Research Statement

Vineet Kumar

My research focus is at the intersection of technology and business & society.[1] Given my background and training in engineering, economics and marketing, I am interested in two areas.

(A) *Digital business models:* My research focuses on long-run strategic choices relevant to digital firms. I develop and use methods based on micro-foundations of agents' preferences, typically empirical structural models. These models yield economically interpretable estimates, enabling counterfactual evaluation of firm or regulatory policies, and technological changes.

(B) *Theory-based machine learning (ML):* I develop and use methods incorporating structured knowledge developed from first principles, to enable human interpretability, and provide representations satisfying desirable properties (e.g. monotonicity of demand curves).

Theory or structured knowledge plays a central methodological role in my research in both areas. I expand on the areas below.

(A) Digital Business Models

Digital firms face unique issues, such as open versus closed source, indirect network effects, or freemium with perpetually free products. My research can broadly be themed as focusing on three different sources of connections: between *products*, *consumers* and *data*.

1) Connections across Products – Product Line

Product design is especially important to digital firms. I detail the connections between products, with a research overview and agenda provided in [P1].

Bundling Hardware and Software: Bundling is a flexible product strategy commonly used across several markets. While past research highlights how bundling can exploit negative correlations in consumers' static valuations across products, its dynamic effects are less understood. In [P4], I develop a forward-looking model to empirically evaluate its long-term impact, with data from videogame consoles (hardware) and games (software). This model shows that bundles pull demand forward, prompting lower valuation consumers to purchase earlier rather than delay – a mechanism identifiable only in a forward-looking context.

Through this model, I explore several strategic questions. First, I find that bundling increases sales volume and shifts purchasing temporally, which is crucial since a larger hardware installed base leads to multiple software purchases made earlier. Second, I demonstrate that mixed bundling outperforms pure bundling. Third, a more positive correlation in valuations between hardware and software results in higher sales through this mechanism.

My results, driven by a dynamic mechanism, contrast with existing research that considers static settings and finds bundling more effective with negative correlation. Additionally, I demonstrate that bundling is more effective when indirect network effects are weaker, a previously unknown

[1] I write in the first person singular, while acknowledging and recognizing collaborations with co-authors.

finding. Overall, my research has significant implications for any context where dynamic factors are important.

Methodologically, I was the first to integrate software and hardware markets using a structural model, instead of using quantity as a proxy for indirect network effects like prior studies. In this model, consumers form expectations about both hardware and software value when purchasing hardware. I also develop a novel identification strategy for correlation in valuations based on the tying ratio, leveraging the fact that consumers purchase one hardware unit, but multiple software units.

Related Methodological Studies: This research prompted me to explore methods for estimating dynamic demand models. Examining the widely used inclusive value approach, I demonstrate that it can result in highly biased estimates of key quantities like elasticities and profits [P5]. Motivated by the need for models that flexibly accommodate dynamics with large state spaces, I develop a new method that can be estimated without reducing dimensionality [P3]. For a wide class of problems involving terminating or renewal choices, I obtain consumer preferences with the simplicity of a linear regression. I prove identification and recover the evolution of unobservable product characteristics, enabling counterfactual analysis. Applying this method, I evaluate the monetary value of product features in the early smartphone market, finding that Bluetooth and Wi-Fi had the biggest impact on Apple's sales compared to other features.

(Digital) Transformation: Technological transformation, linking old products to new ones, is crucial and inherently risky for firms. Empirical structural studies of transformation impact are rare. In [P13], I examine the implications of the shift from an older technology (physical or slower) to newer technology (digital or faster), using a panel from a firm employing the "Netflix" model. I determine consumer preferences for viewing (physical) content. Typically, quality improvements are expected to increase consumer value and firm revenue too. However, I show that with service improvements, this logic may not hold. The reason is that such improvements also reduce the differentiation between the product versions (plans). In the limit (zero service time), just one plan would be offered, reducing the firm's ability to price discriminate and capture surplus. My counterfactual analysis with reduced service time reveals that while created value increases, differentiation across versions decreases, potentially collapsing. This mechanism demonstrates that even transformations creating more value can lead to lower profits and revenue. This risk is prevalent in many transformations, especially digital ones. Firms must therefore understand, evaluate, and develop new versioning strategies to mitigate these transformation risks.

