

Filtered Coupling Measures for Variable Selection in Sparse Vector Autoregressive Modeling

Seungkyu Lee^a · Changryong Baek^{a,1}

^aDepartment of Statistics, Sungkyunkwan University

(Received May 18, 2015; Revised June 8, 2015; Accepted July 12, 2015)

Abstract

Vector autoregressive (VAR) models in high dimension suffer from noisy estimates, unstable predictions and hard interpretation. Consequently, the sparse vector autoregressive (sVAR) model, which forces many small coefficients in VAR to exactly zero, has been suggested and proven effective for the modeling of high dimensional time series data. This paper studies coupling measures to select non-zero coefficients in sVAR. The basic idea based on the simulation study reveals that removing the effect of other variables greatly improves the performance of coupling measures. sVAR model coefficients are asymmetric; therefore, asymmetric coupling measures such as Granger causality improve computational costs. We propose two asymmetric coupling measures, filtered-cross-correlation and filtered-Granger-causality, based on the filtered residuals series. Our proposed coupling measures are proven adequate for heavy-tailed and high order sVAR models in the simulation study.

Keywords: sparse vector autoregressive model, partial spectral coherence, Granger causality

1. 서론

벡터자기상관회귀 모형은 정상성을 따르는 다차원 벡터들의 선형종속관계를 연구하는데 널리 사용되는 모형이다. 따라서 각 변수의 시간에 따른 종속 관계(temporal dependence)뿐만 아니라 변수들 사이의 종속 관계(interdependence)까지도 동시에 모형으로 설명할 수 있는 장점이 있다. 벡터자기상관회귀 모형은 Sims (1980) 이후 특히나 계량경제분야에서 높은 예측력을 가지는 모델로 알려져 많은 연구자들의 관심을 받았다. 하지만, 벡터자기상관회귀 모형은 차원이 높아짐에 따라 추정해야할 모수가 빠르게 증가하므로 불안정한 추정으로 인하여 예측력의 저하와 해석의 어려움을 동반하는 문제를 가지고 있다. 예를 들어 Bernanke 등 (2004)의 분석의 경우에 최대 6개의 변수만을 이용하여 분석을 하고 있다.

차원이 증가함에 따라서 추정하여야할 모수가 제곱에 비례하여 증가하는 단점을 보완하기 위해서 지난 수십 년 동안 많은 연구가 진행되어 왔다. Valdés-Sosa 등 (2005) 및 Haufe 등 (2010)을 비롯한 많은 실증데이터 분석 논문에서 VAR 모형의 계수들이 0에 가까운 값을 가짐이 밝혀졌고 이를 모형으로 고려하

This research was supported by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1006025).

¹Corresponding author: Department of Statistics, Sungkyunkwan University, Sungkyunkwan-ro, Jongno-gu, Seoul 110-745, Korea. E-mail: crbaek@skku.edu

여 그러한 계수들을 정확하게 0으로 둔 희박벡터자기상관회귀모형(sparse VAR models; sVAR)이 높은 차원에서의 VAR 모형의 결점을 보완할 수 있는 적절한 방법으로 등장하였다. 그 결과 다수의 연구가 어떠한 방법으로 희박성, 즉 계수를 어떻게 0으로 만들 것인지에 대해서 고민하게 되었고 현재까지 크게 두 가지 방향으로 연구가 진행되어 왔다.

첫째는 LASSO를 이용한 회귀분석에서의 변수 선택 기법을 사용하여 희박벡터자기상관모형(sparse VAR models; sVAR)을 추정하는 방법이다. 이 경우 VAR모형의 0이 아닌 계수의 개수에 벌점을 부여함으로써 희박성을 부여한다. 예를 들어 Hsu 등 (2008), Song과 Bickel (2011), Davis 등 (2015) 등의 논문들이 LASSO를 활용한 sVAR 모형 추정에 대해서 공부하였다.

둘째는 네트워크에 기반을 둔 통계분석기법에서의 아이디어를 기반으로 두 변수를 한 쌍으로 이루어서 상관관계를 측도(coupling measure)한 뒤 이를 바탕으로 변수 선택에 이용하는 방법이다. 이 방법은 Davis 등 (2015)에서 구체적으로 연구되었으며 2단계 접근법을 제시하고 있다. 1단계에서는 부분 스펙트럼 일관성(partial spectral coherence; PSC)에 기반을 두고 모든 가능한 변수 조합에 대해 상관관계를 측도한 다음, 상관관계가 높은 순서에 따라 순차적으로 VAR의 계수를 0으로 두지 않고 나머지 변수에 대해서는 0으로 두는 제한적 VAR 모형(restricted VAR models)을 적합한 뒤 BIC를 통해서 가장 좋은 모형을 찾는다. 두 번째 단계는 첫 번째 단계의 미세한 조정단계로 좀 더 희박한 모형을 만들어 주기 위해서 1단계에서 선택된 모형을 기준으로 t -통계량의 절댓값의 크기에 따라서 전진 변수 선택을 하는 것이다. Davis 등 (2015)는 시뮬레이션 및 다양한 실증 자료 방법을 통해서 PSC에 기반을 둔 방법이 효율적으로 sVAR 모형을 선택함을 보였다.

본 논문은 보다 더 정확하고 효율적으로 1단계 변수 선택 과정을 수행할 통계량을 제안하는 것에 목적이 있다. 부분 스펙트럼 일관성의 경우 Davis 등 (2015)에서 이미 지적하였듯이 방향성을 고려하지 않는 대칭적인 특성을 가지고 있다. 즉 VAR(1) 모형의 i 번째 행 j 번째 열에 해당하는 계수를 $A(i, j)$ 라고 한다면 VAR모형의 계수행렬에서는 대칭성을 가정하지 않지만 PSC에 기반을 둔 변수 선택법에서는 $A(i, j)$ 가 0이 아니라면 $A(j, i)$ 도 반드시 0으로 선택하게 되어 VAR모형의 계수행렬이 비대칭행렬일 수 있다는 점과 상충된다. 이는 곧 추정해야 할 모수의 수를 최대한 줄이고자 하는 sVAR모형의 효율성을 크게 저하시키는 요인이 될 수 있다. 또한 특정 상황에서 급격한 성능저하를 보이는 경우가 발생하는 것이 모의실험을 통해 확인되어 이러한 경우에도 강건(robust)한 결과를 보이는 통계량을 제시하고자 한다.

