

Robust estimation of sparse vector autoregressive models

Dongyeong Kim^{a,b}, Changryong Baek^{1,a,b}

^aDepartment of Statistics, Sungkyunkwan University; ^bDepartment of Fintech, Sungkyunkwan University

Abstract

This paper considers robust estimation of the sparse vector autoregressive model (sVAR) useful in high-dimensional time series analysis. First, we generalize the result of Xu *et al.* (2008) that the adaptive lasso indeed has robustness in sVAR as well. However, adaptive lasso method in sVAR performs poorly as the number and sizes of outliers increases. Therefore, we propose new robust estimation methods for sVAR based on least absolute deviation (LAD) and Huber estimation. Our simulation results show that our proposed methods provide more accurate estimation in turn showed better forecasting performance when outliers exist. In addition, we applied our proposed methods to power usage data and confirmed that there are unignorable outliers and robust estimation taking such outliers into account improves forecasting.

Keywords: robust estimation, vector autoregression, adaptive lasso

1. 서론

데이터의 수집, 저장 기술이 발달하고 그 중에서도 시간에 따라 관측되는 데이터의 양이 방대해짐에 따라, 고차원 시계열 자료를 분석하는 것에 대한 중요성이 커지고 있다. 경제 지표, 주가와 같은 데이터 뿐만 아니라 모바일 기기, 센서 등에서 자동적으로 수집되는 데이터의 양이 폭발적으로 증가하게 되었다. 이에 따라 시계열 자료의 차원이 커지는 문제와 동시에, 다수의 자료와는 다른 경향으로 관측되는 이상점(outlier)의 발생 또한 빈번하다.

다변량 시계열 자료 분석을 위해 가장 보편적으로 사용되는 모형은 시간의존구조를 모형화하는 벡터자기회귀모형(vector autoregressive model; VAR)이다. 벡터자기회귀모형은 다변량 시계열 자료의 의존구조를 잘 설명하며 추정 또한 용이하여 널리 사용되고 있다. 하지만 차원이 증가함에 따라 추정해야 할 모수가 급격하게 증가하며 이는 예측력의 저하와 해석의 어려움을 가져온다. 이를 해결하기 위해 Davis 등 (2016)은 lasso 방법론을 이용하여 불필요한 계수에 대해 정확히 0으로 설정하여 추정할 계수의 숫자를 줄이는, 즉 희박벡터자기회귀모형(sparse VAR model; sVAR) 모형을 제안하였다. 추후 Lee와 Baek (2016)을 통해 adaptive lasso를 사용하여 lasso 추정량의 일치성을 보완하는 방법론이 개발되었다.

반면, 시계열 모형을 실제 자료에 적용하는 과정에서 이상점이 존재하는 경우에 모형 선택(model selection)이나 계수 추정에서 어려움을 겪거나 예측 성능이 저하되는 현상이 잘 알려져 있으며 대표적으로 Chen과 Liu (1993)에서 보고된 바 있다. 일반적인 iid 상황에서 Xu 등 (2008)은 lasso 추정량이 어느 정도의 강건함(robustness)을 가짐을 이론적으로 밝혔다. 하지만 고차원 시계열에서의 연구 결과는 보고된 바 없어 본 연구는 먼저 lasso에 기반한 시계열 모형인 sVAR 모형의 추정에 있어서도 이상점의 영향을 많이 받는지, 또 어느

This work was supported by the Basic Science Research Program from the National Research Foundation of Korea (NRF-2022R1F1A1066209).

¹ Corresponding author: Department of Statistics and Department of Fintech, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, 03063 Seoul, Korea. E-mail: crbaek@skku.edu

정도의 강건함을 보이는지 모의실험을 통해서 살펴보고자 한다. 그 외에 *iid* 자료에서 이상점에 강건한 추정량을 얻기 위해 Wang 등 (2007)은 최소절대편차(least absolute deviation; LAD)를 이용한 lasso 추정량을, Lambert-Lacroix와 Zwald (2011)는 Huber 함수를 이용한 Huber lasso를 제안한 바 있다. 본 논문은 고차원 시계열 모형인 sVAR 추정에 있어서 lasso 방법론이 얼마나 강건한지 또한 다른 로버스트한 추정방법과의 비교를 통해 각 방법들의 장단점을 비교분석해 보고자 한다.

본 논문의 구성은 다음과 같다. 제 2장은 sVAR 모형에 대해 설명하고 이상점에 강건한 다른 추정 방법들에 대해 설명한다. 제 3장에서는 모의 연구를 통해 추정 방법에 따른 sVAR 모형의 성능을 비교한다. 이를 바탕으로 4장에서는 전력사용량 데이터에 대해 설명하고, 제시된 방법론을 적용하여 예측 성능을 비교한다. 마지막으로 제 5장에서는 본 논문의 결론과 추가 논의 사항에 대해서 다룬다.

2. 방법론

2.1. 다변량 시계열과 Vector Autoregressive Models (VAR)

다변량 시계열 자료 $Y_t = (Y_{t,1}, Y_{t,2}, \dots, Y_{t,k})^T$ 에 대해서 차원이 k 이고 차수 p 를 갖는 VAR(p) 모형은

$$Y_t = \mu + \sum_{j=1}^p A_j Y_{t-j} + Z_t, \quad t = p+1, \dots, T$$

으로 주어진다. 이 때, 행렬 A_1, \dots, A_p 는 크기가 $k \times k$ 인 실수 행렬로 자기회귀 계수들을 나타내며, $\{Z_t\}$ 는 평균 $\mathbf{0} = (0, 0, \dots, 0)^T$, 분산-공분산 행렬 Σ_Z 를 갖는 다변량 정규분포를 따르는 k 차원의 *iid* 확률 변수이다. Y_t 는 인과 과정(causal process)임을 가정하며, Z_t 는 $Y_s (s < t)$ 와 독립임을 가정한다. 추가적으로 일반화를 위해 $\{Y_t\}$ 의 평균은 $\mathbf{0}$ 을 가정한다. 즉, $\mu = \mathbf{0}$ 를 가정한다.

VAR 모형의 계수 행렬 A_1, \dots, A_p 는 다변량 시계열 데이터의 시간 의존 구조(temporal dependence structure)를 설명한다. 주어진 T 개의 시계열 자료를 통해 주어진 계수 행렬을 추정하는 것이 목적이다. 하지만, 모든 계수들을 모수화한다면, $k^2 p$ 개의 계수를 추정해야 하며 이는 차원인 k 가 커짐에 따라 추정해야 할 모수가 급격하게 늘어나는 것을 의미하며 불안정한 추정과 예측 성능, 그리고 시간 의존 구조의 해석을 어렵게 한다. 이를 해결하기 위해서 많은 계수들을 0으로 두는 희박벡터자기회귀모형이 제시되었다. 일반적인 회귀 모형에 벌점(penalty)을 가하는 방식인 lasso는 Tibshirani (1996)에 의해 제안되었으며 이를 벡터자기회귀모형에 적용하여 시간 의존 구조를 설명할 수 있는 형태로 Davis 등 (2016)에서 제안하였다.

