Penalization method for sparse time series model

Heewon, Park
Chuo University, Department of Mathemetics
1-13-27, Kasuga, Bunkyo-ku
Tokyo 112-8851, Japan
E-mail: heewonn@gug.math.chuo-u.ac.jp

Fumitake, Sakaori
Chuo University, Department of Mathemetics
1-13-27, Kasuga, Bunkyo-ku
Tokyo 112-8851, Japan
E-mail: sakaori@math.chuo-u.ac.jp

1. Introduction

Vector autoregressive (VAR) model has been widely used in many field, e.g., genetics, science, economics and finance. Especially, in genetics, VAR model with high dimensional setting is studied extensively. In order to find a suitable VAR model and improve the forecasting accuracy, capability of true model selection is important matter in several areas.

In recent years, numerous studies on penalization method have been going on for true model selection of VAR model. Haufe et al. (2008) proposed several sparse approaches (e.g. ridge regression and multiple test, granger causality test, lasso, group lasso). Ren and Zhang (2010) proposed the subset selection methods for VAR model using a adaptive lasso in order to overcome lasso's problem by different amount of penalty to each coefficient. Shimamura et al. (2009) proposed a recursive regularization for VAR model for $n \leq p$ case using an elastic net which has a combination penalty term the ridge and the lasso.

However, existing penalization methods can not reflect properties of time series model with lagged variables. In order to improve the forecasting accuracy of univariate time series model, Park and Sakaori (2011) proposed a lag weighted lasso which reflects deceasing variable effect as the lag increase. The lag weighted lasso assigns different penalties depend on not only coefficient size but also lag effect. The superiority of the lag weighted lasso is identified in forecasting accuracy and true model selection for univariate time series model.

In this study, we extend the lag weighted lasso from univariate to multivariate time series model. We also propose a lag weighted penalization methods for VAR model namely a lag weighted ridge regression and a lag weighted elastic net. And we show that superiority of the lag weighted penalization methods in forecasting accuracy and true model selection for VAR model.

2. VAR models

We consider a stationary VAR(L) model :

$$(1) \qquad \boldsymbol{Y}_t = \nu + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}\beta^l_{j'j}Y_{jt-l} + e_t,$$

where ν is a $p \times 1$ vector of intercept term, $\boldsymbol{\beta}^l$'s are $p \times p$ coefficient matrices, $\boldsymbol{B} = (\boldsymbol{\beta}^1, ..., \boldsymbol{\beta}^L)$, $t = 1, ..., n$, $\boldsymbol{Y}_t$ is a $p \times 1$ vector, $\boldsymbol{Z}_t = (\boldsymbol{Y}_1^T, .., \boldsymbol{Y}_{t-L}^T)^T$, and e_t is a $p \times 1$ white noise with nonsingular covariance matrix Σ.

In this study, we consider the centralized VAR(L) model (1) with same symbols of $\boldsymbol{Y}_t$s and $\boldsymbol{Z}_t$s denoting their centralized values. The matrix version of (1) without intercept is

(2) $$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Z} + e,$$

with $\boldsymbol{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_n)$, $\boldsymbol{Z} = (\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)$, and $e = (e_1, ..., e_n)$. The vector version (2) is

(3) $$\mathbf{y} = \mathbf{z}^T\boldsymbol{\beta} + \mathbf{e},$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\boldsymbol{\beta} = vec(\mathbf{B})$, $\mathbf{z} = \mathbf{Z} \bigotimes I_p$, $\mathbf{e} = vec(e)$, and $\mathbf{e} \sim N(0, \sigma^2 I)$.

3. Lag weighted penalization method for VAR model

In general time series model, consideration of lag effects is essential because effect of variable depends on length of lag period. In other words, variable effect declines as lag increases even variable having strong effect. For this reason, many studies on reflecting the lag effect have been developed, and reveal that consideration of lag effect raise the forecasting accuracy. In line with this thinking, Park and Sakaori (2011) proposed the lag weighted lasso for univariate time series model. In this study, we expand the lag weighted lasso in VAR model and we propose the lag weighted penalization method for VAR model having same idea of the lag weighted lasso.

