.

Lachapelle, T. 2016. That OpenTable Deal. *Bloomberg*. Retrieved from https://www.bloomberg.com/gadfly/articles/2016-11-18/priceline-s-troubled-opentable-deal.

Lee, C., Menon, A., & Tabakovic, H. (2018). *Using machine learning to predict high-impact general technologies [Working paper]*. Ithaca, NY: Cornell University.

Lettau, M., & Pelger, M. (2018). *Estimating latent asset-pricing factors (No. w24618)*. Cambridge, MA: National Bureau of Economic Research.

Lin, M., Ke, X., & Whinston, A. B. (2012). Vertical differentiation and a comparison of online advertising models. *Journal of Management Information Systems*, *29*(1), 195–236.

Liu, CZ, Au, YA, Choi, HS. 2012. An empirical study of the freemium strategy for mobile apps- evidence from the google play market. International Conference on Information Systems Proceedings. Retrieved from https://aisel.aisnet.org/icis2012/proceedings/EconomicsValue/14/.

Lomas N. 2014. Monument Valley, ustwo's sumptuous Escher-inspired iOS game, lands globally on April 3. *TechCrunch*. Retrieved from https://techcrunch.com/2014/03/26/monument-valley-price-and-launch-date/.

Massa, L., Tucci, C. L., & Afuah, A. (2017). A critical assessment of business model research. *Academy of Management Annals*, *11*(1), 73–104.

McDonald, R., & Eisenhardt, K. M. (2019). Parallel play: Startups, nascent markets, and effective business-model design. *Administrative Science Quarterly*. https://doi.org/10.1177/0001839219852349.

Menon, A., Nave, G., & Bhatia, S. (2019). *Emotional expressions predict risky decisions by S&P 500 executives [Working paper]*. Philadelphia, PA: University of Pennsylvania.

Ott, T. E., & Eisenhardt, K. M. (2020). Decision weaving: Microfoundations of strategy formation in nascent markets. *Strategic Management Journal*.

Piskorski, M. (2007). *LinkedIn*. Boston, MA: Harvard Business School Case Clearing House.

Puranam, P., Shrestha, Y. R., He, V. F., & von Krogh, G. (2018). *Algorithmic induction through machine learning: Opportunities for management and organization research [Working Paper]*. INSEAD: Fontainebleau, France.

Raj, M., & Seamans, R. (2019). Primer on artificial intelligence and robotics. *Journal of Organization Design*, *8*(1), 1–14.

Rathje, J., & Katila, R. (2020). *Supervised machine learning and quasi experimental designs: Machine learning-matching methods for causal inference in strategy research [Working Paper]*. Stanford, CA: Stanford Technology Ventures Program.

Reuters. 2017. BRIEF-WebMD sees FY 2017 revenue about $710 million to $730 million. *Reuters*. Retrieved from: https://www.reuters.com/article/brief-webmd-sees-fy-2017-revenue-about-7/brief-webmd-sees-fy-2017-revenue-about-710-million-to-730-million-idUSASA09M5W.

Rietveld, J. (2018). Creating and capturing value from freemium business models: A demand-side perspective. *Strategic Entrepreneurship Journal*, *12*(2), 171–193.

Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being good or being known- an empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal*, *48*(6), 1033–1049.

Sawers, P. 2019. Spotify grows users 30% in Q3 2019, premium subscribers reach 113 million. *VentureBeat*. Retrieved from: https://venturebeat.com/2019/10/28/spotify-grows-users-30-in-q3-2019-premium-subscribers-reach-113-million/.

Shapiro, C., & Varian, H. R. (1998). *Information rules: A strategic guide to the network economy*. Cambridge, MA: Harvard Business Press.

Snihur, Y., & Zott, C. (2019). The genesis and metamorphosis of novelty imprints: How business model innovation emerges in young ventures. *Academy of Management Journal*. https://doi.org/10.5465/amj.2017.0706.

Statista. 2018. Annual revenue of King.com from 2010 to 2016 (in million U.S. dollars). Retrieved from: https://www.statista.com/statistics/288974/king-annual-revenue/ [18 January, 2018].

Udemy, 2016. A year of Udemy: 2016 in review. *Udemy*. Retrieved from: https://about.udemy.com/udemy-for-business/a-year-of-udemy-2016-in-review/. [30 January, 2018].

Valinsky J. 2016. Snapchat quietly kills in-app purchase option. *Digiday*. Retrieved from: https://digiday.com/media/snapchat-shelves-snap-replays/.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Verhage, J. 2019. Robinhood is worth $7.6 billion after new funding round. *Bloomberg*. Retrieved from: https://www.bloomberg.com/news/articles/2019-07-22/robinhood-is-worth-7-6-billion-after-new-funding-round.

Yin, P. L., Davis, J. P., & Muzyrya, Y. (2014). Entrepreneurial innovation: Killer apps in the iPhone ecosystem. *The American Economic Review*, *104*(5), 255–259.

Zott, C., Amit, R., & Massa, L. (2011). The business model: Recent developments and future research. *Journal of Management*, *37*, 1019–1042.