Methodologically, I develop a dynamic structural model with forward-looking consumers, specifying a multiplicative error structure (rather than discrete choice logit) to avoid selection of dominated plans. I incorporate a rich specification of heterogeneity, crucial for the mechanism examined. I propose a novel strategy for identification of switching costs without price variation, based on content enhancement that increases gains from switching. This approach is broadly applicable, such as for app subscriptions.

Open Source: Open source software presents a puzzle: free-riding can produce high-quality products. My model shows how this is sustained in equilibrium [P14]. I examine product strategy, where contributions to product features made by developers or firms are a public good available to all competitors (e.g. Linux, Android or Meta's Llama LLM). Developers signal their capabilities by making feature contributions to open source. Firms build on features and differentiate on complementary dimensions like usability. A higher degree of open source contributions can enhance the differentiation value of usability. I show that, contrary to past research and industry leaders' beliefs, allowing free-riding can lead to higher product quality.

Other Related Work: Freemium is the most popular digital business model (e.g. app stores, cloud storage, SaaS). I undertake a deep dive into designing freemium for a storage service [P16], where referrals are rewarded with an improvement to the free product. I evaluate the tradeoff between growth and monetization, and find that excessive generosity can counterintuitively hurt both outcomes. I also examine the use of time as a versioning and monetization strategy (wait for free) in [P17]. I provide an overview of issues in Freemium for a general audience in [P10]. I also examine the case of a retailer with a physical and digital channel, and evaluate when it is optimal for the firm to match its own price across channels [P9].

2) Connections across Data – Linking Purchase and Usage

I study the connection between types of data: purchase data and usage data. In digital settings, usage data is more readily available and valuable for insights into consumer preferences. However, most marketing and economics studies involve only purchase data. I connect purchase and usage data across several product settings, including [P16, P13]. In [P2], I show how usage data is conceptually distinct and critically important for identifying demand, not just estimating it. Specifically, I demonstrate that without price variation, nonparametric identification of the Willingness to Pay (WTP) distribution for subscriptions is achievable by leveraging usage data, but impossible otherwise. Thus, combining high-frequency usage data with purchase data allows for a conceptual leap in identification of the valuation (WTP) distribution, which was previously thought impossible without price variation. Crucially, this result does not rely on specific parametric functional forms or distribution of shocks. I model the stream of usage utilities over time, connecting the (expectations of) aggregated utility to the purchase decision. I first estimate usage preferences from usage data along with exogenous factors impacting usage, then combine expected usage value with purchase data to obtain the WTP distribution. I conduct counterfactual analyses, such as product design. The framework is flexible in accommodating a large class of usage utility models, making it widely applicable.

3) Connections across Consumers – Networks

Word-of-mouth and referral effects can significantly impact business models. However, my research in this area is more broadly applicable to a wide class of network interventions, e.g. public health. In network interventions, highly-connected individuals are useful to leverage as seeds for interventions. I investigate privacy-sensitive methods for leveraging network structure to obtain higher-degree nodes in *unknown* networks [P15]. The literature focuses on obtaining the entire network structure (not privacy-friendly!), and does not offer provable guarantees, which is challenging in networks. I address both issues. I demonstrate two strategies (ego-based and alter-based), based on the friendship paradox, that ask individuals to nominate random friends. These strategies are simple, mathematically distinct, and offer theoretical guarantees of identifying high-degree individuals. I also discover a new network property (inversity) that perfectly determines which strategy is more effective, based on network structure.

I empirically study whether using friendship paradox strategies can achieve greater product adoption in [P12]. Using a model of communication and adoption, estimated with data on real networks, I evaluate counterfactuals using various seeding strategies. I find that ego-based friendship seeding outperforms random seeding, and even achieves higher adoption than leader-based seeding. The results are robust across different specifications and networks. Methodologically, I introduce a novel nonparametric identification strategy that uses the adoption trajectory's shape to identify the differential impact of leaders, even with leader-only seeding. This approach offers a privacy-friendly approach to seeding, contrasting with the extant literature focus on maximizing user data required.

(B) Theory-based Machine Learning

My research in machine learning (ML) integrates structured knowledge (theory) to develop new ML methods, facilitating insights into consumer responses.[2]

There are several important research questions where ML is essential. While marketing often involves unstructured data like text, visual design, music, and videos, most research has focused on structured numerical data due to availability of standardized analysis methods and data. Traditional non-ML quantitative methods struggle to capture the nuances of unstructured data or generate novel text or images.

