Tao Wang
Department of Economics
University of Victoria
Victoria, BC V8P 5C2

Research Statement*

1 Overview

I am an econometrician whose research develops robust econometric and machine-learning methods for complex data. A central theme of my work is that the conditional mean, although fundamental, is often not the only or most informative object for empirical analysis. In many economic, financial, educational, public-health, and climate applications, data are asymmetric, heavy-tailed, multimodal, censored, spatially correlated, temporally dependent, or distributed across heterogeneous populations. In these settings, mean-based methods can be unstable under extreme observations and may obscure the most representative features of the conditional distribution. My research develops econometric tools for such environments, with a particular focus on the conditional mode, which represents the most likely response value given covariates. Because the mode is tied to the local shape of the conditional distribution, it provides a natural and interpretable target when the empirical question concerns typical outcomes, dominant response patterns, robust prediction, or distributional structure. More broadly, my work contributes to robust distributional learning, including conditional modes, modal volatility, forecast distributions, distributional dependence, and scalable learning procedures with explicit econometric targets.

My research agenda has three connected lines. *The first develops modal regression as a flexible econometric framework for settings in which the conditional mean is incomplete or fragile*. This line begins with semi- and nonparametric models for conditional mode estimation and extends them to core econometric environments, including fixed-effects panel data, varying-coefficient models, functional and semi-functional covariates, measurement error, random truncation, spatial dependence, and nonstationary regressors. The goal is to make the conditional mode a practical and theoretically justified object for empirical researchers working with nonlinear, incomplete, dependent, or otherwise nonstandard data. *The second line extends mode-based and distributional methods to dynamic economic environments*. Many time series in economics, finance, public health, and climate applications exhibit persistence, volatility clustering, heavy tails, structural change, and evolving predictive relationships. My work in this area develops robust methods for nonlinear modal regression under dependence, modal and robust volatility modeling, online forecast combination, near-unit-root estimation, and distributional Granger causality. *The third line connects these econometric ideas with modern machine learning*. Contemporary data are often too large, high-dimensional, sequential, distributed, or heterogeneous for classical nonparametric methods to be applied directly. My work develops scalable methods for distributed mode learning, online kernel-based mode learning, optimal subsampling for functional data, robust transfer learning, graph-based item response modeling, and distributional learning. Across these areas, my objective is not simply to import

*Updated on May 30, 2026.

machine-learning tools into econometrics, but to design learning procedures with clear statistical targets, robustness properties, and theoretical guarantees.

The contribution of this agenda is not merely to replace the conditional mean with the conditional mode. Rather, my work develops econometric methods for empirical settings in which the relevant object is distributional, robust, or locally representative rather than purely average. In many applications, the key question is not only how a covariate changes the mean outcome, but how it changes the most likely outcome, the shape of the conditional distribution, the behavior of volatility, or the predictive distribution over time. This perspective is important for applied economics because heterogeneity, tail behavior, dependence, and incomplete observation are not minor technical complications; they often determine the substantive empirical conclusion. Methods focused only on average effects may understate distributional instability, obscure dominant response patterns, or produce inference that is overly sensitive to a small number of extreme observations. My research provides tools for these situations by treating robustness and distributional structure as central features of the econometric problem.

2 Modal Regression and Robust Nonparametric Econometrics

My first main research area develops modal regression as a general framework for robust semi- and nonparametric econometrics. The main econometric challenge is that the conditional mode is a natural and interpretable target, but it is technically more demanding than the conditional mean; it is defined through the maximizer of a conditional density, so estimation requires learning local distributional shape rather than only conditional expectation. My work develops models, estimators, and asymptotic theory that make this target usable in flexible empirical settings.

