

자원·환경경제연구 제 12 권 제 4 호  
Environmental and Resource Economics Review  
Volume 12, Number 4, December 2003 : pp. 687~708

## VAR와 그래프이론을 이용한 시계열의 인과성 분석\* — 미국 대두 가격 사례분석 —

박호정 \*\* · 윤원철 \*\*\*

〈차례〉

- |         |           |
|---------|-----------|
| I. 서 론  | III. 모형추정 |
| II. 모 형 | IV. 결 론   |

### I. 서 론

경제활동과 관련된 통계자료가 풍부해짐에 따라 이의 활용도가 날로 높아져 가고 있다. 농업, 금융, 에너지, 환경 등과 관련된 다양한 종류의 시계열자료는 과거의 패턴을 이해하는 것뿐만 아니라, 예측을 통해 미래의 정보를 제공한다는

\* 저자들은 논문의 개선에 유익한 조언을 주신 익명의 심사위원들께 심심한 사의를 표한다. 남아있는 오류가 있다면 이는 전적으로 저자들의 책임을 밝혀 둔다.

논문 투고일(2003년 11월 25일)~심사 완료일(2003년 12월 9일).

\*\* 에너지경제연구원 책임연구원(제1저자).

\*\*\* 한양대학교 경제금융학부 조교수(교신저자).

점에서도 정책설정에 유익한 도움을 준다. 특히, 다변량 시계열 분석은 외부충격에 대한 내생변수들간의 동태적 반응을 분석할 수 있다는 점에서 폭넓게 이용되고 있다.

다변량 시계열 모형으로는 벡터자기회귀(vector autoregressive: VAR)와 오차수정모형(vector error correction model: VECM)이 가장 광범위하게 쓰이는데, 시계열간의 공적분 유무에 따라 후자 또는 전자의 모형이 선택된다. VAR와 VECM 모형은 시계열 변수의 개수가 증가함에 따라 추정해야 하는 파라미터의 개수도 급증하는 특징을 갖기 때문에 연립추정식 자체의 관찰만으로는 관련 파라미터에 대한 직관적인 해석을 얻기 어려울 때가 많다. 따라서, 대부분의 연구자는 외생적인 확률충격이 내생변수에 미치는 영향 등을 살펴보기 위해 충격반응분석(impulse response analysis)과 예측오차 분산분해(forecast error variance decomposition) 등에 의존한다. 이 과정에서 이른바 콜레스키분해를 이용하여 오차항간의 이분산의 문제를 제거하는데, 이러한 방법에 관한 주된 비판은 변수의 나열순서에 따라 충격반응 및 예측오차 분산분해의 결과가 달라진다는 것이다. Cooley and Dwyer (1998)가 지적하는 바와 같이, VAR(또는 VECM)는 연구자의 주관적인 경험과 관찰에 의해 상정된 모형구조에 의해 영향을 받기 때문에, 전제되어 있는 구조모형과 독립적으로 해석할 수 없다는 한계를 지닌다.

변수의 임의나열 문제를 해결하고자 Sims (1986), Bernanke (1986)는 구조 VAR 모형(structural VAR model)을 제시하였다. 변수간 당기에 주고받는 영향을 상정할 때 경제이론에 부합하는 인과관계를 부과한다는 것이 구조 VAR 모형의 기본 발상이다. 하지만, 비록 경제이론을 이용하여 구조적 확률오차를 식별한다 하더라도 연구자의 선형적인 지식(a priori knowledge)에 의존한다는 점에서 여전히 임의성이 존재한다. 더욱이, 거시경제 분석에서처럼 이론이 풍부하고 잘 정립되어 있는 경우가 아니면 모형의 전제가 인과관계를 미리 규정하는 오류를 범하게 된다. 이러한 이유로 VAR 모형은 흔히 이론이 없는 (atheoretical) 모형이라는 지적을 받는다(Cooley and LeRoy, 1985).

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

물론 모형의 식별에 상관하지 않는 Pesaran and Shin (1997)의 ‘일반화된 충격반응함수’와 같은 방안이 제안될 수 있으나, Swanson and Granger (1997)는 변수의 임의적인 선택을 배제하기 위한 또 다른 방법으로 그래프 이론이 유용함을 보여 주었다. 특히 그래프 이론은 내생변수간의 동시적 인과관계 (contemporaneous causality)를 규명해 준다는 장점이 있기 때문에, Swanson and Granger (1997) 이후 방향지시 비순환 그래프(directed acyclic graph : DAG)를 중심으로 경제변수의 구조적 관계를 식별하는 연구가 활발히 진행되었다. 그 중 Spirtes, Glymour and Scheines (2000)가 개발한 알고리즘(PC 알고리즘)이 특히 VAR 및 VECM의 모형식별과 연관하여 자주 활용되고 있다.<sup>1)</sup>

논문에서 DAG에서 도출되는 인과성은 Granger 인과성에서 말하는 개념과 상이하다는 점에 유의할 필요가 있다. 후자는 시간상의 인과성(time sequence causality)으로서, 즉 어떤 원인(cause)이 효과(effect)에 대해 시간적으로 선행 한다(precede).<sup>2)</sup> 반면, 그래프 분석에 의존한 인과성은 시간의 흐름에 따른 인과관계를 의미하지 않고, 동시적인 인과성을 말한다.