Zuzul, T., & Tripsas, M. (2019). Start-up inertia versus flexibility: The role of founder identity in a nascent industry. *Administrative Science Quarterly*. https://doi.org/10.1177/0001839219843486.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Tidhar R, Eisenhardt KM. Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strat Mgmt J*. 2020;41: 1245–1273. https://doi.org/10.1002/smj.3142

APPENDIX A—MEASURES

Revenue model

We measure whether the product is a *sales channel* with a single binary variable, *Transaction* = 1 if the product is a conduit to sales of a nonapp (e.g., physical) product like *Domino's Pizza* and *Target*. We exclude these apps because they are not products. We measure *third-party* free revenue model with the binary variable *ThirdParty* = 1 (otherwise 0) if the product is free (i.e., up-front price is zero), offers no up-sells, *and* is not a sales channel. We conducted online searches to confirm that third parties are the primary source of revenues, if any. We measure bundled freemium revenue model with the binary variable *BundledFreemium* = 1 if the product is free up-front and has fewer than four up-sells. Four is the median number of up-sells. We measure fragmented freemium revenue model with the binary variable *FragmentedFreemium* = 1 if the product is free up-front and has four or more up-sells. We measure paid revenue model with the binary variable *Paid* = 1 if the product has an up-front price. When products offer temporary free trials (e.g., N.Y. Times offers a free month) or are subscriptions, we code these as *paid*.

Independent Variables

For easy interpretation and avoiding over-fitting, we use binary predictors (Varian, 2014). We measure quality signals with three binary variables. *Marketing* indicates whether the firm conducts marketing activities that coincide or precede product release. *Marketing* = 1 if the product is referenced in any media reports within 6 months of launch. *Design* is a binary variable = 1 if the developer has at least one other paid popular product, indicating prior design success and design activities that support that success. *OfflineBrand* = 1 if the focal product has a closely related offline product that already exists. We hand-collected these three measures with online

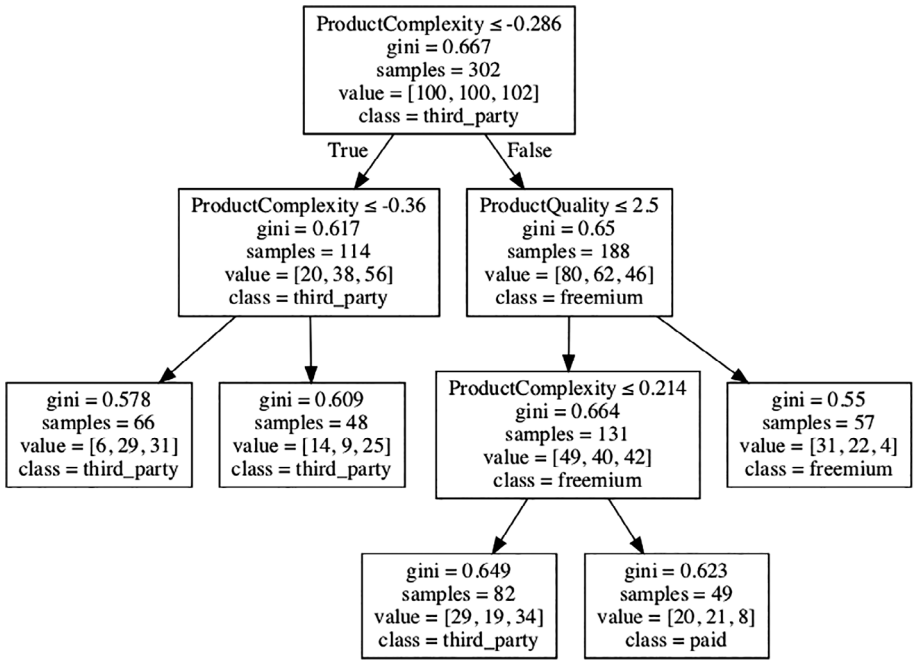

FIGURE A1 Decision tree (unpopular products): Classifies revenue models. Test accuracy (i.e., percent correctly classified) = 47%. Optimal hyperparameters: Minimum samples at leaf = 40, minimum samples at a split = 5, maximum depth = 3. Training accuracy = 47%. Cross-validation accuracy = 39% (standard deviation of 9%). Free is third party only

searches and used Mechanical Turk (2 workers) as robustness checks. In the case of disagreements, we re-examined the data to adjudicate.

We measure product quality with a single variable, *ProductQuality* = average of all user ratings (out of five stars) from the product's App Store profile page. To avoid over-fitting, we round to the nearest integer. We measure user resources with a single binary variable, *UserResources* = 1 if the product collects information or other resources (e.g., financial assets) about users that are valuable for third-parties. Examples are user preferences (e.g., *OpenTable*), problems (e.g., *WebMD*), and spending (e.g., *Mint*). We use a binary measure to provide a conservative, robust measure. We use two independent raters for the full sample (1,009 products). Inter-rater reliability = 91%. The raters resolved any disagreements by discussion and if needed, a third independent rater. We measure product complexity with *ProductComplexity*. This measure = size of the product (in megabytes), normalized for the product's market category to remove any category effects (e.g., *Games* on average bigger than *Finance* products). To normalize each size variable, we subtract from it the mean size of the given product's category and divide by the standard deviation in product sizes of the given product's category.