Edge Deletion Tests in Graphical Models for Multivariate Time Series

Marco Reale* Christopher J. Price* Shaodan Lin† Rory Ellis*

Abstract

Multivariate autoregressive moving average models can be represented as graphical models with nodes representing either the autoregressive or the moving average components and the edges representing a significant partial correlation among them. Absent edges imply a parsimonious structure of the time series models. Different strategies for testing the presence of edges are presented, assessed and compared.

Key Words: Sparsity, graphical lasso, parsimonious models

1. Multivariate time series models

The dynamics of multiple dependent time series can be represented, under appropriate conditions, through multivariate time series models of the autoregressive moving average type. In a quite general form they can be described by the equation

$$\Phi_0 X_t - \sum_{i=1}^{p} \Phi_i X_{t-i} = A_t - \sum_{j=1}^{q} \Theta_j A_{t-j} \tag{1}$$

where the behavior of m time series at a current time t is explained by the other contemporaneous time series in X_t, by the lagged time series X_{t-i} up to a lag p and by the current and lagged values innovations A_{t-j} up to lag q. Here X_{t-i} and A_{t-j}, $i = 0, \ldots, p$, $j = 0, \ldots, q$ are m-dimensional random vectors while Φ_i and Θ_i are $m \times m$ fixed coefficient matrices. Model (1) is a structural autoregressive moving average of order p and q (SVARMA(p,q)) and dimension m. More details about this model can be found in multivariate time series books such as [10] and [13].

In this article we discuss methods that can be applied to model (1), however we consider, just out of illustration, one commonly used subclass: the structural vector autoregression of order p, (SVAR(p)) defined by the equation

$$\Phi_0 X_t = \sum_{i=1}^{p} \Phi_i X_{t-i} + A_t \tag{2}$$

In model (2) we express the current variables at time t (X_t) as a linear combination of the other contemporaneous variables and all the lagged variables up to a lag p ($X_{t-1}, \ldots, X_{t-p}$). It is typically assumed that Φ_0 can be rearranged into a triangular matrix in order to avoid simultaneity.

Even assuming the triangular structure of the contemporaneous coefficient matrix, when all the other coefficients are non-zero (i.e. a saturated model), the resulting model is far too complex most of the times. We therefore compare methods to reduce the complexity by obtaining sparse models where a significant number of parameters assume a zero value.

*School of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

†Statistics New Zealand, 120 Hereford St, Christchurch 8011, New Zealand

2. Graphical modeling context

In particular we consider the problem of parsimonious structural vector autoregressive structures in the context of graphical modeling. This is a very convenient perspective as in a SVAR a zero coefficient corresponds to a vanishing partial correlation.

A graphical model is a representation of the relationship between several random variables through a graph $G = (V, E)$ where each vertex in the set V represents a random variable and each edge in the set E implies the presence of a relationship between the adjacent random variables. Some excellent monographs on graphical modeling are [5], [7] and [16].

In graphical modeling we distinguish between two common type of graphs: one is the *conditional independence graph* (CIG) and it is an undirected graph; the other one is the *directed acyclic graph* (DAG). In both graphs the presence of an edge between vertices implies a conditional dependence, however a missing edge has a different interpretation in the two graphs: it means conditional independence in a CIG and marginal independence in a DAG.

It is possible to move from a concept of marginal independence and hence from a DAG to a CIG, through *moralization* [8].

Time series models can be seen as graphical models with vertices representing the time series $X_{t-i},\ i = 0, \ldots, t-1$, where flow of time provides a natural hierarchy that suits DAG structures.

As an illustration we consider the DAG of a saturates SVAR(2) in Figure 1. In this case all the contemporaneous variables in $X_t = \{X_{1,t}, X_{2,t}, X_{3,t}\}$ depend on the past variables in X_{t-1} and X_{t-2}. All the edges between past variables and current variables are directed consistently with time while the red edges between current variable are at this stage undirected as the contemporaneity does not suggest any direction.