근간에 다수의 관련 논문들이 나오고 있는데, 그 중 Haigh and Bessler (2003)는 국내 주요 곡물시장간 가격의 발견과정(price discovery process)을 이해하기 위해 VECM-DAG 모형을 사용하였다. 유사한 방법을 이용하여 Bessler and Yang (2003)은 주가지수의 국가간 변동 및 상호 영향을 분석하였다. Jeon (2002)은 아시아 금융위기 당시 동북아시아 여러 금융당국의 화폐정책 반응을 분석하기 위해 VAR-DAG 이론을 이용하였으며, Haigh, Park and Bessler (2003)는 미국 재무성의 30년 만기 채권의 거래시 장내거래의 음향크기가 채권가격 변동성과 거래량에 미치는 영향을 살펴보고자 VAR-DAG를 이용하였다.

1) 경제모형에서의 인과성 식별문제를 특별주제로 다룬 「Journal of Econometrics」의 Vol. 39, No.1-2 (1988)는 경제이론, 철학 등 다양한 측면의 광범위한 논의를 제공하고 있다. 이 후 Pearl (1995), Swanson and Granger (1997) 등의 노력에 의해 변수들의 인과관계를 구축하기 위한 그래프이론의 기초가 정립되었다.

2) 전기의  $x_{t-1}$ 이 금기의  $y_t$ 에 영향을 미치면,  $x_t$ 는  $y_t$ 의 Granger 원인이 된다고 한다.

본 논문은 VAR와 DAG를 접목하는 방법론을 소개하는데 주안점을 두되 응용사례로 미국 대두의 메릴랜드 지역 베이시스를 분석한다. 지역 단위의 베이시스 연구는 가격위험 및 생산위험에 노출되어 있는 개별 농가의 위험을 줄이고, 적정 수준의 재고 결정에 영향을 미치는 매우 중요한 정보로 활용될 수 있다 (윤병삼, 2002).

논문의 구성은 다음과 같다. 다음 제Ⅱ절에서 VAR 및 DAG 모형에 대한 방법론을 소개한 후, 제Ⅲ절에서 메릴랜드의 대두 현물가격과 시카고 선물가격을 활용한 예를 제공한다. 결론은 제Ⅳ절에서 맺도록 한다.

## Ⅱ. 모형

### 1. VAR 모형<sup>3)</sup>

$K$ 개의 시계열자료가 있으며, 이의 벡터를  $y_t = [y_{1t}, \dots, y_{Kt}]'$ 로 표기하자. 시차가  $p$ 일 때 벡터자기회귀 VAR( $p$ )의 축약형 모형을 다음 식 (1)과 같이 나타낼 수 있다.

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t \quad (1)$$

잔차항 벡터  $\varepsilon_t = [\varepsilon_{1t}, \dots, \varepsilon_{Kt}]'$ 는  $\varepsilon \sim N_m(0, \Sigma_\varepsilon)$ 의 다변량 정규분포를 따른다. 여기서,  $\Sigma_\varepsilon$ 은 시간불변 공분산행렬(time invariant covariance matrix)을 나타낸다.

VAR 모형은 시계열 분석에 매우 유용한 도구이지만, 추정 파라미터의 개수

---

3) 논의를 단순화하기 위해 VAR 시계열 분석에서 중요한 계절성 등의 주제는 본 논문에서 다루지 않는다. 다만, 단위근과 적정시차 선정에 대해서는 뒤에서 간략히 소개하기로 한다.

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

가 많은 관계로 모형을 통해 내생변수들에 대해 직관적인 정보를 제공하는데 한계가 있기 때문에 연구자들은 아래에서 소개되는 충격반응 분석을 자주 사용한다. 계수행렬  $A_i$ 가  $|z| \leq 1$ 에 대해 다음 식 (2)를 만족하면 시계열 (1)은 안정적이라고 한다.

$$\det(I_k - A_1 z - \cdots - A_p z^p) \neq 0 \quad (2)$$

안정성의 조건을 만족하면 VAR 모형을 이동평균형태로 전환시킴으로써 잔차항의 변화에 대한 시계열 변수의 반응크기를 측정할 수 있다. Wold 정리에 의해 식 (1)을 축차연산하면 다음과 같이 무한차수 벡터이동평균모형으로 나타낼 수 있다.

$$y_t = \Phi_0 \varepsilon_t + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \cdots \quad (3)$$

여기서,  $\Phi_0 = I_K$ 이며,  $\Phi_s$ 는  $K \times K$  계수행렬로서  $\Phi_s = \sum_{j=1}^s \Phi_{s-j} A_j$ 와 같다.<sup>4)</sup>

