
Uncertainty Assessment and False Discovery Rate Control in High-Dimensional Granger Causal Inference

Aditya Chaudhry¹ Pan Xu² Quanquan Gu²

Abstract

Causal inference among high-dimensional time series data proves an important research problem in many fields. While in the classical regime one often establishes causality among time series via a concept known as “Granger causality,” existing approaches for Granger causal inference in high-dimensional data lack the means to characterize the uncertainty associated with Granger causality estimates (e.g., p-values and confidence intervals). We make two contributions in this work. First, we introduce a novel asymptotically unbiased Granger causality estimator with corresponding test statistics and confidence intervals to allow, for the first time, uncertainty characterization in high-dimensional Granger causal inference. Second, we introduce a novel method for false discovery rate control that achieves higher power in multiple testing than existing techniques and that can cope with dependent test statistics and dependent observations. We corroborate our theoretical results with experiments on both synthetic data and real-world climatological data.

1 Introduction

The advent of “big data” in recent years has generated countless opportunities for the prediction of real world phenomena with unprecedented accuracy and at unprecedented scale. Statistical methods for prediction exploit associations in existing data to predict some response variable. However, the task at hand is often not to predict the response variable from pre-existing data, but rather to determine how a change in one or more of the explanatory variables will *cause* changes in the response variable.

¹Department of Mathematics, University of Virginia, Charlottesville, VA 22904, USA ²Department of Computer Science, University of Virginia, Charlottesville, VA 22904, USA. Correspondence to: Quanquan Gu <qg5w@virginia.edu>.

In statistics, causality is often established by means of a controlled, randomized experiment. Nevertheless, controlled, randomized experiments are often infeasible, leaving researchers with only access to observational data. This situation arises routinely when working with time series data. Areas that must cope with this obstacle frequently include genetics (Shojaie & Michailidis, 2010) and neuroscience (Seth et al., 2015). The natural question that arises is: how can one determine which factors cause changes in a certain response variable using only data in which all variables change simultaneously? Causal inference seeks to address this problem.

The classic method for causal inference among time series is a concept from econometrics known as *Granger causality*, named after Nobel Prize winning econometrician Clive Granger (Granger, 1969). Granger causality formalizes the intuitive notion that in a causal system, the cause must precede the effect, and the cause must hold some unique information that helps predict the effect. For example, let X_1, \dots, X_T and Y_1, \dots, Y_T be two stationary one-dimensional time series. We can model time series Y using the following auto-regressive model (Stock & Watson, 2011):

$$Y_t = \sum_{i=1}^p a_i Y_{t-i} + \epsilon_t, \quad (1.1)$$

where a_1, \dots, a_p are the coefficient parameters for the regression, $p < T$ is the *maximal lag* of the model, and ϵ_t is the error term. To determine whether or not X is a *Granger cause* of Y , we also form a second auto-regressive model:

$$Y_t = \sum_{i=1}^p b_i Y_{t-i} + \sum_{i=1}^p c_i X_{t-i} + \omega_t, \quad (1.2)$$

with coefficient parameters $b_1, \dots, b_p, c_1, \dots, c_p$, and error term ω_t . In this classical regime, where the number of observations exceeds the number of variables ($T - p > 2p$), one can fit both of these models with ordinary least squares (OLS). We can conduct an F-test between models (1.1) and (1.2) as well as hypothesis tests on coefficients $c_i, 1 \leq i \leq p$, to determine if the extra information encompassed by previous values of X significantly aids in the prediction of

future values of Y . If this pair of models passes the F-test and at least one of the coefficient hypothesis tests at some significance level α , then we may reject the null hypothesis that X is not a Granger cause of Y (Granger, 1969).

Although the concept of Granger causality has existed for decades, Granger (1969) only rigorously treated the bivariate case. However, as noted by Arnold et al. (2007), Eichler (2006) provided one framework for multivariate analysis by applying graphical models to Granger causal inference.

Multivariate Granger causal inference relies on hypothesis testing of model coefficients in a fitted vector-autoregressive (VAR) model (Lutkepohl, 2007). VAR models are fit with OLS. In the high-dimensional regime, where the number of parameters exceeds the number of observations ($T - p < pd$, where d is the number of time series in the VAR model), OLS estimation is impossible. Hence, one must employ regularized regression methods. Perhaps the most well-known such method is the Lasso (Tibshirani, 1996), which encourages sparsity in the coefficient parameter vector via an ℓ_1 penalty. To conduct Granger causal inference in the high-dimensional regime, Arnold et al. (2007) proposed the ‘‘Lasso Granger’’ estimator, which we fully specify in (3.3). Unfortunately, since the limiting distribution of the underlying Lasso estimator is not normal (Knight & Fu, 2000) and intractable in general (Javanmard & Montanari, 2014), one cannot construct confidence intervals or compute test statistics for hypothesis tests of Lasso Granger coefficient point estimates. Thus, existing methods for high-dimensional Granger causal inference do not allow for the assessment of uncertainty. Uncertainty characterization proves an important, and often necessary, element of research in the natural sciences. Therefore, uncertainty assessment techniques would augment the versatility of high-dimensional Granger causal inference methods, and drive their wider adoption by the scientific community.

Another issue in high-dimensional causal inference is how to limit the number of false positives generated when testing a large number of explanatory variables without sacrificing identification of the true causal effects. That is, the researcher wants to attain high power while still maintaining a low type I error rate. To this end, false discovery rate (FDR) control (Benjamini & Hochberg, 1995) proves an important part of any method for high-dimensional causal inference. Unfortunately, existing FDR control methods cannot cope with the two challenges posed by our setting: dependent test statistics and dependent observations. These methods thus prove unsuitable in many practical applications.

In this paper, we make two contributions. First, we propose a novel asymptotically unbiased estimator for high-dimensional Granger causal inference inspired by Javanmard & Montanari (2014). We leverage this estimator’s unbiasedness to construct confidence intervals and p-values

for coefficient point estimates. In this way, we allow, for the first time, uncertainty characterization in high-dimensional Granger causal inference. Second, we propose a novel FDR control technique that can cope with dependent test-statistics and dependent observations. In addition to surmounting these theoretical obstacles to existing methods, our FDR control technique also achieves higher power in multiple testing than existing methods. Additionally, the proof techniques we use to extend high-dimensional results from the independent and identically distributed (i.i.d.) setting to our time series setting are of independent interest. Specifically, to establish the asymptotic unbiasedness and normality of our estimator, we appeal to Talagrand’s generic chaining (Talagrand, 2006) and martingale theory. We further employ martingale theory, along with empirical process theory, to prove the asymptotic validity of our FDR control procedure.

