Mingyung Kim

727.4 Jon M. Huntsman Hall
The Wharton School, University of Pennsylvania

Mobile: (267) 366-8520
Email: mingyung@wharton.upenn.edu
Website: www.mingyung-kim.com

EDUCATION

The Wharton School, University of Pennsylvania
Ph.D. in Marketing, May 2024 (Expected)
Committee: Eric T. Bradlow (co-chair), Raghuram Iyengar (co-chair),
Christophe Van den Bulte, Ryan Dew

University of California, Berkeley
M.A. in Statistics, May 2017

Yonsei University (Seoul, South Korea)
B.B.A. and B.A. in Applied Statistics, February 2015
Ranked 1st in Applied Statistics and Business Administration

RESEARCH INTERESTS

Marketing analytics, Statistical methods, Machine learning
Network/Graph-based problems in marketing (e.g., data aggregation, market structure)

PUBLICATIONS AND MANUSCRIPT UNDER REVIEW

Kim, Mingyung, Eric T. Bradlow, Raghuram Iyengar (2023), "A Bayesian Dual-Network Clustering Approach for Selecting Data and Parameter Granularities," **Job Market Paper**, Under review at ***Marketing Science***.

Kim, Mingyung, Eric T. Bradlow, Raghuram Iyengar (2022), "Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights," ***Marketing Science***, 41(4):848-866.

Kim, Mingyung, Jeeyeon Kim, Jeonghye Choi, Minakshi Trivedi (2017), "Mobile Shopping through Applications: Understanding Application Possession and Mobile Purchase," *Journal of Interactive Marketing*, 39(1):55-68.

WORKING PAPER AND WORK IN PROGRESS

"The Cure is Worse than the Disease: Individual-Level Fixed Effects in Hazard Models Induce Spurious Peer Effects," with Christophe Van den Bulte.

"Graph Representation Learning for Inferring Market Structure," with Ryan Dew.

PRESENTATIONS

- Marketing Science Conference, Miami, FL, June 2023
  "A Bayesian Dual-Network Clustering Approach for Selecting Data and Parameter Granularities"
- AMA-Sheth Foundation Doctoral Consortium (Small Group Research Presentation), Austin, TX, July 2022
  "Selecting Data Granularity and Model Specification using the Scaled Power Likelihood with Multiple Weights"
- Wharton-INSEAD Doctoral Conference, Virtual, November 2020
  "Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights"
- Marketing Science Conference, Virtual, June 2020
  "Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights"
- Wharton Marketing Departmental Seminar, Philadelphia, PA, February 2020
  "Selecting Data Granularity Using the Power Likelihood"
- Wharton-INSEAD Doctoral Conference, Philadelphia, PA, November 2019
  "Selecting Data Granularity Using the Power Likelihood"

SELECTED HONORS AND SCHOLARSHIPS

ISMS Doctoral Consortium Fellow, 2021–2023
AMA-Sheth Foundation Doctoral Consortium Fellow, 2022
Wharton Marketing Graduate Fellowship, 2017–2022
Ilju Educational Foundation Scholarship, 2017–2021
Kwanjeong Educational Foundation Scholarship, 2015–2017
UC Berkeley Statistics Graduate Fellowship, 2015-2016

TEACHING INTERESTS

Marketing Analytics and Big Data
Statistical and Machine Learning Methods in Marketing

TEACHING EXPERIENCE

Graduate Student Instructor:

Introduction to Marketing: Recitation (The Wharton School, Undergraduate, Fall 2020)
Concepts in Computing with Data: Recitation (UC Berkeley, Undergraduate, Spring 2017)
Introduction to Statistical Computing: Recitation (UC Berkeley, Graduate, Fall 2016)

Teaching Assistant:

Pricing Policy (The Wharton School, MBA/EMBA, Jagmohan Raju, 2020–2023)
Quantitative Marketing (Yonsei University, Undergraduate/Graduate/MBA, 2011–2014)

SELECTED COURSEWORK (* AUDIT)

University of Pennsylvania

Marketing

Empirical Models in Marketing (Eric T. Bradlow, Raghuram Iyengar)
Empirical Models in Marketing* (Ryan Dew)
Measurement and Data Analysis in Marketing (Christophe Van den Bulte)
Economic/OR Models in Marketing (Jagmohan Raju, Ron Berman)
Research Methods in Marketing (Wes Hutchinson, Robert Meyer)

Statistics and Computer Science

High-Dimensional and Non-Parametric Bayesian Statistics (James Johndrow)
Advanced Topics in Machine Learning (Shivani Agarwal)
Deep Learning for Data Science (Konrad Kording)
Advanced Topics in Statistics: Deep Learning (Edgar Dobriban)
Bayesian Machine Learning* (Jacob R. Gardner)
Graph Neural Network* (Navid NaderiAlizadeh, Charilaos Kanatsoulis)

