

# Predictive Customer Analytics

*MGT 100 Week 8*

[Kenneth C. Wilbur](#)

*Professor of Marketing and Analytics*

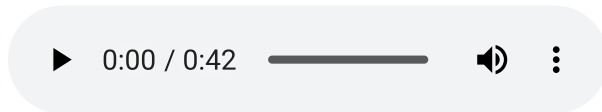
*University of California, San Diego*

[Daniel Yavorsky](#)

*SVP-Analytics*

*GBK Collective*

This version: May 2026 | License: CC BY 4.0 | We use javascript to track readership.  
We welcome reuse with attribution. Please share widely.



[\[download theme song\]](#)

# Predictive Customer Analytics

- Importance of Customer Acquisition
- Market size: How many customers experience the core need?
- Diffusion: How does the served market change over time?
  - Model based on **Bass (1969)**
- CLV: How profitable are customer relationships?

# Importance of Customer Acquisition

- **Einav et al. (2021)** analyzed all US Visa CC transaction data
  - >\$1T spent in 32B transactions by 428MM cards at 1MM stores from 2016-19
  - Assume card~customer

$$\text{BrandRev.} = \sum \text{spend} \equiv \sum \frac{\text{stores}}{1} \frac{\text{cards}}{\text{stores}} \frac{\text{transactions}}{\text{cards}} \frac{\text{spend}}{\text{transactions}}$$

- Research question: How well does each factor correlate with brand revenue?
- Regressed log revenue on log RHS with merchant and year fixed effects

Which of the four factors do you expect to explain most of the variation in revenue? Why?

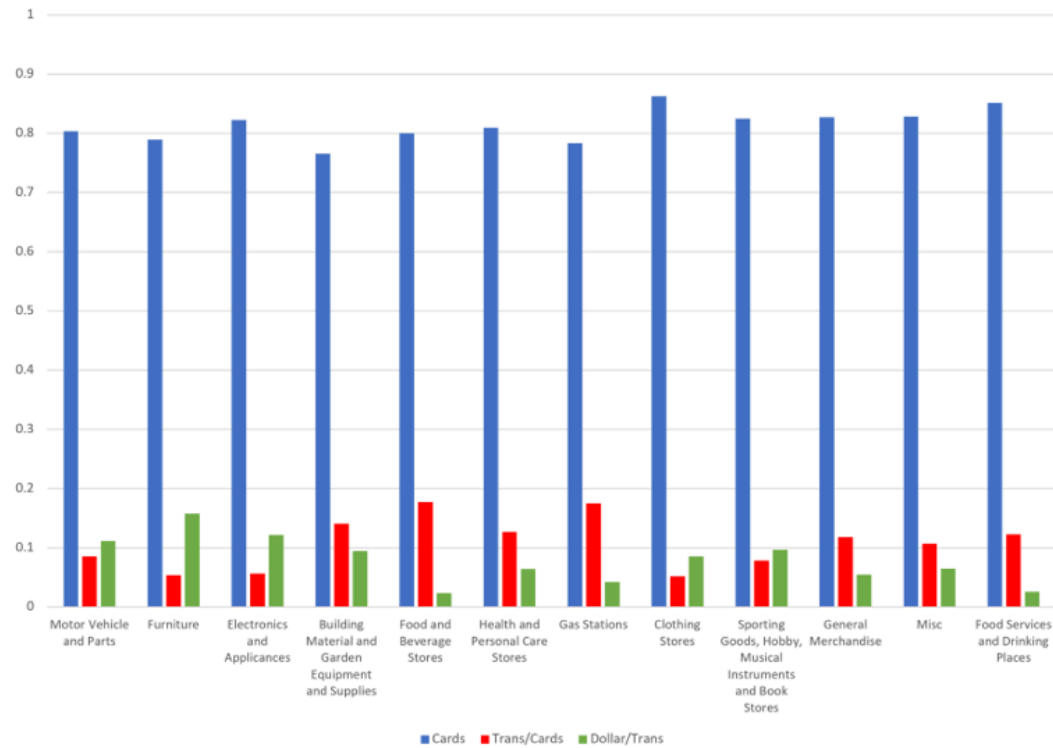
Table 2: Decomposing Sales in Offline Retail

	Stores	Cards/Store	Trans/Card	Dollar/Trans
A. Across Merchants ( <i>N</i> = 939, 304)	0.098 [ 0.123]	0.689 [ 0.761]	0.036 [ 0.231]	0.177 [ 0.547]
B. Within Merchants over Time ( <i>N</i> = 3, 846, 501)	0.164 [ 0.804]	0.681 [ 0.975]	0.098 [ 0.942]	0.056 [ 0.971]
C. Across Stores within Merchants ( <i>N</i> = 1, 926, 714)		0.841 [ 0.972]	0.077 [ 0.809]	0.082 [ 0.933]
D. Within Stores over Time ( <i>N</i> = 7, 746, 218)		0.817 [ 0.995]	0.134 [ 0.969]	0.049 [ 0.987]

Note: All standard errors are less than 0.001. R-Squared values are reported in square brackets. Across Merchant and Across Store within Merchant decompositions are based on 2019 data. Across Merchant regressions include NAICS fixed effects. Within Merchants over Time and Within Stores over Time are based on 2016–2019 data. Within Merchants over Time regressions include merchant and year fixed effects. Across Stores within Merchants regressions include merchant fixed effects. Within store over Time regressions include store and year fixed effects. See [Online Appendix B](#) for robustness with respect to a longer panel of merchant/store data.

The correlation holds remarkably well across different levels within the data. Which factor best predicts revenue? Was it what you expected?

