Research Statement of Zhipeng Liao

My research field is econometric theory and applied econometrics. My work in econometrics is motivated by empirical economic problems where standard methods may not be applicable or may produce misleading results. Examples include estimators that are difficult to compute or have slow convergence rates, tests that over-reject, and confidence intervals that under-cover the true parameters. Specifically, the topics I have been working on include model selection and model averaging, estimation and inference of semi/nonparametric models, adaptive estimation of high-dimensional models, and inference of irregularly identified econometric models.

1. Model Selection and Model Averaging

Model selection in high-dimensional models is very challenging for traditional procedures (e.g., information-criterion methods). Since there are many models to investigate, these procedures become computationally intensive or even intractable in practice. In contrast, the least absolute shrinkage and selection operator (LASSO) is attractive because it conducts model selection in a one-step penalized estimation. Liao (2013) provides a simple method which employs LASSO to perform moment selection and efficient estimation simultaneously in the GMM framework. Liao (2013)’s method is particularly useful when there are many moment conditions to select. Cheng and Liao (2015) generalize Liao (2013) and allows for a divergent number of moment conditions. It revises the adaptive penalty in Liao (2013) so that only informative and valid moment conditions are selected. Liao and Phillips (2015) propose a LASSO approach to perform simultaneous estimation of the cointegrating rank and autoregressive order in cointegrated error correction models. Cheng, Liao and Schorfheide (2016) provides an adaptive group-LASSO estimator that consistently determines the numbers of pre- and post-break factors, and the stability of factor loadings in high-dimensional factor models with possible structural instability. A novel feature of the LASSO estimator is its robustness to unknown break dates because existing procedures either over-estimate the number of factors by neglecting the breaks or require known break dates.

In economic studies, there are often competing theories for one phenomenon. Even when there is only one theory, it can rarely pin down an empirical model to take to the data. Liao and Shi (2020) develop an easy-to-implement test to select from two models according to their population quasi-log likelihood. The test determines the statistical significance of the quasi-log likelihood difference and, when the difference is significantly different from zero, draws the directional conclusion that one model is better than the other. We show that our test has uniform size control regardless of the model relationships (i.e., nested or non-nested), and nontrivial power against root-n local alternatives. We also provide valid post model selection confidence intervals of the model parameters.

Instead of conducting model selection, one may combine estimators from different models and construct an averaging estimator. Cheng, Liao and Shi (2019) propose a GMM averaging estimator which uses a data dependent weight to combine a conservative GMM estimator based on valid moment conditions and an aggressive GMM estimator based on both valid and possibly misspecified moment conditions. We develop a uniform asymptotic approximation to the finite-sample risk and demonstrate the advantage of this averaging estimator over the post model selection estimator.

2. Estimation and Inference of Semi/nonparametric Models

The sieve method is an attractive approach in estimating semi/nonparametric models because of its similarity to the parametric estimation in finite samples. Part of my work in this field is on showing
the asymptotic normality of the sieve plug-in estimator of possibly irregular (i.e., slower than root-n estimable) functionals and providing associated inference procedures. Chen, Liao and Sun (2014) studies the estimation and inference of semi/nonparametric time series models. We show that, even when the sieve score process is not a martingale difference, the asymptotic variances of plug-in sieve M estimators of irregular functionals are the same as those when observations are independently distributed. However, ignoring the temporal dependence leads to inaccurate inference in finite samples. We propose an easy-to-compute and more accurate inference procedure based on a "pre-asymptotic" sieve variance estimator that captures temporal dependence of unknown forms. Chen and Liao (2014) complements Chen, Liao and Sun (2014) by providing consistent variance estimators of plug-in sieve M estimators of possibly irregular functionals, and showing that, even for hypothesis testing of irregular functionals, the sieve likelihood ratio statistic is asymptotically Chi-square distributed. Buchinsky, Li and Liao (2021) propose sieve minimum distance (SMD) estimation of semiparametric models in applications that combine different data sets. We establish the asymptotic normality of SMD estimators and provide consistent estimators of their asymptotic variances and hence Gaussian-based inference procedures. Hahn, Liao and Ridder (2018) studies two-step sieve M estimation of general semi/nonparametric models, where the second step involves sieve estimation of unknown functions that may use the nonparametric estimates from the first step as inputs, and the parameters of interest are functionals of unknown functions estimated in both steps. We provide a general theory on the asymptotic normality of plug-in two-step sieve M estimators, consistent estimators of asymptotic variances of two-step sieve estimators and hence Wald-type procedures. Hahn, Liao and Ridder (2021) reconsider the assumptions that ensure the identification of the production function in Olley and Pakes (1996). We show that an index restriction plays a crucial role in the identification, especially if the capital stock is measured by the perpetual inventory method. The index restriction is not sufficient for identification under sample selectivity. The index restriction makes it possible to derive the influence function and the asymptotic variance of Olley-Pakes estimator.