3.1 Penalization method for VAR model

In the practice of statistical modeling, variable selection is a important matter, and it is often desirable to have accurate forecasting model with a sparse representation since recent data sets are usually high dimensional with large number of predictors[2]. For these reasons, several methods for variable selection and estimation by penalization are proposed as following:

(4) $$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\{\|\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\|^2\} + \lambda \mathrm{pen}(\boldsymbol{\beta})\},$$

where $\mathrm{pen}(\boldsymbol{\beta})$ is a penalty function. The penalization methods superior to forecasting accuracy and true model selection by imposing the penalty on object function of OLS.

Recently, the adaptive lasso which proposed for overcome lasso's problem is in the spotlight. The penalty of the adaptive lasso is defined by

(5) $$\mathrm{pen}(\beta) = \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{\hat{j'}jl}|\beta^l_{j'j}|,$$

where $w_{\hat{j'}jl} = 1/|\beta^{\hat{l}}_{j'j}|^\gamma$, $\gamma > 0$ and the least square estimators or the ridge estimators can be used as $\beta^{\hat{l}}_{j'j}$. The adaptive lasso imposes different amount of penalty to each coefficient by weights unlike the lasso. By penalty with weight, the adaptive lasso enjoy oracle properties for linear model $n \geq p$.

In case of $p \geq n$, the lasso and the adaptive lasso select at most n variables. Also the lasso and the adaptive lasso can not to reflect grouped effect, it means that if model has group of variables among which correlations are very high, then the lasso tends to select only one variable from the group and does not care which one is selected[8]. To overcome the lasso's problem, Zou and Hastie (2005) proposed a elastic net having the penalties which is combination the ridge and the lasso. However the elastic net can't enjoy the oracle properties. To achieve oracle properties, Ghosh (2007) proposed a adaptive elastic net which assigns different weights to different coefficient on lasso's penalty term. The penalty of the adaptive elastic net is defined by

(6) $$\mathrm{pen}(\beta) = \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{\hat{j'}jl}|\beta^l_{j'j}| + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta^l_{j'j})^2,$$

where $w_{\hat{j'}jl} = 1/|\beta^{\hat{l}}_{j'j}|^{\gamma}$.

3.2 Lag weight for VAR model

In this section we introduce lag weighted penalization methods to estimate sparse VAR model. In general time series model, response variable is explained by lagged variable, and variable effect declines as the lag increases even variable having strong effect. Shrestha (2008) assumed that response variable is explained by cumulative and extended lag effects of explanatory variables. Also they claimed that $t-$th weighted explanatory variable is expressed as follows under the assumption:

$$X_t = \sum_{l=0}^{L} w_l X_{t-l}, \tag{7}$$

where $w_l = \alpha(1-\alpha)^l$, $0 < \alpha < 1$, that is, w_l is a geometrically decreasing weight up to $L-$th lag. From idea of $w_l = \alpha(1-\alpha)^l$, Park and Sakaori (2011) proposed the lag weighted lasso reflecting lag effect for univariate time series model. In this study, we expend the lag weighted estimation in VAR model:

$$\boldsymbol{\beta}^{lwtla} = \arg\min_{\boldsymbol{\beta}}\{\|\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\|^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{\hat{j'}jl}|\beta^{l}_{j'j}|\}, \tag{8}$$

where weight is composed with two part, the part of coefficient size and the part of lag effect. The lag weighted lasso has 3-type of adaptive weight as follows:

- reflect only lag effects:

$$w^{(1)}_{j'jl} = \frac{1}{[\alpha(1-\alpha)^l]^{\gamma}}, \tag{9}$$

- reflect coefficient and lag effects and γ affects to only coefficient effects:

$$w^{(2)}_{j'jl} = \frac{1}{(|\hat{\beta}_{j'jl}|)^{\gamma}}\frac{1}{\alpha(1-\alpha)^l}, \tag{10}$$

- reflect coefficient and lag effects, γ affects to coefficient and lag effects:

$$w^{(3)}_{j'jl} = \frac{1}{[|\hat{\beta}_{j'jl}|\alpha(1-\alpha)^l]^{\gamma}}, \tag{11}$$

From these weights, the lag weighted lasso can reflect properties of time series model that variable effect declines as the lag increases. In other words, estimators of variable having small $\alpha(1-\alpha)^l$ and $\hat{\boldsymbol{\beta}}$(variable in distant past and has small effect) is estimated in small or this variable is delete in model. The estimators of the lag weighted lasso are shown in Table 1.