우선, 벡터자기회귀(VAR) 모형의 추정을 위하여 Lütkepohl (2005)에 따라 다음과 같은 행렬

$$\mathbb{Y} = \mathbb{A}\mathbb{X} + \mathbb{Z} \quad (2.1)$$

로 모형을 표현할 수 있다. 수식 (2.1)은 $\mathbb{Y} = (Y_{p+1}, \dots, Y_T)$, $\mathbb{A} = (A_1, \dots, A_p)$, 그리고 $\mathbb{X} = (X_p, \dots, X_{T-1})$ 이며 각 X_t 는

$$X_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}$$

이고 $\mathbb{Z} = (Z_{p+1}, Z_2, \dots, Z_T)$ 이다. 이를 일반적인 회귀 모형의 형태로 변환하기 위해 식 (2.1)을 벡터화(vectorize)하여 표현하면

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{z}$$

와 같이 표현된다. 이와 같은 형태는 일반적인 회귀식의 형태로 표현되기 때문에 계수 추정이 용이하다. $\mathbf{Y} = \text{vec}(\mathbb{Y})$, $\mathbf{X} = (\mathbb{X}^T \otimes I_K)$, $\boldsymbol{\alpha} = \text{vec}(\mathbb{A})$, 그리고 $\mathbf{z} = \text{vec}(\mathbb{Z})$ 이다. 기본적인 회귀 추정량인 최소제곱법으로 이루어진 추정량(ordinary least squares; OLS)은

$$\hat{\boldsymbol{\alpha}}^{\text{OLS}} := \underset{\boldsymbol{\alpha}}{\text{argmin}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|_2^2 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \text{vec} \left(\mathbb{Y} \mathbb{X}^T (\mathbb{X} \mathbb{X}^T)^{-1} \right)$$

형태로 주어진다.

Davis 등 (2016)을 확장하여 Lee와 Baek (2016)은 시계열 자료간의 의존성 구조를 고려하고 성근추정일치성(sparsistency)를 위해 adaptive lasso (AL) 추정량을 다음과 같이 정의하였다.

$$\hat{\boldsymbol{\alpha}}^{\text{AL}} := \underset{\boldsymbol{\alpha}}{\text{argmin}} \left\| \left(I_T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \mathbf{Y} - \left(\mathbb{X}^T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \boldsymbol{\alpha} \right\|_2^2 + \lambda \|\mathbf{w}^T \boldsymbol{\alpha}\|_1.$$

즉, 성근추정일치성을 위해서 adaptive lasso를 고려하였고 만약 $\mathbf{w} = \mathbf{1}$ 인 경우에는 lasso 추정량이 된다. 또한 분산-공분산 행렬의 역함수의 제곱근 행렬을 곱해줌으로써 시간 의존 구조를 고려한 모수를 추정을 제안하였고, 추정은 $\boldsymbol{\alpha}$ 와 Σ_Z 를 반복적으로 추정해나가는 것을 제안하였다. Σ_Z 의 초기값으로는 OLS, MLE 추정량 등을 사용할 수 있으며 예로 MLE 추정량의 형태는

$$\hat{\Sigma}_Z = \frac{1}{T} (\mathbb{Y} - \hat{\mathbb{A}}\mathbb{X})(\mathbb{Y} - \hat{\mathbb{A}}\mathbb{X})^T$$

와 같다.

하지만, adaptive lasso는 성근추정일치성을 만족하지만 여전히 큰 계수값들에 대해서 편차가 존재하는 문제점이 있다 (Fan과 Li, 2001). 이러한 편차를 줄이기 위해 Baek 등 (2017)에서는 2단계 추정 방법을 사용한다. 첫 번째 단계에서는 앞서 설명한 방식으로 0이 아닌 계수의 위치를 파악한다. 두 번째로는 0이 아닌 계수들만 추정하는 제약적 일반화된 최소 제곱법(constrained least squares)을 사용하여 최종적인 계수를 추정한다. 제안한 2단계 추정 방법은 Shin 등 (2022)에서 표본 크기가 작지 않을 때 SCAD 및 MCP와 같은 비볼록함수를 사용하는 방법론과 비교하여 뒤쳐지지 않음을 모의 실험을 통해 밝혔다. 또한 페널티 모수인 λ 를 적절하게 결정하는 것이 실증자료 분석에서 중요하다. Baek과 Park (2021)은 시계열 자료의 의존적 구조를 고려한 block-CV 방법을 제안하고 그 유용성에서 검증하였다. 예를 들어, $T = 100$ 이고 10-CV의 경우 시간의 순서에 따라 (X_1, \dots, X_{10}) , (X_{11}, \dots, X_{20}) , \dots , (X_{91}, \dots, X_{100}) 로 자료를 분할하여 CV를 계산하게 된다.

2.2. Robustness of adaptive lasso

이상점에 강건한 방법론들을 알아보기에 앞서 Xu 등 (2008)에 따라 lasso 추정량과 강건함(robustness)과의 관계를 알아보기로 한다. 이상점이 존재하는 형태를 표현한 회귀의 형태는 lasso의 특수한 형태 중 하나로 표현될 수 있다는 것이다. 이상점이 존재하는 형태를 식으로 표현한 것은

$$\min_{\boldsymbol{\alpha}} \left\{ \max_{\Delta \mathbf{X} \in \mathcal{U}} \|\mathbf{Y} - (\mathbf{X} + \Delta \mathbf{X}) \boldsymbol{\alpha}\|_2^2 \right\} \quad (2.2)$$

이며 여기서 \mathcal{U} 는 불확실성 집합이다. 식 (2.2)는 어떤 잠재적인 오염으로 인해 관측된 자료가 손상된 상황을 나타낸 것이다. 식 (2.2)에 사용된 목적함수는 가장 오염이 큰 상황에서 최적해를 찾는 문제이며 일반적으로 최소극대화문제(min-max problem)로 불리는 형태이다. \mathcal{U} 로 표현된 불확실성 집합은 각각의 변수들에 대해 오염의 정도에 대한 제한된 범위를 가지고 있으며 식으로 표현한 형태는

$$\mathcal{U} \triangleq \{(\delta_1, \dots, \delta_m) \mid \|\delta_i\|_2 \leq c_i, i = 1, \dots, m\}$$

이며 주어진 $c_i \geq 0$ 에 대해 성립한다. 이러한 조건을 가진 오염된 데이터를 최적화하는 식을 전개한다면

$$\max_{\Delta \mathbf{X} \in \mathcal{U}} \|\mathbf{Y} - (\mathbf{X} + \Delta \mathbf{X}) \boldsymbol{\alpha}\|_2^2 = \|\mathbf{Y} - \mathbf{X} \boldsymbol{\alpha}\|_2^2 + \sum_{i=1}^m c_i |\alpha_i| \quad (2.3)$$

과 같은 식을 얻을 수 있다. 식 (2.3)에서 $c_i = c$ 로 설정하고 각 변수들에 대해 표준화를 진행한다면 우변을 최소화하는 형태는

$$\boldsymbol{\alpha}^{\text{lasso}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \|\mathbf{Y} - \mathbf{X} \boldsymbol{\alpha}\|_2^2 + c \sum_{i=1}^m |\alpha_i| \right\}$$

이며 잘 알려진 lasso의 형태와 같다. 더 자세한 증명은 Xu 등 (2008)에서 확인할 수 있다.