Given their growing capabilities, ML models are increasingly used in academia and industry. However, they are typically opaque black boxes, leading to significant issues. First, these complex models (e.g., ChatGPT) have billions of parameters, whose interpretations are unknown, and their training data is often undisclosed. Second, they are atheoretical, lacking a true understanding of consumers (e.g. the same models are applied in medicine). Third, they are not interpretable – we don't know how they work, why and when they fail.[3]

My focus is on addressing these challenges by developing ML methods based on theory or structured knowledge.[4] My research is focused around three aspects: (a) develop methods to incorporate structured knowledge (from marketing, economics and other fields) into ML models, (b) provide model and data transparency, and (c) improve explainability and interpretability along with performance. Several questions arise: What concepts and ideas can we bring to ML? How can we incorporate theory into complex deep learning models with unstructured data? How can we demonstrate impact? I detail these critical challenges across various studies detailed below.

In [P18], I aim to automatically obtain and quantify interpretable visual characteristics of products, and consumer preferences for these characteristics. Visual appearance is crucial in many

[2] My interest in ML was sparked during my PhD classes in ML, including an impactful one from a pioneer in the field, Tom Mitchell, who wrote a foundational text. Tom has kindly acknowledged my input in the second edition of his book. I continued my learning, attening Neurips and ML conferences over the years.

[3] Algorithms for self-driving vehicles illustrate these points, and the resulting impact. See: The Hidden Autopilot Data That Reveals Why Teslas Crash (WSJ, 07/30/2024).

[4] Other researchers are also integrating academic theories into ML, e.g. in physics [O1].

categories, but is high-dimensional and challenging to interpret. Prior research either required human experts or focused on non-interpretable characteristics. My theory-based method automatically discovers and quantifies visual characteristics *without expert input*. Theory is crucial in two ways: first, in specifying an objective function that incentivizes low-dimensional, orthogonal representations based on product design properties; second, by leveraging the idea that products have distinct brand-specific "visual signatures" (e.g., BMW cars or LV handbags). By transforming the learning problem from unsupervised to supervised using brand and other functional characteristics, I achieve higher performance and greater human interpretability than extant methods. Moreover, we can automatically generate counterfactual visual designs without experts, representing a significant methodological leap not thought possible for products.

Music and Emotion: Music significantly drives consumer emotions but is underexplored. In [P7], I explain why listeners feel specific emotions when listening to music, breaking from the previous black box approach. I use theory from multiple fields on consonance and dissonance and their connection to the listener's emotion: (a) the mathematics of sound waves and (b) psychology of human music perception. This theoretical framework underpins the creation of novel non-contiguous consonance filters within a convolutional neural network, unlike traditional contiguous filters in traditional (visual) ML. These filters allow us to visualize how specific music features impact listener responses, enhancing explainability. I also develop an application using emotional congruence or contrast for contextual ad targeting, offering a privacy-friendly approach without using consumer data.

Learning Unknown Demand Curves: Learning Unknown Demand Curves: To learn an unknown demand curve through experimentation, I develop a reinforcement learning model with nonparametric multi-armed bandits (MAB) [P19]. Classic A/B testing is inefficient as it explores all prices equally. MABs offer a more sophisticated "learning while earning" approach, but they are atheoretical. Economic theory features downward sloping demand curves, yet incorporating this notion into MABs is challenging.[5] I find that monotonicity provides dual benefits. First, it enhances algorithm performance by leveraging informational externalities across prices (arms). Second, it guarantees a monotonic demand curve. This guarantee is crucial for automated pricing decisions, as non-monotonicity can lead to unrealistic high prices and algorithmic failures.

Overall, my ML-based research incorporates theory to enhance ML capabilities in performance, interpretability, and providing representations satisfying desirable properties. These methods yield valuable insights into consumer and firm choices, fostering a symbiotic knowledge creation process. All my ML research is transparent, with open-source code available for examination, critique, and further development, thus improving stakeholder trust and acceptance of AI systems.

[5] Exceptions include Veblen goods.

Teaching

I have developed and taught elective courses at the master's level and contributed to the executive MBA and executive education programs. My teaching combines lectures on principles with Socratic discussions of case studies. In the *Digital Strategy* course, I explore digital business models (e.g., Dropbox), evaluate models used by disruptors and complementors, investigate platform success drivers and barriers, and examine digital transformation and emerging technologies. This course has been utilized in custom executive education programs.