The concept of the lag weighted lasso can be also applied to other lag weighted penalization methods. The estimators of the lag weighted elastic net are shown in Table2.

Practically, the choice of the tuning parameters is important. In this study, we use the data in the last couple of periods as the test data to choose the optimal set of tuning parameters $(\alpha, \gamma, \lambda)$ minimizing the prediction error RPE(relative prediction error) as follows:

$$\text{RPE} = E[(\hat{\mathbf{y}} - \mathbf{z}^T\boldsymbol{\beta})^2]/\sigma^2. \tag{12}$$

4. Simulation study

Table1. Estimates of the lag weighted lasso

weight	lag weighted lasso estimates
$w^{(1)}$	$\hat{\boldsymbol{\beta}}^{wtla(1)} = \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'jl}^{\hat{(1)}}\vert\beta_{j'j}^{l}\vert\}$
	$= \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{[\alpha(1-\alpha)^l]^\gamma}$
$w^{(2)}$	$\hat{\boldsymbol{\beta}}^{wtla(2)} = \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'jl}^{\hat{(2)}}\vert\beta_{j'j}^{l}\vert\}$
	$= \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{(\vert\hat{\beta}_{j'jl}\vert)^\gamma[\alpha(1-\alpha)^l]}$
$w^{(3)}$	$\hat{\boldsymbol{\beta}}^{wtla(3)} = \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'jl}^{\hat{(3)}}\vert\beta_{j'j}^{l}\vert\}$
	$= \arg\min_{\boldsymbol{\beta}}\{\Vert\mathbf{y} - \mathbf{z}^T\boldsymbol{\beta}\Vert^2 + \sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{(\vert\hat{\beta}_{j'jl}\vert[\alpha(1-\alpha)^l])^\gamma}$

Table2. Estimates of the lag weighted Elastic net

weight	lag weighted elastic net estimates
$w^{(1)}$	$\hat{\boldsymbol{\beta}}^{wtela(1)} = \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'j}^{\hat{1}}\vert\beta_{j'j}^{l}\vert + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$
	$= \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{[\alpha(1-\alpha)^l]^\gamma} + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$
$w^{(2)}$	$\hat{\boldsymbol{\beta}}^{wtela(2)} = \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'j}^{\hat{2}}\vert\beta_{j'j}^{l}\vert + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$
	$= \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{(\vert\hat{\beta}_{j'jl}\vert)^\gamma[\alpha(1-\alpha)^l]} + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$
$w^{(3)}$	$\hat{\boldsymbol{\beta}}^{wtela(3)} = \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} w_{j'j}^{\hat{3}}\vert\beta_{j'j}^{l}\vert + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$
	$= \arg\min_{\beta}\{\Vert y - z^T\beta\Vert^2 + \lambda_1\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p} \frac{\vert\beta_{j'j}^{l}\vert}{(\vert\hat{\beta}_{j'jl}\vert[\alpha(1-\alpha)^l])^\gamma} + \lambda_2\sum_{l=1}^{L}\sum_{j'=1}^{p}\sum_{j=1}^{p}(\beta_{j'j}^{l})^2$

In this section we present simulation study to compare the performance of lag weighted penalization methods. We consider following model for data generation.

$$\text{Model}: \; Y_t = \beta^1 Y_{t-1} + \beta^2 Y_{t-2} + \beta^3 Y_{t-3} + e_t,\; e_t \sim N(\mathbf{0}, \textstyle\sum_e), \text{ where } \textstyle\sum_e = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We assumed two situations for generalization:

- Situation 1: The effect of variable decreases as the lag increases.

