Samuel Levy

Tepper School of Business
4765 Forbes Ave
Pittsburgh, PA 15213

www.levysamuel.com
+1 (412) 773-2627
samlevy@cmu.edu

Education

08/2018 – present **Carnegie Mellon University**
Ph.D. in Marketing - Minor in Statistics and Machine Learning
M.S. in Marketing (May 2020)
Dissertation committee: Alan Montgomery (Chair, Tepper), Tim Derdenger (Tepper), Joy Lu (Tepper), Kannan Srinivasan (Tepper), Asim Ansari (Columbia, Outside reader), Fred Feinberg (University of Michigan, External committee member)
Dissertation title: *Essays on Bayesian Machine Learning in Marketing*

08/2016 – 06/2018 **Tilburg University**
M.S. in Marketing (Research Track) - *Cum Laude*
Tilburg School of Economics and Management

08/2010 – 06/2014 **Ecole Normale Superieure Paris-Saclay, Université Paris 1 Panthéon-Sorbonne**
B.S. in Economics - *Cum Laude*

Research Interests

Substantive: Customer Analytics, Choice Modeling, Privacy.
Marketing Contexts: Telecommunications Industries, Consumer Packaged Goods, B2B, Charitable Organizations.
Methodological: Bayesian Econometrics, Bayesian Nonparametrics, Machine Learning, Deep Generative Modeling.

Papers

Under Review

"Privacy Preserving Data Fusion"
with Longxiu Tian and Dana Turjeman. *Revise and resubmit at Marketing Science.*
Awarded Israel Science Foundation (equivalent of $30,000 per year, 2023-2026). Acceptance rate: 26%.

Working Papers

"Digital Twins: A Generative Approach for Counterfactual Customer Analytics" *(Job Market Paper)*

"Understanding Consumer Expenditure Through Gaussian Process Choice Models"
with Alan Montgomery.

Work in Progress

"Understanding the Dynamics of Appeals Scales to Infer Potential to Donate"
with Joy Lu and Alan Montgomery.

"Multiview Topic Model For Purchase Prediction"
with Dokyun Lee, Daniel McCarthy, and Alan Montgomery.

Awards and Honors

2024 **Finalist, 2023 ISMS Doctoral Dissertation Proposal Competition**.
2023 **ISMS Doctoral Consortium Fellow**.
2022 **AMA-Sheth Foundation Doctoral Consortium Fellow**.

2021 **External Grant, Corporate Sponsor**. co-PIs: Prof. Alan Montgomery & Prof. Katia Sycara

2020 **Dean's Research Fund, Tepper School of Business, ($2,000)**

2019 **Wharton Customer Analytics Initiative, Collaborative Data Grant (Unique Team Selected)**

2018 - 2024 **William Larimer Mellon Ph.D. Fellowship, Carnegie Mellon University**

2017 **Koopmans Scholarship (Dean's scholarship, € 12,000), Tilburg University**

2017 **Fellow, Quantitative Marketing & Structural Economics Workshop**

2016 **French *Agrégation* in Economics and Business Administration (Top 1% Nationwide)**

2011 **Full Scholarship - Ecole Normale Superieure Paris Saclay (€ 63,000)**

Conference Presentations

2023 "Understanding Consumer Expenditure through Gaussian Process" with Alan Montgomery. INFORMS Marketing Science Conference, University of Miami.

"Privacy Preserving Data Fusion" with Longxiu Tian and Dana Turjeman. INFORMS Marketing Science Conference, University of Miami.

"Understanding the Dynamics of Appeals Scales to Infer Potential to Donate" with Joy Lu and Alan Montgomery. INFORMS Marketing Science Conference, University of Miami.

2022 "Understanding the Dynamics of Appeals Scales to Infer Potential to Donate" with Joy Lu and Alan Montgomery. Marketing Dynamics, Georgia State University.

"Understanding Consumer Expenditure through Gaussian Process" with Alan Montgomery. Joint Statistical Meetings - Marketing Section, Washington DC.

"Understanding Consumer Expenditure through Gaussian Process Choice Models " with Alan Montgomery. INFORMS Marketing Science Conference (Virtual).

