Proceedings of Machine Learning Research vol 120:1–10, 2020

# Sparse and Low-bias Estimation of High Dimensional Vector Autoregressive Models

**Trevor Ruiz** TREVOR.RUIZ@OREGONSTATE.EDU
**Sharmodeep Bhattacharyya** SHARMODEEP.BHATTACHARYYA@OREGONSTATE.EDU
*Department of Statistics, Oregon State University, Corvallis, OR*

**Mahesh Balasubramanian** MBALASU2@ASU.EDU
*Compiler Microarchitecture Lab, Arizona State University, Tempe, AZ*

**Kristofer Bouchard** KEBOUCHARD@LBL.GOV
*Biological Systems and Engineering Division, Lawrence Berkeley National Labs, Berkeley, CA*

## Abstract

Vector autoregressive (VAR) models are widely used for causal discovery and forecasting in multivariate time series analysis. In the high-dimensional setting, which is increasingly common in fields such as neuroscience and econometrics, model parameters are inferred by $L_1$-regularized maximum likelihood (RML). A well-known feature of RML inference is that in general the technique produces a trade-off between sparsity and bias that depends on the choice of the regularization hyperparameter. In the context of multivariate time series analysis, sparse estimates are favorable for causal discovery and low-bias estimates are favorable for forecasting. However, owing to a paucity of research on hyperparameter selection methods, practitioners must rely on *ad-hoc* methods such as cross-validation (or manual tuning). The particular balance that such approaches achieve between the two goals — causal discovery and forecasting — is poorly understood. Our paper investigates this behavior and proposes a method (UoI-VAR) that achieves a better balance between sparsity and bias when the underlying causal influences are in fact sparse. We demonstrate through simulation that RML with a hyperparameter selected by cross-validation tends to overfit, producing relatively dense estimates. We further demonstrate that UoI-VAR much more effectively approximates the correct sparsity pattern with only a minor compromise in model fit, particularly so for larger data dimensions, and that the estimates produced by UoI-VAR exhibit less bias. We conclude that our method achieves improved performance especially well-suited to applications involving simultaneous causal discovery and forecasting in high-dimensional settings.

**Keywords:** multivariate time series, vector autoregressive models, high dimensional data, sparsity, LASSO, Granger causality, graphical models, bootstrap, Union of Intersections

## 1. Introduction

Multivariate time series data are now ubiquitous across scientific fields and increasingly high-dimensional. In neuroscience, for instance, intra-cortical electrophysiology produces simultaneous recordings of neural activity as measured by large arrays of hundreds to thousands of electrodes at sampling rates exceeding 24kHz (Buzsáki et al. (2012)). Data with analogous structure is generated in neuroscience from electroencephalography (EEG) (Astolfi et al. (2007)), and various other sources (Brown et al. (2004); Pillow et al. (2008); Bassett et al. (2011)). In econometrics and fincance, multivariate time series data is used for forecasting, macroeconomic studies, and structural analysis (Sims (1980); Fackler and Krieger (1986); Forni et al. (2005); Stock and Watson (2002); Tsay (2005)).

Similar data are arising on increasing scales in environmental science and geosciences (Karpatne et al. (2018); Triacca et al. (2013)), epidemiology (Cunniffe et al. (2015)), and sociology (McFarland et al. (2016)). Such growing data dimensionalities and length of recordings present opportunities for scientific discovery alongside methodological and computational challenges for data analysis.

Vector autoregressive (VAR) models provide a flexible framework for forecasting, structural analysis (finding a unique process parametrization under constraints on the error term), impulse response analysis (describing the propagation of a 'shock' or erratic event throughout the system), and estimation of various types of causality (Lütkepohl (2005)). Furthermore, VAR model parameters are conceptually straightforward to estimate, although computationally scaling to large systems remains a challenge.