Figure 1: Decomposing Merchant Sales Growth by Industry



Note: This figure displays the coefficients of the “Within Merchant over time” decomposition by industry. The regressions are run with Visa data from 2016 through 2019, and include merchant and year fixed effects.

The correlation holds remarkably well across industries, illustrating its generality. What does this imply for prioritization among these four variables?

# Predictive Analytics Help to Inform Decisions

- Weather forecasts
  - What to do, what to wear
- Stock prices
  - Buy or sell, how much, when
- Safety
  - Where to live, how to commute
- Lifespan
  - Schooling, savings/spending, work/retire
- Product quality
  - Buy now/later, return, warranty, insurance

Data always describes the past, but decisions affect the future. Predictive analytics help to convert past data  $X$  into decisions  $A$  to maximize expected utility of outcomes:

$$\arg \max E[U(A, Y)|X]$$

Examples include election forecasts, product recommendations, and fraud prevention



# Understanding Predictive Analytics

- Correlations alone can enable powerful predictions
  - Causal drivers may help, but aren't necessary
- Predictive analytics are not designed to be prescriptive or diagnostic
  - Predictive analytics are oft misused & overinterpreted
- Predictive Analytics typically have wide error bars
  - Often forgotten or ignored, even by those who know better
  - CIs should account for data variability, parameter estimation error, and model uncertainty
- Numbers oft misinterpreted, e.g., 92% means 11 out of every 12

Predictive analytics is seductive, as astrology or “picks of the week” suggest. Where have you seen people misunderstand or overinterpret a prediction?

# Remember Dray?



Dray has a powerful predictive analytics engine between his ears. He knows what will happen next after someone puts on their shoes and picks up the leash. But he doesn't always know why—he is less good at diagnostic analytics

# Market Size

- Market size ( $N$ ): # of people who might pay to address the core need in a given time period
  - Alternatively measured in \$, units or volume
  - Noisy but helps inform potential returns to investments
  - Typical investor's first question: How big is the market? \$100B market differs from a \$100MM market
  - How will you know if you got the right answer?
  - What happens if you overestimate market size?
- “Marketing myopia:” Neglecting nontraditional competitors, e.g. Zoom v. Uber or Carnival v. Whistler

Market size counts both purchasers and non-purchasers who share the core need. Past purchasers can often be counted, but non-purchasers usually can't

# Market Size

- 3 ways to estimate:
  - “Top Down” Total Addressable Market (TAM):  
How many people have the core need?
  - “Bottom up” Served Available Market (SAM):  
How many people currently pay to solve the core need?
    - ⇒  $TAM = SAM + \text{Unserved}$
  - Analyst estimates
- Best practice: Use all three, triangulate, gauge sensitivity

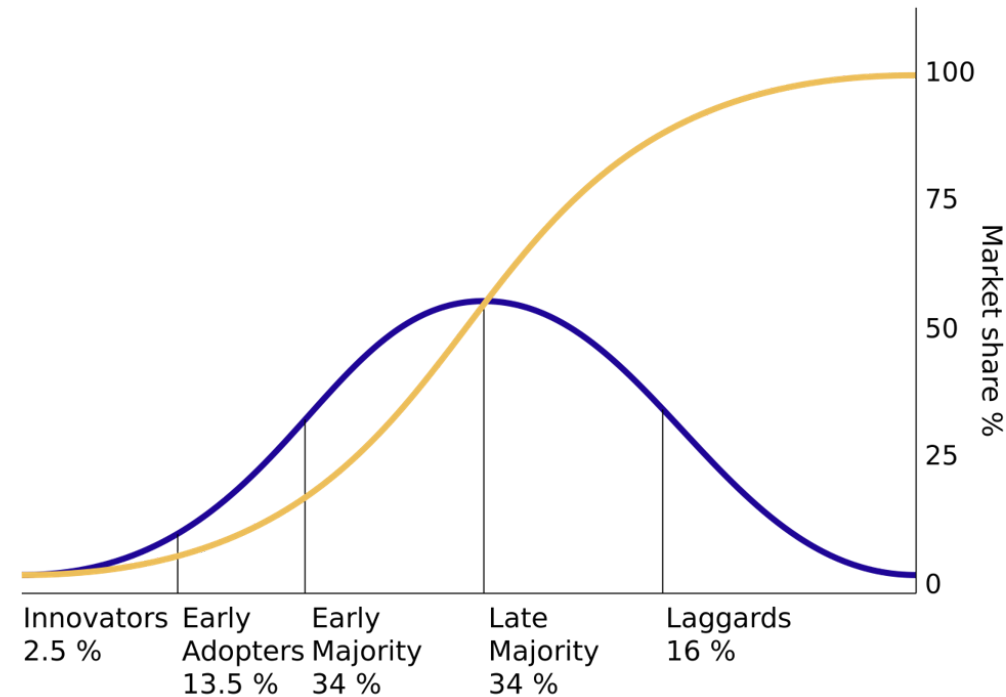
Each method has different blind spots – top-down requires estimating Unserved, bottom-up ignores Unserved, analysts have incentives.

# Case study: US Mattress Market

- USA population: ~340 million
- Assumption:  $TAM \approx SAM$  (why? pros, cons?)
- Assumption: Avg mattress lasts 7 years (pros, cons?)
- Market size  $\approx$  47.1 million people annually
- Average mattress price: \$283, across all bed sizes
- Market size  $\approx$  \$13.3B/year

Market size predictions usually require assumptions. “Sensitivity analysis” shows how much results differ when you vary the assumptions. What changes if mattresses last 5 years instead of 7? How much do those analyst estimates cost?