A central paper in this line is *Semiparametric Partially Linear Varying Coefficient Modal Regression* (Ullah et al., 2023, **Journal of Econometrics**). The paper addresses a tradeoff that appears repeatedly in applied econometrics, i.e., fully nonparametric models are flexible but can be hard to interpret and estimate precisely, while parametric models are interpretable but vulnerable to misspecification. We resolve this tension by combining linear components with varying-coefficient components under a modal regression objective, allowing some effects to remain directly interpretable while other effects vary flexibly with observed characteristics. In *Nonparametric Estimator for Conditional Mode with Parametric Features* (Wang, 2024a, **Oxford Bulletin of Economics and Statistics**), I consider a related problem from another angle. The paper uses available parametric features to guide nonparametric mode estimation without changing the target parameter, showing how structural information can improve conditional mode estimation while preserving the robustness of the mode as the object of inference.

A second part of this research studies modal inference when the response distribution is only partially observed or when the covariate structure is more complex than in standard cross-sectional regression. Many empirical data sets are not simple random samples with fully observed scalar outcomes. In *Kernel Mode-Based Regression under Random Truncation* (Wang and Yao, forthcoming-a, **Statistica Sinica**), the observed sample is distorted by random truncation, so the estimator must correct for the observation mechanism before recovering the conditional mode. In *Robust Semi-Functional Censored Regression*

(Wang, 2026, **Journal of Multivariate Analysis**), the response is censored and the covariates may be functional. In *Semi-Functional Varying Coefficient Mode-Based Regression* (Wang, 2025a, **Journal of Multivariate Analysis**), I study models in which covariates may be curves or trajectories rather than finite-dimensional variables. This setting is important for modern empirical work involving functional predictors such as time paths, signals, or repeated measurements. In *Optimal Subsampling for Functional Quasi-Mode Regression with Big Data* (Wang, 2025b, **Journal of Computational and Graphical Statistics**), I address the computational side of this problem by developing subsampling methods that retain the inferential target while reducing computational burden in large functional data sets.

A third part of this work addresses measurement error, spatial dependence, nonstationarity, and panel-data heterogeneity. Since modal regression depends on the conditional density, covariate measurement error can distort not only regression coefficients but also the estimated shape of the response distribution. In *Semiparametric Modal Regression with Varying Coefficients and Measurement Error* (Ullah and Wang, 2026, **Journal of Statistical Planning and Inference**), we study modal inference when covariates are contaminated. In *B-Splines Modal Estimation under Measurement Error with Deconvolution* (Wang, forthcoming-a, **The Canadian Journal of Statistics**), I develop spline-based deconvolution methods for modal estimation. Spatial and nonstationary data raise additional issues because observations may be correlated across locations or evolve with persistent regressors. *Nonparametric Spatial Mode-Oriented Regression* (Wang and Yao, 2025a, **Electronic Journal of Statistics**) and *Nonparametric Spatial Modeling towards the Mode* (Wang and Yao, forthcoming-b, **Statistica Sinica**) develop modal methods that account for spatial dependence. *Kernel Mode-Based Varying Coefficient Models with Nonstationary Regressors* (Wang and Yao, 2026a, **Journal of Nonparametric Statistics**) addresses mode-based varying-coefficient estimation with nonstationary regressors, while *Adaptive Robust Estimation for Varying Coefficient Models* (Wang and Yao, 2026b, **Computational Statistics** & **Data Analysis**) develops adaptive robust procedures for varying-coefficient models. This agenda also enters a core empirical framework in economics through *Modal Regression for Fixed Effects Panel Data* (Ullah et al., 2021, **Empirical Economics**). Fixed-effects panel models are widely used to control for unobserved heterogeneity, but they usually focus on conditional means. This paper develops a modal alternative, allowing researchers to estimate the most likely conditional response while accounting for unit-specific unobserved effects.

This line of work develops modal regression from a specialized robust estimator into a broader econometric framework. Across nonlinear models, incomplete outcomes, functional covariates, measurement error, spatial dependence, nonstationarity, big data, and panel-data heterogeneity, the central goal is to make the conditional mode both theoretically tractable and empirically usable. The resulting methods provide applied researchers with robust alternatives to mean-based econometrics when average effects do not adequately describe the dominant response pattern or the local shape of the conditional distribution.

