RESEARCH STATEMENT
ELIANA CHRISTOU

My research interests consist of sufficient dimension reduction, quantile regression (QR) and its applications, and functional data analysis (FDA). I have organized my research into six sections below.

1 Sufficient Dimension Reduction

Overview. Statistical analysis seeks to determine the relationship between a response variable Y and a p-dimensional set of predictors $\mathbf{X}$. As a preliminary step, an investigator will seek a graphical analysis to gain a better understanding about the relationship and guide the choice of an initial model. However, when the dimension of the predictor variable $\mathbf{X}$ is large, straightforward graphical analysis and preliminary model fitting procedures can be challenging. In situations like these, *dimension reduction techniques* are necessary, i.e., techniques that reduce the number of predictor variables under consideration without any loss of information on the regression. Formally, we seek to find a $p \times d$, $d \leq p$, matrix $\mathbf{A}$ such that $Y \perp\!\!\!\perp \mathbf{X}|\mathbf{A}^\top\mathbf{X}$, implying that Y and $\mathbf{X}$ are independent given $\mathbf{A}^\top\mathbf{X}$. In other words, the high-dimensional $p\times 1$ predictor vector $\mathbf{X}$ can be replaced with the low-dimensional $d\times 1$ predictor vector $\mathbf{A}^\top\mathbf{X}$ without any loss of information. Therefore, the goal is to find such a matrix $\mathbf{A}$, and specifically, to find the subspace that is generated by $\mathbf{A}$. Many research areas can benefit from dimension reduction, such as biology, computer science, finance, neuroscience and artificial intelligence, among others.

Contribution. There are a lot of methods proposed in the literature that focus on identifying the smallest subspace that achieves the greatest dimension reduction. For example, sliced inverse regression (SIR; Li 1991) focuses on modeling the conditional mean, while sliced average variance estimation (SAVE; Cook and Weisberg 1991) focuses on modeling the conditional variance in the process of finding the smallest subspace. However, these methods have certain limitations, such as their sensitivity to extreme values. Although many methods have been proposed to overcome these limitations, they are iterative in nature. While iterations add to the computational complexity of the procedure, a more serious issue is that of convergence. To overcome this problem, I proposed a *non-iterative* and *robust against extreme values* dimension reduction technique, called the *sliced inverse median regression* (SIMR) in Christou (2020a). Simulation studies and a real data application suggest that the proposed methodology can be used as an alternative robust dimension reduction technique under the presence of extreme values. This work was supported, in part, by funds provided by the University of North Carolina at Charlotte *(Faculty Research Grant)*.

Future Work. The new reduced sufficient predictor $\mathbf{A}^\top\mathbf{X}$ defines linear combinations of the predictor variable $\mathbf{X}$ that contain all the information necessary for the regression. However, the coefficients that define those combinations include all the predictor variables, where it might be the case that some predictors are unnecessary. To ease the interpretation of the new sufficient predictors and to increase the efficiency of the resulting linear combinations, we can extend my above work by producing sparse linear combinations $\mathbf{A}^\top\mathbf{X}$, i.e., linear combinations where some coefficients are zero. This will be considered in a future work with my collaborator Dr. Andreas Artemiou from Cardiff University, UK.

2 Quantile Regression

Overview. There are many research areas where extremes are important, such as in meteorology for extreme weather conditions, in finance for extreme returns, or in medical studies for low- or high-risk individuals. To study the extreme parts of data, we generally consider specific quantiles and use QR. However, applications of QR can become challenging when dealing with high-dimensional data, making it necessary to use dimension reduction techniques. The techniques discussed in Section 1 focus on the entire conditional distribution and often provide more information than necessary when interest lies on a specific conditional quantile. Therefore, the focus is on developing methodologies that identify the fewest linear combinations of the predictor $\mathbf{X}$ that are necessary to estimate the conditional quantile, i.e., identifying a $p \times d_\tau$ matrix $\mathbf{B}_\tau$, $d_\tau \leq p$, such that $Y \perp\!\!\!\perp Q_\tau(Y|\mathbf{X})|\mathbf{B}_\tau^\top \mathbf{X}$, where $Q_\tau(Y|\mathbf{X})$ denotes the τth conditional quantile of Y given $\mathbf{X}$. As in Section 1, the focus is on estimating the subspace that is generated by $\mathbf{B}_\tau$ and which will define the new sufficient predictors $\mathbf{B}_\tau^\top \mathbf{X}$.

Contribution. There is limited work on dimension reduction techniques that focus on conditional quantiles. For that reason, most of my research is devoted to this area.