The rest of this paper is organized as follows. Section 2 contextualizes our contributions in the existing literature. Section 3 sets up the problem of high-dimensional Granger causal inference. Section 4 presents our novel de-biased estimator and FDR control procedure. Section 5 establishes our main theoretical results, which we corroborate empirically in Section 6. Section 7 concludes the paper.

2 Related Work

As mentioned above, Clive Granger examined bivariate Granger causality in 1969 (Granger, 1969). Advances in the area of graphical models provided a strong framework for multivariate causal inference in general (Pearl, 2009). Graphical models were first applied specifically to Granger causal inference by Eichler (2001) and Eichler (2006), and have provided a foundation for more complex models.

However, these methods rely on OLS estimation, which is impossible in the high-dimensional regime. Meinshausen & Bühlmann (2006) applied Lasso to the estimation of high-dimensional graphical models. Arnold et al. (2007) then applied the method proposed by Meinshausen & Bühlmann (2006) to multivariate Granger causal inference, and introduced the estimator of primary interest for this work: the Lasso Granger estimator. The Lasso Granger estimator yields a coefficient vector in which non-zero coefficients indicate conditional Granger causes of the response variable.

Classical methods for uncertainty analysis prove impossible for the Lasso Granger estimator. Recent work (Lee et al., 2013; Lockhart et al., 2014; Taylor et al., 2014) in the area of high-dimensional inference has made great strides toward addressing this issue. Early work focussed on constructing p-values and confidence intervals for Lasso coefficients via the bootstrap (Chatterjee et al., 2013; Liu et al., 2013a). However, later work found that these methods perform poorly compared to more recent methods (Dezeure et al., 2015), especially in non-i.i.d. settings (Karoui & Purdom, 2016).

Perhaps the most promising work in high-dimensional inference has emerged from the perspective of bias correction (Bühlmann et al., 2013; Zhang & Zhang, 2014). Subsequent work by Van de Geer et al. (2014) and Javanmard & Montanari (2014) introduced a method to de-bias the Lasso solution to yield asymptotically valid confidence intervals and hypothesis tests for coefficient point estimates. Nevertheless, these existing methods assume that the observations forming the design matrix are independent, and so cannot tackle causal inference among time series. Our method applies the Lasso de-biasing technique to the original Lasso Granger estimator. We overcome the inability of existing methods to cope with non-i.i.d. data by using Talagrand’s generic chaining (Talagrand, 2006) and the martingale technique to derive the asymptotic distribution of our novel de-biased Lasso Granger estimator.

Hypothesis testing in the high-dimensional setting raises the need for procedures to address the multiple testing problem. FDR control is one such way to control type I error in multiple testing. Our setting poses two challenges to existing FDR control procedures. First, the most-widely used methods, such as the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), assume the test statistics under consideration are independent. While Benjamini & Yekutieli (2001) proposed a slight variation on the Benjamini-Hochberg procedure that could control FDR under “positive regression dependency” (e.g., when the covariance matrix of the explanatory variables is strictly positive), in our setting where the explanatory variables interact in complex ways, the test statistics will not satisfy this property. This version of the Benjamini-Hochberg procedure achieves only low power in the presence of a general dependence structure (Romano et al., 2008), and is thus unsuitable for our setting. Recent methods from the area of graphical models, which explicitly model the dependency of explanatory variables, have made progress in addressing the case of dependent test statistics (Xie et al., 2011; Liu et al., 2013b). However, these methods still encounter the second challenge of our setting: dependent observations arising from time series data. To control FDR for dependent observations, one must resort to assumption-free methods, such as the Bonferroni technique, that achieve low power in practice. We propose a FDR control procedure that can cope with dependent test statistics and observations, and that achieves high power.

Notation We denote matrix $\mathbf{A} = [A_{i,j}] \in \mathbb{R}^{m \times n}$ and column vector $\mathbf{v} = [v_i] \in \mathbb{R}^T$. We write the ℓ_p norm of vector \mathbf{v} as $\|\mathbf{v}\|_p = \sum_{i=1}^T |v_i|^p$. Furthermore, $\|\mathbf{v}\|_\infty$ denotes the max-norm of vector \mathbf{v} : $\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq T} |v_i|$. Additionally, $\|\mathbf{v}\|_0 = \text{supp}(\mathbf{v})$ designates the cardinality of the support (the set of all non-zero entries) of \mathbf{v} . We represent the the max-norm of matrix \mathbf{A} as $\|\mathbf{A}\|_\infty = \max_{i,j} |A_{i,j}|$. The minimum and maximum singular values of \mathbf{A} are denoted by $\sigma_{\min}(\mathbf{A})$ and $\sigma_{\max}(\mathbf{A})$, respectively. $\Phi(x) \equiv$

$(1/\sqrt{2\pi}) \int_{-\infty}^x e^{-t^2/2} dt$ refers to the cumulative distribution function of the standard normal distribution. For a random variable X and a sequence of random variables X_n , we write $X_n \xrightarrow{P} X$ if X_n converges in probability to X , and $X_n \xrightarrow{D} X$ if X_n converges in distribution to X . For sequences of random variables X_n and Y_n , we say $X_n \asymp Y_n$ if X_n has the “same asymptotic order” as Y_n , that is, if both sequences bound each other up to some universal multiplicative constant.

3 Granger Causality and its Estimator

In this section, we set up the problem of high dimensional Granger causal inference. Denote the design matrix \mathbf{X} , the number of parameters d , the number of observations T , and the maximal lag p . For a given design matrix $\mathbf{X} = [X_{t,j}] \in \mathbb{R}^{T \times d}$ define the sample covariance matrix $\widehat{\Sigma} = \mathbf{X}^\top \mathbf{X} / T \in \mathbb{R}^{d \times d}$. The j -th column of \mathbf{X} represents time series \mathbf{X}_j , $1 \leq j \leq d$. We can further denote lagged versions of each column in the design matrix with $\widetilde{\mathbf{X}}_{t,j} = (X_{t-p,j}, X_{t-p+1,j}, \dots, X_{t-1,j})^\top \in \mathbb{R}^p$. Note that $X_{t-1,j}$ represents the observation immediately before $X_{t,j}$ in time series \mathbf{X}_j . In Granger causal analysis, the response variable is one of the explanatory variables. Hence, we can model an arbitrary variable $X_{t,j}$, with $1 \leq j \leq d$ and $p+1 \leq t \leq T$, by using the lagged values of all explanatory variables as predictors:

$$X_{t,j} = \sum_{i=1}^d \boldsymbol{\theta}_i^{j* \top} \widetilde{\mathbf{X}}_{t,i} + \epsilon_{t,j}. \quad (3.1)$$