식 (3)에서  $\varepsilon_t$ 의 한 단위 변화인 확률충격(random shock 또는 innovation)이  $s$ 기의  $y$ 에 미치는 영향을  $\partial y_{t+s} / \partial \varepsilon_t = \Phi_s$ 로부터 구할 수 있다. 보다 구체적으로 살펴보면,  $\varepsilon_t$ 의 첫 번째 항이  $\delta_1$ 만큼 변하고, 두 번째 항이  $\delta_2$ , 그리고  $K$ 번째 항이  $\delta_K$ 만큼 변한다고 하자. 그러면 이들 변화가  $y_{t+s}$ 에 미치는 영향은 아래와 같다.

$$\Delta y_{t+s} = \frac{\partial y_{t+s}}{\partial \varepsilon_{1t}} \delta_1 + \frac{\partial y_{t+s}}{\partial \varepsilon_{2t}} \delta_2 + \cdots + \frac{\partial y_{t+s}}{\partial \varepsilon_{Kt}} \delta_K = \Phi_s \delta \quad (4)$$

4) 무한합의 수렴을 위해서는  $\lim_{s \rightarrow \infty} \Phi_s = 0$ 을 만족하여야 한다.

여기서,  $\delta' = [\delta_1, \dots, \delta_K]$ 이다. 결국 확률오차인  $\varepsilon_t$ 에 의한  $y_t$ 의 충격반응 요인은  $\Phi_s$ 의 항들에 의해 측정되는데, 충격반응은 일회성의 확률오차가 금기 및 다음 기의 내생변수에 미치는 영향을 추적한다는 점에서 예측오차 충격반응 (forecast error impulse response)이라고 불리기도 한다.

식 (4)에서 문제는 VAR 모형의 추정에서 나온 공분산행렬  $\Sigma_\varepsilon$ 이 대각행렬이 아니라는 점이다. 즉,  $\varepsilon_{it}$ 와  $\varepsilon_{jt}$ 가 계열상관뿐만 아니라 동시상관 (contemporaneous correlation)까지 보일 수 있으므로, 미래 어느 시점에 변수의 변화의 순 효과를 측정할 수 없다. 따라서, 오차항간에 서로 동시상관하지 않도록 오차항 벡터를 직교화할 필요가 있으며, 이를 위해 일반적으로는 다음에서 소개되는 콜레스키 분해(Cholesky decomposition)가 사용된다.

$E[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon$ 는 대칭 양부호행렬이므로 항상 다음과 같이 나타낼 수 있다.

$$\Sigma_\varepsilon = ADA'$$

여기서  $D$ 는 대각행렬이고,  $A$ 는 역행렬이 존재하는 하방삼각행렬로서 대각행렬원소가 1로 이루어져 있다. 이와 같은 전환을 콜레스키 분해 또는 삼각분해 (triangular factorization)라 한다. 그 다음, 행렬  $A$ 를 이용하여 직교잔차항 벡터를  $u_t = A^{-1}\varepsilon_t$ 와 같이 정의한 후, 벡터곱에 기대값을 취하면 다음 식 (5)를 얻는다.

$$E[u_t u_t'] = E[A^{-1}\varepsilon_t \varepsilon_t' (A^{-1})'] = A^{-1}(ADA')(A^{-1})' = D \quad (5)$$

$D$ 는 정의에 의해 대각행렬이므로 오차항간에 직교화가 이루어졌으며, 따라서 동시상관을 배제한 충격반응을 분석할 수 있다.<sup>5)</sup> 하지만,  $Au_t = \varepsilon_t$ 를 행렬 형태로 전개한 다음 식 (6)에서 보듯이 오차항이  $u_{1t} = \varepsilon_{1t}$ ,  $u_{2t} = \varepsilon_{2t}$  –

---

5) 직교충격반응 분석에 대한 자세한 설명은 김명직 · 장국현 (2002, pp. 382~385)을 참조하기 바란다.

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

$a_{21}u_{1t} = \varepsilon_{2t} - a_{21}\varepsilon_{1t}$  와 같이 축차적으로 정의되기 때문에 변수의 나열순서에 따라 충격반응의 분석 결과가 달라진다.

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{K1} & a_{K2} & a_{K3} & \cdots & a_{KK} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \vdots \\ u_{Kt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \vdots \\ \varepsilon_{Kt} \end{bmatrix} \quad (6)$$

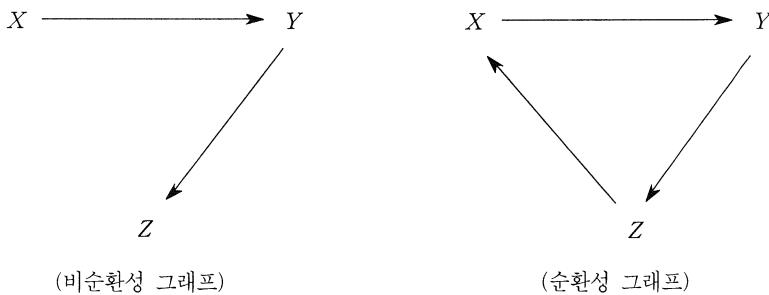
즉,  $y_{1t}$ 는 다른 변수들에 대해 전혀 영향을 받지 않으나,  $y_{2t}$ 는  $y_{1t}$ 에 영향을 받고,  $y_{3t}$ 는  $y_{1t}$ 와  $y_{2t}$ 에 영향을 받는다는 식의 인과순위(causal ordering)가 전제되어야 한다.

