Research Statement

Mario Giacomazzo
Arizona State University, Tempe, AZ

Background

The original intent of my Bachelor's degree was to build a teaching career in secondary mathematics education. My scholastic achievement and loving professors diverted my path to higher education. Unsure of my mathematical interests after a rigorous, but limited, small college experience, I focused many PhD applications to pure and applied math departments. In the end, I am thankful I chose the only statistics program to which I applied.

My diverse coursework has given me a vast background in statistics. Whether or not a course was theoretical or applied did not matter. My early curiosity pushed me to learn. I have extensively studied many methods from both Frequentist and Bayesian viewpoints. Also, I was able to see statistical approaches in many other applied areas such as econometrics, clinical trials, bioinformatics, biostatistics, and industrial engineering.

At my core, I have a teacher's heart. The reward from independent studying and research is collected when I get the opportunity to communicate my understanding to others. Through many statistical consulting opportunities, I practiced both doing and explaining statistics. Graphical output and technical reports allowed me to turn a researcher's data into a story. Walking clients through statistical approaches used for inference required deeper understanding of the material and advanced my interpersonal skills.

Despite all my work under the academic bubble, my internship at the Health Service Advisory Group (HSAG) provided valuable experience unobtainable from coursework. At HSAG, I analyzed national home health care patient measure data, provided by the Centers for Medicare and Medicaid Services, to identify high and low performing facilities. Factor analysis was used to create metrics to quantify quality of care. Then, generalized linear models provided reasons for the discrepancies seen between facilities. With records in the millions, I developed comfortability analyzing "big data". This environment required efficient and well documented SAS code. Weekly tasks and technical reports had to be accomplished under strict deadlines. By the end, their trust in me opened the opportunity to introduce more advanced methods to their statistical library. The confidence gained performing under pressure empowered me in future independent research.

Current

The bulk of my Ph.D. research has been allocated to time series analysis under both Frequentist and Bayesian frameworks. The primary objective in time series analysis is to identify the model that describes the underlying dynamics and minimizes future forecasting error. The foundation of time series analysis was established to answer applied questions in finance and economics where current monetary policy is developed based on predicted future behavior. The two applications to which I have dedicated considerable time have been outside finance and economics.

Climate change has had an adverse effect on the biodiversity in rivers. Both the distribution and quantity of various freshwater species have slowly been modified to combat increasing water temperatures. Maintainers of these ecosystems use predictive models to make proactive decisions to mitigate these changes. Acknowledging seasonal patterns and air-water relationships help in forecasting the water temperatures multiple days in advance.

Traffic management systems monitor real-time traffic characteristics to proactively reduce congestion. The stress rising populations have on urban networks makes short-term forecasting of key traffic variables (volume, speed, occupancy, etc.) a necessary part in the constant management of traffic quality. Corresponding data naturally exhibits weekly seasonal patterns, space-time dependencies, nonlinear dynamics, and heteroscedasticity.

The naïve method is to forecast using most recent information; however, prevalence of dynamical changes, nonstationarity, and heteroskedasticity require more complicated models. Nonlinear regime switching time series models can be used to model the aforementioned phenomena. Specifically, I have extensively studied threshold autoregressive (TAR) and smooth transition autoregressive (STAR) models estimated using classical and Bayesian techniques. These models are essentially weighted averages of linear autoregressive models where weights are determined by known information about an endogenous or exogenous transition variable. The key difference between these models is in the weight function. TAR models are characterized by abrupt jumps across regimes, whereas STAR models exhibit slower regime changes.

In Giacomazzo & Kamarianakis (2017a), Bayesian regularization was used to estimate linear and nonlinear time series designed to forecast maximum water temperatures. Multiple simulation studies provide efficacy that Bayesian regularization evoked through modified priors can estimate a more flexible STAR or TAR model. Hierarchical global-local prior specifications to autoregressive coefficients lead to pseudo-model selection through shrinkage estimation. Bayesian Lasso and Bayesian Horseshoe are popular examples of Bayesian shrinkage priors. Simple modifications are made to allow for regime-specific shrinkage. Also, adoption of a Dirichlet prior handles the uncertainty in choosing a transition variable while leading to a more encompassing STAR model.

Bayesian shrinkage usually does not completely constrain insignificant parameters to 0. Also, when the initial dimension of the parameter space is large, some significant coefficients may be overshrunk to combat overfitting. Giacomazzo & Kamarianakis (2017b) combines Bayesian shrinkage with predictive variable model selection techniques to quickly obtain parsimonious models designed for short-term forecasting of traffic occupancy in a small urban network. Using threshold autoregressive distributed lag (TARDL) models, a three step methodology induces regime-specific shrinkage and selection. First, TARDLs are fitted using Bayesian horseshoe priors for sparse estimation. Next, a forward stepwise procedure selects the optimum sub-model at each level of flexibility based on minimization of Kullback-Leibler (KL) divergence from the posterior predictive distribution from the full reference model. Finally, the best model is chosen based on regime-specific forecasting accuracy for select horizons.

Future

Future proposed ideas are specific topics identified through time dedicated in my doctoral research. In general, I have developed a passion for time series problems. Using mathematical statistics and probability to predict the future is both mysterious and beautiful to me.

Most research pertaining to Bayesian shrinkage methods involves cross-sectional data. Explanatory covariates in time series models are highly collinear by nature. I plan to further study and compare Bayesian shrinkage and model selection methods in the handling of multicollinearity. I want to conduct a theoretical and empirical exploration of Bayesian posterior estimates on achievement of oracle properties.

The optimal model for forecasting usually changes over time. Linear and nonlinear models may produce different forecasts that vary in accuracy across regimes. Forecasts from these models can often be combined to produce predictions better than the individual forecasts. I plan to advance research in linear and nonlinear combination schemes using both static and dynamic weights. Bayesian approaches in this realm can utilize new information to update posterior weights to increase forecasting accuracy.

Until now, I have primarily used Bayesian methods to handle regime switching time series models incorporating at most 3 regimes. The simplification has allowed me to analyze my proposed methodologies using popular Bayesian software i.e. JAGS, BUGS, STAN integrated through R. Uncertainty in the number of regimes can be included in MCMC algorithms, but comes at a high computational cost.

Whether the sample size or number of parameters is large, Bayesian MCMC posterior sampling algorithms can be impractical due to slow mixing across chains. I plan to actively advance my understanding in both computer science and machine learning to seek for improvements in correcting the inefficiencies. In the current climate, Bayesian handling of the "big data" problem needs to be addressed.

In terms of workplace, I do not have a strong inclination toward academia or industry. Ultimately, I desire to work in a collaborative environment where I am able to provide statistical support on challenging interdisciplinary applied projects. Opportunities to publish in peer reviewed journals and to present statistical research at conferences is highly desired.

References

Giacomazzo, M. and Kamarianakis, Y. (2017a). Bayesian Shrinkage Estimates of Logistic Smooth Transition Autoregressions. Submitted to *Communications in Statistics: Simulation and Computation.*

Giacomazzo, M. and Kamarianakis, Y. (2017b). Sparse Bayesian Estimation of Linear and Nonlinear Models for Short-term Forecasting of Traffic Occupancy. Plan to Submit to *International Journal of Forecasting.*