따라서, lasso 추정량이 식 (2.2)와 같은 오염된 데이터를 최소제곱 하여 최적화하는 문제와 같다는 점은 lasso 추정량이 이상점에 강건하다는 특성을 갖는다는 것이다. 이것이 중요한 첫 번째 이유는 강건함이라는 특성이 정규화(regularize)와 이상점으로서의 보호라는 두 특성의 연결 고리를 만들어 주기 때문이다. 이는 불확실성 집합을 다르게 고려함에 따라 lasso를 더 일반화할 수 있으며 정규화방식을 선택함에 있어서 원칙적으로 결정할 수 있도록 한다. 두 번째 이유는 강건함은 lasso의 해의 다양한 특성을 연구하는 수단으로 이용하는 강력한 성질이 될 수 있다는 점이다. lasso의 해의 강건함은 그 해가 왜 희박한지에 대한 설명도 할 수 있다. 이러한 장점으로 인해 lasso 추정량과 강건함의 관계에 대한 인식은 큰 의미를 준다고 할 수 있다.

2.3. Other robust estimation methods for sVAR

본 논문에서 고려한 강건한 추정방법은 다음과 같다. 첫 번째 방법으로 Wang 등 (2007)에서 제안한 LAD-lasso다. LAD-lasso는 목적함수의 제곱 오차를 L_1 -오차, 즉 절대값으로 대체하여 두꺼운 꼬리 또는 이상점에 영향을 적게 받게 된다. 추정량은 다음과 같이 표현된다.

$$\hat{\boldsymbol{\alpha}}^{\text{LAL}} := \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\| \left(I_T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \mathbf{Y} - \left(\mathbb{X}^T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \boldsymbol{\alpha} \right\|_1 + \lambda \|\mathbf{w}^T \boldsymbol{\alpha}\|_1.$$

하지만, 절대값 오차는 오차의 크기가 1보다 작은 경우에는 오히려 제곱 오차에 비해서 큰 벌점을 주는 것과 같기 때문에 제곱 오차보다 비효율적일 수 있다. 이러한 비효율성을 해결하기 위해 Lambert-Lacroix와 Zwald (2011)는 Huber 손실 함수를 사용한 Huber adaptive lasso (HAL) 추정법을 제시하였고, sVAR 모형에는 다음과 같이 적용된다.

$$\hat{\boldsymbol{\alpha}}^{\text{HAL}} := \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} 1^T \mathcal{H}_M \left(\frac{\left(I_T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \mathbf{Y} - \left(\mathbb{X}^T \otimes \Sigma_Z^{-\frac{1}{2}} \right) \boldsymbol{\alpha}}{s} \right) + \lambda \|\mathbf{w}^T \boldsymbol{\alpha}\|_1.$$

여기에서 Huber 손실 함수 (Huber, 1981)는 양의 실수 M 및 벡터 $\mathbf{z} = (z_1, \dots, z_n)^T$ 에 대해서 각 성분(componentwise) 별로

$$\mathcal{H}_M(\mathbf{z}) = \begin{cases} z_i^2, & |z_i| \leq M, \\ 2M|z_i| - M^2, & |z_i| > M \end{cases}$$

값을 주는 벡터 함수이다. HAL에 쓰인 $s > 0$ 은 분포에 대한 척도 모수(scale parameter)이다. 따라서 HAL은 sM 보다 작은 오차에 대해서는 제곱을 하게 되고, 이보다 더 큰 오차에 대해서는 선형으로 표현되는 것을 알 수 있어 Huber 손실 함수는 M 이 커지면 최소제곱법의 형태와 같고 M 이 작아지면 LAD와 같은 형태가 된다. Huber (1981)는 이상점이 존재하지 않는 정규분포를 따르는 자료에 대해서 효율성을 잃지 않으면서도 강건함을 유지하는 모형을 위해 $M = 1.345$ 을 제안하였다. 하지만, 이렇게 M 에 대한 결정을 하더라도 척도 모수 s

에 대한 추정에 대한 문제는 여전히 존재 한다. 본 논문에서 s 에 대한 선택은 Huber (1981)에 따라 우선 단변량 시계열에 대해서 각각 $r_{t,i} = Y_{t,i} - \hat{Y}_{t,i}$ 를 구하여 $r_{t,i}$ 에 대해서

$$\hat{s}_i = \text{median}_{1 \leq i \leq T} \{r_{t,i}\}$$

의 값을 반복적으로 진행하여 s_i 를 구하게 된다. 이를 토대로 최종적으로 사용할 s 는

$$\hat{s} = \text{median}_{1 \leq i \leq k} \{\hat{s}_i\}$$

이다.

3. 모의 실험

본 장에서는 이상점이 존재하는 시계열 자료에 대해서 AL, LAL, HAL 추정 방법을 사용하여 sVAR 모형을 추정하였을 때 추정 성능과 예측 성능을 비교하고자 한다. 실험에 사용한 모형은 6차원 VAR(1) 모형으로

$$Y_t = AY_{t-1} + Z_t, \quad Z_t \sim iid \text{MVN}(\mathbf{0}, \Sigma_Z)$$

으로 주어지며 AR계수 및 이노베이션(innovations) 벡터 $\{Z_t\}$ 의 분산 공분산 행렬은

$$A = \begin{pmatrix} 0.8 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & -0.3 & 0.0 \\ 0.6 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.6 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} \delta^2 & \delta/4 & \delta/6 & \delta/8 & \delta/10 & \delta/12 \\ \delta/4 & 1 & 0 & 0 & 0 & 0 \\ \delta/6 & 0 & 1 & 0 & 0 & 0 \\ \delta/8 & 0 & 0 & 1 & 0 & 0 \\ \delta/10 & 0 & 0 & 0 & 1 & 0 \\ \delta/12 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

을 사용하였다.

또한 이상점에 대한 영향력을 관찰하기 위하여 다변량 시계열 자료에 이상점을 추가한 가법형(additive) 오염자료(contaminated data)를 생성하였다. 이상점이 추가된 데이터 $Y_{c,t} = (Y_{c,t,1}, \dots, Y_{c,t,2})^T$ 를 생성하는 식은

$$Y_{c,t,i} = Y_{t,i} + \eta \cdot P_{t,i} \cdot \text{sign}(Y_{t,i}), \quad i = 1, 2, \dots, 6$$

이다. 여기서 η 는 이상점의 크기를 나타내며 $P_{t,i}$ 는 성공확률이 p 인 iid 베르누이 분포를 따른다. 모의 실험 모형을 단순하게 만들기 위해서 이상점은 모든 차원에 대해서 동시에 나타나는 모형을 사용하였다. 다만, 논문의 간결성을 위해서 자세히 보고하지는 않겠지만, 각 차원에 대해서 이상점이 동시에 발생하지 않는 경우라 할지라도 모의 실험의 결과와 경향성은 달라지지 않았다.

