

# Customer-Base Analysis Using Repeated Cross-Sectional Summary (RCSS) Data

Proposal for  
2008 Russell Ackoff Doctoral Student Fellowships  
for Research on Human Decision Processes and Risk Management

By  
**Kinshuk Jerath**  
Doctoral Candidate in Marketing  
Wharton School  
University of Pennsylvania

September 2007

## Descriptive Summary

A number of researchers have developed models of repeat buying behavior that can be used as a basis for computing quantities of managerial interest such as customer lifetime value. The Pareto/NBD is an example of such a model.

These models typically require that the analyst have access to individual-customer-level data. However, this is not always practical. While many reporting systems are able to create simple data summaries for a fixed period of time (e.g., a quarterly histogram of number of purchases), the process of extracting raw individual-level data can be a time-consuming task. The mere process of physically transferring raw customer-level data can be risky, as a number of privacy-related news stories attest. And in many countries, data protection laws can complicate the process of transferring raw data to the analyst, especially when the analyst is located in another country.

This research project proposes to explore the possibility of estimating these models using repeated cross-sectional summaries of the transaction data (e.g., four quarterly histograms). Such summaries are easy to create, are easy to distribute, and are free of any privacy concerns. In particular, I propose to examine i) how much "information" is lost when fitting the model to these repeated cross-sectional data summaries instead of the raw individual-customer-level data, and ii) the number of cross-sections required to minimize information loss.

I propose to carry out a comprehensive simulation study covering a vast spectrum of market scenarios characterized by various levels of customer base penetration and purchase frequency and also validate the results from the simulation on real datasets. Preliminary results indicate that across most of these datasets, the model fit (and parameter values) associated with the use of RCSS

data can closely match the model fit (and parameter values) associated with individual-customer-level data. This is a finding of great practical significance for firms grappling with data sufficiency issues due to privacy laws and their own IT incapacibilities.

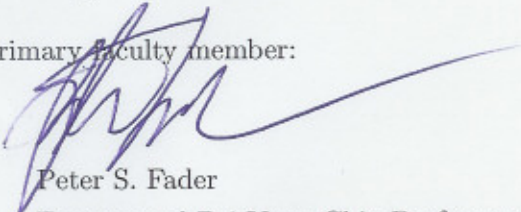
## Budget

1. Attending and presenting research at the INFORMS Marketing Science Conference in Vancouver, Canada from June 12–14, 2008. Estimated expenses: \$1500.
2. Attending and presenting research at the AMA Advanced Research Techniques Forum in Asheville, NC, USA from June 15–18, 2008. Estimated expenses: \$1000.

Other anticipated resources of research and travel funding for the above requirements: none.

## Faculty Members

Primary faculty member:



Peter S. Fader

Frances and Pei-Yuan Chia Professor, Professor of Marketing  
Wharton School  
University of Pennsylvania  
faderp@wharton.upenn.edu

External faculty member:

Bruce G.S. Hardie  
Associate Professor of Marketing  
London Business School  
bhardie@london.edu