2021 "Understanding Consumer Expenditure through Gaussian Process Choice Models" with Alan Montgomery. Joint Statistical Meetings - Marketing Section (Virtual).

"Understanding Consumer Expenditure through Gaussian Process Choice Models" with Alan Montgomery. INFORMS Marketing Science Conference (Virtual).

2020 "Multiview Topic Model For Purchase Prediction" with Dokyun Lee, Daniel McCarthy, and Alan Montgomery. Joint Statistical Meetings - Marketing Section (Virtual).

"Multiview Topic Model For Purchase Prediction" with Dokyun Lee, Daniel McCarthy, and Alan Montgomery. INFORMS Marketing Science Conference (Virtual).

Teaching

Instructor (Tepper School of Business)

Summer 2023 **70381: Marketing I**
Level: Undergraduate.
Instructor Rating: 4.63/5.0
School Average: 4.21/5.0

Teaching Assistant (Tepper School of Business)

Fall 2020 **70381: Marketing I**
Level: Undergraduate. Instructor: Prof. Hui Li

Summer 2020 – Fall 2020 **70467: Machine Learning for Business Analytics**
Level: Undergraduate. Instructors: Prof. Benjamin Moseley, Prof. Andrew Li

Fall 2020 **47747: Bayesian Statistics**
Level: Ph.D. Instructor: Alan Montgomery

Summer 2019, Summer 2020 **Math Skills Workshop**
Level: MBA

Selected Ph.D. Coursework

2018 – 2020 **Marketing**
Structural Modeling and Quantitative Methods (with Hui Li)
Analytical Modeling (with Kannan Srinivasan)
Bayesian Statistics in Marketing (with Alan Montgomery)
Multivariate Data Analysis (with Alan Montgomery)
Consumer Behavior (with Jeff Galak)

2018 – 2020 **Machine Learning and Statistics**
Machine Learning (with Ziv Bar-Joseph and Pradeep Ravikumar)
Probabilistic Graphical Models (with Eric Xing)
Convex Optimization (with Ryan Tibshirani)
Probability and Statistics (with Jing Lei)
Advanced Probability Overview (with Alessandro Rinaldo)
Economining (with Dokyun Lee and Zachary Lipton).

2018 – 2020 **Economics**
Microeconomics (with Bertan Turhan)
Econometrics I (with David Childers)
Econometrics II (with Robert Miller)
Game Theory (with Onur Kesten).

Work Experience

Fall 2021 – Spring 2022 **Research Assistant at the University of Michigan**
Data Collection and Modelisation. Advised by Prof. Fred Feinberg.

Languages

English (fluent), French (native).

Computing

Programming Languages: R, Python, JAX, PyTorch, Julia, Stan, Pyro, NumPyro, SQL.

References

Alan Montgomery (Advisor)
Professor of Marketing
Head, Ph.D. Program
Carnegie Mellon University
alm3@andrew.cmu.edu
+1 (412) 268-4562

Fred Feinberg
Handleman Professor of Management
and Professor of Statistics
University of Michigan
feinf@umich.edu
+1 (734) 764-4711

Joy Lu
Assistant Professor of Marketing
Carnegie Mellon University
tonglu@andrew.cmu.edu
+1 (925) 216-6850

Selected Abstracts

Digital Twins: A Generative Approach for Counterfactual Customer Analytics (Job Market Paper)

This research provides a novel methodology, Digital Marketing Twins, that automatically extracts latent features from individual-level brand survey responses to inform a statistically-principled, deep generative model of customer-side brand affinity and firm-side performance factors. The proposed model enables marketers to find drivers of individual-level brand affinity, as opposed to traditionally observed metrics that must be analyzed in aggregation. The framework serves a counterfactual purpose at the customer level. The generative part of the model *completes* the distribution of survey responses over time, and across firms – thereby addressing the archetypal missing data problem – by imputing customer responses in counterfactual regimes. The proposed prescriptive framework also proposes policy optimization through customer surveys, using Bayesian optimization, which efficiently identifies "paths of least resistance" among customer responses to service-quality questions – a search that otherwise would represent a complexity of $\mathcal{O}(n^d)$.