모의 실험에 대한 결과로는 RMSE, 영이 아닌 계수의 수, ME 이렇게 3가지 지표를 사용하였다. 첫 번째로, root mean square error (RMSE)는 추정량의 불일치도를 나타내기 위해 사용하였으며 경험적 기대값 $\mathbb{E}(\cdot)$ 에 대해서

$$\text{RMSE} = \mathbb{E} \|\hat{\alpha} - \alpha\|_2$$

로 정의된다. 두 번째로는 0이 아닌 계수의 개수(non-zero-coef)이며 이는 성능에 대한 척도가 될 수는 없지만, 결과를 해석하는데에 도움을 주는 지표이다. 세 번째로는, misspecification error (ME)이며 0인 계수들을 정확히 0으로 추정하였는지에 대한 지표이다. 식은 다음과 같으며

$$\text{ME} = \frac{\mathbb{E} \|\mathbf{1}_{\{\hat{\alpha}=0\}} - \mathbf{1}_{\{\alpha=0\}}\|_2^2}{k^2},$$

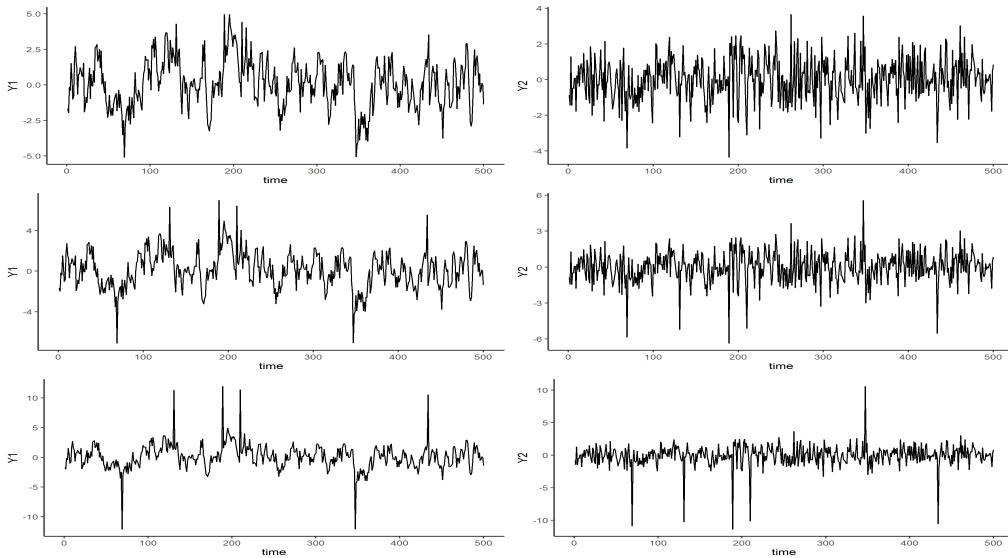


Figure 1: Time plot of simulated data with outliers.

여기서 $1_{(\alpha=0)}$ 은 해당원소가 0일 때는 1의 값을 갖는 지시함수이다.

모의 실험에서는 비희박 모형인 OLS 추정량, lasso 추정량, adaptive lasso (AL) 추정량, LAL 및 HAL 추정량을 비교하였다. 모의 실험에 사용한 표본수 $N = 200, 500, 1,000$ 이며 이노베이션 벡터의 분산-공분산 모수 $\delta = 1$ 을 사용하였으며 이상점의 크기를 결정하는 모수 $\eta \in \{3, 5, 10\}$ 을 사용하였다. $\eta = 3$ 인 상황은 이상점의 크기가 작은 상태로 잡음(noise)이 추가된 상황으로 생각할 수 있으며, $\eta = 5$ 에서 $\eta = 10$ 으로 가면서 이상점의 영향력이 커지는 상황이다. 이상점 발생 빈도를 결정하는 베르누이 분포의 확률 $p \in \{0.01, 0.02, 0.05\}$ 을 사용하였다. 성능 측도 계산에 필요한 경험적 기대값은 500번 반복을 통해서 구하였다.

Figure 1은 $N = 500$ 과 $p = 0.01$ 인 상태에서 $\eta \in \{3, 5, 10\}$ 에 따라 변화하는 모습을 시각화 한 것으로 위 패널에서 아래 패널로 내려올수록 이상점의 크기가 커지는 것을 확인할 수 있다. 첫 번째 열은 Y_1 이고 두 번째 열은 Y_2 이며 아래로 갈수록 이상점의 영향이 큰 상황이다. Y_1 은 분산 공분산 형태에 따라 다른 시계열 자료의 영향을 더 많이 받기 때문에 분산이 더 큰 모습을 볼 수 있으며, Y_2 에서는 Y_1 에 비해 분산이 작기 때문에 이상점의 여부에 대해 잘 확인할 수 있다. 앞서 설명하였듯, 이상점을 같은 시점에 발생시키고자 하였으며 Y_1 과 Y_2 의 같은 시점에 이상점이 존재하는 것을 확인할 수 있다.

먼저, LAL 및 HAL의 성근추정일치성을 만족하는지 살펴보고 위해서 이상점이 존재하지 않은 상황에서 ($\eta = 0$)의 추정에 대한 성능 비교이다. Table 1에서는 공분산 모수 $\delta = 1$ 로 고정된 상태에서 $N = 200, 500, 1,000$ 에 따라 모형간의 성능에 대해서 알아보았다. AL, LAL 및 HAL의 경우 표본수가 증가할수록 RMSE 및 ME 값이 줄어드는 것을 관찰할 수 있어 성근추정일치성을 만족함을 알 수 있다. 따라서 로버스트 추정 방법인 LAL과 HAL의 경우 이상점이 없는 경우라 할지라도 참모형을 잘 추정함을 알 수 있다.

또한 오염된 자료에서도 표본수가 증가할수록 성근추정일치성을 만족하는지 살펴보기 위해서 $p = 0.02$, $\eta = 3$ 으로 고정시킨 상태에서 표본수가 증가함에 따라 RMSE 및 ME값이 줄어드는 지 살펴보았다. Table 2에서 살펴보듯이 Lasso를 비롯한 로버스트 추정방법인 LAL 및 HAL 모두 참모형에 가까운 추정값을 주는 것을 확인할 수 있었다.

