Samuel J Thomas

Data Scientist
Address: 70 W Clear Lake Ln, Westfield IN 46074
samuel.thomas522@gmail.com +1 317 696 9214 linkedin.com/in/samueljthomas1/ | *Updated:* September 23, 2020

Professional Experience

Used k-means and Gaussian Mixture Models (GMM) to identify optimal locations for wholesaler territory coverage. This analysis is being used to place wholesalers in future coverage scenarios.

Used a generalized linear mixed effect model to determine capacity requirements for internal wholesalers. This analysis is being used to determine coverage responsibilities for the future state of the salesforce.

Developed and implemented a test/control framework for a sales initiative on product models of mutual funds.

Served as lead analyst in developing statistical models for the AFSG budget process. Models are used to develop hiring plans to support a >1,500-person service center and by Global Finance for revenue projections. Accurate forecasting is critical to the financial health of the organization.

Developed a machine learning algorithm (based on C5) to identify retirement plans at risk for attrition. At-risk plans, as identified by the model, are twice as likely to leave as non-risk plans.

Developing R Shiny app to streamline the AFSG long-range forecasting process. Reduction of 4 FTE in analyst time

Using spacy NLP library in Python to explore open text messages from salesforce data

Serving as Math/Theory contributor in Capital Group Data Science guild

Capital Group Internal Consulting: statistical support/consulting for Client Analytics, Business Management, Work Management, Legal/Compliance, Human Resources, Quality, Retirement Plan Services, and American Funds Service Company.

Developed a mathematical model to estimate the steady-state account volumes based on queuing theory (M/M/∞ queue).

Used PyMC3 (Python) to fit a generalized linear model using MCMC techniques (Hamiltonian Monte Carlo). Model identified economic factors associated with Plan Premier Participant counts.

Identified contact rate influencers using Generalized Least Squares (GLS) models. Daily seasonal factors were fit using Fourier series. Discussion with Data Science Interest group contributed to increased interest in leveraging AFS data for analysis for broader CG applications.

Developed a custom optimization model in Python to estimate the number of Shareholder Services associates needed if processing work was outsourced. Model influenced decision to retain processing work in-house.

Developed a regression model to match the automated pricing results from a website. The model was developed using statistical software and translated to Excel for the client's use.

Developed a predictive model based on machine learning algorithms for an environmental sensor application with over 400 variables.

Created an automated forecasting model to predict fuel demand in various locations in the UK. This model was used to anticipate geographical fueling needs for a trucking company.

Used text mining to analyze open-responses to survey questions from a call center. This analysis was used to identify drivers of caller satisfaction.

Evaluated search paths commonly used by CSRs to find scripts using sequencing analysis in R. This analysis was used to identify opportunities to improve script searching capabilities.

Identified opportunities for quality and cost savings for the 1-800-MEDICARE call center through data mining and statistical methods.

Academic Experience

Co-instructor for PBHL-B 646 Advanced Generalized Linear Models with Wanzhu Tu, Spring 2019, Indiana University.

Developed a convolutional neural network model in Tensorflow to classify minerals based on spectral imaging from the planet Mars.

Employment

Data Scientist

Capital Group Companies 2008, 2010–present

Develop predictive models to assist business planning for a service center of 1,500 associates. Serves as internal statistical subject matter expert.

Principal

Revelant Technologies 2008–2010

Leveraged data to identify opportunities for quality and efficiency improvement opportunities for the 1-800-MEDICARE service center.

Associate Actuary

Milliman 2007–2008

Assistant Actuary

WellPoint/Anthem 2005–2007

Integration Engineer

Hurco Companies 2000–2005

Plant Engineer

Delphi Packard Electric 1995–2000

Teaching

Co-instructor, Biostatistics

Indiana University 2019

PBHL-B 646 Advanced Generalized Linear Models

Adjunct Professor, Mathematics

Ivy Tech University 2006–2014

Courses taught:

- Math 211 Calculus
- Math 136 College Algebra
- Math 135 Finite Math

Education

Indiana University, Ph.D. Biostatistics — February 2021

Purdue University, M.S. Mathematics — 2006

University of Notre Dame, B.S. Electrical Engineering — 1995

PUBLICATIONS

Thomas, S. and Tu, W. (2020). *Hamiltonian Monte Carlo*. In Wiley StatsRef: Statistics Reference Online (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J.L. Teugels). doi:10.1002/9781118445112.stat08243

Thomas, S., & Tu, W. (2020). *Learning Hamiltonian Monte Carlo in R*. arXiv preprint arXiv:2006.16194.

Green, Brice and Thomas, Samuel, *Inference and Prediction of Stock Returns using Multilevel Models* (August 31, 2019). Available at SSRN: https://ssrn.com/abstract=3411358 or http://dx.doi.org/10.2139/ssrn.3411358

In Progress

Thomas, S. and Tu, W. 2020. *Fitting Bayesian Multivariate Generalized Additive Models in R*.

Thomas, S. and Tu, W. 2021. *Semiparametric Regression Application to Furosemide*.

Whitepaper

Thomas, S. *mlts: an R package to forecast multi-level time series*

Talks

A Bayesian Analytical Software Based on Hamiltonian Monte Carlo. Regenstrief Institute, 12/4/2019.

Using Fourier Series to Model Daily Seasonal Patterns of Redemptions. Capital Group Companies, Data Science Interest Group, 2018

Improving Capacity and Financial Planning, a Guide to Business Forecasting with Alteryx. Inspire 2016 Alteryx Conference, San Diego, CA.

Predicting At-Risk Plans Using the C5 Algorithm. Capital Group Companies, Data Science Interest Group, 2015

UseR 2012 at Vanderbilt University. UseR 2012 Vanderbilt University.

Software Packages

hmclearn: An R package to fit statistical models with Hamiltonian Monte Carlo. https://cran.r-project.org/web/packages/hmclearn/index.html

mvgamHMC: An R package to fit semiparametric regression models using Hamiltonian Monte Carlo. Development in progress.

mlts: An R package to automatically develop forecasts and perform cross-validation for bottoms-up forecast models. Internal package for Capital Group Companies.

Interests

- Statistical Computation
- Bayesian Analysis: Markov Chain Monte Carlo
- R, Python, C++, SQL, Azure Databricks, Tensorflow (Deep Learning)