Regularised Cov. Estimation

Jacopo Lussetti

2025-11-01

This report is aiming at replicating the methodology by Basu and Michailidis [2015] on estimating covariance matrices and their precision matrices for high-dimensional data.

Previous studies on the covariance estimator(Chen et al. [2013]) found out that the functional dependence measure scale with the spectral radius $\rho(A)$. We define functional dependence measure scale quantify the magnitude of time dependency between current and past observations. More formally,

$$\theta_{i,w,j} = \|Z_{ji} - Z'_{ji}\|_w = \left(\mathbb{E}|Z_{ji} - Z'_{ji}|^w\right)^{\frac{1}{w}} \tag{1}$$

where Z_{ji} are stationary process, and $Z'_{ji} = g(\mathcal{F}'_{\rangle})$ with the filtration being defined as $\mathcal{F}^i(\dots e_{-1}, e'_0, e_1, \dots e_i)$, for e'_0 is an independent copy of the original innovation. Additionally, we also consider the Short-Range Dependence:

$$\Theta_{mw} = \max_{1 \le j \le p} \sum_{\ell=m}^{\infty} \theta_{i,w,j} < \infty \tag{2}$$

In example 2.2 on stationary linear processes, he considers $\mathbf{z}_i = \sum_{m=0}^{\infty} A_m e_{i-m}$ with $\Sigma_e = \mathbb{E}(\mathbf{e}_i \mathbf{e}_i^T), \Sigma_{z_i} = \sum_{m=0}^{\infty} A_m \Sigma_e A_m^T$. We assume that e_{ij} i iid with mean 0, variance 1, $A_i = (a_{i,jk})_{1 \le j,k \le q} \quad \text{s.t.} \max_{j \le p} \sum_{k=1}^{p} a_{i,jk}^2 = \mathcal{O}(i^{-2-2\gamma})$. Then by the Rosenthal's theory:

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} X_i\right|^p\right] \le C_p \left(\sum_{i=1}^{n} \mathbb{E}[|X_i|^p] + \left(\sum_{i=1}^{n} \mathbb{E}[X_i^2]\right)^{p/2}\right)$$

In this setting, for the jth component we have $Z_{ji} = \sum_{m=0}^{\infty} \sum_{k=1}^{p} a_{m,jk} e_{k,i-m}$, and the coupled version differs only in the innovation at time 0:

$$Z_{ji} - Z'_{ji} = \sum_{k=1}^{p} a_{i,jk}(e_{k0} - e'_{k0}).$$

Applying Rosenthal's inequality to the independent summands $X_k = a_{i,jk}(e_{k0} - e'_{k0})$ gives

$$\mathbb{E}\left|\sum_k X_k\right|^w \le C_w \left(\sum_k |a_{i,jk}|^w + \Big(\sum_k a_{i,jk}^2\Big)^{w/2}\right),$$

and therefore

$$\theta_{i,w,j} = \|Z_{ji} - Z'_{ji}\|_w \le C_w \Big(\sum_{k=1}^{p} a_{i,jk}^2\Big)^{1/2}.$$

Under the assumption $\max_{j \le p} \sum_{k=1}^{p} a_{i,jk}^2 = \mathcal{O}(i^{-2-2\gamma})$, we obtain $\theta_{i,w,j} = \mathcal{O}(i^{-1-\gamma})$ and hence

$$\Theta_{mw} = \max_{1 \le j \le p} \sum_{\ell=m}^{\infty} \theta_{\ell,w,j} = \mathcal{O}(m^{-\gamma}),$$

A special case of equation 1 is VAR(1) process $\mathbf{z}_i = A\mathbf{z}_{t-1} + \mathbf{e}_i$, where A is a real matrix with spectral norm $\rho(A) < 1$, and the functional dependence measure $\theta_{i,2q,j} = \mathcal{O}(\rho(A)^i)$

Exercise 1

#Data Generation To simulate the stochastic regression model, we will generate data from our code that allows to randomly generate an uppper triangle coefficient matrix. Then we will simulate the data with a VAR(1) model. Then we will proceed to use a Gaussian VAR(1) model \

$$X^t = AX^{t-1} + \epsilon^t, \quad \epsilon^t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 I_{Kp}), \quad \text{and} \quad \text{diag}(A) = 0.2$$

We first define a series of equations that we will use throughout the simulation. To ensure that l2-norm of the coefficient matrix is equal to a target value, we apply the following procedure. We will now replicate figure 1 from Basu and Michailidis [2015]. He provided the following values for the spectral radius $\rho(A)$ and fixed diagonal elements of the generating coefficient matrix A to generate the VAR(1) process.