In particular, large datasets require high-dimensional process models. In this context, parameters are estimated with sparsity-inducing constraints, which has motivated research on sparse estimation of high-dimensional vector autoregressive model parameters (Song and Bickel (2011); Fan et al. (2011); Han et al. (2015); Qiu et al. (2016); Basu and Michailidis (2015); Hall et al. (2016)). Interesting sparsity constraints also arise in related literature on joint estimation of multiple Gaussian graphical models (Guo et al. (2011); Danaher et al. (2014)).

Sparse estimation methods for time series models typically rely on $L_1$-regularized maximum likelihood estimation. However, it is known that this technique can result in overfitting — specifically overly-dense estimates — and excessive bias (Bühlmann and van de Geer (2011); Meinshausen and Bühlmann (2010)) in high-dimensional regression and precision matrix estimation, and it is likely that these problems persist in the time series context. Nonetheless, few alternatives to RML are available to date in multivariate time series analysis for high-dimensional data.

This paper offers a two-fold contribution to the above-cited work: (i) we propose an inference procedure (UoI-VAR) that improves on $L_1$-regularized maximum likelihood estimation (RML) of high-dimensional vector autoregressive models by leveraging the Union of Intersections (UoI) statistical machine learning framework (Bouchard et al. (2017)); and (ii) we provide simulation- and application-based support for the algorithm. Together, these results indicate that UoI-VAR will enhance inference in VAR models across application domains.

## 2. Background

This section begins with a brief introduction to estimation of VAR(D) processes (we defer to Lütkepohl (2005) for a thorough review of the subject and closely follow the notation therein) and continues with a description of the Union of Intersections framework, summarizing key ideas elaborated in more detail in Bouchard et al. (2017).

### 2.1. Vector autoregressive models

Formally, the stochastic process $\{X_t : \Omega \to \mathbb{R}^M\}_{t\in\mathbb{Z}}$ is a vector autoregressive process of order $D$ (VAR(D)) if for all $t \in \mathbb{Z}$

$$X_t = \nu + \sum_{d=1}^{D} A_d X_{t-d} + \epsilon_t \ , \quad \begin{cases} \mathbb{E}\epsilon_t & = 0 \\ \mathbb{E}\epsilon_t\epsilon_t' & = \Sigma \\ \mathbb{E}\epsilon_t\epsilon_s' & = 0, \quad \forall t \neq s \end{cases} \tag{1}$$

where $\Sigma \in \mathbb{R}^{M\times M}$ is positive definite and $A_1, \ldots, A_D$ satisfy $\det\left(I - \sum_{d=1}^{D} A_d z^d\right) \neq 0$ for all $|z| \leq 1$. The latter condition ensures that the process is well-defined, stationary and stable.

Given an observed time series of length $T$, denoted $\{x_t \in \mathbb{R}^M\}_{t=0}^{T}$, a VAR(D) model for the data can be expressed in the form of a multivariate multiple regression $Y = UB + E$ where the response is denoted by $Y = (x_T\ x_{T-1}\ \cdots\ x_D)^T$ and comprises the observations beginning from time $D$, the error terms are denoted by $E = (\epsilon_T\ \epsilon_{T-1}\ \cdots\ \epsilon_D)^T$, and the linear predictor $\boldsymbol{UB}$ on the right-hand side denotes

$$\left(\begin{array}{c|c|c|c} 1 & x_{T-1}^T & \cdots & x_{T-D}^T \\ 1 & x_{T-2}^T & \cdots & x_{T-D-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{D-1}^T & \cdots & x_0^T \end{array}\right) \left(\begin{array}{c} \nu^T \\ A_1^T \\ \vdots \\ A_D^T \end{array}\right)$$

The classical estimation technique is to estimate $B$ with ordinary least squares (OLS) by $\hat{B} = (U'U)^{-1}U'Y$, and then estimate $\Sigma$ by $\hat{\Sigma} = \frac{1}{T-1}(Y - U\hat{B})(Y - U\hat{B})'$. The equivalence of this procedure with maximum likelihood estimation assuming $\epsilon_t$ are Gaussian is well-established.