우리는 먼저 PSC가 좋은 성능을 보이는 이유가 다차원 시계열 자료에서 선택된 한 쌍을 제외한 나머지 변수의 효과를 제거한 잔차를 이용하기 때문임을 모의실험을 통해 밝힌다. 또한, 비대칭성을 유지하는 새로운 측도로서 필터링된 잔차에 기반을 둔 교차상관관계(Filtered-cross-correlation; FCC) 및 그래인저 인과관계(Filtered-Granger-causality; FGC)를 제안하고 모의실험을 통해 우리가 제안한 두 측도가 꼬리가 두터운 이노베이션을 통해 생성된 sVAR 모형이나 차수가 높은 모형에서도 0이 아닌 변수를 매우 정확하게 선택함을 보였다.

이 논문은 다음과 같이 구성되어 있다. 제 2장에서는 벡터자기상관회귀 모형과 희박벡터자기상관회귀 모형에 대해 살펴본 후 변수 선택 과정을 수행할 수 있는 기존의 상관통계량들을 소개한다. 제 3장에서는 모의실험을 통해 기존 방법론의 한계를 살피고 나머지 변수들의 효과를 제거하는 필터링을 통해 성능을 높일 수 있음을 밝힌다. 이를 바탕으로 비대칭성을 고려한 커플링 측도로서 FCC와 FGC를 제 4장에서 소개하고 모의실험을 통해서 제안한 커플링 측도의 유용성을 제 5장에서 살펴며 제 6장에서는 결론을 다루었다.

2. 희박 벡터자기상관회귀 모형(sVAR)

2.1. 벡터자기상관회귀 모형(VAR)

차원이 K 인 다변량 시계열 자료 Y_1, \dots, Y_n 에 대해 차수 p 를 갖는 VAR(p)모형은

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + Z_t, \quad t = 1, \dots, T$$

으로 주어진다. 여기에서 Z_t 는 평균 0, 분산-공분산 행렬 Σ 인 K 차원의 i.i.d. 정규분포를 따르는 확률변수이다. 또한 Y_t 는 인과과정(causal process)임을 가정하며 Z_t 는 $Y_s, s < t$ 와 독립임을 가정한다. 행렬 A_1, \dots, A_p 는 크기가 $K \times K$ 인 실수 행렬들로 AR계수를 나타내며 이는 곧 매개변수인 A_1, \dots, A_p 를 추정하기 위해서는 $K^2 p$ 개의 모수에 대한 추정을 해야 함을 의미한다. 따라서 차원 K 가 커질수록 추정해야 하는 매개변수의 개수가 $O(K^2)$ 으로 증가하여 모형이 매우 복잡해질 뿐만 아니라 추정이 매우 불안하게 되어 예측력을 현저히 저하시킨다. 때문에 절약성의 원리에 따라 간결한 모형을 선호하게 되는데 희박 벡터자기상관회귀 모형은 이러한 목적에서 벡터자기상관회귀 모형의 결점들을 보완할 수 있어 추정의 효율성과 모형의 예측력을 증가시킬 수 있다. 그러나 여기서 어떠한 변수를 제거해야 할지, 정확히 어느 위치에서 0 계수(zero coefficient)를 가지게 하여 불필요한 정보를 제거할지에 대해서는 자명하지 않다.

효율적인 변수 선택을 위해 Davis 등 (2015)은 부분 스펙트럼 일관성에 기반을 둔 2-단계 접근 sVAR 모형 적합 방법을 제시하였다. 1단계에서는 부분 스펙트럼 일관성(PSC)에 기반을 두고 VAR 모형의 0이 아닌 (i, j) 번째 계수(nonzero coefficient)를 선택한다. 계수행렬의 대각원소의 값들, 즉 자기 자신과의 상관성은 항상 포함시킨 채 $i \neq j$ 인 쌍에 대해서 PSC 통계량이 큰 값에서부터 작은 값으로 정렬한 다음, 순차적으로 AR계수를 0이 아닌 계수로 설정한 후, BIC를 통해 VAR 모형의 차수 p 와 0이 아닌 계수의 개수 M 을 선택한다. 2단계에서는 1단계를 거친 AR 계수들 중에서도 대각 원소를 비롯한 여전히 0의 값을 가질 수 있는 계수들이 존재하므로 t -통계량의 절댓값을 구한 뒤 전진 변수 선택을 실시하여 개선과 추정을 실시한다.

본 논문에서는 1단계, 즉 0이 아닌 (i, j) 번째 계수(nonzero coefficient)를 선택하는 변수 선택 과정의 성능을 높이는 데에 초점이 있다. 특히 부분 스펙트럼 일관성의 특성을 파악하여 다른 통계량과의 접목 혹은 수정의 방법을 통해 성능을 개량하는 것에 집중하고자 한다.

2.2. 변수 선택을 위한 상관 통계량

2.2.1. 부분 스펙트럼 일관성 관측벡터 Y_t 의 i 번째 원소를 $Y_{t,i}$, j 번째 원소를 $Y_{t,j}$ 라 쓰고 i, j 번째 원소를 제거한 나머지 벡터를 $Y_{t,-ij}$ 라고 하자. 먼저 $Y_{t,i}$ 와 $Y_{t,j}$ 의 조건부 상관관계를 구하기 위해서는 $Y_{t,-ij}$ 의 선형효과를 제거해 주어야 한다. $Y_{t,i}$ 에서 $Y_{t,-ij}$ 의 효과를 제거한 최적 선형 결합은 다음과 같이 구할 수 있다.

$$\left(D_{h,i}^{opt} \in \mathbb{R}^{K-2}, h \in \mathbb{Z} \right) := \operatorname{argmin}_{D_{h,i} \in \mathbb{Z}} E \left(Y_{t,i} - \sum_{h=-\infty}^{\infty} D_{h,i} Y_{t-h,-ij} \right)^2.$$

이를 통해 얻어진 잔차는

$$\epsilon_{t,i} = Y_{t,i} - \sum_{h=-\infty}^{\infty} D_{h,i}^{opt} Y_{t-h,-ij} \tag{2.1}$$

으로 주어지며 비슷한 방법으로 $\epsilon_{t,j}$ 를 구할 수 있다. 따라서 $Y_{t,-ij}$ 의 효과가 제거된 $Y_{t,i}$ 와 $Y_{t,j}$ 가 조건부 무상관(conditionally uncorrelated)하다면 $\epsilon_{t,i}$ 와 $\epsilon_{t,j}$ 는 모든 시차에서 무상관성을 지닌다. 즉 모든 정수 h 에 대하여 $\gamma_{ij}^\epsilon(h) := \text{Cov}(\epsilon_{t+h,i}, \epsilon_{t,j}) = 0$ 이다. 이는 진동수 영역(frequency domain)으로 바꾸어서 생각하면 두 잔차의 교차 스펙트럼 밀도함수

$$f_{ij}^\epsilon(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{ij}^\epsilon(h) e^{-ik\omega}, \quad \omega \in (-\pi, \pi]$$