This research applies Digital Marketing Twins methodology to the competitive landscape of the U.S. wireless telecommunications retail market, leveraging a unique dataset of large-scale quarterly brand surveys from all three major carriers (AT&T, T-Mobile, and Verizon) from 2020 to 2022. It optimizes over the learned generative model from the multi-firm brand surveys to provide marketing policy recommendations according to individual-level counterfactual responses and different carriers. Empirically, this approach reveals latent asymmetries in competition in terms of brand affinity, together with a nonlinear increase in brand affinity for certain types of drivers, such as satisfaction with network speed, but a nonlinear decrease in brand affinity for customers who report greater likelihoods of changing plans, providers, or devices, relative to their current wireless services.

Privacy Preserving Data Fusion (with Longxiu Tian and Dana Turjeman)

Data fusion combines multiple datasets to make inferences that are more accurate, generalizable, and useful than those made with any single dataset alone. However, data fusion poses a privacy hazard due to the risk of revealing user identities. We propose a privacy preserving data fusion (PPDF) methodology intended to preserve user-level anonymity while allowing for a robust and expressive data fusion process. PPDF is based on variational autoencoders and normalizing flows, together enabling a highly expressive, nonparametric, Bayesian, generative modeling framework, estimated in adherence to differential privacy – the state-of-the-art theory for privacy preservation. PPDF does not require the same users to appear across datasets when learning the joint data generating process and explicitly accounts for missingness in each dataset to correct for sample selection. Moreover, PPDF is model-agnostic: it allows for downstream inferences to be made on the fused data without the analyst needing to specify a discriminative model or likelihood a priori. We undertake a series of simulations to showcase the quality of our proposed methodology. Then, we fuse a large-scale customer satisfaction survey to the customer relationship management (CRM) database from a leading U.S. telecom carrier. The resulting fusion yields the joint distribution between survey satisfaction outcomes and CRM engagement metrics at the customer level, including the likelihood of leaving the company's services. Highlighting the importance of correcting selection bias, we illustrate the divergence between the observed survey responses vs. the imputed distribution on the customer base. Managerially, we find a negative, nonlinear relationship between satisfaction and future account termination across the telecom carrier's customers, which can aid in segmentation, targeting, and proactive churn management. Overall, PPDF will substantially reduce the risk of compromising privacy and anonymity when fusing different datasets.

Understanding Consumer Expenditure Through Gaussian Process Choice Models (with Alan Montgomery)

Consumers change their choice as expenditures within a category increase. Traditional choice models usually make restrictive structural assumptions to specify the expenditure elasticity. This imposed functional form of utility strongly influences the range of estimable substitution patterns across goods. Consumers with highly nonlinear preferences may have consumption thresholds in which buying patterns dramatically change when price or budget changes. Understanding these thresholds with a flexible utility-based model could lead to improved pricing and promotion decisions. Using Gaussian process priors on utility functions, we relax the functional form of both inside goods and outside good, within the context of constrained utility maximization. We estimate a general direct utility choice model for simultaneous purchases within a product category. We build a hierarchical model by borrowing information from a parametric functional form that constitutes an informative prior at the individual level. Our model captures non-linear rates of satiation and precise baseline preferences that traditional non-homothetic parametric models fail to capture by assuming a given functional form of utility. The proposed model automatically detects non-linear patterns of consumption from the data and provide a more precise statistical inference.

Understanding the Dynamics of Appeals Scales to Infer Potential to Donate (with Joy Lu and Alan Montgomery)

We explore the role of direct solicitations with suggested amounts, or "appeals scales", in charitable giving. While it is known that these scales can influence immediate donation behavior, little has been studied about their dynamic impact over time. Leveraging transaction and marketing data from a major U.S. university, we explore how different suggested amounts, the frequency of direct solicitations, and their timing influence the willingness to donate over time. We demonstrate that these suggested amounts impact individual distributions for donation willingness, estimated through a Bayesian dynamic hierarchical model of latent reference prices. Given the inherent issue of endogeneity, we conduct a series of randomized controlled trials. We randomize various parameters of the appeals scales, including the endpoints (the lowest amounts asked), and the steepness (the dollar interval between each ask in a mailer). The primary objective of this research is to develop an optimal rule for suggested amounts, the number of direct solicitations, and timing, aiming to maximize revenue from donations.

Last updated: January 23, 2024