# Diffusion curves

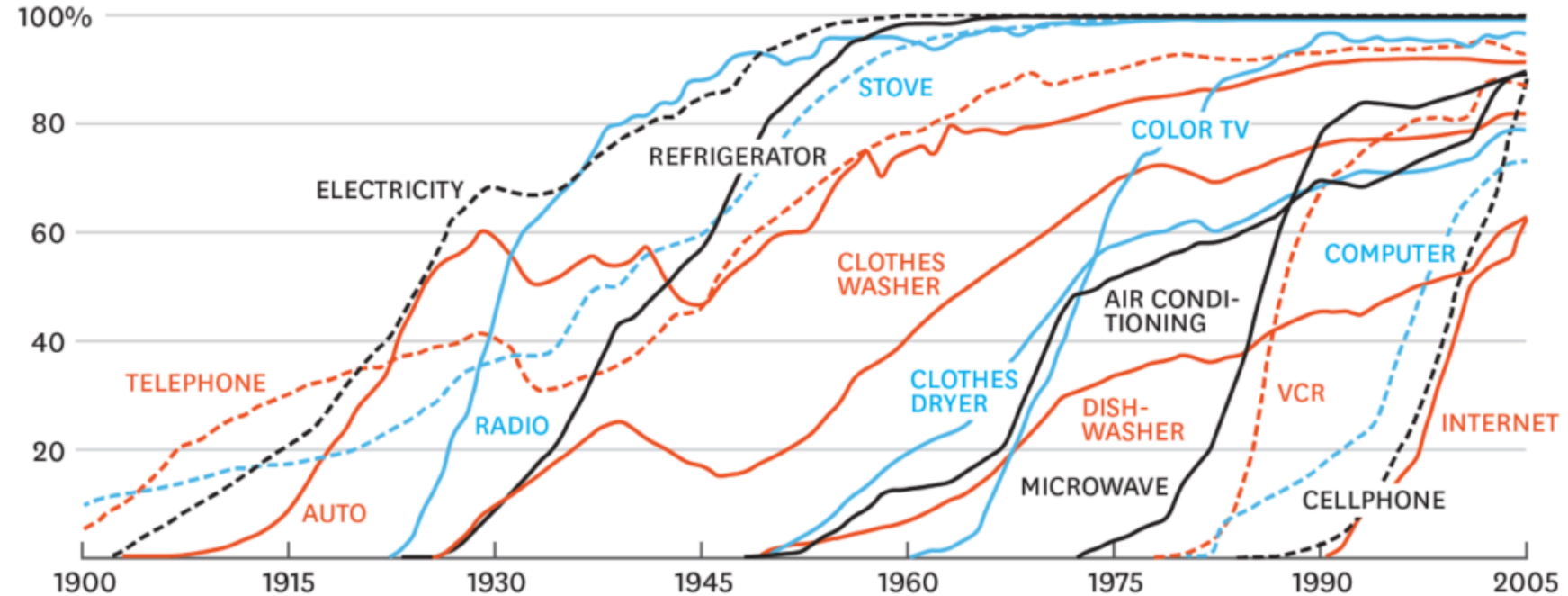


Diffusion curves show how SAM changes over time. Adoption is often an S-curve: slow start, accelerating takeoff, saturation, eventual decline. Blue is the density, Yellow is the CDF. Sometimes a “Dip” phase appears between early adopters and early majority; sometimes a “Decline” phase follows maturity. Also appears in epidemiology, with adoption recast as infection and influence recast as contagion

# Diffusion curves

## CONSUMPTION SPREADS FASTER TODAY

PERCENT OF U.S. HOUSEHOLDS



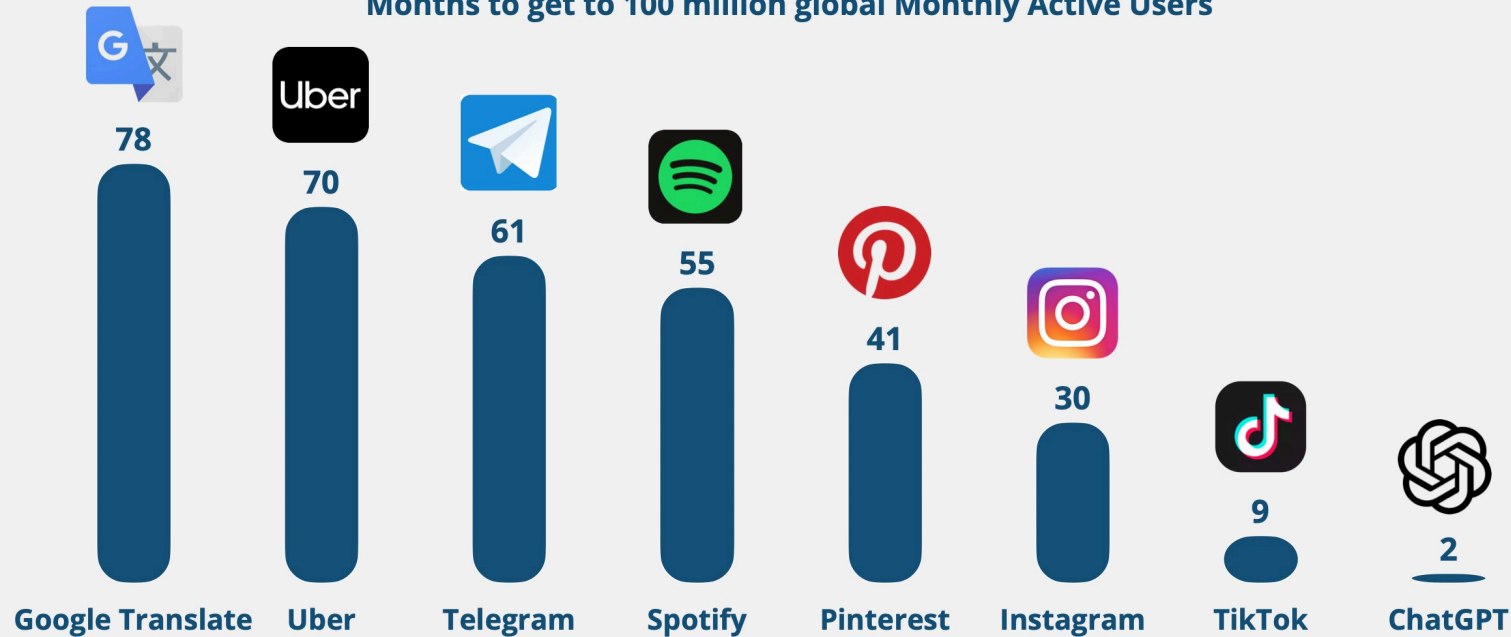
SOURCE NICHOLAS FELTON, THE NEW YORK TIMES