When $M$ is large and $A_1, \ldots, A_D$ are sparse, $B$ is instead estimated using $L_1$-regularized maximum likelihood (RML) by

$$\hat{B} = \arg\min_{B} \{-\ell(B; X) + \lambda P(B)\} \tag{2}$$

where $\ell(B; X)$ denotes the log-likelihood of $B$ given data $X = (x_1 \cdots x_T) \in \mathbb{R}^{M\times T}$, and $P(B) = \|\text{vec}B\|_1$ is the sparsity-inducing regularization term applied to the vectorized parameter matrix $B$.[1]

### 2.2. Union of Intersections algorithmic framework

The Union of Intersections (UoI) is a statistical machine learning framework that separates sparse parameter selection (an 'intersection step') from parameter estimation (a 'union' step). The advantages of this approach are established for several sparse learning techniques, including regression, classification, and matrix decomposition (Bouchard et al. (2017); Ubaru et al. (2017)). The intersection step first infers several candidate parameter support sets (sets of nonzero parameter locations). This is accomplished through inferring support sets for a range of sparsity-inducing regularization strengths on bootstrap samples, and applying intersection operations across the bootstrap samples. In the union step, estimates are calculated without regularization for each candidate support on bootstrap samples, and predictive quality is evaluated on separate samples. Finally, the estimates that optimize predictive quality are averaged (a union operation with respect to the selected candidate supports) to produce the final output.

## 3. Methods

Our method extends the UoI algorithmic framework to estimate sparse VAR(D) models. The resulting estimation algorithm separates the estimation procedure into an intersection step utilizing RML for sparse parameter selection and a union step utilizing OLS for parameter estimation. The intersection

1. Due to the equivalence between maximum likelihood and ordinary least squares, $\|Y - UB\|_F^2$ can be substituted for $-\ell(B; X)$ in Equation (2) and LASSO regression on the vectorized problem (obtained by column-stacking the response $Y$) can be used to find the solution with fast, numerically stable, and widely available algorithms (Friedman et al. (2007, 2010)).

step estimates a collection of candidate support sets for the transition matrices $A_1, \ldots, A_D$ for a VAR(D) model, and the union step produces the final estimates $\hat{A}_1, \ldots, \hat{A}_D$ by combining the supports that optimize forecasting quality. These steps are given in detail separately as Algorithms 1 and 2.

**Algorithm 1** Intersection step

**Input:**
data $\{x_t \in \mathbb{R}^M\}_{t=0}^T$
regularization path $\lambda \in \mathbb{R}^K$
number of bootstrap samples $B_1$
thresholding parameter $s$
**for** $b = 1$ to $B_1$ **do**
draw bootstrap sample $\{x_t^*\}_{t=0}^T$
**for** $k = 1$ to $K$ **do**
$\tilde{B}_k \leftarrow \arg\min_B \{-\ell(B\ ;\ x_0^*, \ldots x_T^*) + \lambda_k P(B)\}$
$S_{b,k} \leftarrow \{(i,j) :\ \tilde{b}_{ij} \neq 0\}$
**end for**
**end for**
$S_k \leftarrow \left\{(i,j) :\ \sum_{b=1}^{B_1} \mathbb{1}\{(i,j) \in S_{b,k}\} \geq sB_1\right\}$
**Output:** Support sets $S_1, \ldots, S_k$

In the intersection step (Algorithm 1), supports of RML estimates are computed for a fixed regularization path $\lambda$ on $B_1$ bootstrap samples. These support sets are aggregated across bootstrap samples separately for each $\lambda_k$ by a thresholded intersection operation. In other words, this procedure selects consistently recurring RML support sets under resampling of the data $B_1$ times, with a specifiable recurrence threshold $s$. The outputs of the procedure form the inputs to the union step (Algorithm 2).

The union step (Algorithm 2) begins by repeating the following procedure $B_2$ times. Training and test bootstrap samples are drawn, and OLS estimates are computed for each candidate support set from the intersection step on the training bootstrap sample. The estimates that minimize a user-defined loss function $f$ on the test set are stored. Each iteration of the procedure is repeated with new boostrap samples, producing $B_2$ parameter estimates. Finally, an average over the $100 \times q\%$ most sparse estimates produces the output. This portion of the algorithm averages OLS estimates selected by bootstrapped cross-validation under resampling of the data $B_2$ times.