In another elective, *Artificial Intelligence: Strategy & Marketing*, my goal is to help students understand the ideas, gain familiarity with the methods and their business applications. I introduce the primary ML methods (unsupervised, supervised, reinforcement and generative), followed by an examination of how organizations obtain value from AI, through case studies. I demonstrate how fairness (and bias) are often designed into algorithms, whether intentionally or not, and the resulting implications for stakeholders. In pedagogical experiments, I have found that assigning students to present (carefully curated) academic papers can be effective, with faculty guidance.

References

[P1] Steve Berry, Ahmed Khwaja, Vineet Kumar, Andres Musalem, Kenneth C Wilbur, Greg Allenby, Bharat Anand, Pradeep Chintagunta, W Michael Hanemann, Przemek Jeziorski, et al. Structural models of complementary choices. *Marketing Letters*, 25:245–256, 2014.

[P2] Cheng Chou and Vineet Kumar. Estimating Demand for Subscription Products: Identification of Willingness to Pay Without Price Variation. *Marketing Science*, 2024.

[P3] Cheng Chou, Tim Derdenger, and Vineet Kumar. Linear Estimation of Aggregate Dynamic Discrete Demand for Durable Goods without the Curse of Dimensionality. *Marketing Science*, 38(5):888–909, 2019.

[P4] Timothy Derdenger and Vineet Kumar. The Dynamic Effects of Bundling as a Product Strategy. *Marketing Science*, 32(6):827–859, 2013.

[P5] Timothy Derdenger and Vineet Kumar. Estimating Dynamic Discrete Choice Models with Aggregate Data: Properties of the Inclusive Value Approximation. *Quantitative Marketing and Economics*, 17(4):359–384, 2019.

[P6] Hortense Fong, Vineet Kumar, Anay Mehrotra, and Nisheeth K Vishnoi. Fairness for AUC via Feature Augmentation. *arXiv preprint arXiv:2111.12823*, 2021.

[P7] Hortense Fong, Vineet Kumar, and K Sudhir. A Theory-based Interpretable Deep Learning Architecture for Music Emotion. *Marketing Science 44(1)*, 2025.

[P8] Soheil Ghili, Vineet Kumar, and Fei Teng. Spatial Distribution of Access to Service: Theory and Evidence from Ridesharing. *Management Science (Accepted)*, 2025.

[P9] Pavel Kireyev, Vineet Kumar, and Elie Ofek. Match Your Own Price? Self-Matching as a Multichannel Retailer's Pricing Strategy. *Marketing Science*, 36(6), 2017.

[P10] Vineet Kumar. Making 'Freemium' Work. *Harvard Business Review*, 92(5):27–29, 2014.

[P11] Vineet Kumar and Kannan Srinivasan. Predicting Customer Value using Clumpiness – Commentary. *Marketing Science*, Mar-Apr, 2015.

[P12] Vineet Kumar and K Sudhir. Can Random Friends Seed More Buzz and Adoption? Leveraging the Friendship Paradox. *Management Science 71(1)*, 2025.

[P13] Vineet Kumar and Yacheng Sun. Designing Pricing Strategy for Operational and Technological Change. *Management Science*, 66(6):2706–2734, 2020.

[P14] Vineet Kumar, Brett R Gordon, and Kannan Srinivasan. Competitive Strategy for Open Source Software. *Marketing Science*, 30(6):1066–1078, 2011.

[P15] Vineet Kumar, David Krackhardt, and Scott Feld. On the Friendship Paradox and Inversity: A Network Property with Applications to Privacy-sensitive Network Interventions. *Proceedings of the National Academy of Sciences*, 121(30), 2024.

[P16] Clarence Lee, Vineet Kumar, and Sunil Gupta. Designing Freemium: Strategic Balancing of Growth and Monetization. *Available at SSRN 2767135*, 2019.

[P17] Peter S Lee, Vineet Kumar, and K Sudhir. Monetizing Serialized Content: How Wait for Free Impacts Paid and Free Consumption. *Working Paper*, 2025.

[P18] Ankit Sisodia, Alex Burnap, and Vineet Kumar. Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. *Journal of Marketing Research 62(3)*, 2025.

[P19] Ian Weaver, Vineet Kumar, and Lalit Jain. Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. *Marketing Science 44(6)*, 2025.

Other References

[O1] George Em Karniadakis et al. Physics-informed machine learning. *Nature Reviews Physics.* 3(6). 2021