HBR.ORG

There are claims that diffusion is speeding up

# Time to Reach 100M Users

Months to get to 100 million global Monthly Active Users



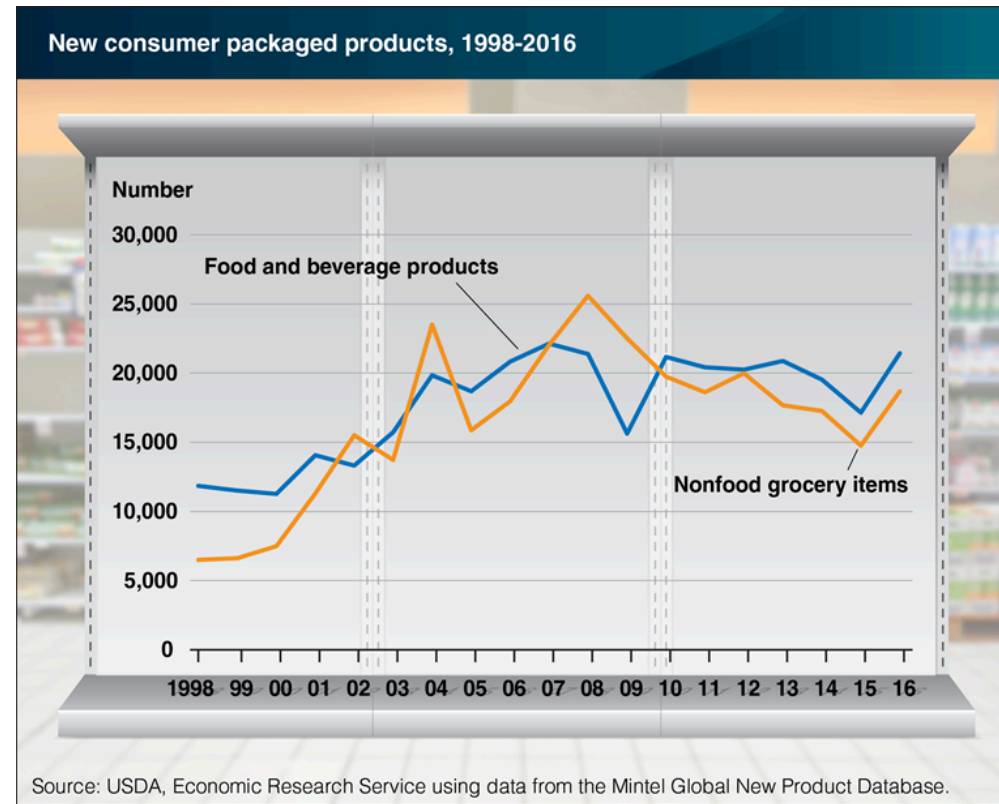
Source: UBS / Yahoo Finance

 @EconomyApp

 APP ECONOMY INSIGHTS

Digitization and social media are plausible mechanisms for accelerating diffusion curves, but these products are not randomly selected. Large majorities of products get lost in the multitude

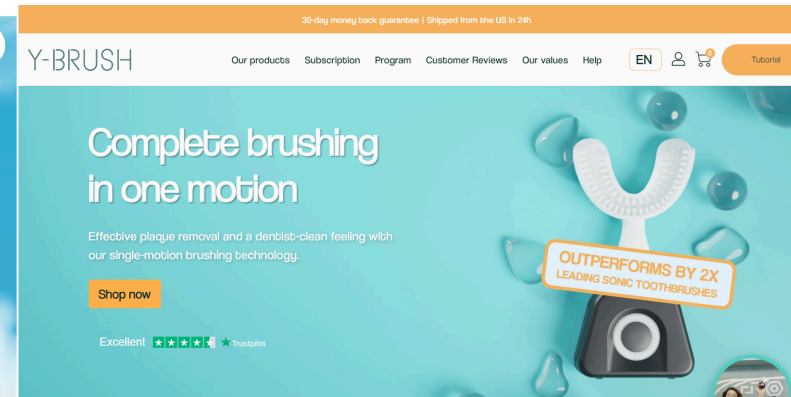
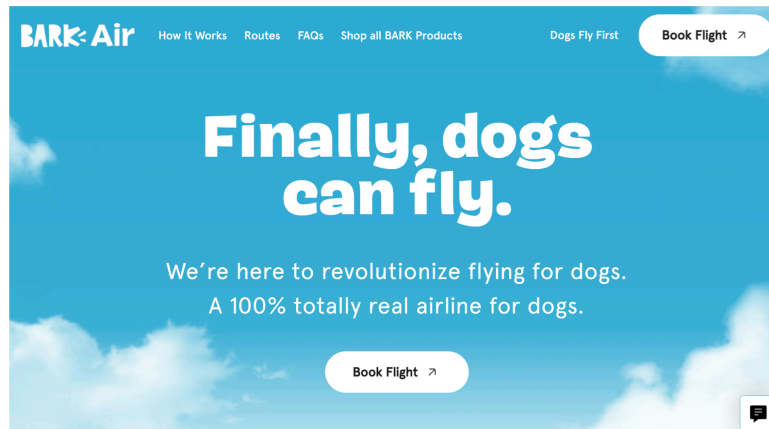
# New CPG Products by Year



In food and drug channels, 35-40k new SKUs launch each year; 90-95% fail completely within six months (Wilbur & Farris 2014). Why is product failure so common, and why do firms keep launching anyway?

