Research Statement

Yaping Wang*
November 4, 2025

I am an econometrician working at the intersection of *high-dimensional econometrics*, *statistical learning*, and *empirical macro-finance*. My research studies how to make economic forecasting and risk measurement reliable when the data are high-dimensional, noisy, and driven by both latent common shocks and idiosyncratic forces. I focus on settings where traditional factor models or mean regression break down, such as weak factor structures, heterogeneous policy responses, and tail risks. To address these challenges, I combine economic structure with modern tools from factor analysis, quantile methods, and statistical learning to understand when factor-based predictions work, when they fail, and how they can be improved.

My research follows this agenda through three related papers. The first and central paper, my Job Market Paper (JMP), develops a high-dimensional *quantile* regression framework that combines factor structures with sparse heterogeneity to improve distributional and tail forecasting. The second paper provides a theoretical foundation for principal component regression (PCR) and explains why PCR-based forecasts can work even the prediction target is generated by a factor model. The third closes the loop by selecting the number of factors endogenously using cross-validation rather than information criteria.

How can we capture both common and idiosyncratic forces in predicting the distribution of economic outcomes?

In my JMP *Bridging Dense and Sparse Models in High-Dimensional Quantile Regression*, I study this question by developing a high-dimensional quantile regression framework that combines latent factors with sparse idiosyncratic components. While factor models are effective at capturing common comovements, they often fail to explain tail risks or distributional effects such as asymmetric responses to monetary shocks or sector-specific vulnerabilities. Conversely, sparse regressions capture heterogeneity but ignore pervasive common forces. This limitation motivates a model in which conditional quantiles depend on both low-dimensional latent factors and a sparse set of idiosyncratic predictors.

I estimate the latent factors from the data via PCA, combine them with high-dimensional idiosyncratic predictors, and apply an ℓ_1-penalized quantile regression only to the idiosyncratic coefficients. I show theoretically that this estimator can consistently recovers both dense and sparse components even when factors are weak and the data exhibit time dependence. In applications, the integrated approach improves distributional forecasts relative

* Department of Economics and Business, Universitat Pompeu Fabra and Barcelona School of Economics; E-mail: `yaping.wang@upf.edu`.

to dense-only and sparse-only benchmarks, uncovering asymmetries that mean regression obscures. Methodologically, it clarifies when factor augmentation reduces omitted-variable bias in high-dimensional quantile models, and economically, it links pervasive shocks with localized transmission channels across the conditional distribution.

Why does principal component regression work so well in macroeconomic forecasting, even when the data do not follow a factor model?

In my second paper, *Performance of Empirical Risk Minimization for Principal Component Regression*, I study PCR as a method for predicting the conditional mean of economic variables. PCR is widely used in macroeconomic forecasting, where a few linear combinations of many predictors often capture most of the predictive content. It works well in practice, but existing theory largely justifies PCR only under factor models. In reality, factors may be weak or noisy, and the true predictive relationship need not lie exactly in the factor space.

I reinterpret PCR as an empirical risk minimization method with spectral regularization. Instead of asking whether PCR recovers the true factors, I ask whether it achieves low prediction error. I show that PCR can attain risk close to the best linear predictor even when no exact factor structure holds. The key insight is that if most predictive variation lies in the leading principal components of predictors, then projecting onto them introduces only small approximation error while reducing estimation variance. I provide nonasymptotic risk bounds that formalize this bias-variance trade-off and show that PCR adapts to both strong and weak factor regimes under weak time dependence. This explains why factor-based methods perform well for *mean* forecasting, while motivating my Job Market Paper, which extends these ideas to *quantile* regression.

This paper is currently in Revise and Resubmit at *Econometric Theory*.

How should we choose the number of factors when prediction, rather than factor recovery, is the objective?

In my third paper, I study factor selection as a supervised learning problem. While the previous papers justify and extend factor-based prediction, both rely on choosing the number of factors. In practice, this decision is usually made using information criteria or fixed rules, even though the objective is predictive performance rather than structural recovery.

I propose selecting the number of factors using cross-validation, where factors are re-estimated in each training sample and the value of k is chosen to minimize out-of-sample prediction error in y. I show that this approach achieves near-oracle predictive risk under both strong and weak factor structures, and extend the analysis to time-series data using blocked cross-validation. This paper closes the loop of my research agenda by making factor-based forecasting not only theoretically valid but also fully data-adaptive.

Future Research.

One direction is to develop *quantile impulse response functions* in high-dimensional settings with latent factors, which would allow researchers to study how monetary or financial shocks affect the entire distribution of outcomes, rather than only the mean, and whether these responses differ across recessions and expansions. This involves combining factor-augmented local projections with external instruments and quantile techniques.

A second direction is to allow factor strength and sparsity patterns to vary over time. During crises, common shocks may become more dominant while idiosyncratic forces weaken, or vice versa. I plan to build predictive models where the number of factors and the relevance of idiosyncratic components are allowed to evolve, and to derive conditions under which estimation and prediction remain valid under such instabilities.

These tools can be applied to questions such as inflation-at-risk, growth-at-risk, or heterogeneous effects of monetary policy across sectors or regions. More broadly, I aim to develop econometric methods that make distributional forecasting and risk analysis more credible in macroeconomic settings, while remaining closely connected to economic theory and policy use.