3 Time Series, Dependence, Volatility, and Distributional Dynamics

My second main research area develops robust methods for dependent and dynamic data. The econometric difficulty is that time series are rarely independent and often display persistence, nonlinear dependence, heavy tails, volatility clustering, and structural change. In financial, macroeconomic, public-health, and climate data, the conditional mean may be unstable or incomplete as a forecasting target, especially when extreme shocks affect averages more than the most likely trajectory. My work develops mode-based and distributional methods for these settings across forecasting horizons and policy contexts, with the broader aim of expanding time-series econometrics beyond mean-based prediction.

In *Nonlinear Modal Regression for Dependent Data with Application for Predicting COVID-19* (Ullah et al., 2022, **Journal of the Royal Statistical Society: Series A**), we study nonlinear modal regression under dependence and apply the method to epidemic prediction. The paper changes forecasting target from average conditional trajectory to most likely conditional trajectory, providing a robust alternative when outcomes are volatile and affected by unusual shocks. This theoretical direction continues in *Nonlinear Kernel Mode-Based Regression for Dependent Data* (Wang, 2024b, **Journal of Time Series Analysis**), which develops kernel mode-based estimation under dependence conditions, and in *Parametric Modal Regression with Autocorrelated Error Process* (Wang, 2025c, **Statistica Sinica**), which studies modal regression when the regression errors are serially correlated. These papers reformulate modal inference for time-series environments where standard independent-data arguments no longer apply.

Volatility modeling provides a natural application because financial time series are heavy-tailed and dominated by rare influential movements. Standard volatility methods summarize risk through conditional means of squared returns or related transformations, but such summaries can be sensitive to extremes. In *Modal Volatility Function* (Ullah and Wang, 2025, **Journal of Time Series Analysis**), we introduce a mode-based volatility function that characterizes the most likely conditional behavior of volatility-related outcomes. In *Mode Meets Mean: A New Robust Volatility* (Wang, forthcoming-b, **Journal of Time Series Analysis**), I connect modal and mean-based perspectives to construct a robust volatility measure. The purpose is not to discard volatility ideas, but to provide volatility tools that remain informative when the distribution of returns is asymmetric, heavy-tailed, or affected by outliers.

My more recent time-series work moves from point prediction to distributional prediction. *Testing Distributional Granger Causality with Entropic Optimal Transport* (Wang, forthcoming-c, **Journal of Time Series Analysis**) asks whether one time series helps predict the entire conditional distribution of another. Classical Granger causality is often formulated through conditional means, but predictive information may appear through variance, skewness, tail behavior, or other distributional features. Entropic optimal transport provides a computationally feasible way to compare conditional distributions while preserving a meaningful geometric interpretation. In *Online Randomized Distributionally Robust Forecast Combination for Dependent Data* (Wang, forthcoming-d, **Journal of Time Series Analysis**), I study forecast combination when observations are dependent and the forecasting environment may change over time. The method adapts sequentially to model uncertainty and dependence. In *On Robust Estimation for Moderate Deviations from a Unit Root* (Wang, forthcoming-e, **TEST**), I study robust estimation in near-unit-root settings, where persistence can make conventional procedures fragile.

This line of work extends robust econometric learning from cross-sectional regression to time-series environments. Across these dynamic settings, the common goal is to develop methods that remain reliable when time-series data are persistent, heavy-tailed, and distributionally unstable. These methods give empirical researchers tools for studying not only the expected path of a process, but also the most likely path, the evolution of volatility, and the predictive content of the full conditional distribution.