가 모든 주파수 ω 에서 0의 값을 가짐을 의미한다. 교차 스펙트럼 밀도함수 $f_{ij}^\epsilon(\omega)$ 은 $Y_{t,-ij}$ 가 주어졌을 때, $Y_{t,i}$ 와 $Y_{t,j}$ 의 조건부 상관을 의미하며 이를 스케일 조정된 것이 바로 부분 스펙트럼 일관성(partial spectral coherence; PSC) (Brillinger, 1981)으로 다음과 같이 주어진다.

$$\text{PSC}_{ij}(\omega) := \frac{f_{ij}^\epsilon(\omega)}{\sqrt{f_{ii}^\epsilon(\omega)f_{jj}^\epsilon(\omega)}}, \quad \omega \in (-\pi, \pi].$$

그런데, $\text{PSC}_{ij}(\omega)$ 는 ω 에 의존하므로 Dahlhaus (2000); Dahlhaus 등 (1997)에서 제시된 바와 같이, 요약통계량을 아래와 같이 정의하며

$$\hat{S}_{ij} := \sup_{\omega} \left| \widehat{\text{PSC}}_{ij}(\omega) \right|^2, \quad i < j, j = 1, \dots, K.$$

Davis 등 (2015)에서는 \hat{S}_{ij} 이 큰 순서에 따라 1단계 변수 선택을 한다.

2.2.2. 그래인저 인과관계 Granger (1969)는 인과관계를 다음과 같이 정의하였다. $P_t(A|\bar{B})$ 를 B_t 의 과거 정보인 \bar{B}_t 를 이용한 A_t 에 대한 최적, 불편, 최소제곱 예측변수(optimum, unbiased, least-squares predictor)라 하자. 예측 잔차는 $\epsilon_t(A|B) = A_t - P_t(A|\bar{B})$ 으로 정의되며, $\sigma^2(A_t|\bar{B}_t)$ 는 예측 잔차의 분산을 뜻한다. U_t 는 $t-1$ 이전의 누적된 모든 정보를 뜻하며 $U_t - Y_t$ 는 특정 Y_t 의 정보가 제거된 모든 정보를 뜻한다. 만약,

$$\sigma^2(Y_i|U) < \sigma^2(Y_i|U - Y_j), \quad i \neq j$$

이면, 즉 Y_j 의 정보가 Y_i 의 예측 오차를 줄이는데 기여하므로 Granger (1969)는 이를 Y_j 가 Y_i 에 인과한다 라고 정의하였다. 본 논문에서는 시차가 있는 그래인저 인과관계, 즉 $Y_{t-h,j}$ 가 $Y_{t,i}$ 에 인과 하는지를 알 수 있는 다음의 정의를 사용하였다.

$$\sigma^2(Y_{t,i}|U) < \sigma^2(Y_{t,i}|U - Y_{t-h,j}), \quad i \neq j.$$

위 정의는 다음과 같은 VAR(h) 모형에서

$$Y_{t,i} = a_0 + a_1 Y_{t-1,i} + \dots + a_h Y_{t-h,i} + b_1 Y_{t-1,j} + \dots + b_h Y_{t-h,j} + \epsilon_t. \quad (2.2)$$

귀무가설

$$H_0 : b_1 = b_2 = \dots = b_h = 0$$

에 대한 F -검정 결과와 같다. 본 논문에서는 $i, j, i \neq j$ 쌍에 대한 F -검정 통계량을 커플링 측도로 사용하였으며 F 값의 크기에 따라서 sVAR모형에 포함될 변수를 선택하였다. 또한 그래인저 인과관계

의 경우 차수 h 의 선택이 중요함이 잘 알려져 있다. 본 논문에서는 데이터에 의하여 적응되어(data adaptive) 결정되는 다음의 방법을 통해서

$$\hat{h} = \min_h \left\{ h \geq 1 \text{ such that } |\hat{\gamma}_{i,j}(h)| \leq \frac{1.96}{\sqrt{n}} \right\} \quad (2.3)$$

시차를 선택하였다.

3. 기존 방법론의 한계 및 필터링의 필요성

기존의 제안된 여러 커플링 측도들이 sVAR 모형에서의 변수 선택을 얼마나 잘 하는지에 대해서 알아보 기 위하여 다음의 모의실험을 진행하였다. 다음의 sVAR(1;6) 모형에 대해서

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \\ X_{t,4} \\ X_{t,5} \\ X_{t,6} \end{pmatrix} = \begin{pmatrix} .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -.3 & 0 \\ .6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .8 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \\ X_{t-1,4} \\ X_{t-1,5} \\ X_{t-1,6} \end{pmatrix} + \begin{pmatrix} Z_{t,1} \\ Z_{t,2} \\ Z_{t,3} \\ Z_{t,4} \\ Z_{t,5} \\ Z_{t,6} \end{pmatrix}, \quad (3.1)$$

이노베이션 벡터인 $(Z_{t,1}, \dots, Z_{t,6})'$ 는 평균이 $(0, 0, 0, 0, 0, 0)'$ 이고 분산 공분산 행렬 Σ_z 이

$$\Sigma_z = \begin{pmatrix} \delta^2 & \delta/4 & \delta/6 & \delta/8 & \delta/10 & \delta/12 \\ \delta/4 & 1 & 0 & 0 & 0 & 0 \\ \delta/6 & 0 & 1 & 0 & 0 & 0 \\ \delta/8 & 0 & 0 & 1 & 0 & 0 \\ \delta/10 & 0 & 0 & 0 & 1 & 0 \\ \delta/12 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.2)$$

으로 주어진 다변량 정규분포 혹은 자유도가 df 인 다변량 t -분포를 따른다고 가정하였다. 이와 더불어 기존의 잘 알려진 상관정도를 측정하는 상관계수(correlation), 교차 상관계수(cross-correlation)를 이용한

$$\widehat{CC}_{i,j} = \max_h \hat{\gamma}_{i,j}^X(h), \quad i, j = 1, \dots, 6,$$

및 부분 교차 자기 상관계수(partial-crosscorrelation; Reinsel (2003))에 기반을 둔

$$\widehat{PC}_{i,j} = \max_h \widehat{PCC}_{i,j}(h), \quad i, j = 1, \dots, 6$$

를 커플링 측도로 사용하여 변수 선택 성능을 비교하였다.

