Samuel J Thomas

Data Scientist
Address: 70 W Clear Lake Ln, Westfield IN 46074

✉ sljt@capgroup.com 📞 +1 317 696 9214 🌐 linkedin.com/in/samueljthomas1/ | *Updated:* December 11, 2020

Education

Indiana University, Ph.D. Biostatistics	February 2021
Purdue University, M.S. Mathematics	2006
University of Notre Dame, B.S. Electrical Engineering	1995

Employment

Data Scientist

Capital Group Companies	2008, 2010–present

Developed machine learning and statistical models to re-design North American Distribution sales territories. These models optimized territory coverage for over 200 sales professionals across the country.

Leads a team of 3 consultants to build analytics for three new departments of the North American Client Group. These departments work with the home offices of financial services firms for product placement and sales initiatives.

Developing a model to estimate the sales lift for a major coverage redesign initiative. This model is based on historical and projected marketshare, based largely on activity data reported through salesforce.com.

Extensive experience with using internal and external data to drive business decisions related to sales analytics. Worked with senior business leadership to build the foundations for reporting and analytics.

Fit a Generalized Linear Mixed Effects model (GLMM) to sales and activity data to identify the most significant drivers of work for Internal Wholesalers (IW). This model was used to assign IW coverage.

Developed the forecasting methodology and implementation for the American Funds Service Company budget process. This process has been adapted and reproduced for multiple departments in the organization.

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

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.stato8243

Thomas, S., & Tu, W. (2020). *Learning Hamiltonian Monte Carlo in R*. arXiv preprint arXiv:2006.16194. Tentatively accepted to The American Statistician pending satisfactory revisions.

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. *An R package for Bayesian Multivariate Generalized Additive Models*.

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

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

Software Packages

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

bayesGAM: An R package to fit semiparametric regression models using Hamiltonian Monte Carlo. https://cran.r-project.org/web/packages/bayesGAM/index.html

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

Skills

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

Talks

A Bayesian Analytical Software Based on Hamiltonian Monte Carlo. Regenstrief Institute, 12/4/2019. https://www.youtube.com/watch?v=sBA3lAoNhto

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.

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.

Additional Professional Experience

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.

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

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

Identified service center 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.