Ackerberg, Chen, Hahn and Liao (2014) characterize the semiparametric efficiency bound for a class of semiparametric models in which the unknown nuisance functions are "exactly identified" via nonparametric conditional moment restrictions, and the finite dimensional parameters are "over-identified" via unconditional moment restrictions involving the nuisance functions. We show that semiparametric two-step optimally weighted GMM estimators achieve the efficiency bound, where the nuisance functions could be estimated via any consistent nonparametric procedures in the first step. Chen and Liao (2015) studies the two-step GMM estimation of semi/nonparametric where an unknown function is estimated in the first step by sieve extremum estimation. We show that although long-run variances of semiparametric two-step GMM estimators may not have explicit expressions, they can be well approximated by sieve variances which have closed-form expressions. Based on sieve variances, we provide auto-correlation robust inference for semiparametric two-step GMM with weakly dependent data.

One of my recent work, Li and Liao (2020) provide a strong approximation theory for mixingale arrays. This result is used to justify the asymptotic validity of a uniform confidence band for series estimators. We show that it can also be used to conduct nonparametric specification test for conditional moment restrictions. We provide a Stata package of our nonparametric specification test, and write a paper (Gao, Li and Liao, 2020) to show how to use this package in practice. Li, Liao and Quaedvlieg (2021) develops a new strong approximation theory which substantially improves the theory of Li and Liao (2020) by allowing the dimension of the mixingale arrays grow much faster with the sample size. In this paper, we apply the new approximation to construct a test for the
conditional superior predictive ability (CSPA) of a family of forecasting methods with respect to a benchmark. By inverting the CSPA tests for a set of benchmarks, we obtain confidence sets for the uniformly most superior method. The usefulness of the method is demonstrated in empirical applications on volatility and inflation forecasting. The new strong approximation is also applied in Horvath, Li, Liao and Patton (2020) to develop a new specification test for the dynamic quantile models. We propose a novel bootstrap method in this paper and show that it significantly outperforms the benchmark asymptotic approximation in finite samples, especially for tail quantiles such as Value-at-Risk (VaR). We use the proposed new test to study the VaR and CoVaR of a collection of US financial institutions.

The strong approximation also has many potentially applications in continuous time series models. For example, it can be used to construct uniform inference for the spot volatility process in financial econometrics. Jacod, Li and Liao (2021) establish a uniform inference theory for the volatility process, through the development of a new coupling result for the spot volatility estimators. We also extend the theory to a multivariate setting for constructing uniform confidence bands for the nonlinear functionals of spot volatility process, such as the beta and idiosyncratic variance processes. Consistent estimation the spot volatility process requires the block size (or bandwidth), say k, diverges with the sample size. However, the inference based on the divergent-k theory may suffer size distortion in finite samples, since k is always finite in practice. Bollerslev, Li and Liao (2021) provide a fixed-k asymptotic theory for the spot volatility process. While the resulting spot volatility estimator is no longer consistent, our fixed-k theory permits the construction of valid uniform confidence bands for the volatility process. An empirically realistically calibrated simulation study underscores the practical reliability of the new inference procedures.

The bootstrap is widely used in practice for inference of semi/parametric econometric models. Asymptotic justification of the bootstrap often takes the form of weak convergence of the bootstrap distribution to some limit distribution. Theoretical literature recognized that the weak convergence does not imply consistency of the bootstrap second moment or the bootstrap variance as an estimator of the asymptotic variance, but such concern is not always reflected in the applied practice. The bootstrap variance is usually used in the empirical studies to conduct hypothesis test. Hahn and Liao (2021) show that such common bootstrap based standard error in fact leads to a potentially conservative inference.