이번 모의실험에서 얼마나 정확하게 0이 아닌 계수를 찾아내는지 살펴보는 방법은 다음과 같다. 먼저 sVAR(1;6) 모형 (3.1)을 생성하여 각각의 제시된 상관 정도를 측정하는 방법에 따라서 상관성의 강도를 순위로 기록한다. 이 상관성의 강도에 따라 참인 모형에서 0이 아닌 모수의 숫자만큼, 또 대칭성을 고려하여 변수 선택을 한다. 위 과정 모두 500번을 반복하여 상대 빈도를 계산하였다. 예를 들어 PSC 및 상관계수를 이용한 경우 모의실험 모형 (3.1)은 (1, 4), (2, 4), (3, 5) 세 곳에서 0이 아닌 변수를 선택 하게 되므로 $\hat{S}_{i,j}$, $i < j$, $i, j = 1, \dots, 6$ 인 쌍에 대해서 계산한 다음에 상위 세 개의 값을 선택하여 빈

도수를 산출한다. (부분) 교차 상관계수에 기반을 둔 CC 및 PC의 경우 상위 6개의 변수를 선택하였고 그라인저 인과관계는 대각원소에 대해서 변수 선택을 하지 못하지만 대칭성을 고려하지 않으므로 모든 $i, j, i \neq j$ 인 쌍에 대해서 통계량을 계산하고 상위 네 개의 값을 뽑아 500번의 반복을 통해 상대 도수를 계산하였다.

각 측도의 변수선택 성능을 요약하기 위한 통계량으로 다음을 고려하였다. 먼저 M 을 영이아닌 계수에 대한 지시행렬로써

$$M = \{M_{ij}\} = \begin{cases} 1, & \text{if } A_k(i, j) \text{ is non-zero for some } k = 1, \dots, p, \\ 0, & \text{otherwise} \end{cases}$$

이라 정의하자. 측도성능 요약 통계량인 MSP(mean squared proportion)은 다음과 같이 정의된다.

$$\text{MSP} = \frac{1}{|\mathcal{I}|} \sum_{i,j \in \mathcal{I}} (M_{ij} - \widehat{M}_{ij})^2,$$

여기에서 \widehat{M}_{ij} 는 커플링 측도에 의해서 (i, j) 번째 원소가 0이 아닌 것으로 추정된 상대도수를 의미하며 인덱스집합 \mathcal{I} 는 각 커플링 측도에 따른 영이아닌 계수로 설정할 수 있는 모든 가능한 순서쌍을 의미한다. 예를 들어 모형 (3.1)의 경우 대칭성을 고려하지 않은 그라인저 인과관계의 경우 대각원소에 위치하지 않은 쌍에 대해서만 계산할 수 있으므로

$$\mathcal{I}(\text{Granger}) = \{(i, j) | i \neq j, i, j = 1, \dots, 6\}$$

으로 주어져 $|\mathcal{I}| = 30$ 이다. 반면 대칭적으로 쌍을 고르는 PSC나 Correlation의 경우에는 역시나 대각원소에 위치한 쌍에 대해서 계산을 할 수 없으므로 $\mathcal{I}(\text{PSC}) = \mathcal{I}(\text{Granger})$ 이나 $\widehat{M}_{ij} = \widehat{M}_{ji}$ 로 주어진다. 일반적으로 $M_{ij} = 1$ 이 $M_{ji} = 1$ 을 의미하지 않으므로 MSP 측도는 대칭성을 고려한 별점함수이다. 비슷하게 CC 및 PC의 경우에는 대각원소에 대해서도 0이 아닌 계수로 둘 수 있으므로

$$\mathcal{I}(\text{CC}) = \{(i, j) | i, j = 1, \dots, 6\}$$

$|\mathcal{I}| = 36$ 이다. 따라서 측도성능 요약 통계량 값이 0에 가까울 수록 좋은 성능을 나타내며 값이 클수록 추정 성능이 좋지 않음을 나타낸다.

Figure 3.1은 표본의 크기가 200이고 평균은 0, 분산 공분산 행렬 (3.2)가 $\delta = 10$ 으로 주어진 다변량 정규분포를 통해 생성된 모형 (3.1)에 대해서 각각의 커플링 측도들에 의한 변수 선택의 결과를 나타낸다. 또한 편의를 위해서 대칭성을 고려하지 않은 PSC 및 상관계수의 경우 변수 선택의 상대 도수를 대칭적으로 표시하였다. 먼저 모든 커플링 측도들이 완벽하게 0이 아닌 계수를 추정해내지는 못하여 오분류가 일어났다. 오분류의 관점에서 PC가 가장 많은 오분류를 하였고 CC 및 상관계수의 경우 그 성능이 나빠지는 않으나 PSC나 그라인저 인과관계를 이용한 것보다는 떨어짐을 알 수 있다. PSC 및 그라인저 인과관계의 경우 0이 아닌 변수를 잘 판별하였지만 PSC의 경우 실제 0인 계수들에 대해서 오분류한 빈도가 상대적으로 그라인저 인과관계보다 높게 나타났으며 그라인저 인과관계에서는 (2, 1)에 위치한 변수에 대해서 0이 아닌 계수로 잘못 가려내는 빈도가 높았다. 이번 모의 실험 결과 PSC 및 그라인저 인과관계가 대체적으로 0이 아닌 계수를 잘 가려냄을 알 수 있다. Table 3.1은 MSP 값을 나타낸 표로 PSC 및 그라인저 인과관계가 가장 작은 값을 나타냈고 상관계수 및 PC를 이용한 방법이 가장 큰 값을 나타내 앞서 논의했던 내용과 일치함을 확인할 수 있다.

먼저 PSC가 좋은 성능을 보이는 이유는 상관계수 및 교차 상관계수를 비롯한 다른 측도와는 다르게 i, j 번째 변수의 커플링 정도를 계산하는데 있어서 다른 변수와의 상관관계를 제거해 주기 때문으로 보인다.