University of California, Berkeley

Seminar in Marketing: Marketing Strategy (J. Miguel Villas-Boas)
Choice Modeling (Przemyslaw Jeziorski)
Statistical Models: Theory and Applications (Philip B. Stark)
Advanced Introduction of Probability and Statistics (Jim Pitman, Yuekai Sun)

REFERENCES

Eric T. Bradlow (Committee Co-Chair)
The K.P. Chao Professor of Marketing
Vice Dean of Analytics at Wharton
Chairperson of Marketing
University of Pennsylvania
ebradlow@wharton.upenn.edu
Office: (215) 898-8255
Mobile: (610) 322-8091

Raghuram Iyengar (Committee Co-Chair)
Miers-Busch, W'1885 Professor of Marketing
Faculty Co-Director of Wharton AI & Analytics for Business
University of Pennsylvania
riyengar@wharton.upenn.edu
Office: (215) 898-2391

Christophe Van den Bulte (Committee)
Gayfryd Steinberg Professor of Marketing
University of Pennsylvania
vdbulte@wharton.upenn.edu
Office: (215) 898-6532

Ryan Dew (Committee)
Assistant Professor of Marketing
University of Pennsylvania
ryandew@wharton.upenn.edu
Office: (215) 898-2534

SELECTED ABSTRACTS

A Bayesian Dual-Network Clustering Approach for Selecting Data and Parameter Granularities (*Job market paper*, with Eric T. Bradlow and Raghuram Iyengar, Under review at *Marketing Science*)

While there are well-established methods for model selection (e.g., BIC, marginal likelihood), they generally condition on an a priori selected data (e.g., SKU-level data) and parameter granularity (e.g., brand-level parameters). That is, researchers think they are doing model selection, but what they are really doing is model selection *conditional* on their choices of data and parameter granularities. In this research, we propose a *Bayesian dual-network clustering* method as a novel way to make these two decisions simultaneously. To accomplish this, the method represents data and parameters as two separate networks with nodes being the unit of analysis (e.g., SKUs). The method then (a) clusters the two networks using a covariate-driven distance function which allows for a high degree of interpretability and (b) infers the data and parameter granularities that offer the best in-sample fit, akin to standard model selection methods. We apply our method to SKU-level demand analysis. The results show that the choices of data and parameter granularities based on our method as compared to those from extant approaches (e.g., latent class analysis) impact the demand elasticities and the optimal pricing of SKUs. We conclude by highlighting the generalizability of our framework to a broad array of marketing problems.

Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights (with Eric T. Bradlow and Raghuram Iyengar, Published in *Marketing Science*)

Firms employ temporal data for predicting sales and making managerial decisions accordingly. To use such data appropriately, managers need to make two major analysis decisions: (a) the temporal granularity (e.g., weekly, monthly) and (b) an accompanying demand model. In most empirical contexts, however, model selection, sales forecasts, and managerial decisions are vulnerable to both of these choices. While extant literature has proposed methods that can select the best-fitted model (e.g., BIC) or provide predictions robust to model misspecification (e.g., weighted likelihood), most methods assume that the granularity is either correctly specified or pre-specify it. Our research fills this gap by proposing a method, the scaled power likelihood with multiple weights (SPLM), that not only identifies the best-fitted granularity-model combination jointly, but also conducts doubly (granularity and model) robust prediction against their potentially incorrect selection. An extensive set of simulations shows that SPLM has higher statistical power than extant approaches for selecting the best-fitted granularity-model combination and provides doubly robust prediction in a wide variety of mis-specified conditions. We apply our framework to predict sales for a scanner dataset and find that similar to our simulations, SPLM improves sales forecasts due to its ability to select the best-fitted pair via SPLM's dual weights.

Graph Representation Learning for Inferring Market Structure (with Ryan Dew)

This paper aims to uncover market structure, with a focus on complementary and substitutable relationships, within a large set of products. While understanding market structure has played a crucial role in designing new products, repositioning existing products, and planning marketing actions such as pricing, extant literature has mostly focused on learning market structure for a small subset of products or at an aggregated level (e.g., brand, category). We seek to overcome this limitation by using a modern graph representation learning technique termed Variational Graph Auto Encoder (VGAE). Specifically, we plan to extend VGAE, which has primarily been used to learn synergistic and antagonistic effects among a large set of molecules in the field of Computational Biology, to learn complementary and substitutable relationships among a large set of products.