Given the intensive use of resampling methods, bootstrap procedures suitable for time series are required to implement this method. One of the most common bootstrap procedures for time series is the moving block bootstrap (Kunsch et al. (1989); Liu and Singh (1992)), in which the original time series is divided into overlapping blocks and the blocks are randomly sampled with replacement to construct a resampling of the original time series. The choice of appropriate block lengths is dependent on the statistical problem (Bühlmann and Künsch (1999); Kreiss and Lahiri (2012)). We chose to draw bootstrap samples of the same length as the original time series, though for large datasets it may be desirable to draw smaller bootstrap samples for computational speed in estimation.

The hyperparameters $B_1$, $B_2$, $s$, and $q$ control the sparsity of the estimates at each stage of the algorithm. Increasing $B_1$ produces greater sparsity among candidate supports, decreasing $s$ creates less sparsity, and doing both simultaneously helps to stabilize the candidate supports against erratic

**Algorithm 2** Union step

**Input:**
data $\{x_t \in \mathbb{R}^M\}_{t=0}^T$
candidate support sets $S_1, \ldots, S_K$ (from intersection step)
number of bootstrap samples $B_2$
loss function $f$
threshold parameter $q$
**for** $b = 1$ to $B_2$ **do**
draw bootstrap samples $\{x_t^{*(1)}\}_{t=0}^T$, $\{x_t^{*(2)}\}_{t=0}^T$
**for** $k = 1$ to $K$ **do**
$\mathcal{B} \leftarrow \left\{B \in \mathbb{R}^{(D+1)\times M} : \mathbb{1}\{b_{ij} \neq 0\} = \mathbb{1}\{(i,j) \in S_k\}\right\}$
$\tilde{B}_k \leftarrow \arg\min_{B\in\mathcal{B}} \left\{-\ell(B\ ;\ x_0^{*(1)}, \ldots, x_T^{*(1)})\right\}$
$f_k \leftarrow f(\tilde{B}_k; x_0^{*(2)}, \ldots, x_T^{*(2)})$
**end for**
$\hat{B}^{(b)} \leftarrow \tilde{B}_J : J = \arg\min_k\{f_k\}$
**end for**
$\hat{B} \leftarrow \frac{1}{|Q|}\sum_{b\in Q} \hat{B}^{(b)}$, where $Q = \{b : \sum_{j=1}^{B_2} \mathbb{1}\{q_j \leq q_b\} \leq qB_2\}$ and $q_b = \sum_{i,j} \mathbb{1}\{\hat{b}_{ij}^{(b)} \neq 0\}$
**Output:** estimate $\hat{B}$

bootstrap samples. Similarly, increasing $B_2$ generally increases density and decreasing $q$ counteracts this effect, and doing both stabilizes the output of the union step. The ability to tune the algorithm using $s, q$ can be especially helpful when computational resources are limited.

## 4. Results

We assessed the performance of UoI-VAR on synthetic and real data relative to RML. This section presents the results of our simulation study followed by the results of our data application.

### 4.1. Simulation study

We conducted a simulation study utilizing synthetic datasets conforming to each combination of process dimensions $M = 5, 10, 20, 40, 80$ and time series lengths $T = 50, 100, 200$. We chose to study performance in the context of estimating VAR(1) processes, since any VAR(D) process can be expressed as a (higher-dimensional) VAR(1) process.[2] The study examined selection accuracy, model fit, and estimation bias for each method.