이상점이 있는 오염된 자료에서 각 방법들에 대한 비교를 하기 위해서 표본수 $N = 500$ 으로 고정시킨 뒤 각기 다른 시나리오에 따른 추정 결과는 Tables 3-5에 요약하였다. 먼저 작은 이상점 크기($\eta = 3$)를 가정한

Table 1: No outlier, $\delta = 1$

N	Type	OLS	Lasso	AL	LAL	HAL
200	RMSE(coef)	0.0615	0.0511	0.0422	0.0270	0.0368
	non-zero-coef	36.0000	16.7720	9.8640	6.5700	8.2120
	ME	0.8333	0.2992	0.1073	0.0167	0.0614
500	RMSE(coef)	0.0381	0.0312	0.0155	0.0128	0.0137
	non-zero-coef	36.0000	16.2080	6.3580	6.0040	6.1060
	ME	0.8333	0.2836	0.0099	0.0002	0.0029
1,000	RMSE(coef)	0.0269	0.0221	0.0089	0.0088	0.0089
	non-zero-coef	36.0000	16.0500	6.0040	6.0000	6.0020
	ME	0.8333	0.2792	0.0001	0.0000	0.0001

Table 2: $p = 0.02$, $\eta = 3$, small outlier

N	Type	OLS	Lasso	AL	LAL	HAL
200	RMSE(coef)	0.0733	0.0620	0.0559	0.0413	0.0504
	non-zero-coef	36.0000	17.2140	11.5520	7.2560	9.4460
	ME	0.8333	0.3115	0.1542	0.0366	0.0957
500	RMSE(coef)	0.0488	0.0420	0.0304	0.0239	0.0257
	non-zero-coef	36.0000	16.4920	7.4020	6.0400	6.3380
	ME	0.8333	0.2914	0.0389	0.0011	0.0094
1,000	RMSE(coef)	0.0369	0.0325	0.0215	0.0208	0.0209
	non-zero-coef	36.0000	16.0520	6.1120	6.0000	6.0060
	ME	0.8333	0.2792	0.0031	0.0000	0.0002

Table 3: $N = 500$, $\delta = 1$, $\eta = 3$, small outlier

p	Type	OLS	Lasso	AL	LAL	HAL
0.01	RMSE(coef)	0.0436	0.0365	0.0229	0.0173	0.0188
	non-zero-coef	36.0000	16.2840	6.9360	6.0200	6.2080
	ME	0.8333	0.2857	0.0260	0.0006	0.0058
0.02	RMSE(coef)	0.0488	0.0420	0.0304	0.0239	0.0257
	non-zero-coef	36.0000	16.4920	7.4020	6.0400	6.3380
	ME	0.8333	0.2914	0.0389	0.0011	0.0094
0.05	RMSE(coef)	0.0599	0.0546	0.0471	0.0410	0.0428
	non-zero-coef	36.0000	16.6240	8.2080	6.1060	6.5640
	ME	0.8333	0.2951	0.0613	0.0031	0.0157

Table 3을 살펴보면 OLS 추정량에 비해서 lasso의 추정량이 더 좋은 성능지표를 줄일 수 있다. 이는 Xu 등 (2008)에서 밝혔듯이 lasso 추정량이 약간의 강건성을 가지고 있음을 보여준다. 하지만, lasso와 비교하면 AL, LAL 및 HAL의 성능이 훨씬 더 우수함을 살펴볼 수 있다. 즉 adaptive lasso를 사용함으로써 성근추정일치성을 가져 성능이 향상됨을 알 수 있다. 또한 이상점의 발생 빈도인 p 가 증가할수록 성능지표가 증가함을 살펴볼 수 있다. 이는 곧 이상점의 빈도가 증가할수록 추정에 어려움을 겪음을 나타내며 고려한 방법론 중에서는 adaptive lasso보다는 LAL 및 HAL이 더 우수한 성능을 보임을 확인할 수 있었다. 이러한 경향은 이상점의 크기를 5로 증가한 Table 4 및 10으로 증가한 Table 5에서도 공통적으로 관찰된다. 또한 발생 빈도 p 를 고정시킨 채 이상점의 크기가 증가할수록 성능지표가 증가함을 관찰할 수 있다. 즉 이상점의 크기 및 빈도가 커질수록

Table 4: $N = 500$, $\delta = 1$, $\eta = 5$, medium outlier

p	Type	OLS	Lasso	AL	LAL	HAL
0.01	RMSE(coef)	0.0573	0.0494	0.0412	0.0299	0.0330
	non-zero-coef	36.0000	16.3820	8.7660	6.1000	6.6320
	ME	0.8333	0.2884	0.0768	0.0029	0.0176
0.02	RMSE(coef)	0.0692	0.0626	0.0570	0.0474	0.0503
	non-zero-coef	36.0000	16.6200	9.9040	6.2500	7.0740
	ME	0.8333	0.2950	0.1084	0.0074	0.0298
0.05	RMSE(coef)	0.0899	0.0862	0.0834	0.0781	0.0804
	non-zero-coef	36.0000	16.3940	10.4600	6.4800	7.8160
	ME	0.8333	0.2888	0.1246	0.0151	0.0508

Table 5: $N = 500$, $\delta = 1$, $\eta = 10$, large outlier

p	Type	OLS	Lasso	AL	LAL	HAL
0.01	RMSE(coef)	0.0981	0.0894	0.0881	0.0764	0.0796
	non-zero-coef	36.0000	15.9480	12.1080	6.5420	7.9860
	ME	0.8333	0.2772	0.1717	0.0264	0.0608
0.02	RMSE(coef)	0.1174	0.1130	0.1123	0.1089	0.1089
	non-zero-coef	36.0000	14.7960	11.6680	6.3960	8.2260
	ME	0.8333	0.2474	0.1639	0.0367	0.0714
0.05	RMSE(coef)	0.1449	0.1453	0.1446	0.1443	0.1439
	non-zero-coef	36.0000	15.6200	11.6380	6.7540	9.3280
	ME	0.8333	0.2750	0.1704	0.0472	0.1043

Table 6: The 1-step-ahead forecasting error comparison

η	p	OLS	Lasso	AL	LAL	HAL
0	0.00	1.0471	1.0404	1.0315	1.0313	1.0307
	0.01	1.0744	1.0694	1.0562	1.0533	1.0542
3	0.02	1.1000	1.0944	1.0809	1.0770	1.0760
	0.05	1.1758	1.1714	1.1612	1.1512	1.1543
5	0.01	1.1238	1.1163	1.1023	1.0923	1.0960
	0.02	1.1739	1.1664	1.1568	1.1404	1.1476
	0.05	1.3317	1.3281	1.3225	1.2998	1.3113
10	0.01	1.3287	1.3104	1.2983	1.2591	1.2174
	0.02	1.3986	1.3819	1.3726	1.3612	1.3672
	0.05	1.6599	1.6453	1.6382	1.5983	1.6204

추정에 부정적 영향을 줄 수 있다. 하지만, LAL 및 HAL 방법의 경우 RMSE 및 ME의 증가폭이 adaptive lasso보다 작아 더 우수한 성능을 보임을 확인할 수 있었다.

추정방법의 성능을 보다 정확하게 비교하기 위해서 Figure 2에 경험적 평균이 아니라 상자그림을 그려보았다. LAL 방법이 평균뿐만 아니라 퍼짐도 작아 본 절에서 고려한 모형에서 가장 좋은 성능을 보임을 알 수 있었다.