Sims (1986)와 Bernanke (1986)는 VAR 모형의 확률오차를 구조적으로 식별하기 위하여 경제이론과 부합하는 구조 VAR 모형을 제안하였다. 구조 VAR 모형은 거시이론 분야에서처럼 통계적으로 검정하고자 하는 이론이 잘 정립되어 있는 경우에는 매우 유용한 방법이다. 하지만, 만일 상이한 경제이론이 적용되거나 한 연구자의 주관적인 경험과 판단이 다른 연구자의 그것과 일치하지 않으면 인과순위가 달라지며, 그 결과 상이한 충격반응 분석결과를 얻게 된다는 점에서 구조 VAR 모형 역시 VAR 모형의 한계를 완전히 극복하지는 못했다고 볼 수 있다. 예를 들어, 본 연구의 응용분야에서처럼 지역간 재화가격의 상호연관에 대해 사전적으로 뚜렷한 인과순위를 정하기 어려운 경우에는 모형의 식별은 전적으로 연구자의 판단에 의존해서 이루어진다. 이와 같은 이유로 그래프이론이 최근 시계열 분석에 도입되기 시작하였다.

## 2. 방향지시 비순환성 그래프

Swanson and Granger (1997)는 VAR 모형의 식별과정에서 변수의 나열순서를 정할 때 주관적인 기준을 배제한 통계적 기준을 마련하기 위하여 그래프

〈그림 1〉 비순환성 및 순환성 그래프



이론을 이용하였다. 이렇게 함으로써 충격반응 및 예측오차 분산분해 분석시 잔차항이 임의의 순서로 나열되는 것을 방지하고 분석의 객관성을 제고할 수 있기 때문이다. 이후, Pearl (2000)과 Spirtes, Glymour and Scheines (2000)는 그래프이론을 이용하여 인과관계를 탐색하는 알고리즘을 개발하였는데, 아래에서는 이를 간단히 소개하기로 한다.

가장 간단한 방향지시 그래프(directed graph)의 예로  $X \rightarrow Y$ 를 생각해 보자. 이 경우에 “ $X$ 는  $Y$ 를 초래한다”라고 표시하기 위해  $X$ 와  $Y$ 를 방향지시화살표로 연결한다. 방향지시 화살표는 연결선 또는 모서리(edge)라고 부르며, 한 연결선을 두고 연결된 변수는 인접(adjacent)하다고 말한다. 만일 그래프가 순환하지 않는 방향지시 그래프들로만 구성되어 있으면 이를 방향지시 비순환성 그래프(directed acyclic graph)라 칭한다. 예를 들어, 〈그림 1〉에서  $X \rightarrow Y \rightarrow Z \rightarrow X$ 는 비순환성 그래프가 되지 못한다.

그래프이론에서는 그림과 아울러 이를 기호화해서 표시하기도 한다. 일반적인 그래프 정의에 의하면 그래프  $G$ 는 순서쌍  $\langle V, E \rangle$ 로 나타내는데, 여기서  $V$ 는 꼭지점(vertex)의 집합,<sup>6)</sup>  $E$ 는 연결선의 집합,  $M$ 은 연결선의 형태를 정의하는 표기(mark)의 집합이다. 만일  $X$ ,  $Y$ ,  $Z$ 의 세 변수가 있다고 하면, 그래

6) 그래프이론에서 꼭지점은 변수(variable)와 동의어로 사용된다.

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

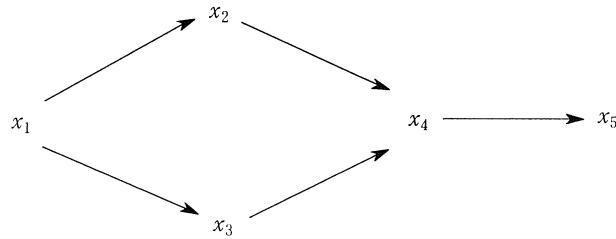
즉  $G$ 는 변수의 집합인  $\{X, Y, Z\}$ 로 구성되며, 나아가 각 변수는 방향지시 화살표로 연결된다. 예를 들어, 그래프  $X \rightarrow Y$ 는 좌측 끝의 순서쌍  $[X, .]$ 와 우측 끝의 순서쌍  $[Y, >]$ 로 나타낼 수 있으며, 이를 함께 묶어  $\{[X, .], [Y, >]\}$  또는 동일하게  $\{[Y, >], [X, .]\}$ 로 표시할 수 있다.  $X \rightarrow Y$ 의 예에서는  $X$  종점에서 표기가 없기 때문에, **空 마크(empty mark)**를 나타내는  $EM$ 을 사용하여  $[X, EM]$ 으로 표기할 수 있다.