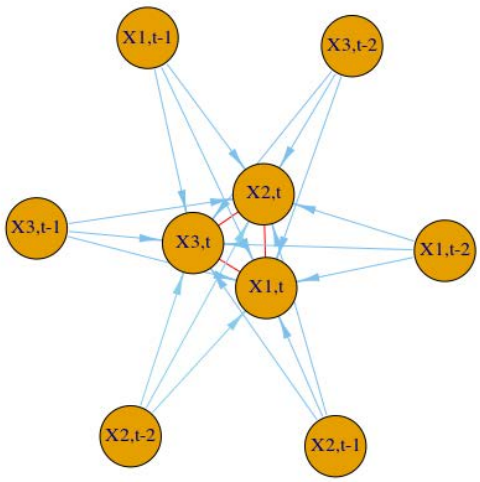

Figure 1: DAG of a saturated SVAR(2)

The aim is to provide a sparse structure for these graphs as well as a direction to the

edges between current variables. Directed acyclic graphs require, because of interpretability, that there is no cycle between contemporaneous variables; this requirement corresponds to the existence of an ordering of the variables such that Φ_0 in Equation 2 is triangular.

In the procedure propose by Tunnicliffe Wilson *et al.* [14] the sparsity of the DAG is obtained primarily through an initial sparse CIG built on the correlation structure of the data. From the CIG, the DAG is later obtained through a process of *demoralization*.

In the next section we will consider alternative methods to achieve sparsity in the conditional independence graph.

3. Sparsity

Graphical modelling provides sparse structural VAR's through sparsity (presence of zeroes) in the precision matrix W. An estimate of the partial correlations τ_{ij} are obtained through a rescaling of the of the entries w_{ij} of the precision matrix:

$$\hat{\tau_{ij}} = \frac{-w_{ij}}{\sqrt{w_{ii}w_{jj}}}$$

A zero in the precision matrix is equivalent to a zero partial correlation or conditional independence under the hypothesis of gaussianity of the time series.

The first approach we cansider is the test proposed by Tunnicliffe Wilson *et al.* [14]. Following this procedure we test the partial correlation with a t-test resulting from the equivalence

$$\tau_c = \frac{t_c}{\sqrt{n-k+t_c^2}}$$

where t_c is a threshold value of a t-distribution with a given number of degrees of freedom and a chosen level of type I error probability. The resulting threshold value of the partial correlation τ_c is used to the st the null hypothesis of $\tau = 0$. It can be shown that this test is equivalent to the classical χ^2 test used in graphical modeling: $-n\ log(1-\tau^2) \sim \chi_1^2$. Both tests suffer from multiple testing and corrections like Bonferroni [4] or Šidák [15] could be used to control the overall error rate. While amelirating the issue of multiple testing, both correction may result as cinservative.

Drton and Perlman [2] [3] propose the *SINful* approach that adopts simultaneous tests for τ_{ij} to control for the overall error rate for incorrect edge inclusion. SINful utilizes the sample partial correlation $\hat{\tau}_{ij}$ to test τ_{ij} and its Fisher's **z** transformation

$$z(\hat{\tau}_{ij}) = \frac{1}{2}\left(\ln\frac{1+\hat{\tau}_{ij}}{1-\hat{\tau}_{ij}}\right)$$

where, as $n \to \infty$, $\sqrt{n-p-1}\cdot(z(\hat{\tau}_{ij}) - z(\tau_{ij})) \sim N(0,1)$. They showed the vector of the $\frac{p(p-1)}{2}$ statistics $z(\hat{\tau}_{ij})$ to have asymptotically a multivariate Normal joint distribution. Eventually simultaneous p-values are partitioned in 3 sets: S for significant, I for indeterminate and N for non-significant.

Meinshausen and Bühlmann [11], in a seminal paper, apply the lasso to the elements of the precision matrix W with a neighborhood approach, where each node is considered as a response variable. The procedure however left open some issues:

- Sparsity is only imposed on the neighborhoods, which becomes an issue if sparsity needs to be considered for the whole partial correlation matrix;
- This method does not consider the symmetric nature of the partial correlation matrix (computationally inefficient);

- It is possible that the neighborhood pursuit approach does not provide sign consistency between regressions.

In order to solve or ameliorate these aspects, other authors have proposed a sequence of papers.