- My first contribution considers a special case of a dimension reduction model, that of a single-index quantile regression (SIQR) model. This model assumes that the predictor variable $\mathbf{X}$ can be replaced by a *single* linear combination of $\mathbf{X}$ without losing any important information necessary to estimate the conditional quantile, i.e., the matrix $\mathbf{B}_\tau$ is a $p \times 1$ matrix. Existing literature on SIQR model includes an iterative algorithm for estimating the vector of the coefficients of the linear combination of $\mathbf{X}$. To avoid iterations and convergence issues in Christou and Akritas (2016) we proposed a *non-iterative algorithm* for estimating the vector of coefficients and derived the asymptotic distribution of the proposed estimator under heteroscedasticity.

- The above work focuses on estimating the coefficients that define the single linear combination of $\mathbf{X}$. However, there are situations where some of these predictors are unnecessary and including them complicates the interpretation of the model, while at the same time decreases the efficiency of the resulting estimator and its predictive ability. Therefore, variable selection plays an important role as it can result in some of the coefficients to be zero and ease the interpretation of the linear combination. For that reason, Christou and Akritas (2018) proposed an algorithm that performs *simultaneous variable selection and parameter estimation* in the context of a SIQR model. The resulting estimator is consistent and satisfies the oracle property.

- As censored survival data occur frequently in biostatistics, environmental sciences, social sciences, and econometrics, we found it natural to extend our methodology to this type of data. Therefore, Christou and Akritas (2019) proposed a *non-iterative* *algorithm* for fitting the SIQR model for *censored data* and derived the asymptotic distribution of the proposed estimator under heteroscedasticity.

- In an effort to develop a methodology that can estimate as many linear combinations of $\mathbf{X}$ as necessary to estimate the conditional quantile, I introduced the *central quantile subspace*, a special case of the T-central subspace introduced in Luo et al. (2014). Therefore, I proposed an algorithm for performing dimension reduction for conditional quantiles, which can also be extended to any statistical functional of the conditional distribution (Christou 2020b). The idea is to produce several vectors that span the central quantile subspace. The initial vector is obtained by the least squares slope resulting from linearly regressing $\mathbf{X}$ on the conditional quantile. This is a very strong result, as it works no matter what is the true relationship between $\mathbf{X}$ and the conditional quantile. Note that if we assume a SIQR model, then this initial vector is all we need to form the single linear combination $\mathbf{B}_\tau^\top \mathbf{X}$; therefore, this methodology provides an improved alternative to that of Christou and Akritas (2016). If more linear combinations are needed, Christou (2020b) proposed a methodology that produces more vectors. The proposed algorithm provides consistent estimates to the directions that span the subspace and can be easily extended to any statistical functional.

- All of the above will not be of much value to the rest of the scientific community if there is not an accessible means of implementing these techniques. As such, I made my work available open source through my `R` package `quantdr` (Christou 2020c), available through the Comprehensive R Archive Network (CRAN).

Future Work. An alternative research area to QR with promising advantages that can be used to investigate the extreme parts of the data is that of *expectile regression (ER)*. Since its inception, research dedicated to ER has been scarce, leading to a disproportionate ratio of QR research to ER research. Studies have shown that estimating quantiles through expectiles using their one-to-one mapping is more efficient for a number of error distributions and provides a more global dependence on the form of the distribution. Therefore, expectiles provide promising alternatives to quantiles and create many future research directions.

Christou (2022+) provides a comprehensive work on dimension reduction techniques for conditional expectiles. Specifically, I introduced the *central expectile subspace* and developed the computational algorithm along with the asymptotic theory of the proposed estimator that defines linear combinations of $\mathbf{X}$ necessary for the conditional expectile. This work is also extended to nonlinear dimension reduction and is currently under review.

The lack of theoretical work around ER leaves room for many future developments. For example, one direction involves the estimation of the financial quantity called *expectile-based Value-at-Risk (EVaR)*; more details on estimating financial risk are given in Section 4.1. Specifically, Value-at-Risk (VaR) is sensitive to the magnitude of extreme losses. In these cases, depending on the tail behavior of the underlying distribution, VaR may be too conservative. A remedy for this has been given by EVaR, which is defined as the maximum possible loss of a given portfolio over a set time period and a given prudentiality level. EVaR can then be thought of as a flexible VaR where the data determines the tail probability instead of specifying it beforehand, as many financial institutions tend to do. In the future, I plan to consider estimating EVaR and provide a comprehensive comparison with VaR.