다음 예로 그래프  $Y \leftarrow X \rightarrow Z$ 를 고려해 보자. 이는  $\langle \{X, Y, Z\}, \{EM, >\}, \{\{X, EM\}, \{Y, >\}\}, \{\{X, EM\}, \{Z, >\}\} \rangle$ 로 기호화할 수 있다.  $X$ 는  $Y$ 와  $Z$ 에 대한 공통의 원인(common cause)이다.  $Y$ 와  $Z$  간의 무조건부 관계(nonconditional association or correlation)는 0이 아니지만,  $X$ 라는 공통원인이 정보로 주어져 있다면  $Y$ 와  $Z$ 의 조건부 관계는 0이 된다(Pearl, 2000, p. 17).

인과성을 나타내는 화살표의 방향을 바꾸어서  $X \rightarrow Y \leftarrow Z$ 를 보자. 기호상  $\langle \{X, Y, Z\}, \{EM, >\}, \{\{X, EM\}, \{Y, >\}\}, \{\{Z, EM\}, \{Y, >\}\} \rangle$ 로 표시 할 수도 있다. 이 경우  $Y$ 는 충돌부(collider)라 부른다.  $X$ 와  $Z$ 의 무조건부 관계는 0이지만,  $Y$ 에 대한 공통효과(common effect)라는 점에서  $X$ 와  $Z$ 의 조건부 관계는 0이 아니다.

인과관계를 구축하기 위해서 DAG는 다음의 기본적인 조건을 만족시켜야 한다. ① **인과충분성 조건**(causal sufficiency condition), ② **마아코브 조건**(Markov condition). 첫 번째의 인과충분성 조건은 최소한 2개 이상의 변수에 영향을 미치는 변수가 분석에서 제외되어서는 안 된다는 것을 의미한다.  $X$ 가  $Y$ 와  $Z$ 의 공통원인일 때  $X$ 를 분석에서 제외했다고 하자. 그러면 외관상  $Y$ 와  $Z$  간에 통계적으로 유의미한 인과관계를 얻을 수 있지만, 이는 실제  $X$ 로부터 기인하였을 가능성이 크다. 따라서  $Y$ 와  $Z$ 의 두 변수에 동시 영향을 미치는  $X$ 가 분석에 반드시 포함되어야만 유사인과관계(spurious causality)를 방지할 수 있다.

〈그림 2〉 그래프 인과관계의 예



마아코브 조건은  $d$ -분리( $d$ -separation)에 의해 설명된다. Pearl (1995)에 의하면, 두 변수  $X$ 와  $Z$  간의 모든 경로를  $Y$ 가 차단하면 “ $Y$ 는  $X$ 와  $Z$ 를  $d$ -분리한다”고 정의하며, 기호로는  $(X \perp Z | Y)_G$ 로 표시한다. 앞서 예의  $X \rightarrow Y \leftarrow Z$ 에서,  $X$ 와  $Z$ 는 비록  $Y$ 에 의해  $d$ -분리되어 있지만,  $Y$ 를 조건부로 취하면  $d$ -연결(d-connected)되어 있다. 반면,  $X \leftarrow Y \rightarrow Z$ 는 공통원인인  $Y$ 를 조건부로 취하기 전에는  $X$ 와  $Z$ 는  $d$ -연결되어 있지만,  $Y$  조건부하에서는  $d$ -분리된다.

인과충분성과 마아코브 조건을 만족하면 방향지시 그래프는 상관계수와 편상관계수(partial correlation)를 이용하여 변수들간의 인과관계를 탐색한다.<sup>7)</sup>

Spirites, Glymour and Scheines (2000)는 변수들간에 통계적으로 무의미한 연결선을 제거함으로써 인과성을 순서화할 수 있음을 증명하였는데, 이 알고리즘의 핵심은 위에서 소개한  $d$ -분리에 기초한 마아코브 성질이다. 이는  $d$ -분리 정의를 이용하면 인접변수의 조건부 관계만을 이용함으로써 결합확률분포를 다음 식 (7)과 같이 분해하여 순차적으로 모든 변수들간의 관계를 정리할 수 있기 때문이다.

$$P(V) = \prod_{v_i \in V} P(v_i | pa_i) \quad (7)$$

7)  $Z$ 조건부하의  $X$ 와  $Y$ 의 상관계수는  $r_{xy|z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}$  와 같다.

여기서  $pa_i$ 는  $v_i$ 의 모변수(parent variable)의 실현된 값을 말한다.

만일 <그림 2>의 그래프가 있다면 여기서  $x_1$ 은  $x_2$ 의 모변수이고,  $x_2$ 와  $x_3$ 는  $x_4$ 의 모변수이며,  $x_4$ 는  $x_5$ 의 모변수가 된다. 따라서 (7)의 결합확률분포를 이용하면 다음과 같다.

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)$$

모변수가 없는  $x_1$ 의 선형적 확률이 외생적으로 주어지면,<sup>8)</sup> 조건부 확률을 순차적으로 계산해 나가기 때문에 DAG는 흔히 베이지안 네트워크(Bayesian network)라고 불리기도 한다.