For each simulation setting (combination of $M, T$), $VAR(1)$ process parameters were generated as follows: $M$ nonzero transition matrix parameters were drawn from a frequency distribution increasing exponenially away from zero in either direction; $\nu = 0$; and $\Sigma = 0.5I$. Thus, the parameters exhibit $1 - 1/M\%$ sparsity. Nonzero parameter positions were randomly allocated to transition matrix positions. Then 50 realizations of each process were simulated, and model estimation was conducted using RML and UoI-VAR. Both methods utilized the same regularization

2. For details, see Lütkepohl (2005). Estimating a VAR(D) model for M-dimensional data can be accomplished by estimating a VAR(1) model for DM-dimensional data after reshaping the original data appropriately. Therefore, estimation of high-order processes is a matter of computational scale.

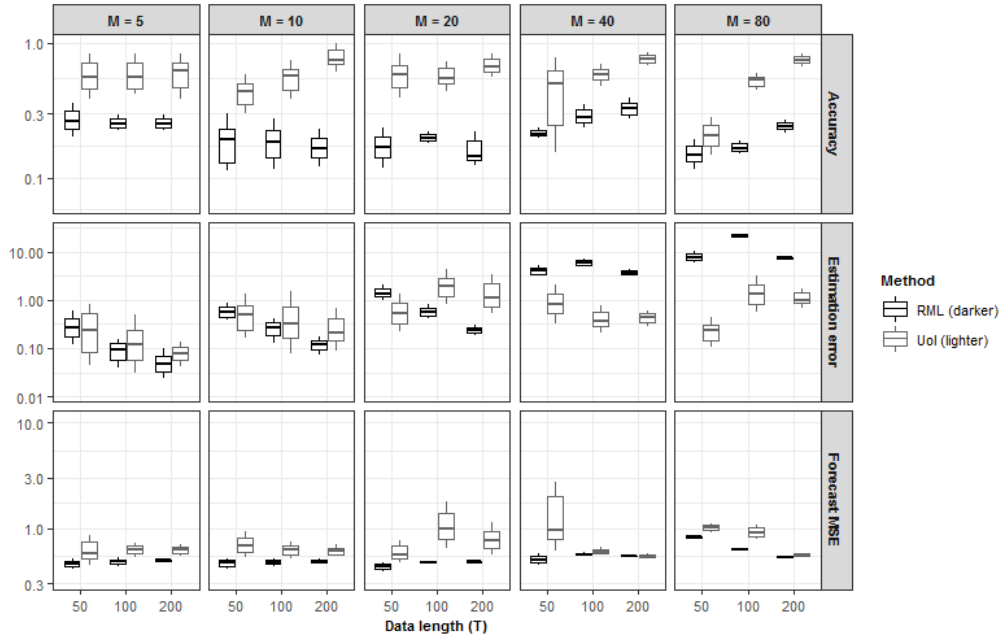


Figure 1: Selection, estimation, and forecasting behavior (rows) observed in the simulation study for different dimensions (columns). Top row, selection accuracy $1 - \frac{FN+FP}{M+FP}$ in terms of false positives (FP) and false negatives (FN); middle row, estimation error computed on the estimated support $\sum_{i,j \in \hat{S}} \|a_{ij} - \hat{a}_{ij}\|_2^2$; bottom row, one-step forecast error $\frac{1}{T-1} \sum_t \|X_{t+1} - \hat{X}_{t+1}\|_2^2$. Each panel displays the corresponding metric against time series length $T$ on the horizontal axis for $UoI$ and RML.

path $\lambda$. For RML, the regularization strength that minimized the average of one-step and four-step forecasting errors over five-fold cross-validation was used for the final estimate. We used the same average of forecast errors for the loss function $f$ in our algorithm, along with hyperparameters $B_1 = 10$, $B_2 = 50$, $s = 1$, and $q = 0.3$. These values of hyperparameters were chosen manually and tend to perform well across a wide range of settings.