3. Adaptive Estimation of High-dimensional Models

Estimation of high-dimensional models usually involves tuning parameters. In practice, cross-validation is a popular method of selecting such parameters. However, theoretical properties of the cross-validated estimators are not known in many cases. Chetverikov, Liao and Chernozhukov (2020) studies LASSO estimators of high-dimensional linear regression models where the tuning parameter of the LASSO penalty is selected by a V-fold cross-validation procedure. For models with Gaussian noise, we show that the cross-validated LASSO estimator achieves the optimal convergence rate up to a logarithm term under the prediction norm. When the noise is not Gaussian, the cross-validated LASSO estimator is shown to have a convergence rate not slower than the square root of the optimal rate up to a logarithm term under the prediction norm. Chetverikov and Liao (2021) studies nonparametric quantile estimators when unknown functions are estimated by the sieve method and the number of sieve functions are selected by a V-fold cross-validation procedure. We show that the cross-validated quantile estimator achieves the optimal convergence rate up to a logarithm term under the L2 norm.
4. Inference of Irregularly Identified Econometric Models

In some econometric models, parameters are globally identified but their local identification conditions may fail. For example, the Jacobian of the moment conditions in GMM models may be zero. The local identification failure leads to a slow rate of convergence of the GMM estimator and a non-standard asymptotic distribution of the J-test statistic. Lee and Liao (2018) find that the zero Jacobian may provide non-trivial local identification for unknown parameters. By exploiting such information in estimation, we provide GMM estimator and over-identification tests with standard properties. These standard properties are attractive because the resulting GMM estimation and inference are more accurate in finite samples.

Hahn and Liao (2018) investigate whether there can exist regular estimators in models characterized by nonparametric instrumental variable (NPIV). We show by several examples that regular estimation is impossible in general for nonlinear functionals. Since regular estimability is often understood to be equivalent to the feasibility of root-n-consistent estimation, our result implies that root-n-consistent estimation is impossible for many estimands in models characterized by NPIV.

One of my recent paper, Cheng, Dou an and Liao (2021) studies evaluating the validity of many influential macro asset pricing models through testing their moment conditions. The main challenging is on how to use the valid but possibly weak information to improve the power of the test and at the same time maintain its size control. We propose a new test which controls size regardless the identification strength of the valid information and becomes the optimal Chi-square test when the valid information is strong. The new test also shows power improvement over the traditional J test when the valid information is weak.

5. Research in the Applied Econometrics

I have been working on several projects in the applied econometrics. Kim, Liao and Tornell (2014) forecasts whether speculators will be increasing or decreasing their positions and use this predicted state to form both directional and point exchange rate forecasts. Empirical forecasting results show that our method is significantly better than the random-walk model. Kim, Kim, Liao and Tornell (2018) presents a heterogeneous-agent asset pricing model where fundamental shocks lead to amplification cycles (bubbles) and the principle of contrarian opinion holds: in equilibrium, informed forward-looking speculators find it optimal to ride the bubble until a time when they switch the sign of their positions. At this switching time, the bubble continues to grow as less-informed speculators become more optimistic (pessimistic). Based on the implications of the model, we propose a forecasting strategy that estimates structural breaks in the bivariate process followed by exchange rates and speculators’ positions. For the six most traded currency pairs, our forecasts outperform the random-walk model over forecasting horizons from 1 month to 12 months.

In my recent work, Liao and Liu (2020) studies the cross-sectional regression which is one of the most popular methods for estimating the linear asset pricing models. The cross-sectional regression is conducted in a two-step procedure and is computationally attractive since the estimators in both steps have the closed-form expressions. On the other hand, one may worry that the cross-sectional regression loses efficiency since the unknown parameters are not estimated simultaneously. In this paper, we show that the optimally weighted cross-sectional regression estimator is as efficient as the one-step estimator which is computationally intensive. Our findings justify the widely used cross-sectional regression methods in the literature.