4 Machine Learning, Scalability, and High-Dimensional Data

My third main research area connects robust econometric ideas with modern machine learning. The motivation is that many contemporary data sets are too large, high-dimensional, sequential, distributed, or heterogeneous for classical nonparametric procedures to be used directly, especially under computational, privacy, and distributional constraints. At the same time, many machine-learning methods are designed primarily for predictive performance and do not always correspond to clearly defined econometric targets. My work in this area develops scalable learning procedures that preserve explicit statistical objects, such as conditional modes, transfer-learning targets, and distributional structure.

In *Distributed Learning for Kernel Mode-Based Regression* (Wang, 2025d, **The Canadian Journal of Statistics**), I study conditional mode estimation when data are distributed across machines or institutions. This problem is relevant when data cannot be centralized because of scale, privacy, or institutional constraints. The paper develops distributed procedures that reduce computational burden while preserving the conditional mode as the target of inference. A related challenge arises when data arrive sequentially rather than being stored as a fixed sample. In many economic, financial, public-health, and online-platform settings, observations are collected over time and full re-estimation after each new observation is computationally inefficient. In *Online Kernel-Based Mode Learning* (Wang and Yao, 2025b, **Journal of Computational and Graphical Statistics**), we develop online updating rules for kernel-based mode estimation. The paper connects mode-based econometrics with streaming-data environments by allowing the estimator to update as new observations arrive while retaining the same population target. This provides a practical framework for robust learning in applications where the data-generating process is observed continuously and may evolve over time.

Transfer learning addresses a different problem, i.e., how to use information from related populations without importing bias from the source distribution. This issue is important in economics and social science because data from one market, region, institution, or population may be informative for another, but the source and target distributions are rarely identical. In *Doubly Debiased Robust Subsampling for Transfer Learning* (Wang and Wong, forthcoming, **Journal of Machine Learning Research**), we develop subsampling and debiasing methods for transfer learning. The central idea is to exploit useful source information while protecting the target analysis against transfer bias. In *Geometry-Misalignment in Distributional Learning* (Wang and Zhong, 2026, **International Conference on Machine Learning**), we study a related foundational issue; the geometry imposed by a learning algorithm may not align with the statistical structure of the distributions being learned. This work connects distributional learning, geometry, and robustness, and reflects my broader goal of designing machine-learning methods around explicit econometric and statistical targets.

My applied machine-learning work also includes *A Machine Learning Strategy for Autism Screening in Toddlers* (Wang et al., 2019, **Journal of Developmental & Behavioral Pediatrics**). This paper develops data-driven screening tools for autism in toddlers and illustrates my interest in methods that are useful for real data and meaningful decision problems. The machine-learning part of my research therefore does not treat prediction as an end in itself. It develops scalable procedures for distributed data, streaming data, transfer learning, and distributional learning while keeping the inferential target interpretable.

5 Additional Econometric Methods and Applications

My work extends robust and distributional ideas to factor modeling, forecast evaluation, and educational measurement. These projects are connected to the main agenda because they address the underlying problem; empirical researchers often work with high-dimensional, heterogeneous, or networked data for which standard mean-based or low-dimensional methods may not capture the relevant structure. Rather than treating these applications as separate from my work on modal and distributional econometrics, I view them as settings in which robustness, interpretability, and explicit econometric targets are essential.

In *Mode-Adaptive Factor Models* (Wang, 2025e, **Scandinavian Journal of Statistics**), I develop factor modeling methods from a mode-oriented perspective. Factor models are widely used in macroeconomics and finance to summarize high-dimensional data, but conventional approaches can be sensitive to heavy tails, asymmetry, and heterogeneous latent structures. The paper asks how common components can be estimated when the central tendency of the data is better represented by the most likely pattern than by the average pattern. By adapting factor modeling to a modal target, this work provides a robust alternative for extracting latent structure from high-dimensional economic and financial data.