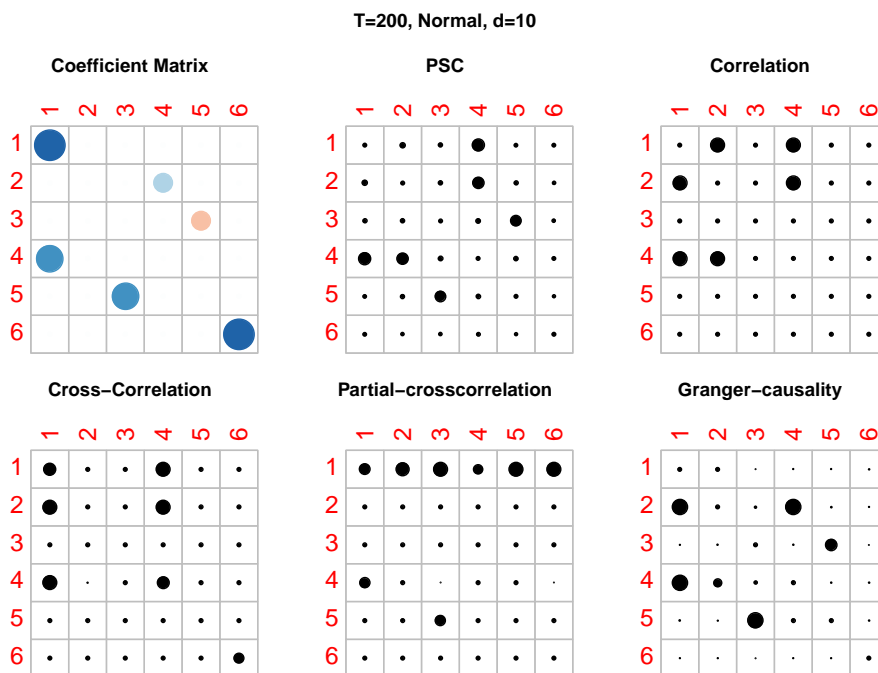


Figure 3.1. The top left matrix represents the true coefficient matrix. Estimated proportion on non-zero coefficients for sVAR(1;6) based on 500 replications.

Table 3.1. MSP for sVAR(1;6) model (3.1)

상관측도	PSC	correaltion	CC	PC	Granger-causality
MSP	0.1254	0.2667	0.1514	0.2674	0.1278

다. 다변량 시계열의 경우 시간에 따른 상관성(temporal dependence)뿐 만 아니라 다른 변수와의 상관성도(interdependence) 존재하기 때문에 PSC의 경우 선택된 한 쌍의 변수이외의 나머지 변수들의 효과를 제거한 잔차를 이용하여 상관관계를 측정하여 좋은 성능을 나타낸 것으로 보인다. 그래인저 인과관계의 경우에도 좋은 성능을 보였는데 수식 (2.2)에서 살펴보듯이 시차 h 사이에 존재하는 시간에 따른 상관성(temporal dependence)에 대한 효과를 모형에 고려하여 적절히 제거해주기 때문으로 보인다.

하지만, PSC의 경우 그 비대칭성 때문에 비효율적인 변수 선택을 함을 관측할 수 있다. 실제 모형의 계수행렬 $A_1(5,3) = .6$ 이고 $A_1(3,5) = -.3$ 으로 모형 (3.1)은 이 두 계수에 대해서 상관관계가 서로 상충되어 있다. 대칭성을 고려하지 않은 그래인저 인과관계는 (5,3) 및 (3,5) 두 위치에서 모두 0이 아닌 변수로 선택을 한 반면 PSC의 경우 그 상대도수가 현저히 낮아짐을 관측할 수 있다. 따라서 대칭성을 고려한 측도가 훨씬 더 좋은 성능을 보임을 확인할 수 있다. 위 모의실험을 토대로 sVAR 모형에서 좋은 성능을 보이는 커플링 측도는 나머지 변수들의 효과를 적절히 제거한 필터링된 자료를 바탕으로 대칭성을 고려하지 않은 측도임을 알 수 있다.

4. 필터를 이용한 커플링 측도

부분 스펙트럼 일관성은 다변량 시계열에서 i, j 번째 자료의 쌍을 제외한 나머지 변수의 효과를 제거함으

로써 기존 통계량보다 정확하게 상관성을 측도하나 방향성을 고려하지 않는 대칭성으로 인해 추정이 잘못된 상황에서는 정확성과 효율성 두 경우가 함께 저하되는 단점을 가지고 있다. 즉 부분 스펙트럼 일관성은 선택된 두 변수 $Y_{t,i}, Y_{t,j}$ 로부터 나머지 변수의 선형 효과를 제거한 수식 (2.1)에서 정의한 잔차벡터 $\{e_{t,i}\}, \{e_{t,j}\}$ 를 이용한 측도이기에 좋은 성능을 보이지만, (i, j) 와 $(j, i), i \neq j$ 의 자료의 쌍들에 대해 대칭적인 통계량을 제시해 주므로 일반적으로 비대칭성을 가지는 VAR 모형에서는 불필요한 위치의 변수까지 포함할 수 있다. 따라서 본 논문에서는 정확성을 높이기 위해서 잔차벡터에 근거하면서도 대칭성을 고려하지 않아 효율성을 동시에 지닌 측도를 다음과 같이 제안한다.

4.1. 필터링된 잔차에 기반을 둔 교차 상관관계(Filtered-Cross-Correlation; FCC)

부분 스펙트럼 일관성을 이용한 측도의 비대칭성은 푸리에 변환과정에서 생기게 된다. 따라서 주파수 영역으로의 변환 전 단계인 조건부 상관성만을 이용하여 변수 선택을 하게 된다면 그 성능을 크게 떨어뜨리지 않으면서도 대칭성을 제거할 수 있을 것이다. 구체적으로 $Y_{t,i}$ 와 $Y_{t,j}$ 의 조건부 상관관계를 구하기 위해서 $Y_{t,-ij}$ 의 선형효과를 제거하여 얻어진 잔차

$$\epsilon_{t,i} = Y_{t,i} - \sum_{h=-\infty}^{\infty} D_{h,i}^{opt} Y_{t-h,-ij}$$

에 대해서 교차공분산을

$$\gamma_{ij}^{\epsilon}(h) = \text{Cov}(\epsilon_{t+h,i}, \epsilon_{t,j})$$

계산하고 이를 표준화시킨 교차상관관계를

$$\rho_{ij}^{\epsilon}(h) = \frac{\gamma_{ij}^{\epsilon}(h)}{\gamma_{ii}^{\epsilon}(h)\gamma_{jj}^{\epsilon}(h)}$$

으로 정의한다. 교차상관관계 $\rho_{ij}^{\epsilon}(h)$ 는 시차 h 에 의존하므로 변수 선택에 쓰이는 FCC 통계량을

$$\text{FCC} := \sup_h |\rho_{ij}^{\epsilon}(h)|$$

으로 정의한다. FCC 통계량에 대한 추정치는 교차상관관계를 표본교차상관관계를 통해 추정함으로써 매우 간단하게 계산할 수 있다.