또한 이상점의 영향에 따라 추정방법에 따라 예측성능이 달라지는 지 모의 실험을 통하여 1-step-ahead 예측오차를 비교하여 Table 6에 보고하였다. 이상점이 없는 경우, $\eta = 0$, 방법론에 따른 예측오차는 미비하여

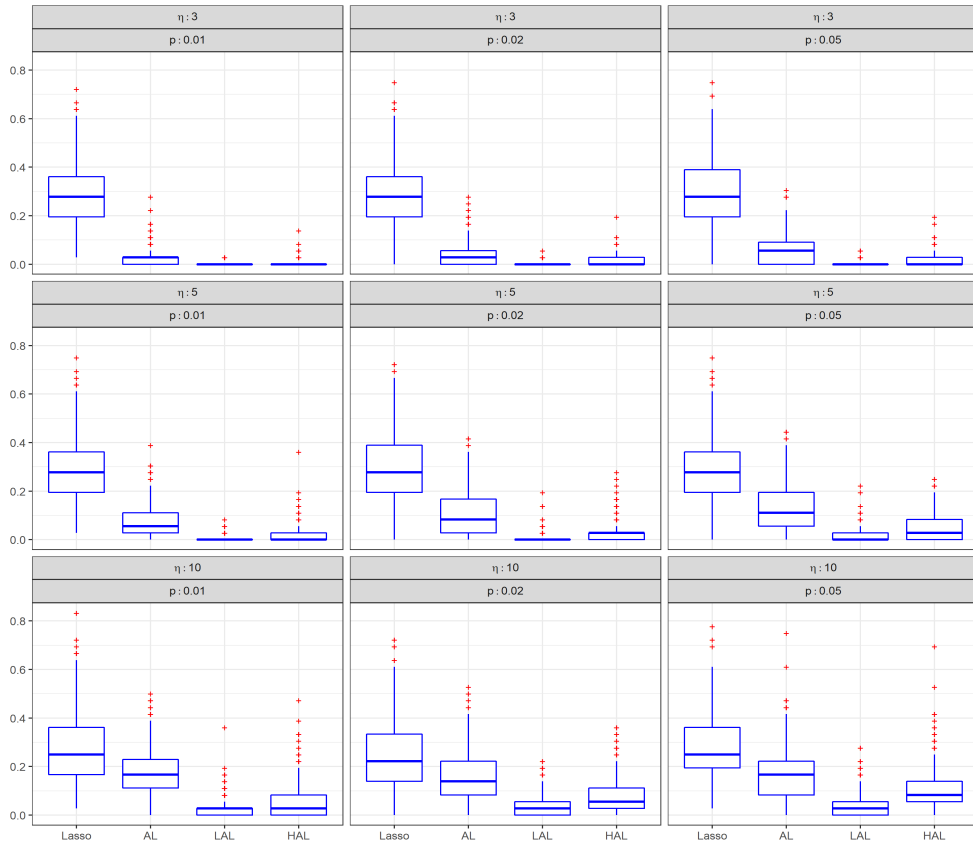


Figure 2: Box plots of misspecification errors.

예측 성능이 비슷함을 알 수 있었다. 이상점의 크기가 커지거나 빈도가 커질 경우 예측 오차는 증가함을 확인할 수 있었다. 즉, 다변량 시계열 모형인 sVAR의 경우라 할지라도 이상점의 영향으로 예측 성능이 저하됨을 다시 확인할 수 있었다. 좀 더 로버스트한 추정방법인 LAL 혹은 HAL이 adaptive lasso보다 근소하게 더 작은 예측오차를 주어 이들 방법론이 더 좋은 성능을 보임을 알 수 있었다. 요약하자면, 고차원 시계열 모형인 sVAR에 모의 실험을 통해 이상점이 추정의 어려움 및 예측 성능을 떨어뜨림을 먼저 확인할 수 있었다. 이상점이 존재함에도 adaptive lasso 방법론은 어느 정도의 강건함을 가지고 성근추정일치성을 만족함을 모의 실험을 통해서 확인할 수 있었다. 또한 이상점의 크기가 커질수록 또 빈도가 증가할수록 성능측도들이 나빠짐을 통해서 추정에 영향을 더 크게 미침을 알 수 있었으며 보다 로버스트한 방법인 LAL 혹은 HAL이 adaptive lasso 보다는 더 강건한 성능을 보임을 확인할 수 있었다. 특히 LAL 모형이 본 논문에서 사용한 시뮬레이션 모형에 대해서는 가장 좋은 성능을 보였다.

4. 실증자료분석

본 장에서는 기존의 실증자료를 바탕으로 로버스트한 sVAR 추정 방법론의 유용성에 대해서 검증하고자 한다. 사용한 자료는 한국에너지공단에서 에너지온실가스기술정보 플랫폼(<https://tips.energy.or.kr/egydata/contest.do>)을 통해 제공하는 비식별화된 건물들의 전력사용량 데이터이다. 전력사용량 데이터

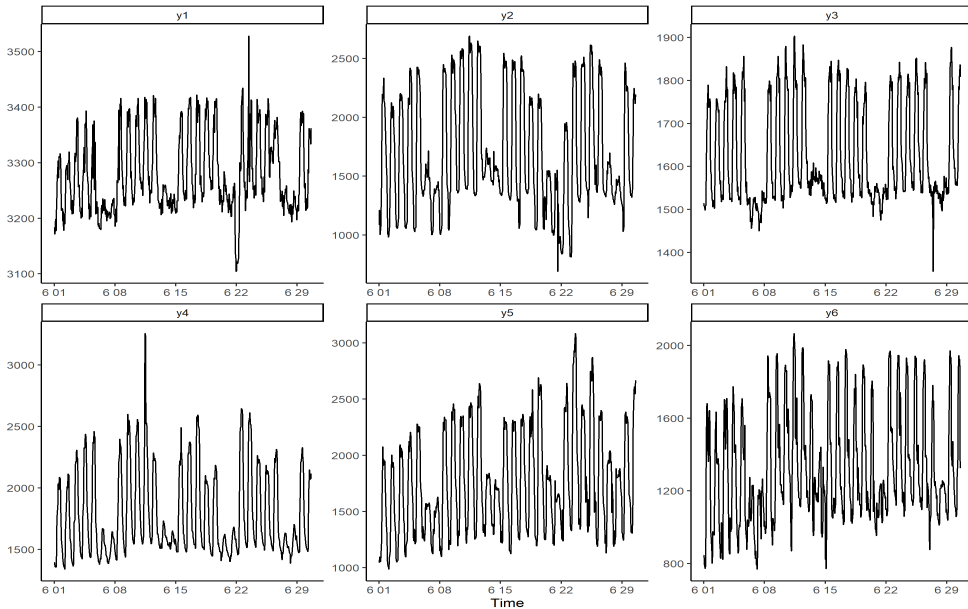


Figure 3: Time plot of electricity usage data.

는 2020년 6월 1일부터 2020년 8월 24일까지 85일간의 전력사용량을 60개의 건물에 대해 1시간 간격으로 시간에 따라 관측한 자료로 차원은 60, 표본의 크기는 2,040인 다변량 시계열자료이다. 분석의 용이성을 위해 본 논문에서는 6개의 건물에 대해서 2020년 6월의 데이터를 학습 데이터로 사용하고 2020년 7월 1일부터 7일까지의 일주일의 자료를 검증 데이터로 사용하였으며 최종적으로 사용된 학습 데이터의 관측값의 개수는 720개, 검증 데이터는 168개이다.