Forecast evaluation raises a different but related econometric challenge. In *Estimation and Testing of Forecast Rationality with Many Moments* (Lee and Wang, 2025, **Macroeconomic Dynamics**), we study forecast rationality when the information set generates many moment conditions. Standard forecast-rationality tests are often designed for low-dimensional settings, but modern forecasting environments may involve many predictors, many instruments, or many restrictions implied by the available information. This paper develops methods for estimation and testing in such many-moment settings, contributing to the econometric literature on forecast evaluation and rational expectations.

My forthcoming work *Graph Neural Item Response Model for Networked Learning Environments* (Wang and Zhong, forthcoming, **Journal of Educational and Behavioral Statistics**) extends this research agenda into educational measurement. Item response models treat persons and items as conditionally independent after accounting for latent traits and covariates. In many learning environments, however, students, items, and response contexts are connected through networks. This paper incorporates graph neural structure into item response modeling while preserving interpretable psychometric components. The goal is to use network information without turning the model into a black-box prediction tool.

These projects show my agenda extends beyond modal regression narrowly defined. The common theme is developing econometric and statistical learning methods for settings in which low-dimensional or mean-based approaches are incomplete. Whether the problem is extracting factors from high-dimensional data, testing forecast rationality with many moments, or modeling networked educational responses, my objective is to build methods that remain interpretable, robust, and useful for empirical research.

6 Future Research Agenda

My future research will continue to develop econometric methods for settings in which the empirical object is distributional, robust, or dynamically evolving. One direction is to deepen the theory of modal and distributional regression. Conditional modes are useful empirical objects, but several inferential questions remain open, including simultaneous inference, high-dimensional modal regression, and causal designs in which treatment effects are better described by changes in the most likely outcome or by changes in the conditional distribution. I am especially interested in developing frameworks that compare and combine means, quantiles, modes, and optimal-transport-based summaries.

A second direction is distributional time-series econometrics. Economic and financial time series often exhibit persistence, volatility clustering, structural breaks, and changing tail behavior. My recent work on modal volatility, distributional Granger causality, online forecast combination, and near-unit-root estimation provides a foundation for studying how the full predictive distribution evolves over time. The goal is to develop methods that allow economists to study not only expected paths, but also the most likely paths, distributional instability, and time-varying predictive relationships.

A third direction is econometric machine learning with explicit targets. Many machine-learning methods are effective prediction tools, but their population targets are often unclear from an econometric perspective. My work on distributed mode learning, online mode learning, transfer learning, graph-based models, and distributional learning points towards a research agenda in which scalability and interpretation are developed together. I aim to build methods that can handle large, heterogeneous, and distributed data while preserving robustness and meaningful econometric interpretation.

These future directions will continue to be shaped by applications in financial volatility, economic forecasting, climate data, public health, and educational measurement. These areas share features that motivate my methodological work: nonlinear dependence, heavy tails, heterogeneity, incomplete observation, and distributional change. My objective is to develop methods that are theoretically grounded, computationally feasible, and useful for empirical researchers facing these forms of complexity.

References

Lee, Tae-Hwy and Tao Wang (2025). Estimation and Testing of Forecast Rationality with Many Moments. *Macroeconomic Dynamics*, 29, e124.

Ullah, Aman and Tao Wang (2025). Modal Volatility Function. *Journal of Time Series Analysis*, 46, 748-773.

Ullah, Aman and Tao Wang (2026). Semiparametric Modal Regression with Varying Coefficients and Measurement Error. *Journal of Statistical Planning and Inference*, 240, 106307.

Ullah, Aman, Tao Wang, and Weixin Yao (2021). Modal Regression for Fixed Effects Panel Data. *Empirical Economics* (Special Issue Honouring B. Baltagi), 60, 261-308.

Ullah, Aman, Tao Wang, and Weixin Yao (2022). Nonlinear Modal Regression for Dependent Data with Application for Predicting COVID-19. *Journal of the Royal Statistical Society Series A*, 185, 1424-1453.