PC 알고리즘은 모든 변수들간의 방향이 설정되어 있지 않는 기초그래프부터 출발한다. 0의 상관계수를 갖는 변수들간의 연결선이 먼저 제거된 후, 남아있는 연결선들 중에서 1계 편상관계수(first order partial correlation)가 0인 연결선들을 제거한다. 여기서 1계 편상관계수라 함은 제3의 변수를 조건부로 한 두 변수들간의 상관계수를 말한다. 그리고 나면 남아있는 연결선들 중에서 2계 편상관계수가 0인 연결선이 순차적으로 제거되는데, 동일한 과정으로 ( $N - 2$ )계 편상관계수에 이르기까지 계속 진행된다. 상관계수가 0이라는 귀무가설을 검정하기 위해 일반적으로 아래의 Fisher- $z$  통계량을 사용한다.

$$z(\rho(i, j|k), n) = \left[ \frac{1}{2} \sqrt{n - |k| - 3} \right] \ln \left[ \frac{1 + \rho(i, j|k)}{1 - \rho(i, j|k)} \right]$$

$n$ 은 추정에 사용된 자료의 개수이며,  $\rho(i, j|k)$ 는 조건부 상관계수이다. 즉,  $k$ 를 조건부로 했을 때의  $i$ 와  $j$ 의 모상관계수(population correlation)를 의미한다.  $|k|$ 는  $k$ 에 있는 변수의 숫자이다. 이상의 PC 알고리즘은 Spirtes를 위

8) 그래프 이론에서는 모변수를 가지지 않는 변수는 외생적으로 가정한다. Spirtes, Glymour and Scheines (2000, p. 12)를 참조하기 바란다.

시한 연구진들이 개발한 소프트웨어 TETRAD II를 사용하면 간편하게 수행될 수 있다. 프로그램의 기본적인 투입자료는 VAR를 추정해서 얻은 분산-공분산 행렬 또는 상관계수행렬로서, 이 투입자료를 바탕으로 TETRAD II는 통계적으로 유의미하지 않는 연결선을 차례로 제거한다.<sup>9)</sup>

### III. 모형추정

#### 1. 자 료

제III절에서는 실제 통계자료를 이용하여 앞에서 소개된 그래프이론을 VAR와 연계해 보도록 한다. 사용된 자료는 미국 메릴랜드 주의 대두 가격과 시카고 선물시장의 대두 선물가격이다. 지역 창고업자(elevator)에서 제시되는 대두의 현물가격은 지역내 수요, 수송비 등의 요소에 의해 각 지역마다 차이가 발생하며, 지역 현물가격에서 선물가격을 제한 베이시스에도 역시 지역간 차이가 발생한다. 각 지역 창고업자의 베이시스 정보는 곡물의 생산 및 판매, 그리고 저장과 같은 농가의 의사결정에 중대한 영향을 미치기 때문에, 지역간 베이시스의 동태적 인과관계를 분석하는 것은 의미가 있을 것이다.

메릴랜드 주에서는 중부 메릴랜드(CMD), 남부 메릴랜드(SMD), 서부 메릴랜드(WMD), 동북부 해안지역(UES), 동남부 해안지역(LES)의 다섯 개의 지역으로 구분하여 현물가격이 제시된다.<sup>10)</sup> 이들 중 대두의 평균 생산량이 가장 많은 지역은 Dorchester, Worcester, Wicomico 카운티를 포함하는 LES이다. 아울러, 외지로 수출되는 대두 물량은 CMD에 위치한 볼티모어 항구로 운반된다.

9) TETRAD II에 대한 자세한 설명은 매뉴얼인 Scheines *et al.* (1994)를 참조하기 바란다.

10) CMD는 Frederick, Carroll, Montogemery, Howard 카운티를 포함한다. SMD는 Prince George와 Anne Arundel 카운티를, WMD는 Washington 카운티와 그 외의 서부지역을 포함한다. UES는 Queen Anne's와 Caroline을 포함한 동북부 해안가이며, LES는 남부해안가 지역을 포함한다(뒤에서 소개되는 <그림 3>의 지도를 참조).

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

베이시스 계산에 사용된 현물가격은 매주 목요일 “Maryland Grain and Livestock Report”에 발표되는 주간 평균가격이다. 본 연구에서는 1990년 1월부터 최근 2003년 6월까지의 CMD, SMD, WMD, UES, LES 각 지역의 가격 정보를 이용하였으며, 현물가격에서 해당 주의 시카고 선물시장의 평균가격을 차감한 베이시스 시계열 자료인 FCMD, FSMD, FWMD, FUFS, FLES를 내생변수로 사용하였다(N= 683).<sup>11)</sup>

### 2. VAR 모형의 추정

모형의 추정에 앞서 시계열의 안정성부터 검정하여야 한다. 각 시계열자료를 가지고  $\Delta y_t = \delta_0 + \delta_1 y_{t-1} + \sum_{s=1}^S \beta_s \Delta y_{t-s} + \epsilon_t$ 를 추정하여 얻은 Augmented Dickey Fuller (ADF) 검정결과가 <표 1>에 나와 있다. 모든 시계열의 ADF 검정값이 유의수준 5~10% 하에 임계치보다 작기 때문에 시계열 불안정성의 귀무가설을 기각한다.