Figure 3은 몇 개 건물에 대한 시계열 그림(time plot)을 나타낸다. 예를 들어 y_1 의 경우 6월 22일 근방, y_3 의 경우 6월 28일 경에 이상점으로 의심되는 점이 뚜렷이 보여 이상점의 영향을 적게 받는 로버스트한 추정 방법을 사용하는 것이 타당해 보인다. 본격적으로 로버스트한 추정방법을 적용하기 전에 시계열 그림에서 살펴보듯이 24시간을 주기로 하는 계절성(seasonality)이 존재함이 보이고, 주중 혹은 주말에 따라 전력사용량에 큰 차이가 있어 평일과 주말의 효과를 모형에 반영할 필요가 있다. 이를 위해서 sVAR 모형에 설명변수를 추가하는 sVARX (sparse vector autoregressive with exogenous variable)을 고려하였다. 설명 변수로는 평일과 주말에 대해서는 가변수(dummy variable)를 사용하였으며 계절성은 $\sin(\cdot)$ 과 $\cos(\cdot)$ 함수를 사용해 설명하는 조화회귀(harmonic regression)를 사용하였다. 이를 수식으로 표현하면

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \beta_1 X_{t,dummy} + \beta_2 X_{t,\sin} + \beta_3 X_{t,\cos} + Z_t \quad (4.1)$$

과 같으며 각 행렬은

$$Y_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ Y_{t,3} \\ \vdots \\ Y_{t,6} \end{pmatrix}, \quad \beta_1 = \begin{pmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \beta_{1,3} \\ \vdots \\ \beta_{1,6} \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} \beta_{2,1} \\ \beta_{2,2} \\ \beta_{2,3} \\ \vdots \\ \beta_{2,6} \end{pmatrix}, \quad \beta_3 = \begin{pmatrix} \beta_{3,1} \\ \beta_{3,2} \\ \beta_{3,3} \\ \vdots \\ \beta_{3,6} \end{pmatrix}$$

Table 7: 1-step forecasting error

	OLS	Lasso	AL	LAL	HAL
MSPE	6265.180	6265.573	6265.522	6255.627	6263.244
MAPE	0.033063	0.033060	0.033051	0.032902	0.033062

로 표현된다. 설명변수가 있는 sVARX 모형에 대한 AL, LAL 및 HAL의 추정 방법은 식 (2.1)과 유사하게 행렬을 통해 표현할 수 있다. 즉

$$\mathbb{Y} = \mathbb{A}\mathbb{X} + \mathbb{Z}$$

에 대해서 $\mathbb{Y} = (Y_2, \dots, Y_T)$, $\mathbb{A} = (A, \beta_1, \beta_2, \beta_3)$, 그리고 $\mathbb{X} = (X_1, \dots, X_{T-1})$ 이며 각 X_t 는

$$X_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ X_{t,dummy} \\ X_{t,sin} \\ X_{t,cos} \end{pmatrix} = \begin{pmatrix} (Y_{1,t} Y_{2,t}, \dots, Y_{6,t})^T \\ (Y_{1,t-1} Y_{2,t-1}, \dots, Y_{6,t-1})^T \\ 1_w \\ \sin(\cdot) \\ \cos(\cdot) \end{pmatrix}, \quad \text{여기서 } 1_w = \begin{cases} 1, & \text{if } t \in \text{weekday,} \\ 0, & \text{if } t \in \text{weekend} \end{cases}$$

와 같다. 이를 다시 일반적인 회귀 모형으로 바꾸기 위해 벡터화를 진행하면

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}$$

이며 $\mathbf{Y} = \text{vec}(\mathbb{Y})$, $\mathbf{X} = (\mathbb{X}^T \otimes I_K)$, $\boldsymbol{\alpha} = \text{vec}(\mathbb{A})$, 그리고 $\mathbf{z} = \text{vec}(\mathbb{Z})$ 을 이용하여 추정을 할 수 있다.

위와 같이 계정성과 주말 효과를 제거한 자료가 정상성을 띠는지 확인하기 위해서 추가적으로 단위근 검정인 ADF 검증을 각 차원에 대해서 진행하였고 모두 기각하여 정상성을 확보했다고 판단하였다. 또 식 (4.1)에서 알 수 있듯이, 실증자료분석에 사용된 모형은 차수가 2인 VAR(2) 모형이다. 이는 BIC를 기준으로 가장 낮은 값을 선택하였다. 예측성능을 평가하기 위해 사용한 지표는 MSPE, MAPE를 사용하였다. MSPE와 MAPE는 각각 다음과 같이 표현된다.

$$\text{MSPE} = \frac{1}{T_0} \sum_{t=T-T_0+1}^T (Y_t - \hat{Y}_t)^2, \quad \text{MAPE} = \frac{1}{T_0} \sum_{t=T-T_0+1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|.$$

여기서 \hat{Y}_t 는 1부터 $t-1$ 시점까지 데이터를 사용해 1-step-ahead 예측한 값이며, T_0 는 검증 데이터의 크기이다. 앞서 언급하였듯, 검증데이터는 7일간의 데이터로 사용하였으며 검증 데이터의 관측값 개수는 168개, 즉 $T_0 = 168$ 이다. 실제로 예측 성능을 비교해본 결과는 Table 7에서 확인할 수 있으며, MSPE의 경우에는 LAL이 가장 좋은 성능을 그 다음으로는 HAL이 좋은 성능을 가졌고 로버스트한 추정방법인 LAL과 HAL이 좋은 성능을 보임을 확인할 수 있었다.

Figure 4는 전력사용량 데이터에 대한 각 모형의 계수 추정 결과를 시각화한 것이다. OLS로 추정한 계수와 비교하였을 때 나머지 모형들은 불필요한 계수를 0으로 추정하며 이는 \hat{A}_2 에서 더 많은 차이를 확인할 수 있었다. AL이 가장 희박하게 추정하는 것을 확인할 수 있었으며 HAL과 LAL이 그 다음으로 희박한 모습을 볼 수 있었으며 사용한 방법에 따라 추정 계수의 값도 미묘하게 차이가 남을 살펴볼 수 있었다.