Here $\boldsymbol{\theta}_i^{j*} \in \mathbb{R}^p$ and $\epsilon_{t,j} \sim N(0, \sigma_j^2)$. Time series \mathbf{X}_i is a conditional Granger cause of time series \mathbf{X}_j (conditioned on the other $d-2$ time series) if $\boldsymbol{\theta}_i^{j*}$ contains any non-zero elements (i.e., $\|\boldsymbol{\theta}_i^{j*}\|_0 > 0$). We can vectorize the sets of all $\boldsymbol{\theta}_i^{j*}$ and all $\widetilde{\mathbf{X}}_{t,i}$, for $1 \leq i \leq d$, as $\boldsymbol{\theta}^{j*} = (\boldsymbol{\theta}_1^{j* \top}, \boldsymbol{\theta}_2^{j* \top}, \dots, \boldsymbol{\theta}_d^{j* \top})^\top \in \mathbb{R}^{pd}$ and $\widetilde{\mathbf{X}}_t = (\widetilde{\mathbf{X}}_{t,1}^\top, \widetilde{\mathbf{X}}_{t,2}^\top, \dots, \widetilde{\mathbf{X}}_{t,d}^\top)^\top \in \mathbb{R}^{pd}$, respectively. Based on (3.1), the Lasso Granger estimator (Arnold et al., 2007) is given by

$$\widehat{\boldsymbol{\theta}}^j = \arg \min_{\boldsymbol{\theta}^j} \frac{1}{2(T-p)} \sum_{t=p+1}^T (X_{t,j} - \boldsymbol{\theta}^{j \top} \widetilde{\mathbf{X}}_t)^2 + \lambda \|\boldsymbol{\theta}^j\|_1,$$

where $\lambda > 0$ is the regularization parameter.

Equivalently, letting $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_{p+1}, \widetilde{\mathbf{X}}_{p+2}, \dots, \widetilde{\mathbf{X}}_T)^\top \in \mathbb{R}^{(T-p) \times pd}$ and $\mathbf{Y}_j = \mathbf{X}_{p+1:T,j}$ represent the lower $T-p$ elements of the j -th column of \mathbf{X} , we can re-express our model in more standard notation as:

$$\mathbf{Y}_j = \widetilde{\mathbf{X}} \boldsymbol{\theta}^{j*} + \boldsymbol{\epsilon}, \quad (3.2)$$

where $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_{(T-p) \times (T-p)})$. We can now re-express

the Lasso Granger estimator as:

$$\hat{\theta}^j = \arg \min_{\theta^j} \frac{1}{2(T-p)} \|\mathbf{Y}_j - \tilde{\mathbf{X}}\theta^j\|_2^2 + \lambda \|\theta^j\|_1. \quad (3.3)$$

For ease of presentation, we will henceforth omit the identifying variable j from θ^{j*} , $\hat{\theta}^j$, and \mathbf{Y}_j , and assume we are referring to some arbitrary response variable. Using the above notation, we can now denote the sample covariance matrix of $\tilde{\mathbf{X}}$ as $\tilde{\Sigma}_n = \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} / (T-p) \in \mathbb{R}^{pd \times pd}$ and the true covariance matrix as $\tilde{\Sigma} = \mathbb{E}[\tilde{\Sigma}_n]$.

4 Asymptotic Inference for Lasso Granger

In this section, we introduce our de-biased Lasso Granger estimator, and construct confidence intervals and p-values for its elements. We will then present our method for false discovery rate control in multiple testing.

4.1 Confidence Intervals and Hypothesis Tests

In deriving a de-biased version of the Lasso Granger estimator, we employ a variation of the Lasso de-biasing procedure proposed by [Javanmard & Montanari \(2014\)](#). In particular, we define the de-biased Lasso Granger estimator $\hat{\theta}^u$ as follows:

$$\hat{\theta}^u = \hat{\theta} + \frac{1}{T-p} \mathbf{M} \tilde{\mathbf{X}}^\top (\mathbf{Y} - \tilde{\mathbf{X}}\hat{\theta}), \quad (4.1)$$

where $\hat{\theta} \in \mathbb{R}^{pd}$ is the parameter vector yielded when computing the Lasso Granger estimator (3.3) for an arbitrary response variable $\mathbf{Y} = \mathbf{Y}_j$. $\mathbf{M} = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{pd})^\top \in \mathbb{R}^{pd \times pd}$ is an estimate of $\tilde{\Sigma}_n^{-1}$, the inverse sample covariance matrix of $\tilde{\mathbf{X}}$, where each \mathbf{m}_i is the solution to the following optimization algorithm:

$$\text{minimize } \mathbf{m}^\top \tilde{\Sigma}_n \mathbf{m} \quad \text{subject to } \|\tilde{\Sigma}_n \mathbf{m} - \mathbf{e}_i\|_\infty \leq \mu, \quad (4.2)$$

where $\mathbf{e}_i \in \mathbb{R}^{pd}$ is the i -th column of $\mathbf{I}_{pd \times pd}$, and the choice of μ will be clear after we deliver theory. Our unbiased estimator $\hat{\theta}^u$, though inspired by [Javanmard & Montanari \(2014\)](#), diverges sharply from their work in several respects. While Javanmard and Montanari use the observed design matrix \mathbf{X} in their estimator, we use the transformed design matrix $\tilde{\mathbf{X}}$. Although in our time series setting the rows of design matrix \mathbf{X} are already dependent, transforming \mathbf{X} to $\tilde{\mathbf{X}}$ exacerbates this dependency and renders the i.i.d. results underpinning Javanmard and Montanari's work unusable. Hence, we appeal to Talagrand's generic chaining ([Talagrand, 2006](#)) and martingale theory to establish our theoretical results about $\hat{\theta}^u$.