USDA (2017) | Wilbur & Farris (2014)

# Which new products will catch on?



New products launch frequently, but most fail, and success is notoriously unpredictable. Addressing a real core need is necessary but far from sufficient. Which of these would you bet on?

# Predicting Diffusion: Bass (1969)

- $M$ : Market size (we'll estimate this)
- $t$ : Time periods
- $A(t)$ : Accumulated sales before time  $t$ 
  - AKA “installed base”
  - $A(0)=0$  by assumption
- $\frac{dA(t)}{dt}$ : number of new adopters in time  $t$
- $R(t)$ : Remaining customers who have yet to adopt,  $R(t) \equiv M - A(t)$

A “differential equation” model relates the level of a variable to its derivative. Fits sigmoid curves well

- Bass (1969) proposed:

$$\frac{dA(t)}{dt} = pR(t) + q\frac{A(t)}{M}R(t)$$

- $p$ : “coefficient of innovation”
- $q$ : “coefficient of imitation”
  - $p$  and  $q$  assumed constant

LHS is growth per unit of time. 1st term on RHS indicates those who would buy regardless of others' choices (“innovators”), which is assumed to be a constant proportion of the unserved market. 2nd term is those whose purchases are influenced by others (“imitators”), increasing in proportion of market that is served

# Estimating Bass model via NLLS

$$\frac{dA(t)}{dt} = pR(t) + q\frac{A(t)}{M}R(t)$$

- This is a first-order diffEQ with analytic solution

$$A(t) = M \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}$$

- If you have sales data by time, you can use Nonlinear Least Squares to estimate  $p$ ,  $q$  and  $M$ , i.e. choosing parameters to minimize square errors  $(LHS - RHS)^2$

You don't have to know how to solve differential equations for this class. Just take my word that the 2nd equation is the solution to the 1st equation.

NLLS is a straightforward generalization of OLS in which the errors can be nonlinear in the parameters. Parameter estimates are still chosen to minimize the sum of square errors, but they don't have closed-form solutions

# Estimating Bass model via OLS

- Or, notice that  $\frac{dA(t)}{dt} = pR(t) + q\frac{A(t)}{M}R(t)$   
 $= p(M - A(t)) + q * \frac{A(t)}{M}(M - A(t))$   
 $= pM + (q - p)A(t) + \frac{q}{M}A(t)^2$   
 $\equiv \beta_0 + \beta_1 A(t) + \beta_2 A(t)^2$
- We can use OLS to regress  $\frac{dA(t)}{dt}$  on a quadratic in installed base  $A(t)$
- Bass model extensions: Multiple markets, hazard models, multiple influence types

If we want, we can solve the system of 3 equations in 3 unknowns to recover  $\hat{p}$ ,  $\hat{q}$  &  $\hat{M}$  from  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ , &  $\hat{\beta}_2$

Bass (2004)

# Models:estimators aren't 1:1?

- Consider 3 OLS estimators:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\min_{\beta}(Y - X\beta)'(Y - X\beta)$$

$$\min_{\beta}[X'(Y - X\beta)]^2$$

“In theory, there’s no difference between theory & practice. In practice, there is.”

Some theoretical models offer multiple estimators. Some have no estimators

Estimators may yield different estimates due to assumptions, numerical properties

Subfields that invent estimators and study their properties: “econometrics,” “data science,” “machine learning,” “statistics”

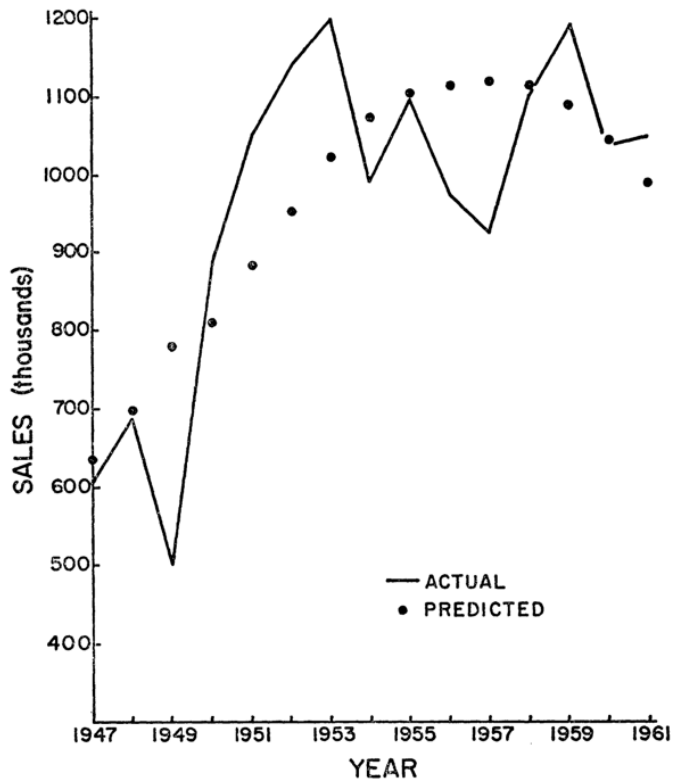


FIG. 5. Actual sales and sales predicted by regression equation (home freezers)