3 Extensions of the Central Quantile Subspace

3.1 Alternative Types of Data

Overview. The dimension reduction techniques described in Section 2 are limited to quantitative predictors $\mathbf{X}$ and do not incorporate other types of data, such as categorical predictors and/or longitudinal data. There are a lot of research areas that include different types of data. For example, medical studies deal with variables such as gender, diagnosis status, education level, or deal with data that involve repeated measurements at different time points. Straightforward application of the proposed dimension reduction techniques to regressions that include, for example, categorical predictors, may not be a proper object of inference of dimension reduction.

Contribution. Christou (2021a) considered partial dimension reduction techniques and introduced the *central partial quantile subspace* in order to facilitate analyzing data involving both quantitative and categorical predictor variables and/or longitudinal data. The idea is to define subpopulations as formed by the different categories of the categorical predictors or the different time points of the longitudinal predictors and perform dimension reduction within each subpopulation. Then, the combination of the central quantile subspaces within subpopulations will form the linear combinations necessary for performing dimension reduction. Simulation studies demonstrate that the computational algorithm is easy to implement and has good finite sample performance

3.2 Nonlinear Dimension Reduction

Overview. All of my previous work focuses on extracting *linear* combinations of $\mathbf{X}$ that contain all the information necessary about the conditional quantile. However, these linear techniques fail to find important *nonlinear* features, which can achieve greater dimension reduction, especially if the data are concentrated on a nonlinear low-dimensional space.

Contribution. Christou et al. (2021) proposed the *first nonlinear generalization* of the work of Christou (2020b). The extension is based on the *'kernel trick'*, a method that transforms the data into a very high-dimensional feature reproducing kernel Hilbert space and then seeks low-dimensional projections by applying a linear algorithm. In other words, linear directions in the feature space correspond to nonlinear directions in the original space. The performance of the algorithm is illustrated through several simulation examples and real data applications, which emphasize visualizations that capture various aspects of the data structure using the extracted nonlinear directions.

To ease the interpretation of the extracted nonlinear combinations, the *transformed dimension reduction* offers a good compromise. For that reason, Christou (2021b) suggested transforming the predictors monotonically and then using the linear dimension reduction techniques on the transformed variables.

4 Applications of Quantile Regression

4.1 Risk Estimation

Overview. Of the many real-world applications of the methodologies that I have developed, a great beneficiary is quantitative finance. Investors who wish to quantify their risk profiles, generally do so through conditional quantiles and various financial risk measures. Economic shocks are, by definition, extreme disturbances to financial markets. Exploring how institutions such as hedge funds or banks allocate their capital requirements accordingly relate directly to understanding the extreme regions of the data. Look no further than the 2000 dot-com bubble, the 2008 subprime mortgage crisis, or the 2016 oil collapse to understand the importance of accurately assessing an investment's exposure to market risk. A common risk measure is *Value-at-Risk* (VaR), which is the maximum possible loss of a given portfolio over a set time period and at a given significance level. Estimating VaR requires the use of statistical techniques that focus on those atypical values of the data, such as QR.

Contribution. Existing methods on VaR estimation are based on first modeling the time series of returns and then using the conditional distribution of the returns to calculate VaR. These parametric models rely on distributional assumptions, typically the normal distribution. However, it is well-known that financial returns exhibit non-normal characteristics including negative skewness, excess kurtosis, and heavy tails. For this reason, more flexible parametric families are sometimes used. In practice, however, no parametric family captures everything, and the dependence structure given by the time series ignores the possibility that the distribution of the innovations may change over time. Therefore, Christou and Grabchak (2019) proposed a *single index quantile regression time series model* (SIQRTS) that applies the non-iterative methodology of Christou (2020b) to the estimation of VaR. The method was used to estimate VaR of four major US market indices and the results show that the SIQRTS often outperforms other common methods.

An alternative to VaR is given by the *Expected Shortfall* (ES), defined as the conditional expectation of the loss given that this loss exceeds the VaR. Although there are several papers considering the estimation of ES, a useful representation of ES given by Taylor (2008) led to a new class of methods for estimating it. Therefore, Christou and Grabchak (2021) considered a novel variant of the ES estimator of Taylor (2008) and showed that the proposed estimator outperforms existing methods for right-skewed return distributions. An extensive simulation study and data analysis is performed to compare the performance of the proposed variant with several more standard parametric and nonparametric estimators of ES used as benchmarks.