Theorem 5.5 in Section 5 below proves that for any $i \in \{1, 2, \dots, pd\}$, the standardized estimate of the i -th element

of $\hat{\theta}^u$ converges in distribution to the standard normal distribution:

$$\sqrt{T-p} \frac{\hat{\theta}_i^u - \theta_i^*}{\sigma [\mathbf{M} \tilde{\Sigma}_n \mathbf{M}^\top]_{i,i}^{1/2}} \xrightarrow{D} N(0, 1). \quad (4.3)$$

Unfortunately, the true noise level, denoted here by σ , is unknown in most real-world applications. Hence, we replace σ with a consistent estimator, denoted $\hat{\sigma}$, yielded by the Scaled Lasso ([Sun & Zhang, 2012](#)):

$$\{\hat{\theta}(\lambda), \hat{\sigma}(\lambda)\} = \left\{ \arg \min_{\theta \in \mathbb{R}^{pd}, \sigma > 0} \left\{ \frac{1}{2\sigma(T-p)} \|\mathbf{Y} - \tilde{\mathbf{X}}\theta\|_2^2 + \frac{\sigma}{2} + \lambda \|\theta\|_1 \right\}, \right. \\ \left. (4.4) \right.$$

where λ is the regularization parameter. [Sun & Zhang \(2012\)](#) prove $\hat{\sigma}$ is a consistent estimator of σ when the penalized loss function is convex. [Sun & Zhang \(2012\)](#) use the i.i.d. assumption to establish convexity. In our non-i.i.d. setting, we establish convexity via a restricted eigenvalue condition for martingale difference sequences. Thus, $\hat{\sigma}$ is consistent in our setting as well. Then by the Slutsky Theorem ([Van der Vaart, 2000](#)), we can replace σ in (4.3) with $\hat{\sigma}$.

One can easily apply (4.3) to construct confidence intervals for θ_i^* , for $1 \leq i \leq pd$. If the significance level is $\alpha > 0$, the $1 - \alpha$ confidence interval for θ_i^* is:

$$I_i = [\hat{\theta}_i^u - \delta(\alpha, T-p), \hat{\theta}_i^u + \delta(\alpha, T-p)], \quad (4.5)$$

where

$$\delta(\alpha, T-p) = \Phi^{-1}(1 - \alpha/2) (\hat{\sigma} / \sqrt{T-p}) [\mathbf{M} \tilde{\Sigma}_n \mathbf{M}^\top]_{i,i}^{1/2}.$$

We prove the asymptotic validity of this confidence interval in Corollary 5.6.

Similarly, we can also conduct hypothesis tests on the individual regression coefficients θ_i^* , for $1 \leq i \leq pd$. In the context of Granger causality, the relevant null and alternative hypotheses are $H_0^i : \theta_i^* = 0$ and $H_a^i : \theta_i^* \neq 0$, respectively. Having zero-coefficients for all variables $p(x-1) < i \leq px$ implies that time series $1 \leq x \leq d$ is *not* a conditional Granger cause of the response time series. Conversely, rejecting H_0^i for any variable $p(x-1) < i \leq px$ amounts to rejecting the null hypothesis that time series x is not a Granger cause of the response time series. We thus consider the following test statistic for $H_0^i : \theta_i^* = 0$:

$$\widehat{Z}_i = \frac{\hat{\theta}_i^u \sqrt{T-p}}{\hat{\sigma} [\mathbf{M} \tilde{\Sigma}_n \mathbf{M}^\top]_{i,i}^{1/2}}. \quad (4.6)$$

Note that under the null hypothesis $\widehat{Z}_i \xrightarrow{D} N(0, 1)$ by (4.3). The hypothesis test at significance level α is thus given by

$$\Psi_Z(\alpha) = \mathbf{1}(-|\widehat{Z}_i| < z_{\alpha/2}), \quad (4.7)$$

where $z_{\alpha/2}$ is the quantile of the standard normal distribution such that $\Phi(z_{\alpha/2}) = \alpha/2$. We reject the null hypothesis if and only if $\Psi_Z(\alpha) = 1$. The p-value for this hypothesis test is

$$P_i = 2(1 - \Phi(|\widehat{Z}_i|)). \quad (4.8)$$

As usual, one would reject H_0^i at a pre-specified significance level α if $P_i < \alpha$. We establish that the type I error of the hypothesis test $\Psi_Z(\alpha)$ converges to the specified significance level, and that the p-value P_i is asymptotically uniformly distributed in Corollary 5.7.

4.2 False Discovery Rate Control

Having established test statistics for individual coefficients of the de-biased Lasso Granger estimator, we now address the issue of FDR control. First, denote the set of coefficient indices i such that $\theta_i^* = 0$ as $\mathcal{H}_0 = \{i | \theta_i^* = 0, 1 \leq i \leq pd\}$. Define the complement of this set as $\mathcal{H}_1 = \{i | \theta_i^* \neq 0, 1 \leq i \leq pd\}$. We define FDR and false discovery proportion (FDP) as follows:

$$\text{FDP}(\nu) = \frac{\sum_{i \in \mathcal{H}_0} \mathbb{1}\{|\widehat{Z}_i| \geq \nu\}}{\max\{\sum_{1 \leq j \leq pd} \mathbb{1}\{|\widehat{Z}_j| \geq \nu\}, 1\}},$$

$$\text{FDR}(\nu) = \mathbb{E}[\text{FDP}(\nu)].$$

When conducting hypothesis tests at significance $0 < \alpha < 1$, we seek the smallest ν such that $\text{FDR}(\nu) \leq \alpha$. In this way, we will be able to reject the null hypothesis as often as possible (i.e., we maximize power) while still guaranteeing that our type I error rate does not exceed α . Thus, the ideal choice of ν is

$$\widehat{\nu} = \inf \left\{ \nu > 0 : \frac{\sum_{i \in \mathcal{H}_0} \mathbb{1}\{|\widehat{Z}_i| \geq \nu\}}{\max\{\sum_{1 \leq j \leq pd} \mathbb{1}\{|\widehat{Z}_j| \geq \nu\}, 1\}} \leq \alpha \right\}. \quad (4.9)$$

Note that the left hand side of the inequality in (4.9) is FDP, whose expectation is FDR. Unfortunately, $\widehat{\nu}$ cannot be computed under the unknown \mathcal{H}_0 (Liu et al., 2013b). However, following Liu & Luo (2014), we use the asymptotic normality of \widehat{Z}_i under the null hypothesis to approximate $\sum_{i \in \mathcal{H}_0} \mathbb{1}\{|\widehat{Z}_i| \geq \nu\}$ by $2(1 - \Phi(\nu))pd$. In multiple hypothesis testing, we use $-\widehat{\nu}$ as the threshold for rejecting the null hypothesis, instead of $z_{\alpha/2}$, in hypothesis test $\Psi_Z(\alpha)$ (4.7). Theorem 5.9 below demonstrates the asymptotic validity of this FDR control method.