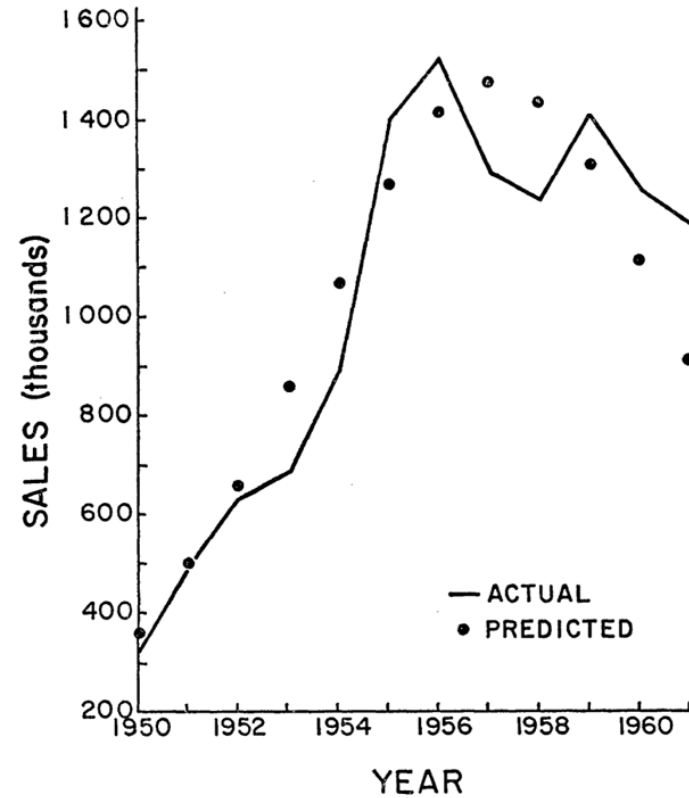


FIG. 8. Actual sales and sales predicted by model (clothes dryers)

These figures are from Bass's original 1969 paper. Compared to having no predictive framework, these were like magic to investors and managers. But look closely— although retrodictions are often near-right on average, every individual retrodiction is wrong

# Rogers' Diffusion Framework

- Diffusion depends on Relative Advantage, Compatibility, Complexity, Trialability, Observability. All from customer's perspective
  - Incredibly influential on managerial practice
  - Provided diagnostics to interpret Bass (1969)'s predictive analytics
  - Prototypes have often been modified to max these 6 dimensions

Everett Rogers was a professor at UNM who spent 50 years studying the “diffusion of innovations,” which now includes over 5k published studies. Rogers' framework is the most widely used to explain and predict new product success/failure. Pick a product and score it on each dimension

Rogers (2010)

# More explanations for diffusion curves

- Conspicuous omission: Customer heterogeneity.  
Preferences, needs, income, or risk attitudes could predict adoption timing
- Conspicuous omission: Marketing efforts like advertising and retail distribution
- Markets typically evolve after introduction
  - Production becomes more efficient & reliable; costs fall; price and customer risk fall
  - New features, technology generations, safety improves
  - Competitors introduce variants targeting unserved customers
  - Network effects, e.g. smartphone compatibility with chargers or accessories
  - Complementors, e.g. Verizon stores, iFixIt, Genius Bar
  - Consumer preferences, e.g. expected reliability rises over time

Bass + Rogers offer the most widely used framework, but neglect important alternate explanations

# Causal Evidence of Social Influence

- “Viral marketing” seeks to identify & seed influencers within social networks
- After Facebook movie app users rated movies, Aral & Walker (2012) randomly notified 1.3MM friends about their movie ratings, then observed app adoption to estimate influence and susceptibility
- Key finding: Social influence causally drives adoption. Influencers are less susceptible, and susceptible users are less influential
- Influence and susceptibility are both heterogeneous:
  - Older users more influential; younger more susceptible
  - Women influence men more than men influence women
  - Influencers with influential friends had most influence in this context
  - Married users tend to be influential; Relationship/Engaged/Complicated users tend to be susceptible; Single users tend to be both influential and susceptible

Classic conundrums in viral marketing: Should you seed the experts, the taste makers, the network-central hubs, or those with track records of relevant influence? When and whom should you gift vs. pay to promote?



**I think my idea on my team is to have the worst ideas so that everyone else can have slightly better ideas. What's one of your worst ideas that you would just never do that's obviously a bad idea?**

I don't know. I'm sure I have a lot of bad ideas. I just don't know which ones are bad before I try them.

**Fair enough.**

The creative process is an interesting thing. We find this a lot at the company when we build features. If you're trying to innovate, a lot of the time, things that end up working seem like bad ideas upfront, and oftentimes, they seem like bad ideas to people who are even really good at innovating. I tell this story that, when we were originally building Discord back in 2015, we were a team of maybe 12 people at the time. Half of the people working on it thought Discord was a bad idea in a 12-person startup. It turned out to be a good idea.

It's really hard to tell beforehand whether innovation is going to work. It's really important to have space to try things, react to it, and innovate and iterate to see where it takes you. Sometimes magic comes out the other end, and a lot of times, you just get duds. But we don't ship those. We try not to ship them.

I love Citron's flat priors and bias for testing. It fascinates me that half the team thought the idea was bad, but they proceeded anyway. I wonder which half he was in?

# Market Size & Career Choice

- Your first job includes a bet on a market
  - Mature market: Big, reputable employers, established career tracks
  - But markets typically decline at some point
  - Growing market: Exciting, high risk, high reward, more opportunity
  - But market may not take off as you expect
  - Will your skillset enable a transition if needed?
- Today's safe option might not be safe tomorrow!
  - Mature markets will decline
  - New markets will grow
  - Consider diffusion trends, not just current market size

I missed a lucrative boat before I found my calling. Should I tell you the story?