Figure 1 summarizes the results of the simulation study. Selection accuracy is a scaled combination of false positives $FP$, the number of positions in the parameter estimates that are in fact zero but estimated as nonzero, and false negatives $FN$, the number of positions that are in fact nonzero but estimated as zero. Across all settings, RML exhibits low accuracy and UoI-VAR exhibits improved accuracy; these behaviors are driven predominantly by false positive rates (not depicted in the figure). The main limitation of UoI-VAR is an increased false negative rate relative to RML when less data are available (shorter $T$ settings). However, this problem diminishes rapidly as $T$ increases, and as a result, for larger $M$, the selection performance of our method improves much faster than RML as $T$ increases. Furthermore, as depicted in the second row of the figure, UoI-VAR achieves dramatically lower estimation errors in large-$M$ settings. Finally, it appears that these improvements come at the cost of a slight decrease in fit; the third row shows forecasting errors on the data to which the models were fit.

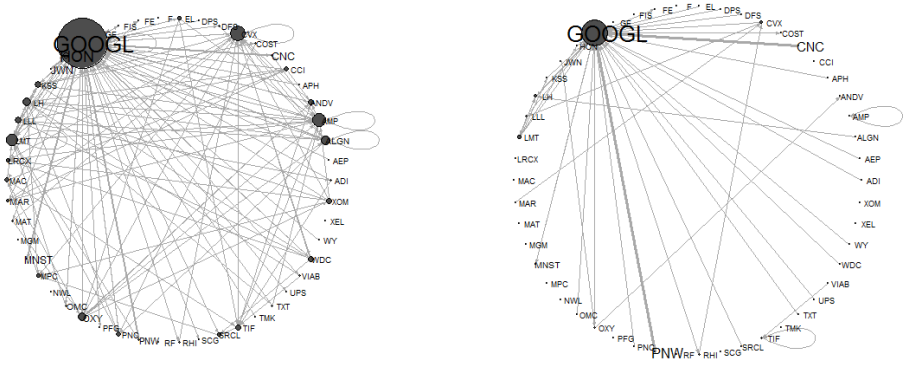


Figure 2: Network visualizations of VAR(1) model parameter estimates for S&P 500 data using RML (left) and our algorithm (right). Node and label size are proportional to degree centrality.

### 4.2. Data analysis

To illustrate an application and further compare each method in the context of a data analysis, a VAR(1) model was used to identify putative causal connections between weekly closes of 50 randomly chosen publicly traded companies listed on the S&P 500 index in 2013-2014. This dataset was chosen due to the absence of benchmark datasets with known ground truth for large multivariate time series; the years 2013-2014 saw a steadily climbing index with no major disturbances. To obtain an approximately stationary process, first-order differences were calculated from the raw series; then, VAR(1) model parameters were estimated from these differences using our method and RML.

The resulting estimates are visualized in Figure 2 as directed graphs comprising nodes representing each vector component and edges indicating the set of nonzero parameters[3]. UoI-VAR identified just 44 edges that describe the dependence of Google's share price on other companies. By comparison, RML identifies 146 edges in which the same pattern UoI-VAR detected is present but obscured by other edges. One-step forecast RMSE averaged over all companies on the same data for the RML method is 8.3993; for our method, 8.4525 (an increase of $0.6\%$ relative to RML). However, the scale of share prices varies widely among the companies; the median per-company forecast RMSE is 4.7807 using RML, and 4.3329 using UoI-VAR (a decrease of $9\%$ relative to RML). We conclude that our method finds more interpretable estimates (sparser graph) while maintaining comparable fit.

## 5. Discussion

This paper proposes a novel method (UoI-VAR) for low-bias and sparse estimation of VAR models, presents simulations that show its advantages, and exemplifies its application in data analysis. The method is flexible, and the hyperparameters allow the analyst to control tolerances for false positives and false negatives without explicitly specifying any *a priori* assumptions about sparsity structure. Promising extensions of this work currently in progress include: (i) development of analogous

3. That is, $G = (V, E)$ where $V = \{1, \ldots, M\}$ and $E = \{(i, j) \in V \times V : A_{ji} \neq 0\}$

methods for estimation of other stochastic processes such as multivariate point processes; (ii) a full theoretical analysis of the algorithm; and (iii) application of the method to scientific datasets.

## Acknowledgments

The authors thank several anonymous reviewers whose comments improved this work.