5 Main Theory

In this section we present our main theoretical results: the test statistic \widehat{Z}_i from (4.6) converges in distribution to the standard normal under the null hypothesis, and the FDR control procedure presented in (4.9) asymptotically controls both FDR and FDP. To begin, we present several definitions.

Definition 5.1. (Vershynin, 2012) A random variable X is *sub-Gaussian* if there exists a constant $C > 0$ such that

$$\mathbb{P}(|X| > t) \leq 2 \exp[-t^2/C^2],$$

for all $t > 0$.

A random vector $\mathbf{X} \in \mathbb{R}^n$ is *sub-Gaussian* if the one-dimensional marginals $\langle \mathbf{X}, \mathbf{v} \rangle$ are sub-Gaussian random variables for all $\mathbf{v} \in \mathbb{R}^n$.

Definition 5.2. (Javanmard & Montanari, 2014) The *sub-Gaussian* norm of a random scalar variable X is:

$$\|X\|_{\psi_2} = \sup_{q \geq 1} q^{-1/2} (\mathbb{E}[|X|^q])^{1/q}.$$

The sub-Gaussian norm of a random vector $\mathbf{X} \in \mathbb{R}^n$ is:

$$\|\mathbf{X}\|_{\psi_2} = \sup_{\mathbf{u} \in S^{n-1}} \|\langle \mathbf{X}, \mathbf{u} \rangle\|_{\psi_2},$$

where S^{n-1} is the unit sphere in \mathbb{R}^n space.

Having established these definitions, we impose two assumptions on the design matrix and the true covariance matrix of the design matrix.

Assumption 5.3. There exist universal constants C_{\min}, C_{\max} such that $0 < C_{\min} \leq \sigma_{\min}(\widetilde{\Sigma}) \leq \sigma_{\max}(\widetilde{\Sigma}) \leq C_{\max}$.

Assumption 5.4. The rows of $\widetilde{\mathbf{X}}$ are sub-Gaussian and the sub-Gaussian norm of each row is bounded by some constant κ so that $\|\widetilde{\mathbf{X}}_i\|_{\psi_2} \leq \kappa$, for $i \in \{1, 2, \dots, T - p\}$.

We use Assumption 5.3 to demonstrate that the restricted eigenvalue condition holds for $\widetilde{\Sigma}_n$ in order to prove the asymptotic unbiasedness of $\widehat{\theta}^u$. Assumption 5.4 plays a role at multiple stages of the proof of Theorem 5.5, including proving the restricted eigenvalue condition for $\widetilde{\Sigma}_n$ and establishing a high-probability bound for the regularization parameter λ . Both of these assumptions prove common in the high-dimensional inference literature.

We leverage Assumptions 5.3 and 5.4 to present the following theorem.

Theorem 5.5. Suppose Assumptions 5.3 and 5.4 are satisfied. Let $s_0 = \text{supp}(\theta^*) \asymp \sqrt{T - p} / \log(pd)$ and $\mu \asymp \sqrt{\log(pd)/(T - p)}$. Then for any element θ_i^u of the de-biased Lasso Granger estimator $\widehat{\theta}^u$ defined in (4.1), we have

$$\sqrt{T - p} \frac{\widehat{\theta}_i^u - \theta_i^*}{\sigma[\mathbf{M}\widetilde{\Sigma}_n\mathbf{M}^\top]_{i,i}^{1/2}} \xrightarrow{D} N(0, 1).$$

Theorem 5.5 immediately yields several useful results. We first demonstrate the asymptotic validity of confidence interval (4.5) for any element of $\widehat{\theta}^u$ in the following corollary.

Corollary 5.6. Denote significance level $\alpha > 0$, and for $1 \leq i \leq pd$, define interval $I_i = [\hat{\theta}_i^u - \delta(\alpha, T - p), \hat{\theta}_i^u + \delta(\alpha, T - p)]$. Here, $\delta(\alpha, T - p) = \Phi(1 - \alpha/2)(\sigma/\sqrt{T - p})[\mathbf{M}\tilde{\Sigma}_n\mathbf{M}^\top]_{i,i}^{1/2}$. Then

$$\lim_{T-p \rightarrow \infty} \mathbb{P}(\theta_i^* \in I_i) = 1 - \alpha.$$

By Corollary 5.6, the asymptotic coverage probability corresponds to the given confidence level. Note that we can replace σ with $\hat{\sigma}$ by the Slutsky Theorem. Similarly, we confirm in the following corollary that the type I error for hypothesis test $\Psi_Z(\alpha)$, introduced in (4.7), matches the given significance level α . Furthermore, we prove that the CDF of the p-value P_i for $\Psi_Z(\alpha)$, which we introduced in (4.8), converges in distribution to a uniform distribution.

Corollary 5.7. With $\Psi_Z(\alpha)$ and P_i defined as above, and significance level $\alpha > 0$, we have:

$$\mathbb{P}(\Psi_Z(\alpha) = 1 | H_0^i) \xrightarrow{(T-p) \rightarrow \infty} \alpha \quad \text{and} \quad P_i \xrightarrow{D} U[0, 1].$$

We now turn our attention to demonstrating the asymptotic validity of the FDR control method we present in Section 4.2. To control FDR we desire the following property:

$$\frac{\sum_{i \in \mathcal{H}_0} \mathbf{1}(|\widehat{Z}_i| \geq \widehat{\nu})}{2|\mathcal{H}_0|(1 - \Phi(\widehat{\nu}))} \xrightarrow{P} 1. \quad (5.1)$$

Unfortunately, in this application, the test statistics \widehat{Z}_i are correlated, rendering the convergence in (5.1) non-trivial. In order to prove (5.1), we will leverage martingale theory, empirical process theory, and the following assumption.

Assumption 5.8. For constant $c > 2$,

$$\sum_{i \in \mathcal{H}_1} \mathbf{1} \left(\frac{|\theta_i^*|}{\sigma \widehat{\Sigma}_{i,i}^{-1/2}} \geq \sqrt{\frac{c \log(pd)}{T - p}} \right) \rightarrow \infty,$$

as $(T - p, pd) \rightarrow \infty$.