4.2. 필터링된 잔차에 기반을 둔 그래인저 인과관계(Filtered-Granger-causality; FGC)

3장의 모의실험에 따르면 그래인저 인과관계 역시 좋은 성능을 보였지만 선택된 i, j 쌍의 자료 이외의 자료에 대한 상관관계를 제거한다면 더 좋은 커플링 측도가 될 수 있음을 살펴보았다. 따라서 조건부 상관관계를 구하기 위해 얻어진 잔차 벡터 $\{e_{t,i}\}, \{e_{t,j}\}$ 에 대해서 시차 그래인저 인과관계를 적용하고 이를 FGC라 정의한다. 즉 $Y_{t-h,j}$ 와 Y_t 는

$$\sigma^2(\epsilon_{t,i}|U) < \sigma^2(\epsilon_{t,i}|U - \epsilon_{t-h,j}), \quad i \neq j$$

이 성립하면 시차 h 에 대해서 필터링된 그래인저 인과관계가 성립한다. 그래인저 인과관계와 마찬가지로 두 잔차벡터를 이용하여 $\text{VAR}(h)$ 를 적합한 후

$$e_{t,i} = a_0 + a_1 e_{t-1,i} + \dots + a_h e_{t-h,i} + b_1 e_{t-1,j} + \dots + b_h e_{t-h,j} + \epsilon_t$$

귀무가설 $H_0 : b_1 = b_2 = \dots = b_h = 0$ 에 대한 F -검정 통계량을 FGC의 상관성 정도로 사용한다. 시차를 선택하는 방법은 그래인저 인과관계에서 사용한 데이터 적응방식인 (2.3)과 마찬가지로 잔차벡터에 대해서 계산한 표본상관계수의 절대값이 $1.96/\sqrt{n}$ 보다 작아지는 처음 h 를 사용한다.

Table 5.1. MSP for sVAR(1;3) model (5.1)

상관측도	PSC	Granger-causality	FCC	FGC
MSP	0.0714	0.0667	0.0021	0.0055

5. 모의실험

본 논문에서 제안한 필터링된 커플링 측도들의 성능을 알아보기 위해서 다음의 모의실험을 하였다. 여러 모의 시험결과 우리가 제안한 필터링된 잔차를 이용한 커플링 측도들이 매우 좋은 성능을 보임을 알 수 있었다. 본 논문에서는 간결한 논문의 발표를 위해서 상대적으로 추정하기 어려운 두 가지 경우, 첫 번째는 두터운 꼬리를 가지는 이노베이션을 사용한 모형, 두 번째는 sVAR(2) 모형을 사용하여 i, j 번째 변수의 상관정도가 불투명한 두가지 경우에 대해서만 결과를 요약 발표한다.

5.1. 모의실험 1

모의실험 1에서는 다변량 정규분포보다 더 두터운 꼬리를 가지는 자유도가 5이고 수식 (3.2)로 주어진 분산-공분산 행렬 Σ_z 를 따르는 다변량 t -분포에서 생성된 데이터를 통해 부분 스펙트럼 일관성, 그래인저 인과관계와 제안된 FCC, FGC의 성능을 비교한다. 모의실험 모형은 차수가 1이고 0이 아닌 계수가 세 개인 sVAR(1;3) 모형으로

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \\ X_{t,4} \\ X_{t,5} \\ X_{t,6} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7 & 0 & 0 \\ -.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .4 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \\ X_{t-1,4} \\ X_{t-1,5} \\ X_{t-1,6} \end{pmatrix} + \begin{pmatrix} Z_{t,1} \\ Z_{t,2} \\ Z_{t,3} \\ Z_{t,4} \\ Z_{t,5} \\ Z_{t,6} \end{pmatrix}, \tag{5.1}$$

이노베이션 벡터 $(Z_{t,1}, \dots, Z_{t,6})'$ 는 평균이 0이고 수식 (3.2)의 분산-공분산 행렬 Σ_z 에서 $\delta = 5$ 로 설정된 다변량 t -분포이다. 다변량 t -분포를 이용한 이유는 우리가 제안한 FCC, FGC가 꼬리가 두터운 모형에서도 성능이 뛰어나도록 보여주기를 위함이다. 지면상 자세한 결과를 보고하지는 않지만 sVAR(1;3) 모형 (5.1)에서 다변량 정규분포를 이용할 경우에서도 FCC, FGC가 PSC나 그래인저 인과관계보다 월등한 결과를 얻었다.

Figure 5.1은 500번의 반복을 통해서 추정된 (i, j) 번째 위치에 대한 변수 선택 빈도를 나타낸다. 먼저 PSC의 경우 (3,1)에 위치한 계수에 대해서는 0이지만 0이 아닌 것으로 잘못 분류하였고 (6,2)에 위치한 0이 아닌 계수에 대해서 0으로 잘못 분류하였다. 그래인저 인과관계의 경우도 역시나 마찬가지로 (3,1), (6,2) 위치한 계수에 대해서 잘못된 변수 선택의 결과를 주어 두터운 꼬리를 가지는 경우 PSC 및 그래인저 인과관계에 기반을 둔 커플링 측도들의 성능이 급격히 떨어짐을 알 수 있다.

하지만 우리가 제안한 i, j 를 제외한 나머지 변수들의 효과를 제거한 필터링된 잔차에 기반을 둔 교차상관계수를 이용한 FCC의 경우 매우 높은 빈도로 0이 아닌 계수들의 위치를 잘 추정하였다. 또한 그래인저 인과관계를 보완한 FGC의 경우에도 0이 아닌 계수를 매우 잘 추정하여 우리가 제안한 방법들이 기존의 PSC 및 그래인저 인과관계를 이용한 방법보다 sVAR 모형의 변수 선택에서 효과적으로 쓰일 수 있음을 알 수 있다. 또한 두터운 꼬리를 가지는 이노베이션에 대해서도 우리가 제안한 방법이 매우 좋은 결과를 보이는 것은 대단히 흥미로운 결과로 보인다. MSP값은 Table 5.1에 요약되어 있으며 FCC 및 FGC가 PSC보다 훨씬 더 작은 값을 가짐을 확인할 수 있다.