α	$\rho(A)$
0.2	0.2
0.2	0.92
0.2	0.96
0.2	1
0.2	1.01
0.2	1.02
0.2	1.03

We will first compute a specific case, in particular for $\langle 0.2, 0.2 \rangle$ to then generalise the code for the other cases. #Second Graphs

for the second graph we have the following situation:

$$\text{VAR}(2) = X_j^t = 2\alpha X_{j-1}^t - \alpha^2 X_{j-2}^t + \xi^t$$

We assume there is no cross-dependence, and the data is centered. and the number of predictors is 500. To ensure that $\Gamma_X(0) = 1$, we need to compute the covariance matrix for the residuals as follow. We consider the definition of Autocovariance of Stable VAR(p) process from Lütkepohl [2005], in particular the *Yule-Walker equations*:

$$\Gamma_Y(0) = \mathbf{A}\Gamma_Y(0)\mathbf{A}^T + \Sigma_V, \text{ where } A = \begin{bmatrix} A_1 & A_2 \\ I_K & 0 \end{bmatrix}, \quad \Sigma_V = \begin{bmatrix} \Sigma_\epsilon & 0 \\ 0 & 0 \end{bmatrix}$$

In our case, we assume that $\Gamma_X(0) = I_K$, so we can rearrange the previous function as follow:

$$\Sigma_V = I_{Kp} - A\Gamma_Y(0)A^T = I_{Kp} - AA^T$$

Unfortunately, for values closer to α, the covariance matrix wont be positive definitive, but we could rely on the fact that there is no cross-dependence, implying that the autocovariance is determined by the scaled residual variance. We considered the VAR(1) representation of the VAR(2) process:

$$X_t = \alpha_i X_{t-1} + \xi_t$$

$$\begin{aligned}
\xi_t &= X_t - 2\alpha X_{t-1} + \alpha^2 X_{t-2} \\
&= \underbrace{(1 - 2\alpha + \alpha^2 L^2)}_{(1-\alpha L)^2} X_t && \text{where } L = \text{Lag Operator} \\
X_t &= (1-\alpha L)^{-2} X_t \\
(1-\alpha L)^{-2} &= \left[\sum_{j=0}^{\infty} (\alpha L)^j\right]^2 \\
&= \sum_{j=0}^{\infty} (j+1)(\alpha L)^j \\
X_t &= \underbrace{\sum_{j=0}^{\infty} (j+1)(\alpha)^j \xi_{t-j}}_{\text{Wold Representation}} \\
\Gamma_X(0) = \text{Var}(X_t) &= \sigma_\xi \underbrace{\sum_{j=0}^{\infty} (j+1)^2 (\alpha)^j}_{(m^2-2m+1)r^m} && \sum_{j=0}^{\infty} r^m = \frac{1}{1-r} \\
\sum_{j=0}^{\infty} m r^{m-1} &= \frac{1}{(1-r)^2} \to m r^m = \frac{r}{(1-r)^2} && \sum_{j=0}^{\infty} m^2 r^m = \frac{r+r^2}{(1-r)^3} \\
\underbrace{\Gamma_X(0)}_{\mathbb{I}} &= \sigma_\xi \frac{(1+\alpha^2)}{(1-\alpha^2)^3} \\
\sigma_\xi &= \frac{(1-\alpha^2)^3}{(1+\alpha^2)} \mathbb{I}
\end{aligned}$$

In this scenario, the assumption \$ = ||A |<1\$ is not applicabale:

$$\begin{aligned}
\|A\| &= \sqrt{\lambda_{max}(A^T A)}, \quad \text{where} \quad A \text{ is the companion matrix} \\
&= \begin{bmatrix} A_1 & A_2 \\ I_K & 0 \end{bmatrix} \\
&= \begin{bmatrix} 2\alpha I_K & \alpha^2 I_K \\ I_K & 0 \end{bmatrix}
\end{aligned}$$

} \ \end{align*} On the top figure, there exhist a cross-sectional dependence defined by the VAR coefficient vectors. We saw at the beginning of the report that the functional dependence scale is of the same order as $\rho(A)$, implying that $\theta \leq c \cdot \rho(A)$ with sufficiently large n. In the picture however we see that even though we fixed $\rho(A)$, by increasing l2 norm of A the estimation error decreases more slowly with increasing sample size n. This indicates that the estimation error is sensitive to the magnitude of the coefficients in the matrix A. Potentially A can slightly exceed 1 as well.

On the second example we have a situation that in some cases the assumption $\|A\| < 1$ might not hold. Assmuing that there is no cross-dependence, the graph shows different convergence speeds depending on α, even though all processes are stable.We notice that for certain level of alpha, we have that with large sample size, the effect of dependence is significantly reduced.

References

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

Xiaohui Chen, Mengyu Xu, and Wei Biao Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021, 2013.

Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Number 978-3-540-27752-1 in Springer Books. Springer, March 2005. ISBN ARRAY(0x677fb3a0). doi: 10.1007/978-3-540-27752-1. URL https://ideas.repec.org/b/spr/sprbok/978-3-540-27752-1.html.