Assumption 5.8 implies that the number of true alternative hypotheses approaches infinity. This property proves important because, as demonstrated by Liu et al. (2014), FDR control is impossible when the number of true alternative hypotheses is fixed. This assumption allows us to present the following theorem:

Theorem 5.9. Assume $pd \leq (T - p)^r$ and $\log(pd) = o(\sqrt{T - p})$ for some $r > 0$. Furthermore, suppose that Assumption 5.8 and the assumptions of Theorem 5.5 hold. Then at significance level α ,

$$\lim_{(T-p, pd)} \frac{\text{FDR}(\widehat{\nu})}{\alpha|\mathcal{H}_0|/(pd)} = 1 \quad \text{and} \quad \frac{\text{FDP}(\widehat{\nu})}{\alpha|\mathcal{H}_0|/(pd)} \xrightarrow{P} 1,$$

as $(T - p, pd) \rightarrow \infty$.

Theorem 5.9 establishes that the FDR control procedure we present in Section 4.2 asymptotically controls both FDR and FDP. Note that the upper bound rate imposed on pd is very mild and will pose no issues in the vast majority of applications. The assumptions of Theorem 5.5 guarantee the asymptotic normality of test statistic \widehat{Z}_i .

6 Numerical Experiments

In this section, we establish the effectiveness of our de-biased Lasso Granger estimator and our FDR control procedure via experimental results. We also demonstrate that our methods outperform existing techniques.

6.1 Synthetic Data

Table 1: Empirical type I errors when testing $H_0 : \theta_i^* = 0$ over 500 simulations for $\alpha = .05$ for “random” and “cluster” transition matrix patterns.

T	d	p	Random	Cluster
150	200	1	.046	.052
150	200	2	.054	.054
250	300	1	.050	.040
250	300	2	.046	.046

In this section, we corroborate our theoretical results and compare our contributions to existing methods with numerical experiments on synthetic data. The data for these experiments are generated by model (3.1). In order to satisfy the assumptions of Theorem 5.5, each θ^{j*} is a sparse vector such that the probability of each element being non-zero is $\sqrt{T - p}/(2pd \log(pd))$ for $1 \leq j \leq d$. We use the R package “flare” (Li et al., 2012) to generate sparse transition matrices, and the “glmnet” package (Friedman et al., 2010) to compute the biased Lasso Granger estimate. We examine multiple different transition matrix patterns (“random” and “cluster”, as generated by the “flare” package) and multiple different configurations of (T, d, p) .

In Table 1, we see that the empirical type 1 error of hypothesis test $\Psi_Z(\alpha)$ (4.7) corresponds to the given significance level across multiple configurations of (T, d, p) . Figure 1(a) corroborates Theorem 5.5 by demonstrating that the empirical distribution of test statistic \widehat{Z}_i under the null hypothesis is the standard normal distribution. Figure 1(a) also illustrates that coefficient point estimates for the biased Lasso Granger estimator do not follow the standard normal distribution. Figure 1(b) validates Corollary 5.7 by demonstrating that the empirical CDF of p-value (4.8) for a true zero parameter is the uniform distribution. Furthermore, Figures 1(c) and 1(d) exhibit that hypothesis test $\Psi_Z(\alpha)$ (4.7) attains higher power than the biased Lasso Granger estimator when testing a single true non-zero parameter. Table 2 demonstrates the accuracy of the de-biased Lasso Granger estimator via computations of the ℓ_1 and ℓ_2 norms of the

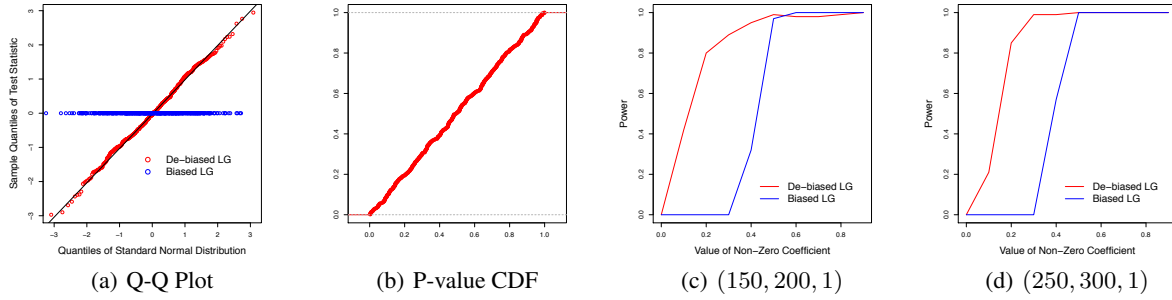


Figure 1: **1(a)** compares the Q-Q plots for test statistic \widehat{Z}_i and the biased Lasso Granger coefficient point estimate $\widehat{\theta}_i$. **1(b)** exhibits an empirical CDF of p-values for \widehat{Z}_i . The Q-Q plots and CDF were computed over 500 realizations of configuration $(T = 250, d = 300, p = 2)$ for a true zero parameter. **1(c)** and **1(d)** compare the power curves for testing $H_0^i : \theta_i^* = 0$ under the biased Lasso Granger (LG) estimator and under hypothesis test $\Psi_Z(.05)$ (4.7) for our de-biased Lasso Granger estimator across 100 simulations for two configurations of (T, d, p) . These simulations all use the “random” transition matrix pattern.

Table 2: Point estimation performance over 100 simulations for “random” and “cluster” transition matrix patterns. The standard deviations are displayed in parentheses.

T	d	p	Random		Cluster	
			L_2	L_1	L_2	L_1
150	200	1	0.35(0.0263)	3.89(0.293)	0.34(0.0230)	3.77(0.261)
250	300	1	0.48(0.0234)	6.67(0.325)	0.49(0.0282)	6.76(0.380)

error vector between $\widehat{\theta}^u$ and θ^* .

Table 3 exhibits that, as suggested by theory, our FDR control procedure outperforms the Bonferroni and Benjamini-Hochberg (B-H) (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001) methods in terms of power, while still maintaining low FDP. While the Bonferroni method generally achieves only low power, the Benjamini-Hochberg method performs poorly in this application because the test statistics exhibit complex dependency, and thus violate a theoretical assumption of the Benjamini-Hochberg method.