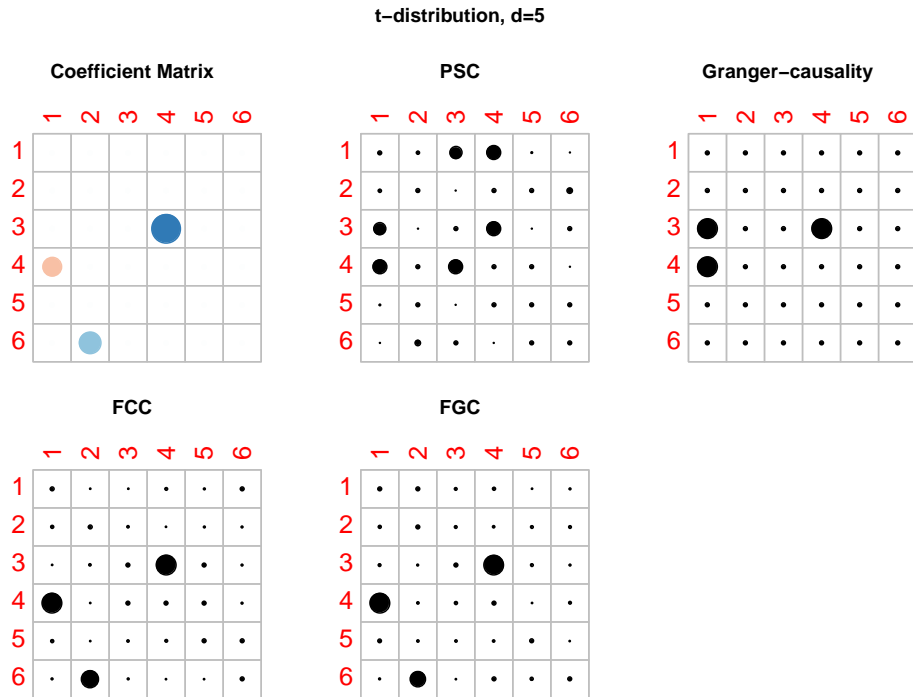


Figure 5.1. Variable selection for sVAR(1;3) model in (5.1) with innovations following multivariate $t(5)$ -distribution.

5.2. 모의실험 2

모의실험 2에서는 AR 계수행렬을 2개를 가지는 경우 즉, VAR(2) 상황에서 부분 스펙트럼 일관성, 그레인저 인과관계와 제한된 FCC, FGC 등의 성능을 비교한다. 모의 실험에 쓰인 VAR(2; 10)모형은 다음과 같이 주어졌으며

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \\ X_{t,4} \\ X_{t,5} \\ X_{t,6} \end{pmatrix} = \begin{pmatrix} .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -.3 & 0 \\ .6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .8 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \\ X_{t-1,4} \\ X_{t-1,5} \\ X_{t-1,6} \end{pmatrix} + \begin{pmatrix} .2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .3 \\ -.3 & 0 & 0 & 0 & 0 & 0 \\ .6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .4 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-2,1} \\ X_{t-2,2} \\ X_{t-2,3} \\ X_{t-2,4} \\ X_{t-2,5} \\ X_{t-2,6} \end{pmatrix} + \begin{pmatrix} Z_{t,1} \\ Z_{t,2} \\ Z_{t,3} \\ Z_{t,4} \\ Z_{t,5} \\ Z_{t,6} \end{pmatrix}, \quad (5.2)$$

이노베이션 벡터 $(Z_{t,1}, \dots, Z_{t,6})'$ 는 평균이 0이고 수식 (3.2)의 분산-공분산 행렬 Σ_z 에서 $\delta = 5$ 를 따르는 다변량 정규분포다.

Figure 5.2은 모의실험 결과를 요약한 그림으로써 맨 왼쪽 위 그림은 VAR(2) 모형 (5.2)에서 0이 아닌 계수들을 모아 하나의 행렬로 나타낸 그림이다. 부분 스펙트럼 일관성(PSC)은 매우 불안정한 추정을 보여 정확도와 효율성 모두 매우 저하되었음을 관찰할 수 있다. 또한 그레인저 인과관계의 경우 (3, 5), (2, 6)의 위치를 추정하는 것에 실패하는 모습을 보여 정확도가 저하되었음을 확인할 수 있다. 한편 FCC는 (1, 4)와 (6, 2)로 잘못 추정하는 빈도가 발생하였지만 여전히 PSC, 그레인저 인과관계에 비

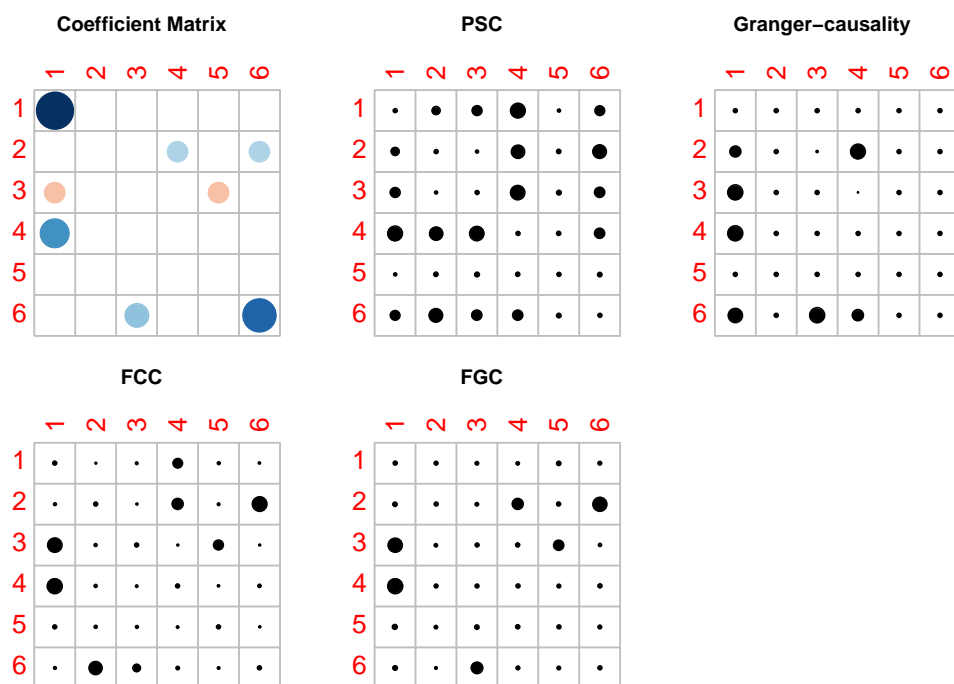


Figure 5.2. Variable selection for sVAR(2; 10) model in (5.2).

