Representative Publications

1. Zhang, J. and **Lin, W.** (2019). Scalable estimation and regularization for the logistic normal multinomial model. *Biometrics*, 75(4), 1098–1108.

   This paper develops a fast algorithm by combining the stochastic approximation EM and Hamiltonian Monte Carlo, along with a condition number regularization method, for scaling up the logistic normal multinomial model to high-dimensional count data.

2. Uematsu, Y., Fan, Y., Chen, K., Lv, J. and **Lin, W.** (2019). SOFAR: Large-scale association network learning. *IEEE Transactions on Information Theory*, 65(8), 4924–4939.

   This paper considers a multivariate regression model for large-scale association learning problems, and develops regularization procedures and optimization algorithms with theoretical guarantees for the case where the coefficient matrix has simultaneously sparse and orthogonal factors.

3. Cao, Y., **Lin, W.** and Li, H. (2019). Large covariance estimation for compositional data via composition-adjusted thresholding. *Journal of the American Statistical Association*, 114(526), 759–772.

   This paper proposes a principled approach to large covariance estimation for compositional data, gives for the first time approximate identifiability conditions for the basis covariance, and unveils an intriguing tradeoff between the curse and blessing of dimensionality.

4. Cao, Y., **Lin, W.** and Li, H. (2018). Two-sample tests of high-dimensional means for compositional data. *Biometrika*, 105(1), 115–132.

   This paper considers a two-sample testing problem for high-dimensional compositional data, formulates a testable hypothesis of compositional equivalence for the basis mean parameters, and develops a test based on the centred log-ratio transformation.

5. **Lin, W.**, Feng, R. and Li, H. (2015). Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *Journal of the American Statistical Association*, 110(509), 270–288.

   This paper proposes a two-stage regularization framework for high-dimensional instrumental variables regression, which extends the classical two-stage least squares method to the case where the numbers of both exposures and instruments can be large.

6. **Lin, W.**, Shi, P., Feng, R. and Li, H. (2014). Variable selection in regression with compositional covariates. *Biometrika*, 101(4), 785–797.

   This paper addresses the high-dimensional regression problem with compositional covariates, proposes a variable selection method by combining the linear log-contrast model and the lasso, and develops the associated theory and optimization algorithm.

7. **Lin, W.** and Lv, J. (2013). High-dimensional sparse additive hazards regression. *Journal of the American Statistical Association*, 108(501), 247–264.

   This paper addresses the problem of variable selection for survival data in a high-dimensional additive hazards model, and develops a wide class of regularization methods with strong theoretical guarantees and efficient implementations.