JOHN H. SCHWENCK

jschwenck12@gmail.com
johnschwenck.github.io
973.610.8220
LinkedIn

EDUCATION

Texas A&M University — **College Station, TX**
The College of Science | Master of Science; ***Statistics*** *– Computational Statistics Track* — *Class of 2021*
**** Relevant Coursework: Distribution Theory, Theory of Inference & Linear Models, Spatial Statistics, Bayesian Computations, Statistical Methods in Finance, Stochastic Calculus*

The Pennsylvania State University — **University Park, PA**
The College of Liberal Arts | Bachelor of Science; ***Economics*** — *Class of 2016*
Smeal College of Business | Bachelor of Science; ***Supply Chain & Information Systems***

WORK EXPERIENCE

Wells Fargo Bank, N.A. — **Charlotte, NC**
Machine Learning Engineer | Artificial Intelligence & Machine Learning (AI/ML) Model Development Center — *July 2021 – Present*

- Centralized research team that develops and deploys AI/ML models across all lines of business within the Consumer space (marketing, pricing, deposit products, call center, and other customer-centric operations)
- Research and translate the latest academic statistical methodologies into applied business solutions
- Collaborate with model validators and regulators to ensure statistical robustness by drafting model development documentation of all framework / architecture, theory, and code for reproducibility
- Maintain the group's internal modeling pipeline Python package that streamlines and automates data aggregation, processing, feature engineering, training/validation/holdout, algorithm execution, diagnostics, leaderboard creation, deployment, and model monitoring
- Projects include:
    - Created various propensity, response, and attrition models for marketing campaign customer targeting
    - Implemented a spatial varying-coefficient model to estimate new locations for prospective branch openings
    - Built an automated Python framework to derive net present value (NPV) curves across customer segments
    - Currently building an NLP-based topic model for our internal "loudspeaker" (i.e. suggestions box) program

Texas A&M Athletics — **College Station, TX**
Statistical Consultant | Sports Science and Analytics — *June 2020 – May 2021*

- Incorporated real-time physiological data from Oura rings and other wearable devices of athletes for functional data regression to estimate various performance metrics and optimize training protocols
- Strategically utilized Men's Basketball COVID-19 gameplay and practice contact tracing data in order to minimize player absence during regular and post-season play by preventing safety protocol violations and avoiding SEC sanctions

South Jersey Industries — **Atlantic City, NJ**
Senior Analyst | Strategic Analytics & Corporate Development — *June 2016 – April 2019*

- Utilized geospatial data and developed an XGBoost model in R to generate pipeline failure probabilities at various locations for gas leak prevention and other potentially fatal situations
- Leveraged R, Python, and SQL to create a company-wide data visualization dashboard in Power BI that integrated departmental performance metrics to drive accountability and transparency in data reporting
- Improved the Trading teams' asset portfolios by developing a quantile-based expected shortfall model using R to predict the likelihood of their natural gas trade volumes receiving a "supply cut" during the pipeline nomination process
- Optimized call center volume by developing a queuing network for call arrival times and a Spatial Poisson Process model to predict the most probable call locations in order to economically dispatch response workers.

RESEARCH EXPERIENCE

Academic Research

Texas A&M – Department of Statistics — **College Station, TX**
Graduate Researcher — *May 2020 – September 2022*

- Conduct research with departmental faculty and collaborative researchers at Johns Hopkins University to develop algorithms through the ***iglu*** R package that detect abnormal rates of change in continuous glucose monitors (cgm) using a variety of machine learning techniques, data visualization, and medical analysis measures
- Creator and author of, ***bp***, the first open-source R package dedicated to measuring and analyzing blood pressure data through a suite of data-processing tools and prognostic metrics from relevant literature | Python equivalent in progress
- Methodological research involved functional data analysis (multilevel fPCA) to analyze sleep duration curves

Penn State Energy Marketing Association (P.S.E.M.A.) — **University Park, PA**
Co-Founder | Director of Macroeconomic Research — *July 2015 – May 2016*

- Founded the organization to bridge the gap between technical and non-technical backgrounds who shared a mutual interest in energy trading and a desire to further their knowledge through cross-disciplinary research
- Directed all research efforts for monthly market outlook newsletters by conducting statistical analyses for each of the four core focus areas (Oil, Gas, Weather, and Economy) and implemented various computational algorithms via R

Independent Research

Riding for Research – A Cycling Trip Across North America — **United States & Canada**
Project Manager | Cyclist — *April 2019 – April 2020*

- Pedaled a 1991 Cannondale M1000 bicycle 6,000 miles from New York City to Seward, Alaska to collect fitness, dietary, and cardiac related data through various health tracking sensors for an initial pilot study
- Analyzed biomedical signals (both covert biomarker signals such as heart rate, blood pressure and sleep, as well as overt signals such as meal and stress logs) to assess the impacts of exercise, stress, sleep, and nutrition
- Collected data hosted on Harvard Dataverse and blood pressure data available in the R package, *bp*

CONFERENCES & PRESENTATIONS

- R/Medicine Conference, R Consortium – Poster Presenter — **Virtual** (due to COVID-19)
  bp: Blood Pressure Analysis in R — *August 2021*

PUBLICATIONS & SOFTWARE DEVELOPMENT

Publications

All code and packages available via GitHub at https://github.com/johnschwenck

- bp: *Blood pressure analysis in R* | Available on GitHub and CRAN
  Schwenck J, Punjabi NM, Gaynanova I (2022). PLOS ONE 17(9): e0268934. https://doi.org/10.1371/journal.pone.0268934
- iglu: *Interpreting Glucose Data from Continuous Glucose Monitors (CGM) in R*
  Authors (CRAN version 3.0.0): Broll S., Shih J., **Schwenck J.**, Hicban M., Buchanan D., Martin M., Chun E., Patel P., Muschelli J., Fernandes N., Urbanek J., Seo J., Meyyappan A., Nguyen N. and Gaynanova I.
- *Riding for Research: Cycling Pilot Study from NYC to Alaska (2019).* **Schwenck J.**, Harvard Dataverse

Publications utilizing John Schwenck's research

- J. A. Handler, C. F. Feied and M. T. Gillam, "*Novel Techniques to Assess Predictive Systems and Reduce Their Alarm Burden*" in IEEE Journal of Biomedical and Health Informatics, 2022, doi: 10.1109/JBHI.2022.3189312.

SKILLS & INTERESTS

Programming: Python, R, SQL
Data Science & Big Data: Jupyter, Spark / Hadoop, H2O, Databricks / Google Cloud Platform (GCP), Git/GitHub
Software: Excel, Power BI, Adobe, LaTeX, Confluence
Research Interests: Spatial statistics, quantile methods, dimensionality reduction, energy markets
Non-academic Interests: Cycling, poker, reading, history, skiing, guitar, NY Giants