$$\beta^1 = \begin{bmatrix} -0.5 & 0.5 & 0.5 & 0.6 \\ 0.4 & -0.6 & 0.3 & 0.4 \\ 0.3 & 0.3 & -0.5 & 0.5 \\ 0.5 & 0.4 & -0.5 & 0.7 \end{bmatrix}, \beta^2 = \begin{bmatrix} 0.4 & 0.2 & -0.1 & 0.2 \\ 0.2 & 0.3 & 0.1 & -0.2 \\ 0.1 & 0.1 & 0.3 & -0.2 \\ 0.3 & 0.1 & 0.4 & 0.5 \end{bmatrix}, \beta^3 = \begin{bmatrix} 0.2 & 0.1 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.1 \end{bmatrix}.$$

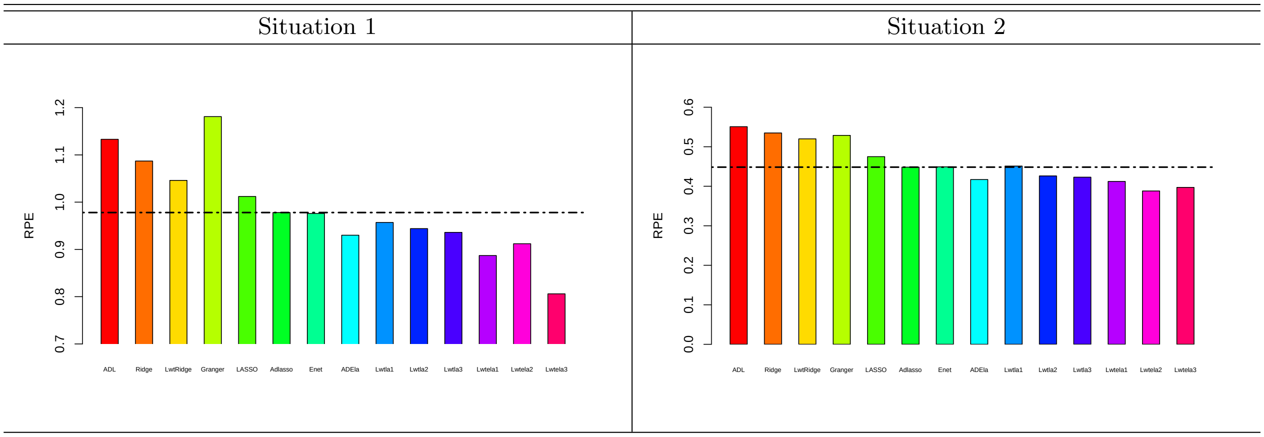
- Situation 2: The effect of variable has no relation to lag period.

$$\beta^1 = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & -0.2 \\ 0 & 0 & 0.3 & 0 \\ 0.1 & 0 & 0 & 0.1 \end{bmatrix}, \beta^2 = \begin{bmatrix} 0.4 & 0 & 0.3 & 0 \\ 0 & 0.3 & 0 & 0.4 \\ 0 & 0 & -0.5 & 0 \\ 0.2 & 0 & 0 & 0.5 \end{bmatrix}, \beta^3 = \begin{bmatrix} -0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0 & 0.7 \end{bmatrix}.,$$

We used the OLS estimator $\hat{\beta}$ in the adaptive weight on all penalization methods. We have used a LARS algorithm for the lasso type penalization methods and a LARS-EN algorithm for the elastic net type penalization methods.

We generate 100 observation (n=100) consisting of the training data of the first 90 observations and test data of the last 10 observations in two situations. And we choose the optimal tuning parameter sets $(\alpha, \gamma, \lambda)$ minimizing RPE. Finally we compute mean RPE in each methods on 100 replications. Figure 1 shows the RPE in each methods.

Figure1. RPE of each methods.



In Figure 1, horizontal line means RPE of the adaptive lasso. We may find that the lag weighted penalization methods outperform than existing penalization methods. Especially the lag weighted elastic net with $w^{(3)}$ and $w^{(2)}$ show the most outstanding performance in situation 1 and in situation 2 respectively. Following Shimamura et al.(2009), we also compare the TP(True Positive), FP(False Positive), TN(True Negative), FN(False Negative) and TDR(True discovery rate : TP/(TP+FP)) to evaluate the capability of correctly selecting subset. Table 3 shows the capability of correctly selecting subset in each methods. We also identify the superiority of the lag weighted penalization methods, especially the lag weighted lasso, for capability of correctly selecting subset in situation 1. In situation 2, however, the lag weighted penalization methods show not good performance. This result is comprehensible from common sense.

