My research is the field of econometric theory and applied econometrics, studying empiri cally-motivated problems where standard inference approaches could be misleading: estimators are inaccurate, tests over-reject true hypotheses, and confidence intervals under-cover. These issues arise due to insufficient information in the data. In particular, I focus on two situations where data insufficiency leads to serious consequences. First, the sample variation that identifies structural parameters is small compared to unmodeled noise in the data. The problem may be due to poor quality of the data or the highly nonlinear nature of many economic models. This phenomenon is called weak identification of the parameters. Second, the sample size is small compared to the number of parameters. In this case, the sample size could be rather large by a traditional standard, but the model to be estimated is much larger than a standard one for modeling flexibility. This is the high-dimensional problem discussed below. A unifying theme of these research topics is to develop *robust* econometric methods that are valid for a large class of dynamic economic models under weak assumptions.

Weak Identification Robust Inference Weak identification occurs when information is scarce in the data and standard econometric theory fails to capture the high degree of uncertainty. One special case is the weak instrument problem, which arises when the instrument induces only a small exogeneous variation in the endogenous regressor. Characterized by a low signal-to-noise ratio, weak identification widely appears in nonlinear economic models, ranging from demand analysis for differentiated products to dynamic stochastic general equilibrium (DSGE) models. Previous solutions either only apply to the linear model or severely sacrifice accuracy (power of the test) for nonlinear models. My research on weak identification provides reliable and efficient methods for a large class of nonlinear models. Andrews and Cheng (2012, Econometrica) developed an asymptotic approximation for the finite-sample distribution under arbitrary identification strength. Cheng (2015, Journal of Econometrics) allowed multiple parameters to have different identification strength and proposed a novel sequential procedure for asymptotic analysis. Andrews and Cheng (2014, Eonometric Theory; 2013 Journal of Econometrics) applied the weak identification robust inference to several models that are important for empirical research. Andrews, Cheng, Guggenberger (2011, CFDP) proposed generic results that can be used to quantify the degree of over-rejection by a standard test and examine the robustness of alternative tests.

High-Dimensional Problems Another main area of my work is model selection and model averaging in high-dimensional problems where the number of unknown parameters is large relative to the sample size, including GMM estimation with many moments and large-scale panel data with a latent structure. I develop asymptotic theories where the number of parameters increases with the sample size and study efficient computation algorithms.

One important high-dimensional context is the inference with large-scale panel data where the number of series N and the number of periods T are both large. Here unobserved factors and factor loadings capture the comovement of macroeconomic and financial time series or unobserved heterogeneity in microeconomic applications. Considering N and T going to infinitely simultaneously, Cheng, Liao, and Schofheide (2015, The Review of Economic Studies) developed a model selection procedure that determines the changes in the number of factors and the changes in factor loadings at an unknown point in time and applied it to study the Great Recession. Cheng and Hansen (2015, Journal of Econometrics) further investigated forecasting in a data-rich environment with model averaging methods.

Another important high-dimensional problem is the estimation of causal effect when a large number of instruments are available and the validity and relevance of these instruments are unknown. Cheng and Liao (2015, Journal of Econometrics) created an information-based LASSO estimator that selects all valid and relevant instruments simultaneously. This is the first procedure that achieves both in a high-dimensional setting.

Other Specification Issues in Dynamic Models Cheng and Phillips (2012, Journal of Econometrics; 2009, Econometrics Journal) pioneered a cointegrating rank estimation method based on information criteria and showed that the estimator is robust to non-Gaussian errors and even allows for temporal dependence and time-varying volatility. This method was applied to study the exchange rates among several major currencies.

In Cheng, Liao, Shi (2015, Working Paper), we proposed an averaging GMM estimator that is robust to model misspecification. This estimator combines a conservative GMM estimator based on valid moment conditions and an aggressive GMM estimator based on both valid and possibly misspecified moment conditions, where the weight is the sample analog of an infeasible optimal weight. We developed a uniform asymptotic approximation to the finite-sample risk and demonstrated the advantage of this averaging estimator over pre-test estimators. We estimated the human capital production function in a life-cycle labor supply model with this averaging approach.

In a structural autoregressive (SVAR) model, Cheng, Han, Montiel Olea (2015, Working Paper) proposed a new identification and estimation strategy by combining information from the variance of reduced-form errors, some external instruments, and a sign restriction. We also considered a joint determination of the VAR lag order and the choice of instrumental variables, and presented empirical studies on the relationship between oil supply and the U.S. economy.

References

[1] "Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities", with Zhipeng Liao and Frank Schorfheide. *The Review of Economic Studies*, 2015, forthcoming.

[2] "Robust Inference in Nonlinear Models with Mixed Identification Strength", *Journal of Econometrics*, 2015, 189(1), 207-228.

[3] "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments", with Zhipeng Liao, *Journal of Econometrics*, 2015, 186(2), 443-464.

[4] "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach", with Bruce E. Hansen, *Journal of Econometrics*, 2015, 186(2), 280-293.

[5] "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure", with Donald W. K. Andrews, *Econometric Theory*, 2014, 30(2), 287-333.

[6] "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure", with Donald W. K. Andrews, *Journal of Econometrics*, 2013, 173(1), 36-56.

[7] "Estimation and Inference with Weak, Semi-strong, and Strong Identification", with Donald W. K. Andrews, *Econometrica*, 2012, 80(5), 2153-2211.

[8] "Cointegrating Rank Selection in Models with Time-Varying Variance", with Peter C. B. Phillips, *Journal of Econometrics*, 2012, 169(2), 155-165.

[9] "Semiparametric Cointegrating Rank Selection", with Peter C. B. Phillips, *The Econometrics Journal*, 2009, 12(1), 83-104.

[10] "Uniform Asymptotic Risk of Averaging GMM Estimator Robust to Misspecification", with Zhipeng Liao and Ruoyao Shi, PIER Working Paper 15-017.

[11] "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests", with Donald W. K. Andrews and Patrik Guggenberger, Cowles Foundation Discussion Paper No. 1813.

[12] "Identification and Estimation of Structural VAR Models with External Instruments", with Xu Han and José Luis Montiel Olea, Working Paper.