Ullah, Aman, Tao Wang, and Weixin Yao (2023). Semiparametric Partially Linear Varying Coefficient Modal Regression. *Journal of Econometrics* 235, 1001-1026.

Wang, Tao (2024a). Nonparametric Estimator for Conditional Mode with Parametric Features. *Oxford Bulletin of Economics and Statistics*, 86, 44-73.

Wang, Tao (2024b). Nonlinear Kernel Mode-Based Regression for Dependent Data. *Journal of Time Series Analysis*, 45, 189-213.

Wang, Tao (2025a). Semi-Functional Varying Coefficient Mode-Based Regression. *Journal of Multivariate Analysis*, 207, 105402.

Wang, Tao (2025b). Optimal Subsampling for Functional Quasi-Mode Regression with Big Data. *Journal of Computational and Graphical Statistics*, 34, 552-566.

Wang, Tao (2025c). Parametric Modal Regression with Autocorrelated Error Process. *Statistica Sinica*, 35, 457-478.

Wang, Tao (2025d). Distributed Learning for Kernel Mode-Based Regression. *The Canadian Journal of Statistics*, 53, e11831.

Wang, Tao (2025e). Mode-Adaptive Factor Models. *Scandinavian Journal of Statistics*, 52, 1206-1238.

Wang, Tao (2026). Robust Semi-Functional Censored Regression. *Journal of Multivariate Analysis*, 211, 105491.

Wang, Tao (forthcoming-a). B-Splines Modal Estimation under Measurement Error with Deconvolution. *The Canadian Journal of Statistics*.

Wang, Tao (forthcoming-b). Mode Meets Mean: A New Robust Volatility. *Journal of Time Series Analysis*.

Wang, Tao (forthcoming-c). Testing Distributional Granger Causality with Entropic Optimal Transport. *Journal of Time Series Analysis*.

Wang, Tao (forthcoming-d). Online Randomized Distributionally Robust Forecast Combination for Dependent Data. *Journal of Time Series Analysis*.

Wang, Tao (forthcoming-e). On Robust Estimation for Moderate Deviations from a Unit Root. *TEST*.

Wang, Tao, Luke E. K. Achenie, Angela Scarpa, Reina S. Factor, Diana L. Robins, and D. Scott McCrickard (2019). A Machine Learning Strategy for Autism Screening in Toddlers. *Journal of Developmental & Behavioral Pediatrics*, 40, 369-376.

Wang, Tao and Weng Kee Wong (forthcoming). Doubly Debiased Robust Subsampling for Transfer Learning. *Journal of Machine Learning Research*.

Wang, Tao and Weixin Yao (2025a). Nonparametric Spatial Mode-Oriented Regression. *Electronic Journal of Statistics*, 19, 4256-4311.

Wang, Tao and Weixin Yao (2025b). Online Kernel-Based Mode Learning. *Journal of Computational and Graphical Statistics*, 34, 1498-1512.

Wang, Tao and Weixin Yao (2026a). Kernel Mode-Based Varying Coefficient Models with Nonstationary Regressors. *Journal of Nonparametric Statistics*, 38, 616-652.

Wang, Tao and Weixin Yao (2026b). Adaptive Robust Estimation for Varying Coefficient Models. *Computational Statistics & Data Analysis*, 222, 108400.

Wang, Tao and Weixin Yao (forthcoming-a). Kernel Mode-Based Regression under Random Truncation. *Statistica Sinica*.

Wang, Tao and Weixin Yao (forthcoming-b). Nonparametric Spatial Modeling towards the Mode. *Statistica Sinica*.

Wang, Tao and Xiaoting Zhong (2026). Geometry-Misalignment in Distributional Learning. *International Conference on Machine Learning (ICML)*.

Wang, Tao and Xiaoting Zhong (forthcoming). Graph Neural Item Response Model for Networked Learning Environments. *Journal of Educational and Behavioral Statistics*.