# Core Needs Data: **Google Trends**

- Best available indicator of unserved market
- Reports search volume indices by keyword, place, time, service
  - Also identifies topics, trending terms & related queries
  - Samples the query database, reports estimates not actual counts
  - Requires a minimum query volume for reporting
  - Free, so it could get sunsetted

Google started Trends back in the 00s because it worried about SEC regulation, given that its search data could predict stock price movements. Seems quaint

[Choi & Varian \(2012\) discuss data usage considerations](#)

# Customer Lifetime Value

- CLV is the most powerful predictive customer analytics metric
  - Expresses the firm's value of an individual customer relationship as the net present value of expected future customer profits
  - Pioneered by catalogue retailers in the 1960s
  - Has spread widely, but not yet everywhere
  - CLV metrics enable quantification & serious discussion of new policies
  - But, prediction uncertainty hinders CLV adoption and usage

CLV can account for customer heterogeneity using same approaches we talked about in week 5

# CLV Example: Housing First

- “In 2005, Utah set out to fix a problem that’s often thought of as unfixable: chronic homelessness. The state had almost 2,000 chronically homeless people. Most of them had mental-health or substance-abuse issues, or both. Utah started by just giving the homeless homes...
- The cost of shelters, emergency-room visits, ambulances, police, and so on quickly piles up. Lloyd Pendleton, the director of Utah’s Homeless Task Force ... said that the average chronically homeless person used to cost Salt Lake City more than \$20,000/year. Putting someone into permanent housing costs the state just \$8,000 [including case managers]...
- Utah’s first pilot program placed 17 people in homes scattered around Salt Lake City, and after twenty-two months not one of them was back on the streets. In the years since, the number of Utah’s chronically homeless has fallen by 74%.”

Republican Utah might not be your first guess as to whom invented Housing First, but its CLV analysis made this dramatic policy shift look like a no-brainer

# Housing First: Looking deeper

- Housing First has certainly not solved homelessness
  - “Chronic” means disabled and unhoused for 1+ yrs, or 4x in 3 yrs
  - ~28% of CA homelessness is chronic (2019)
  - UT originally claimed 90% reduction, then revised their metric definitions
  - Reliable efficacy metrics are rare
  - Housing First has been implemented haphazardly
  - UT built new apartments. CA cities mostly use shelters, SROs, vouchers
  - Key Q: Require wraparound services? E.g. Addiction treatment, etc
  - Key Q: Does Housing First somehow encourage homelessness?
- I claim: Quantification enables bold policy shifts
  - U.S. HUD adopted Housing First as preferred approach to homelessness in 2014
  - CLV quantifies policy costs and benefits & enables ex-post evaluations
  - We then can use data to refine CLV estimates and policies

Basing policy choices on CLV requires us to quantify and clarify our assumptions, and shows how we expect policies to impact customers. Quantifying disagreements facilitates next steps toward resolutions. What are possible next steps toward evaluating how Housing First policy attributes affect homelessness?

# Calculating CLV

- $T$ : planning horizon
- $m_t$ : contribution margin of serving customer  $i$  in time  $t$
- $r$ : retention probability that customer buys in  $t + 1$
- $i$  is the cost of capital
- $CLV = \sum_{t=0}^T \frac{m_t r^t}{(1+i)^t}$ 
  - Assumes start-of-year calculation, including yet-to-be-realized margins in year  $t = 0$ ; alternate formula assumes end-of-year calculation starting from  $t = 1$
- $m_t$  and  $r$  observable in past data; future values are predictions

$T$  is often set at 3 years, but Bezos says Amazon's 7-year horizon lets them justify investments competitors can't match. The horizon, retention, and discount rate are all judgment calls. Prediction uncertainty increases with  $T$

# CLV Example

- A tennis club charges an annual fee of \$300
- The average club member spends \$100 per year (concessions, etc.)
- The average contribution margin on add'l expenditures is 60%
- Historically, 80% of the members rejoin the club in any given year. The club's cost of capital is 15%
- What is the club's CLV per member for  $T=1$ ,  $T=2$ , or  $T=3$ ?

# Using CLV for Customer Acquisition

- Marketing campaigns should be profitable if Avg. Customer Acquisition Cost (CAC) < CLV
  - Caveat: So long as acquired customers have  $CLV \geq \text{avg CLV}$  of existing customers
  - Often, managers impose a “fudge factor” as a speedbump
- Suppose the tennis club has a chance to pay \$20k for a billboard. It will be seen by 100K people with an expected conversion rate of .1%. Should we do it?
  - To decide, lay out decision rule  $MR \geq ff * MC$  and calculate
- Similar “break-even” calculations possible for any fixed cost that impacts future profits, e.g. product launch, store openings, price promotions