- Peng *et al* [12] proposed a *Sparse Partial Correlation Estimation* (SPACE) method to deal with lack a symmetricity.
- Friedman *et al* [6] proposed the GLASSO method with an algorithms that provides a maximum likelihood estimates of the elements of the precision matrix.
- Cai *et al* [1] proposed a consistent and computationally efficient method (CLIME) by banding the precision matrix.
- Liu and Wang [9] proposed a method (TIGER) insensitive to the tuning parameter λ so the whole dataset can be used for estimation.

In the next section we compare these different methods through a simulation exercise.

4. Simulation

We simulated two different SVAR models of different complextities. For each one of them different sample sizes were considered: 1000, 2000, 5000, 10000, 20000, 50000 data points. The simpler model is a SVAR(2) with 3 time series, while the more complex is a SVAR(3) model with seven time series.

Table 1: Results of a simulations of a SVAR(2) for 1000 and 10000 data points. The table reports the true positive rate, the true negative rate, the false positive rate and the false negative rate of different methods.

Sample size	Method	TPR	TNR	FPR	FNR
1000	Tua	0.9564	0.9380	0.0620	0.0436
	TBo	0.9293	0.9950	0.0050	0.0707
	SIN	0.929	0.998	0.002	0.071
	GLASSO	0.9	0.8889	0.1111	0.1
	SPACE	0.9943	0.983	0.017	0.0057
	CLIME	0.7836	0.299	0.701	0.2164
	TIGER	0.5557	0.9	0.1	0.4443
10000	Tua	1	0.9520	0.0480	0
	TBo	0.9993	0.9970	0.0030	0.0007
	SIN	0.999	0.997	0.003	0.001
	GLASSO	0.9	1	0	0.1
	SPACE	0.9821	1	0	0.0179
	CLIME	0.7779	0.439	0.561	0.2221
	TIGER	0.5379	0.784	0.216	0.4621

The different methods were compared considering the correctly identified (TP), the correctly identified missing edges (TN), the edges erroneously identified (FP) and the erroneously identified missing edges (FN). Four common ratios were calculated out these

measures: the true positive rate (TPR), the true negative rate (TNR), the false positive rate (FPR) and the false negative rate (FNR). These ratios are defined as:

$$TPR = \frac{TP}{TP + FN};\ TNR = \frac{TN}{TN + FP};\ FPR = \frac{FP}{FP + TN};\ FNR = \frac{FN}{FN + TP}$$

As a representation of all the simulations we report the results in the case of the simpler SVAR(2) with 1000 and 10000 data points. The methods compared are: the t-test of Tunnicliffe Wilson *et al.* (Tua); the t-test with Bonferroni correction (TBo); SINful (SIN); GASSO; SPACE; CLIME and finally TIGER. The results are shown in Table 1.

5. Conclusions

The two main conclusions from this simulations results are first that the t-test with a Bonferroni correction gives results equivalent to the SINful approach and second, that the above methods seem to perform better than the other methods. The last conclusion should however be taken with caution due to the need of fine tuning of other methods.

References

[1] L. W. Cai, T. and X. Luo. A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(404):594–607, 2011.

[2] M. Drton and M. Perlman. Model selection for gaussian concentration graphs. *Biometrika*, 91(3):591–602, 2004.

[3] M. Drton and M. Perlman. A sinful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200, 2008.

[4] O. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.

[5] D. Edwards. *Introduction to Graphical Modelling*. Springer, Boca Raton, 2000.

[6] H. T. Friedman, J. and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[7] S. Lauritzen. *Graphical Models*. Chapman & Hall, Boca Raton, 1996.

[8] S. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988.

[9] H. Liu and L. Wang. Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *Electronic Journal of Statistics*, 11(1):241–294, 2017.

[10] H. Lütkepohl. *New Introduction to Multiple Time Series*. Springer, Berlin, 2005.

[11] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34(3):1436–1462, 2006.

[12] W. P. Z. N. Peng, J. and J. Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746, 2012.

[13] R. Tsay. *Multivariate Time Series Analysis*. Wiley, Hoboken, 2014.

[14] G. Tunnicliffe Wilson, M. Reale, and J. Haywood. *Models for Dependent Time Series*. Chapman & Hall, Boca Raton, 2016.

[15] Z. Šidaḱ. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):628–633, 1967.

[16] J. Wittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, 1990.