Table 5.2. MSP for sVAR(2; 10) model (5.2)

상관측도	PSC	Granger-causality	FCC	FGC
MSP	0.2604	0.1813	0.1621	0.1041

해서는 좋은 성능을 유지하는 것을 확인할 수 있었으며 놀랍게도 FGC는 모든 모수의 위치를 잘 추정하여 다른 통계량들에 비해 가장 좋은 정확성과 효율성을 보였다. 따라서 이번 모의실험을 통해서 우리가 제안한 필터링된 잔차를 이용한 그래인저 인과관계 FGC가 차수가 높은 sVAR 모형에서도 효율적으로 변수를 선택함을 알 수 있다. Table 5.2에 요약된 MSP 통계량을 통해서도 FGC가 가장 작은 오차를 보여 가장 효율적이며 기존의 PSC 및 그래인저 인과관계보다 더 좋은 성능을 보임을 확인할 수 있다.

6. 결론

회박벡터자기회귀모형은 매우 큰 다차원의 시계열 벡터들 간의 선형 종속관계를 연구할 때 효율적인 변수 선택 방법으로 잘 알려진 모형이다. 예를 들어 Davis 등 (2015)는 부분 스펙트럼 일관성에 기반을 둔 회박벡터자기회귀모형을 사용할 것을 제안하고 있다. 본 논문에서는 PSC가 변수 선택에서 다른 측도들보다 뛰어난 성능을 보이는 이유는 선택한 한 쌍의 변수를 제외한 나머지 변수에 의해 생성되는 상관관계를 효과적으로 제거한 잔차에 기반을 두기 때문임을 모의실험을 통해서 밝혔다. 하지만 PSC의 경우 sVAR 모형 계수들의 비대칭성을 무시하기 때문에 효율적인 변수 선택과정에 불리한 요소로 작용하며 두터운 꼬리를 가지는 자료의 경우 그 효율이 급격히 저하되는 것을 살펴보았다.

따라서 PSC의 좋은 효율과 비대칭성을 고려하기 위한 커플링 측도로서 나머지 변수들의 선형효과를 제

거한 잔차에 기반을 둔 교차상관계수인 FCC와 개선된 그래인저 인과관계를 이용한 FGC를 제안하였다. 우리가 제안한 두 커플링 측도는 모의실험 하에서 두터운 꼬리를 가지거나 혹은 sVAR 모형의 차수가 2 이상인 복잡한 모형에서도 0이 아닌 계수들을 기존의 측도보다 훨씬 정확하게 찾아냈으며 특히 FGC의 경우 가장 좋은 성능을 보였다. 본 논문에서 제안한 FCC와 FGC는 sVAR 모형뿐만 아니라 네트워크 모형을 비롯한 여러 다차원 모형으로 쉽게 응용이 가능하므로 표본의 개수는 작으나 추정해야 할 모수가 많은 경우에서의 변수 선택에서 매우 유용하게 쓰일 것으로 본다.

References

- Bernanke, B. S., Boivin, J. and Elias, P. (2004). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. Technical report, National Bureau of Economic Research.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, **36**, SIAM.
- Dahlhaus, R. (2000). Graphical interaction models for multivariate time series, *Metrika*, **51**, 157–172.
- Dahlhaus, R., Eichler, M. and Sandkühler, J. (1997). Identification of synaptic connections in neural ensembles by graphical models, *Journal of Neuroscience Methods*, **77**, 93–107.
- Davis, R. A., Zang, P. and Zheng, T. (2015). Sparse vector autoregressive modeling, *arXiv:1207.0520*.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37**, 424–438.
- Haufe, s., Müller, K. R., Nolte, G. and Krämer, N. (2010). Sparse causal discovery in multivariate time series, *Journal of Machine Learning Research : Workshop and Conference Proceedings*, **6**, 97–106.
- Hsu, N.-J., Hung, H.-L. and Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, **52**, 3645–3657.
- Reinsel, G. C. (2003). *Elements of Multivariate Time Series Analysis*, Springer Science & Business Media.
- Song, S. and Bickel, P. J. (2011). Large vector autoregressions. *arXiv:1106.3915*.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica: Journal of the Econometric Society*, 1–48.
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L. and Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression, *Philosophical Transactions of the Royal Society B*, **360**, 969–981.

필터링된 잔차를 이용한 희박벡터자기회귀모형에서의 변수 선택 측도

이승규^a, 백창룡^{a,1}

^a성균관대학교 통계학과

(2015년 5월 18일 접수, 2015년 6월 8일 수정, 2015년 7월 12일 채택)

요약

벡터자기회귀모형은 다차원의 시계열 자료간의 선형종속 관계를 연구하는데 효율적인 모형이다. 하지만 차원이 높아질 경우 추정해야할 모수가 급격히 증가하여 추정이 불안정해지고 예측력의 저하 및 해석의 어려움을 동반하는 문제를 가지고 있다. 이를 보완하기 위해서 많은 계수를 0으로 두는 희박벡터자기회귀모형이 제안되었고 고차원 시계열 분석에서 유용함이 밝혀졌다. 이 논문에서는 희박벡터자기회귀모형 추정에 있어서 어떠한 계수를 0으로 두어야 하는지를 판단해주는 한 쌍의 변수에 대한 상관 정도를 추정해주는 커플링 측도를 제안한다. 먼저 이 논문에서는 부분 스펙트럼 일관성에 기반을 둔 커플링 측도를 사용한 변수 선택의 경우 다른 변수의 효과를 제거한 잔차에 기반을 두었기에 좋은 효율성을 보임을 밝힌다. 하지만 부분 스펙트럼 일관성의 경우 벡터자기회귀모형 계수의 비대칭성을 고려하지 못한다는 단점이 있어 이를 보완하고자 필터링을 통해 다른 변수의 효과를 제거한 잔차에 기반을 둔 동시에 비대칭성을 가지는 커플링 측도들, 필터링된 잔차를 이용한 교차 상관성과 그래인저 인과관계를 제안한다. 모의실험을 통해 우리가 제안한 방법론들이 두터운 꼬리를 가지거나 높은 차수의 희박벡터자기회귀모형의 경우에도 매우 정확하게 0이 아닌 변수를 선택함을 보인다.

주요용어: 희박벡터자기회귀모형, 부분상관스펙트럼, 그래인저 인과관계

¹교신저자: (110-745) 서울특별시 종로구 성균관로 25-2, 성균관대학교 통계학과. E-mail: crbaek@skku.edu