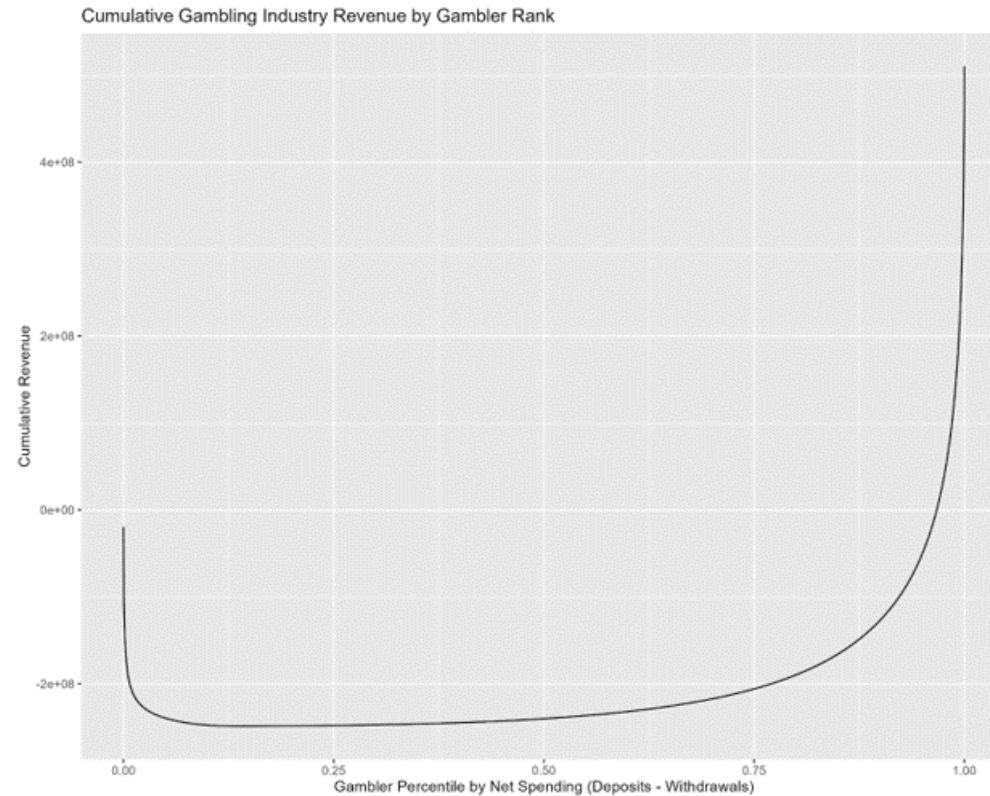
Most common “fudge factor” values: 3 or 7. Why?

# CLV Metrics in Practice

- CLV popularity rose alongside CRM data systems
  - Firms used CLV to set customer experiences: high-CLV flyers got upgraded, high-CLV lodgers got better rooms, high-CLV shoppers got sterling return service and attention, high-CLV callers got shorter wait times and more consideration
- In the 00s, consulting firms published claims that 20-30% of customers were unprofitable. Firing the least-profitable customers sounded attractive
  - American Express offered some cardholders \$300 to cancel their cards. Best Buy stopped notifying some shoppers about upcoming promotions. Banks used minimum balances and teller fees to drive away some account holders. Amazon stops selling to frequent returners. Uber terminates driver and rider accounts if they violate policy
- Customers talk to each other; firing customers is a brand risk

Firing customers can be direct/confrontational or indirect/quiet. Better yet, redesign product menus and reprice so that unprofitable customers either become profitable or self-select out

MIT Sloan – “Firing the Right Customers Is Good Business” (2024)



This graphic shows total gambling industry net revenue as a function of gambler net spending percentiles. 5% of gamblers cost online gambling operators 20% of their revenue, whereas the most profitable 3% of customers generate half of net revenue. Operators block and limit frequent winners, so pros use “beards”

Funt (2022) | Rose-Berman (2025) | McCarthy et al. (2026)

# CLV Cautions & Risks

- CRM data may be incomplete, disconnected, error prone
  - E.g., can you connect customer identity across different credit card #s?
  - Think about how identity mismeasurement affects retention rate measurement
- CLV models are predictive analytics, not prescriptive or diagnostic
  - CLV usually contains no customer purchase model, because it's purely predictive
  - Heterogeneous CLV-based policies may be self-fulfilling: treat someone as unprofitable, they may act that way
- Guiding Principles
  - Firm & customer both benefit by increasing customer value provided. CLV metrics for value creation benefit everyone
  - Firm & customer have opposing interests in price. CLV for margin extraction may generate perverse customer incentives
  - Customer dissatisfaction may reveal CLV flaws; requires careful attention
  - Hence many firms now measure point-of-sale satisfaction, rather than blaming customers for unprofitability

# Example: CLV for Pricing

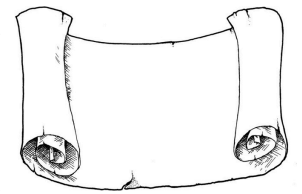
- Assume price is \$100, margin \$50, retention 50%, discount rate 10%, horizon T=1:  $CLV = 50 + .5 * 50 / 1.1 = \$72.72$
- Suppose you consider increasing price to \$120, holding all else constant: Then,  $CLV = 70 + .5 * 70 / 1.1 = \$101.82$
- Should you raise the price?
  - What is this analysis missing?
  - How could we resolve that problem?
  - Hint: CLV is predictive analytics, not diagnostic or prescriptive

When can CLV analysis optimize price?

Palumbo (2023) | Pfeiffer and Farris (2004)

# Class script

- Let's estimate the Bass model using OLS and NLLS, compare their properties, and simulate SAM evolutions over different time horizons



Wrapping up

# Competition

Suppose you consider a 28-period time horizon, with true  $p = .03$ , true  $q = .1$ , and  $M = 435$ . Simulate data from these parameters. Now suppose you only had data for periods 4-12, supposing that measurements for periods 1-3 are unavailable. Use OLS and NLLS to estimate the Bass model based only on data from periods 4-12. Then use both estimated models to make out-of-sample predictions for periods 13-28. Calculate mean square prediction errors in periods 13-28 for both models. Visualize the two MSPEs to illustrate which estimator generalized to the non-training data better.



# Recap

- Customer acquisitions best predict revenues
- Market size estimates how many people share a core need
- Diffusion models predict how served market changes
- CLV metrics can quantify and enable novel policies

# Going further

- Innovation Diffusion and New Product Growth Models
- Customer-Base Valuation in a Contractual Setting with Heterogeneity
- Exploring the Distribution of Customer Lifetime Value
- Firing your best customers