다음 단계로 적정 시차  $p$ 를 결정한다. 임의의  $m = 0, \dots, p_{\max}$ 에 대한 VAR ( $m$ )을 추정해서  $p$ 를 결정하는 가장 보편적인 방법으로는 SC (Schwartz Criterion), AIC (Akaike Information Criterion)와 HQC (Hannan-Quinn Criterion)가 있다. <표 2>의 결과를 보면, SC와 HQC를 기준으로 한 적정시차는 각기 1과 2로서 비슷하게 나왔으나, AIC는 10의 시차를 허용하였다. Lütkepohl and Breitung (1997, p. 315)이 지적한 바와 같이, AIC는 점근적 (asymptotic)으로 시차를 과대 추정하는 경향이 있는 반면 SC는 최소한도로 시차를 규정하여 필요한도 내에 가장 최소의 파라미터를 추정도록 해주기 때문에 본 연구에서는 SC 기준을 따르기로 한다.

<표 3>은 각 지역의 베이시스간의 인과성 검정결과를 보여준다. 비록 VAR 모형은 각 변수들간의 동태적 관계를 보여 주지만, 추정된 다수의 파라미터에

11) RATS 5.01을 사용하였으며, 사용된 프로그램과 데이터는 요청시 저자로부터 구할 수 있다.

〈표 1〉 ADF 검정 결과

시계열	FCMD	FSMD	FWMD	FUES	FLES
ADF 검정값	-8.7308(2)	-7.1503(4)	-10.1981(1)	-8.5153(3)	-12.7370(1)

주: 임계치는 -2.86(유의수준 5%), -2.57(유의수준 10%)임.

( ) 안은 Schwartz Criterion에 의해 선정된 시차임.

〈표 2〉 적정시차의 결정

시 차	SC	AIC	HQC
0	27.1010	27.0433	27.0767
1	24.6501*	24.3039	24.5040
2	24.7571	24.1225	24.4893*
3	24.9535	24.0305	24.5640
⋮			
9	26.3028	23.6357	25.1780
10	26.5727	23.6119*	25.3242
11	26.8345	24.5794	25.4621

주: 각 통계량은  $\log|\Sigma u(m)| + (C/T)mK^2$ 에 의해 구할 수 있음.

여기서  $C$ 는 SC, AIC, HQC에서 각각  $\log T$ , 2,  $2\log T \log \log T$ 를 나타냄.

\* 각 기준에 의해 적정시차를 결정해 주는 임계치임.

대한 직관적인 해석이 용이하지 않기 때문에, 변수들간의 충격반응을 분석하는 것이 필요하다. 이에 사전적으로 인과순위를 전제하지 않도록 DAG를 이용하여 동시적 인과관계 구조를 파악하도록 한다. DAG 분석을 위한 TETRAD II의 투입자료로 위의 VAR(1) 분산-공분산행렬로부터 구한 상관계수행렬 (8)을 이용한다.

$$\hat{\Sigma}_\epsilon = \begin{bmatrix} 1.0000 & & & \\ 0.8576 & 1.0000 & & \\ 0.8834 & 0.8951 & 1.0000 & \\ 0.8612 & 0.9703 & 0.9233 & 1.0000 \\ 0.8351 & 0.8918 & 0.9242 & 0.9330 & 1.0000 \end{bmatrix} \quad (8)$$

〈표 3〉 VAR 추정결과

	FCMD	FSMD	FWMD	FUES	FLES
FCMD(1)	0.6490* (0.0680)	-0.0400 (0.0592)	0.0677 (0.0646)	-0.0002 (0.0591)	0.0048 (0.0646)
FSMD(1)	-0.0757 (0.1540)	0.4615* (0.1342)	0.1801 (0.1463)	-0.0343 (0.1339)	0.0572 (0.1463)
FWMD(1)	0.0763 (0.0710)	0.0101 (0.0619)	0.7694* (0.0675)	-0.0565 (0.0618)	-0.0630 (0.0675)
FUES(1)	-0.1939 (0.1850)	-0.3290** (0.1611)	-0.4339* (0.1757)	0.3302** (0.1609)	-0.1654 (0.1757)
FLES(1)	-0.1272 (0.1072)	0.1705** (0.0934)	-0.2345** (0.1018)	0.0655 (0.0932)	0.4689* (0.1018)
상수항	10.8805* (2.1548)	1.8647 (1.8778)	13.3521* (2.0466)	7.1873* (1.8739)	14.8205* (2.0472)
D-W 통계량	2.1373	2.0975	2.0739	2.1126	2.1042

주 : ( ) 안은 표준오차. \*와 \*\*는 각기 1%, 5%의 수준에서 통계적으로 유의함을 나타냄.