Another common method that is used for modeling the volatility of financial returns is that of *stochastic volatility* (SV) models, which are very flexible and able to capture additional stylized features of the returns, such as skewness, excess kurtosis, and leverage effects. Although a lot of research effort has focused on option pricing for SV models, much less attention has been paid to the problem of calculating risk. However, this problem is not trivial as calculating ES for SV models requires working with the product of two random variables, where the distribution of the product may be quite complicated. Therefore, Grabchak and Christou (2021) introduced the first work on estimating ES for SV models and proposed *two Monte Carlo methods* that are easy to implement and have a competitive performance.

Recently, we introduced a new method for estimating the ES. This method uses the idea of composite quantile regression (CQR), which is a method that allows for the estimation of multiple quantiles at once. The idea relies on estimating ES by averaging over multiple VaR estimates and therefore, it is natural to use a CQR model. To the best of our knowledge, this is the *first work* that uses CQR for risk estimation. This work was the result of an invited contribution to Econometrics and Statistics (Christou and Grabchak 2022). This work was supported, in part, by funds provided by the University of North Carolina at Charlotte *(Faculty Research Grant)*.

Future Work. VaR describes a risk measure for an individual institution. However, in times of financial crisis, losses of certain financial institutions can threaten the financial system as a whole. To capture this dependency, we will use a measure of systemic risk, called conditional VaR (CoVaR), which is defined as the value at risk of the financial system conditional on an institution begin under distress. Therefore, my collaborator Dr. Michael Grabchak from UNC Charlotte and I will propose various techniques for estimating CoVaR and capture the association between an institution's stress and the overall risk in the financial system.

4.2 Sports Analytics

Overview. March Madness is the annual National Collegiate Athletic Association (NCAA) Division I men's basketball tournament and is one of the most popular sporting events in the United States in terms of viewership and bets placed. A common way for fans to bet or compete during the tournament is to enter in a *bracket competition* and predict the outcome of all 63 games before any of them has been played. Existing methods to complete a bracket consist of different ratings methods, such as the seed number, or models that focus on estimating win probabilities. A different approach for calculating the probability of a win is to utilize binary QR, where a more complete picture of a probability is obtained by averaging over multiple quantile levels.

Contribution. My student, Kimberly Mays, and I proposed the *average binary quantile regression (ABQR)* method that uses the dimension reduction method of Christou (2020b) for the estimation of the conditional quantiles and then estimates the probability of a win by averaging across a grid of conditional quantiles (Mays and Christou 2022+). The proposed method has a competitive performance and often outperforms other commonly used methods, especially for predicting upset games. This work is currently under review.

Future Work. The proposed method focuses on estimating the win probability of each team no mater what their opponent is. Future work will consider incorporating this information by update the win probabilities accordingly. Additionally, the proposed method can be applied to other single-elimination tournaments, such as the NCAA Division I women's basketball tournament, the College Football Playoff, or the Football Association Cup. Finally, we can consider a different 'success' for the binary response Y and calculate its probability, such as a team's probability of covering a spread of interest.

5 Dimension Reduction for Functional Data

Overview. There are situations where the variables of interest are considered a smooth stochastic process and require the use of functional data analysis (FDA). FDA methods are particularly useful in the medical field as the data are complex (images, videos, high frequency, scans, etc.) and require innovative analyses. Whereas certain medical-specific data types such as functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) have historically relied on more traditional methods, such as correlation analysis and temporal linear modeling, leveraging FDA allows for a more accurate representation of the data at each point along the continuum interval.

Contribution. I have been recently began to work in the area of FDA and appreciate its many advantages to various real-world problems. My collaborators, Dr. Eftychia Solea from CREST-ENSAI and Dr. Jun Song from Korea University, and I are currently working on a *robust alternative of sliced inverse regression for multivariate functional data*. Their many years of experience in the field of FDA have contributed significantly in my own understanding of the subject. The proposed method is motivated by the fact that existing dimension reduction methods for functional data rely on the covariance operator of the random function $\mathbf{X}$, which might be sensitive to heavy-tailed data that are also contaminated with unusual observations. Therefore, we propose to replace the covariance operator with the robust spatial sign Kendall's tau covariance operator of $\mathbf{X}$. Simulation results show the advantage of the proposed methodology over the functional SIR when data exhibit skewness or unusual patterns. This work is under preparation.

Future Work. Although there is an extensive amount of work on functional mean regression, the research area of functional QR remains under-explored. Assuming that the response variable is scalar and the predictor is a random function (scalar-on-function), I will develop a method that replaces the infinite dimensional predictor with few finite predictors without losing any important information necessary to estimate the conditional quantile. The proposed methodology will offer an attractive alternative to investigate highly skewed and contaminated functional data. Moreover, as many medical data sets include categorical predictors (gender, diagnosis status, etc.), I will also extend the methodology under the presence of such predictor types. The proposed work will be applied to fMRI data in order to better understand the brain structure of healthy children and children with attention deficit hyperactivity disorder (ADHD), and to EEG data in order to investigate the brain performance of non-alcoholic and alcoholic individuals. This work will be supported by the *National Science Foundation (Launching Early-Career Academic Pathways in the Mathematical and Physical Sciences)*, Award # 2213140, $213,462, 2022-2024.