Lastly, Figure 2 demonstrates that our de-biased Lasso Granger estimator paired with our FDR control procedure outperform the original biased Lasso Granger estimator in terms of precision and recall. Define sets $TP = \{i \in \mathcal{H}_1 | \mathbb{1}(\theta_i^* \text{ identified as non-zero})\}$ and $FP = \{i \in \mathcal{H}_0 | \mathbb{1}(\theta_i^* \text{ identified as non-zero})\}$, so precision is $|TP| / \max\{|TP| + |FP|, 1\}$, and recall is $|TP| / |\mathcal{H}_1|$. Note that precision is equivalent to $1 - \text{FDP}$ and recall is equivalent to power. We calculate precision and recall at each point along the Lasso-path of the regularization parameter to generate the curves in Figure 2. These curves demonstrate that our de-biased Lasso Granger estimator and FDR control procedure achieve higher recall than the original Lasso Granger estimator without sacrificing precision. Thus, not only does our method provide the interpretability and flexibility of uncertainty characterization, it also achieves higher power

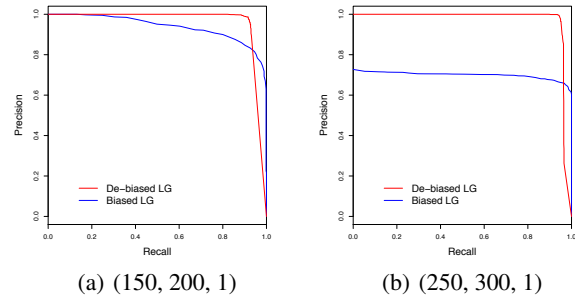


Figure 2: Precision-recall curves for our de-biased Lasso Granger (LG) estimator with our FDR control procedure at $\alpha = .05$ versus the original biased Lasso Granger estimator over 50 realizations of the noted configuration (T, d, p) .

than the original Lasso Granger estimator while maintaining low FDP. Therefore, our method proves more suitable for high-dimensional Granger causal inference.

6.2 Real Data

To demonstrate the applicability of our method to real-world data, we consider the climatological data set made available by Lozano et al. (2009). This data set contains monthly observations for seventeen climatological variables (e.g., temperature, precipitation, CO₂, CH₄, etc.) for 128 grid points spanning the continental United States (latitudes 32.975 to 45.475 and longitudes -84.75 to -117.25) from 1990 to

Table 3: Empirical FDP and power values for the indicated methods for $\alpha = .05$ over 50 simulations for the “random” and “cluster” transition matrix patterns.

T	d	p	Transition Matrix Pattern	FDR Control		B-H		Bonferroni	
				FDP	Power	FDP	Power	FDP	Power
150	200	1	Random	.002	.91	.005	.32	0	.31
250	300	1	Random	.002	.92	.003	.49	0	.43
150	200	1	Cluster	.007	.85	.008	.36	0	.33
250	300	1	Cluster	.002	.99	0	.34	0	.33

2002. Following the setup from Lozano et al. (2009), we enforce stationarity by deseasonalizing the data using the R package “deseasonalize” (McLeod & Gweon, 2013). We model the monthly temperature change of each grid point as a linear model of the first three lagged values of all explanatory variables in the surrounding 3×3 grid. Thus, for each of the 81 interior grid points, we obtain design matrices with dimensions $T = 13 \times 12 = 156$ (13 years of monthly data), $d = 17 \times 9 = 153$ (17 climatological variables observed at 9 grid points), and $p = 3$. For each of these design matrices, we use the R package “glmnet”(Friedman et al., 2010) to produce the biased Lasso Granger estimate from (3.3), and then apply (4.1) to construct the de-biased Lasso Granger estimate.

For each grid point, we test the significance of the three lagged values of monthly changes in Carbon Dioxide (CO2) emissions for that grid point to determine if local CO2 emissions are a Granger cause of temperature changes when conditioned on many other climatological variables. Recall that an explanatory variable is a Granger cause of the response variable if and only if any of the coefficients for any of the lags prove significant. We use the Bonferroni method, the Benjamini-Hochberg (B-H) procedure, and our FDR control method from Section 4.2 to control for multiple testing. At significance level $\alpha = .05$, the Bonferroni and Benjamini-Hochberg methods found that CO2 emissions are a Granger cause of monthly temperature changes for 10 of the 81 grid points, whereas our FDR control method discovered 13 such grid points. We thus corroborate the findings of Lozano et al. (2009), who employed graphical Granger modeling methods to establish Granger causality between CO2 emissions and temperature changes, and those of many climate researchers who have found increased CO2 emissions to “cause” higher temperatures. We also find empirical evidence that our FDR control method achieves higher power than the Bonferroni and Benjamini-Hochberg methods. Figure 3 displays the results of this simulation.

7 Conclusion

In this paper, we propose a novel unbiased estimator for conducting Granger causal inference in the high-dimensional

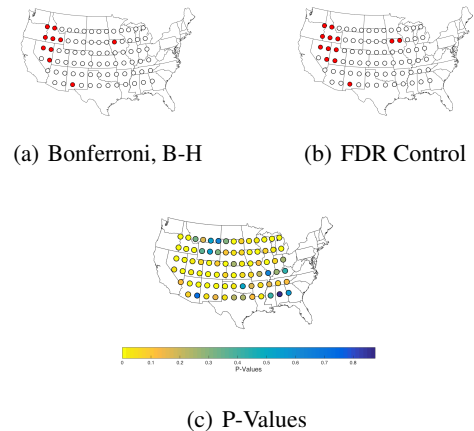


Figure 3: Maps of all interior grid points. Grid points where local CO2 emissions register as a Granger cause of temperature changes appear in red in 3(a) and 3(b). Test statistic critical values are chosen via the method indicated in captions of these two maps. 3(c) displays p-values for the coefficients corresponding to CO2 emissions in the regression for each grid point.

regime. We introduce test statistics and confidence intervals for our estimator, thereby accomplishing the previously impossible task of uncertainty characterization in high-dimensional Granger causal inference. Additionally, we introduce a novel method for false discovery rate control that achieves higher-power in multiple testing than existing alternatives in our setting. Lastly, we validate our theoretical results with experiments on both synthetic data and real-world climatological data. Future extensions of our work may include generalizations of our method to cope with non-Gaussian noise and non-linear causality.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments. This research was sponsored in part by the National Science Foundation IIS-1618948 and IIS-1652539. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