## References

Laura Astolfi, Febo Cincotti, Donatella Mattia, M Grazia Marciani, Luiz A Baccala, Fabrizio de Vico Fallani, Serenella Salinari, Mauro Ursino, Melissa Zavaglia, Lei Ding, et al. Comparison of different cortical connectivity estimators for high-resolution eeg recordings. *Human brain mapping*, 28(2):143–157, 2007.

Danielle S Bassett, Nicholas F Wymbs, Mason A Porter, Peter J Mucha, Jean M Carlson, and Scott T Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641–7646, 2011.

Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.

Kristofer Bouchard, Alejandro Bujan, Farbod Roosta-Khorasani, Shashanka Ubaru, Mr Prabhat, Antoine Snijders, Jian-Hua Mao, Edward Chang, Michael W Mahoney, and Sharmodeep Bhattacharya. Union of intersections (uoi) for interpretable data driven discovery and prediction. In *Advances in Neural Information Processing Systems*, pages 1078–1086, 2017.

Emery N Brown, Robert E Kass, and Partha P Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature neuroscience*, 7(5):456–461, 2004.

Peter Bühlmann and Hans R Künsch. Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310, 1999.

Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data*. Springer, 1 edition, 2011.

György Buzsáki, Costas Anastassiou, and Christof Koch. The origin of extracellular fields and currents – eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13:407–420, June 2012.

Nik J Cunniffe, Britt Koskella, C Jessica E Metcalf, Stephen Parnell, Tim R Gottwald, and Christopher A Gilligan. Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10, 2015.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397, 2014.

James S Fackler and Sandra C Krieger. An application of vector time series techniques to macroeconomic forecasting. *Journal of Business & Economic Statistics*, 4(1):71–80, 1986.

Jianqing Fan, Jinchi Lv, and Lei Qi. Sparse high-dimensional models in economics. *Annu. Rev. Econ.*, 3(1):291–317, 2011.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100 (471):830–840, 2005.

Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007. doi: 10.1214/07-AOAS131.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

Jian Guo, Elizaveta Levina, George Michailidis, and Ji Zhu. Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15, 2011.

Eric C Hall, Garvesh Raskutti, and Rebecca Willett. Inferring high-dimensional poisson autoregressive models. In *Statistical Signal Processing Workshop (SSP), 2016 IEEE*, pages 1–5. IEEE, 2016.

Fang Han, Huanran Lu, and Han Liu. A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16:3115–3150, 2015.

Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

Jens-Peter Kreiss and Soumendra Nath Lahiri. *Bootstrap methods for time series*, volume 30 of *Time Series Analysis: Methods and Applications*, chapter 1, pages 3–26. Elsevier, 2012. doi: 10.1016/B978-0-444-53858-1.00001-6.

Hans R Kunsch et al. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989.

Regina Y Liu and Kesar Singh. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225:248, 1992.

Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, 1 edition, 2005.

Daniel A McFarland, Kevin Lewis, and Amir Goldberg. Sociology in the era of big data: The ascent of forensic social science. *The American Sociologist*, 47(1):12–35, 2016.

Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society*, 72(4):417–473, 2010.

Jonathan Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan Litke, E. Chichilnisky, and Eero Simoncelli. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature*, 454(7207):995–999, August 2008.

Huitong Qiu, Fang Han, Han Liu, and Brian Caffo. Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):487–504, 2016.

Christopher A Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.

Song Song and Peter J Bickel. Large vector auto regressions. *arXiv preprint arXiv:1106.3915*, 2011.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002.

Umberto Triacca, Alessandro Attanasio, and Antonello Pasini. Anthropogenic global warming hypothesis: testing its robustness by granger causality analysis. *Environmetrics*, 24(4):260–268, 2013.

Ruey S Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.

Shashanka Ubaru, Kesheng Wu, and Kristofer E Bouchard. Uoi-nmfcluster: A robust nonnegative matrix factorization algorithm for improved parts-based decomposition and reconstruction of noisy data. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*, pages 241–248. IEEE, 2017.