6 Miscellaneous Work

There are many research areas that can benefit from statistical analysis and I am always willing to help. For example, Dr. Moutaz Khouja and Dr. Antonis Stylianou from the Belk College of Business at UNC Charlotte, considered a multi-item newsvendor optimization problem, where I was responsible for the development of the statistical algorithm (Khouja et al. 2020). Other works include five conference proceedings, one of them voted as the Best Paper for Construction Division by the American Society for Engineering Education during the 126th Annual Conference and Exposition (Tymvios and Christou 2019) and another one was written with my student Rawan Al-Shaer (Al-Shaer et al. 2020); see my Curriculum Vitae (CV) for a complete list of the conference proceedings.

References

[1] Al-Shaer, R., Spring, J., and Christou, E. (2020). Learning the Associations of MITRE ATT&CK Adversarial Techniques, *IEEE Conference on Communication and Network Security (CNS)*, Virtual Conference, June 29 - July 1, 2020.

[2] Christou, E. (2020a). Robust Dimension Reduction using Sliced Inverse Median Regression, *Statistical Papers* 61:1799–1818.

[3] Christou, E. (2020b). Central quantile subspace, *Statistics and Computing* 30:677–695.

[4] Christou, E. (2020c). quantdr: Dimension Reduction Techniques for Conditional Quantiles, *R package*, Version 1.2.2.

[5] Christou, E. (2021a). Sufficient Dimension Reduction for Conditional Quantiles with Alternative Types of data, *Journal of Statistical Computation and Simulation* 92(2):300–317.

[6] Christou, E. (2021b). Transformed Central Quantile Subspace, *Statistics* 55(2):350–366.

[7] Christou, E. (2022+). Dimension Reduction Techniques for Conditional Expectiles, *under review*.

[8] Christou, E., and Akritas, M.G. (2016). Single index quantile regression for heteroscedastic data, *Journal of Multivariate Analysis* 150:169–182.

[9] Christou, E., and Akritas, M.G. (2018). Variable selection in heteroscedastic single index quantile regression, *Communication in Statistics - Theory and Methods* 47(24):6019–6033.

[10] Christou, E., and Akritas, M.G. (2019). Single index quantile regression for censored data, *Journal of Statistical Methods and Applications* 28(4):655–678.

[11] Christou, E., and Grabchak, M. (2019). Estimation of value-at-risk using single index quantile regression, *Journal of Applied Statistics* 46(13):2418–2433.

[12] Christou, E., and Grabchak, M. (2021). Estimation of Expected Shortfall using Quantile Regression: A Comparison Study, *Computational Economics* https://doi.org/10.1007/s10614-021-10164-z

[13] Christou, E., and Grabchak, M. (2022). Risk Estimation With Composite Quantile Regression, *Econometrics and Statistics* https://doi.org/10.1016/j.ecosta.2022.04.004

[14] Christou, E., Settle, A., and Artemiou, A. (2021). Nonlinear dimension reduction for conditional quantiles, *Advances in Data Analysis and Classification* 15:937–956.

[15] Cook, R.D., and Weisberg, S. (1991). Li, K.C.: Comment on 'Sliced inverse regression for dimension reduction'. *Journal of the American Statistical Association* 86:328–332.

[16] Grabchak, M., and Christou, E. (2021). A Note on Calculating Expected Shortfall for Discrete Time Stochastic Volatility Models, *Financial Innovation* 7(43).

[17] Khouja, M., Christou, E., and Stylianou, A. (2020). A Heuristic Approach to In-Season Capacity Allocation in a Multi-product Newsvendor Model, *Omega* 95, 102252.

[18] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 86(414):316–327.

[19] Luo, W., Li, B., and Yin, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest, *The Annals of Statistics* 42(1):382–412.

[20] Mays, K., and Christou, E. (2022+). March Madness Prediction using Quantile Regression, *under review*.

[21] Taylor, J.W. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics* 6(2):231–252.

[22] Tymvios, N., and Christou, E. (2019). Tenure: Perceptions of Requirements and Impediments for Civil Engineering & Construction Disciplines, *126th ASEE Annual Conference and Exposition*, Tampa, FL, June 15-19, 2019.