Table3(a). Capability of correctly selecting subset : Situation 1.

	Granger	Lasso	Adatpive lasso	Elastic net	Adaptive elastic net	lwtlasso			lwtela		
						$w^{(1)}$	$w^{(2)}$	$w^{(3)}$	$w^{(1)}$	$w^{(2)}$	$w^{(3)}$
TP	**0.946**	0.867	0.740	0.884	0.816	0.776	0.727	0.711	0.775	0.250	0.696
FP	0.919	0.777	0.352	0.810	0.852	0.363	**0.265**	0.256	0.790	0.583	0.583
TN	0.081	0.223	0.648	0.190	0.148	0.638	0.735	**0.744**	0.210	0.417	0.417
FN	**0.054**	0.133	0.260	0.116	0.184	0.224	0.273	0.289	0.225	0.750	0.304
TDR	0.507	0.527	0.677	0.522	0.489	0.682	0.733	**0.735**	0.495	0.300	0.544

5. Conclusion and remark

In this study, we extend the lag weighted lasso in VAR model and we proposed the lag weighted penalization methods. The lag weighted penalization methods impose the difference penalty on difference variable and length of lag. Our simulation studies show that the lag weighted penalization methods outperform than existing penalization methods in situation that variable effect decreases as the lag increases.

Table3(b). Capability of correctly selecting subset : Situation 2.

	Granger	Lasso	Adatpive lasso	Elastic net	Adaptive elastic net	lwtlasso			lwtela		
						$w^{(1)}$	$w^{(2)}$	$w^{(3)}$	$w^{(1)}$	$w^{(2)}$	$w^{(3)}$
TP	0.625	0.719	0.580	**0.753**	0.708	0.733	0.594	0.594	0.566	0.635	0.490
FP	**0.339**	0.623	0.363	0.604	0.601	0.542	0.347	0.351	0.432	0.554	0.344
TN	**0.661**	0.377	0.637	0.396	0.399	0.458	0.653	0.649	0.568	0.446	0.656
FN	0.375	0.281	0.420	**0.247**	0.292	0.267	0.406	0.406	0.434	0.365	0.510
TDR	**0.649**	0.536	0.615	0.555	0.541	0.575	0.631	0.629	0.567	0.534	0.588

Improvement of forecasting accuracy and capability of correctly selecting subset on VAR model is a good news for several areas. Especially, the lag weighted penalization method will bring advance in genetic area using VAR model with high dimensional data.

However proposed the lag weighted penalization methods are not suitable when effect of variable has no relation to lag. And we show the superiority of the lag weighted penalization methods by only simulation studies. Therefore future study will be required to theoretical approach(e.g. oracle properties) and consider the generalized weights which can apply in any situation.

REFERENCES

[1] S. Ghosh. Adaptive elastic net: An improvement of elastic net to achieve oracle properties.

[2] D. Hong and F. Zhang. Weighted elastic net model for mass spectrometry imaging processing. *Mathematical Modelling of Natural Phenomena*, 5:115–33, 2010.

[3] H. Park and F. Sakaori. Lag weigted lasso. *submitted*, 58:267–288, 2001.

[4] Y. Ren and X. Zhang. Subset selection for vector autoregressive provesses via adaptive lasso. *Statistis and Probability letters*, 80:1705–1712, 2010.

[5] T. Shimamura, S. Imoto, R. Yamaguchi, A. Fujita, M. Nagasaki, and S. Miyano. Recursive regularization for inferring gene networks from time course gene expression profiles. *BioMed central*, 2009.

[6] S. L. Shrestha. Time series modeling of respiratory hospital admissions and geometrically weighted distributed lag effects from ambient particulate air pollution within kathmandu valley, nepal. *Environ Model Assess*, 12:239–251, 2007.

[7] N. K. Stefan Haufem Klaus-Robert Miller, Guido Nolte. Sparse causal discovery in multivatiate time series. *Worlshop and Conference Preceedings*, 2008.

[8] H. Zuo and T. Hastie. Regularization and variabel selection via the elastic net. *Journal of the Royal Statistical Society*, 67:301–320, 2005.