5. 결론 및 논의점

희박벡터자기회귀모형은 다변량 시계열 자료에 대한 관계를 파악하거나, 예측을 진행할 때에 효율적인 방법으로 알려져 있다. 본 논문에서는 이상점이 존재하는 경우에 sVAR 모형의 추정에 어떤 영향을 받는지 살펴 보았다. Xu 등 (2008)이 iid 자료에서 밝혔듯이 adaptive lasso 방법이 고차원 시계열 모형인 sVAR에서도 어느

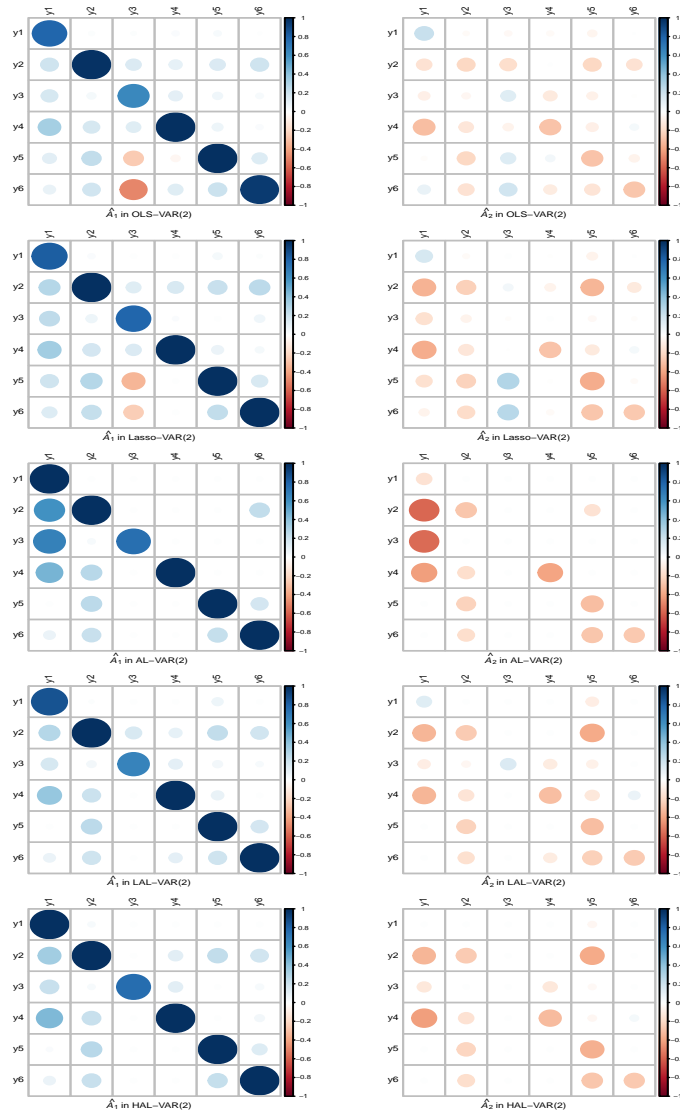


Figure 4: Coefficient estimates from the $s\text{VAR}(2)$ at lags 1 and 2, respectively.

정도의 강건함을 가지고 있음을 확인할 수 있었다. 하지만, 이상점이 많아지고 벗어남이 커질수록 효율성이 급격히 저하됨을 모의 실험을 통해 확인할 수 있었다. 반면, 최소절대편차와 Huber 함수를 adaptive lasso에 적용한 LAL 및 HAL 방법이 기존 adaptive lasso에 비해서 이상점이 존재하는 경우에도 0인 계수들을 정확하게 0으로 찾아내는 비율이 훨씬 높았으며 예측 성능 또한 우수함을 확인할 수 있었다. 또한 실증자료분석에서 일반적인 데이터에 이상점으로 의심가는 부분들이 존재할 수 있음을 확인하였고, 실제로 결과에서도 LAL 및 HAL 방법이 더 좋은 예측 성능을 보였다.

물론 로버스트 추정방법인 LAL 및 HAL에도 보완점은 있다. 우선 HAL의 경우 튜닝 모수인 M 에 대한 선택이 중요한데 이에 대한 보다 통계적인 방법론에 대한 필요성이 요구된다. 또한 adaptive lasso를 확장하여

SCAD와 같이 비볼록 벌점함수를 가지는 강건한 추정 방법에 대한 연구도 흥미로운 추가 연구가 될 수 있을 것으로 보인다.

References

- Baek C, Davis RA, and Pipiras V (2017). Sparse seasonal and periodic vector autoregressive modeling, *Computational Statistics & Data Analysis*, **106**, 103–126.
- Baek C and Park M (2021). Sparse vector heterogeneous autoregressive modeling for realized volatility, *Journal of the Korean Statistical Society*, **50**, 495–510.
- Chen C and Liu LM (1993). Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association*, **88**, 284–297.
- Davis RA, Zang P, and Zheng T (2016). Sparse vector autoregressive modeling, *Journal of Computational and Graphical Statistics*, **25**, 1077–1096.
- Fan J and Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.
- Huber PJ (1981). *Robust Statistics*, Wiley, New York.
- Lambert-Lacroix S and Zwald L (2011). Robust regression through the Huber’s criterion and adaptive lasso penalty, *Electronic Journal of Statistics*, **5**, 1015–1053.
- Lee SG and Baek C (2016). Adaptive lasso in sparse vector autoregressive models, *The Korean Journal of Applied Statistics*, **29**, 27–39.
- Lütkepohl H (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Shin AJ, Park M, and Baek C (2022). Sparse vector heterogeneous autoregressive model with nonconvex penalties, *Communications for Statistical Applications and Methods*, **29**, 53–64.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267–288.
- Wang H, Li G, and Jiang G (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso, *Journal of Business & Economic Statistics*, **25**, 347–355.
- Xu H, Caramanis C, and Mannor S (2008). Robust regression and lasso, *Advances in Neural Information Processing Systems*, **21**, 103–126.

Received May 11, 2022; Revised June 13, 2022; Accepted June 22, 2022

희박 벡터 자기 회귀 모형의 로버스트 추정

김동영^{a,b}, 백창룡^{1,a,b}

^a성균관대학교 통계학과, ^b성균관대학교 핀테크융합전공

요 약

본 논문은 고차원 시계열 자료에 이상점이 존재하는 경우 희박벡터자기회귀모형(sparse VAR; sVAR)의 모수를 강건하게 추정하는 방법에 대해서 연구하였다. 먼저 Xu 등 (2008)이 독립인 자료에서 밝혔듯이 adaptive lasso 방법이 sVAR 모형에서도 어느 정도의 강건함을 가짐을 모의 실험을 통해 알 수 있었다. 하지만, 이상점의 개수가 증가하거나 이상점의 영향력이 커지는 경우 효율성이 현저히 저하되는 현상도 관찰할 수 있었다. 따라서 이를 개선하기 위해서 최소절대편차(least absolute deviation; LAD)와 Huber 함수를 기반으로 별점화시키는 adaptive lasso를 이용하여 sVAR 모형을 추정하는 방법을 본 논문에서는 제안하고 그 성능을 검토하였다. 모의 실험을 통해 제안한 로버스트 추정 방법이 이상점이 존재하는 경우에 모수 추정을 더 정확하게 하고 예측 성능도 뛰어남을 확인했다. 또한 해당 방법론들을 전력사용량 데이터에 적용한 결과 이상점으로 의심되는 시점들이 존재하였고, 이를 고려하여 강건하게 추정하는 제안한 방법론이 더 좋은 예측 성능을 보임을 확인할 수 있었다.

주요용어: 로버스트추정, 벡터자기회귀모형, adaptive lasso

이 논문은 한국연구재단의 지원을 받아 수행된 기초연구 사업임 (NRF-2022R1F1A1066209).

¹교신저자:(03063)서울특별시 종로구 성균관로 25-2, 성균관대학교 통계학과 및 핀테크융합전공.

E-mail: crbaek@skku.edu