References

- Arnold, Andrew, Liu, Yan, and Abe, Naoki. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 66–75. ACM, 2007.
- Benjamini, Yoav and Hochberg, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.
- Benjamini, Yoav and Yekutieli, Daniel. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pp. 1165–1188, 2001.
- Bühlmann, Peter et al. Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242, 2013.
- Chandrasekaran, Venkat, Recht, Benjamin, Parrilo, Pablo A, and Willsky, Alan S. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.
- Chatterjee, A, Lahiri, SN, et al. Rates of convergence of the adaptive lasso estimators to the oracle distribution and higher order refinements by the bootstrap. *The Annals of Statistics*, 41(3):1232–1259, 2013.
- Dezeure, Ruben, Bühlmann, Peter, Meier, Lukas, Meinshausen, Nicolai, et al. High-dimensional inference: Confidence intervals, p -values and r -software hdi. *Statistical Science*, 30(4):533–558, 2015.
- Eichler, Michael. Graphical modelling of multivariate time series. 2001.
- Eichler, Michael. Graphical modeling of multivariate time series with latent variables. 2006.
- Friedman, Jerome, Hastie, Trevor, Simon, Noah, Tibshirani, Rob, Hastie, Maintainer Trevor, and Matrix, Depends. Package ‘glmnet’. *Journal of Statistical Software*, 33:1, 2010.
- Gordon, Yehoram. On milman’s inequality and random subspaces which escape through a mesh inn. In *Geometric Aspects of Functional Analysis*, pp. 84–106. Springer Berlin Heidelberg, 1988.
- Granger, Clive WJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.
- Hall, Philip and Heyde, Christopher Charles. *Martingale limit theory and its application*. Probability and mathematical statistics : a series of monographs and textbooks. Academic Press, Boston, San Diego, New York, 1980. ISBN 0-12-319350-8. URL <http://opac.inria.fr/record=b1089826>.
- Javanmard, Adel and Montanari, Andrea. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1): 2869–2909, 2014.
- Johnson, Nicholas, Sivakumar, Vidyashankar, and Banerjee, Arindam. Structured stochastic linear bandits. *arXiv preprint arXiv:1606.05693*, 2016.
- Karoui, Noureddine El and Purdom, Elizabeth. Can we trust the bootstrap in high-dimension? *arXiv preprint arXiv:1608.00696*, 2016.
- Knight, Keith and Fu, Wenjiang. Asymptotics for lasso-type estimators. *Annals of statistics*, pp. 1356–1378, 2000.
- Lee, Jason D, Sun, Dennis L, Sun, Yuekai, and Taylor, Jonathan E. Exact post-selection inference with the lasso. *arXiv preprint arXiv:1311.6238*, 2013.
- Li, Xingguo, Zhao, Tuo, Yuan, Xiaoming, and Liu, Han. An r package flare for high dimensional linear regression and precision matrix estimation. *R Package Vigette*, 2012.
- Liu, Hanzhong, Yu, Bin, et al. Asymptotic properties of lasso+ mls and lasso+ ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics*, 7:3124–3169, 2013a.
- Liu, Weidong and Luo, Shan. Hypothesis testing for high-dimensional regression models. 2014.
- Liu, Weidong, Shao, Qi-Man, et al. Phase transition and regularized bootstrap in large-scale t -tests with false discovery rate control. *The Annals of Statistics*, 42(5):2003–2025, 2014.
- Liu, Weidong et al. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978, 2013b.
- Lockhart, Richard, Taylor, Jonathan, Tibshirani, Ryan J, and Tibshirani, Robert. A significance test for the lasso. *Annals of statistics*, 42(2):413, 2014.
- Lozano, Aurelie C, Li, Hongfei, Niculescu-Mizil, Alexandru, Liu, Yan, Perlich, Claudia, Hosking, Jonathan, and Abe, Naoki. Spatial-temporal causal modeling for climate change attribution. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 587–596. ACM, 2009.
- Lutkepohl, Helmut. *New Introduction to Multiple Time Series Analysis*. Springer, 2007. ISBN 3540262393.

- McLeod, A. I. and Gweon, Hyukjun. Optimal deseasonalization for monthly and daily geophysical time series. *Journal of Environmental Statistics*, 4(11), 2013. URL <http://www.jenvstat.org/v04/i11>.
- Meinshausen, Nicolai and Bühlmann, Peter. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pp. 1436–1462, 2006.
- Mendelson, Shahar, Pajor, Alain, and Tomczak-Jaegermann, Nicole. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17(4):1248–1282, 2007.
- Negahban, Sahand N., Ravikumar, Pradeep, Wainwright, Martin J., and Yu, Bin. A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 11 2012. doi: 10.1214/12-STS400.
- Pearl, Judea. *Causality*. Cambridge university press, 2009.
- Romano, Joseph P, Shaikh, Azeem M, and Wolf, Michael. Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test*, 17(3):417–442, 2008.
- Seth, Anil K, Barrett, Adam B, and Barnett, Lionel. Granger causality analysis in neuroscience and neuroimaging. *The Journal of Neuroscience*, 35(8):3293–3297, 2015.
- Shojaie, Ali and Michailidis, George. Discovering graphical granger causality using the truncating lasso penalty. *Bioinformatics*, 26(18):i517–i523, 2010.
- Stock, James and Watson, Mark. *Introduction to Econometrics (3rd edition)*. Addison Wesley Longman, 2011. Professor Stock receives royalties for this text.
- Sun, Tingni and Zhang, Cun-Hui. Scaled sparse linear regression. *Biometrika*, pp. ass043, 2012.
- Talagrand, Michel. *The generic chaining: upper and lower bounds of stochastic processes*. Springer Science & Business Media, 2006.
- Talagrand, Michel. *Upper and lower bounds for stochastic processes: modern methods and classical problems*, volume 60. Springer Science & Business Media, 2014.
- Taylor, Jonathan, Lockhart, Richard, Tibshirani, Ryan J, and Tibshirani, Robert. Post-selection adaptive inference for least angle regression and the lasso. *arXiv preprint*, 2014.
- Tibshirani, Robert. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- Van de Geer, Sara, Bühlmann, Peter, Ritov, Ya’acov, Dezeure, Ruben, et al. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.
- Van der Vaart, Aad W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In Eldar, Y.C. and Kutyniok, G. (eds.), *Compressed Sensing: Theory and Applications*, pp. 210–268. Cambridge University Press, 2012.
- Xie, Jichun, Cai, T Tony, Maris, John, and Li, Hongzhe. Optimal false discovery rate control for dependent data. *Statistics and its interface*, 4(4):417, 2011.
- Zaitsev, A Yu. On the gaussian approximation of convolutions under multidimensional analogues of sn bernstein’s inequality conditions. *Probability theory and related fields*, 74(4):535–566, 1987.
- Zhang, Cun-Hui and Zhang, Stephanie S. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B*, 76(1):217–242, 2014.