### 3. 그래프 분석

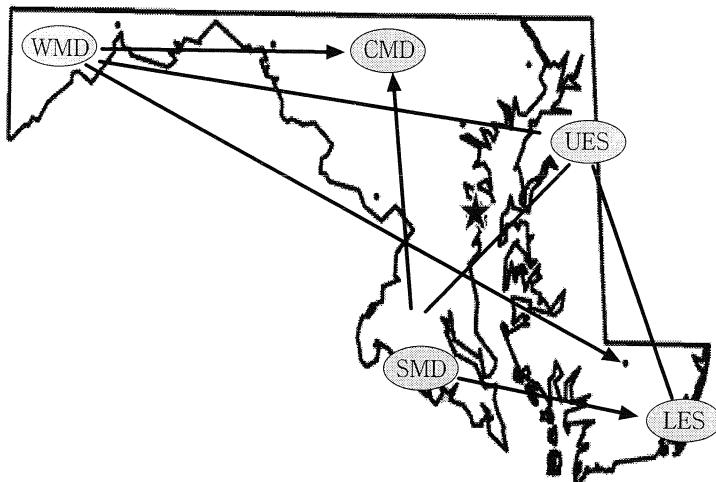
DAG의 PC 알고리즘은 먼저 방향지시가 하나도 되어 있지 않은 그래프 (completely undirected graph)부터 시작하여 상관계수를 기준으로 인과성이 없는 연결선을 하나씩 제거한다. ‘상관계수가 0이다’는 귀무가설을 기각하기 위해서는 적정 유의수준을 채택하여야 하는데, 본 연구의 분석결과 제거되는 연결선의 개수는 유의수준에 민감하지 않은 것으로 나타났다.<sup>12)</sup> 즉, 유의수준을 1%

12) 대부분의 경우 5% 및 10%의 유의수준을 채택한다. Spirtes, Glymour and Scheines (2000)는 몬테카를로 시뮬레이션을 통해, 표본의 수가 100보다 작을 경우에는 유의수준 20%, 표본수가 100 이상이면 5% 내지 10%가 바람직하다는 사실을 보여주었다.

〈표 4〉 TETRAD에 의한 연결선 제거

제거 연결선	상관계수	p-값
SMD-WMD	$\rho_{SMD, WMD \perp UES} = -0.0085$	0.8247
CMD-LES	$\rho_{CMD, LES \perp WMD, UES} = -0.0413$	0.2864
CMD-UES	$\rho_{CMD, UES \perp WMD, SMD} = -0.0307$	0.4277

〈그림 3〉 DAG 분석결과(10% 유의수준)



주: UES-WMD, UES-SMD, UES-LES 간에는 화살표의 방향이 유의하지 않아 무방향지  
시선(undirected graph)으로 표시하였음.

에서 10%까지 상향조정해도 제거되는 연결선은 CMD-LES, CMD-UES,  
SMD-WMD로 동일하였다.

구체적으로 살펴보면, 알고리즘의 첫 번째 탐색작업으로서 무조건부 상관계수  
를 이용하였을 때에는 어떠한 연결선도 제거할 수 없는 것으로 나타났다. 다음  
1계 조건부 상관계수를 보자. SMD-WMD의 상관계수는 UES를 조건부로 했  
을 때  $-0.0085$ 이지만, 그 p-값이 0.8247이기 때문에 통계적으로 유의하지 않았

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

다(<표 4> 참조). 이는 그래프이론에 관한 앞 절의 소개에서 보듯이 UES가 SMD와 WMD의 베이시스에 공통원인임을 의미한다. 다음 2계 조건부 상관계수를 이용하여, CMD-LES의 상관계수가 WMD, UES를 조건부로  $p$ -값이 0.2864이기 때문에 통계적으로 유의하지 않다고 판단된다. 마찬가지 방법으로 CMD-UES의 상관계수 역시 WMD, SMD의 2계 조건부로  $p$ -값이 0.4277이기 때문에 제거한다. 3계 조건부 상관계수에서 추가적으로 제거된 연결선은 없었다. <그림 3>은 DAG의 최종 분석결과로서 메릴랜드의 주 지도를 바탕으로 각 지역 베이시스의 동시적 인과성을 보여주고 있다.

<그림 3>의 DAG 결과를 보면, CMD는 WMD와 SMD의 정보흡수원 (information sink) 역할을 하며, LES는 WMD와 SMD의 정보채널 (information channel) 역할을 한다는 것을 알 수 있다. UES-WMD, UES-SMD, UES-LES에 대해서는 유의한 방향지시선을 이끌어낼 수는 없었지만 완전 독립적이지는 않기 때문에 화살표 없는 일직선으로 연결하였다.

## 4. 예측오차 분산분해 및 충격반응분석

그래프의 인과성이 동시적 인과성이라면 예측오차분해 및 충격반응은 단기 인과성(short term causality)을 위한 것이다. 위에서 얻은 인과관계를 사용하여 변수를 Bernanke 방식에 의해 순서화한 후, 예측오차 분산분해와 충격반응 분석을 시도하였다.

예측오차 분산분해는 동태적으로 어떤 변수가 타 변수와 비교하여 통계적으로 외생적 또는 내생적인가를 식별하는데 유용한 정보를 준다. 만일, 어떤 변수의 예측오차 분산의 상당부분이 그 변수 자체의 확률오차에 의해 설명되면 그 변수는 통계적으로 외생적인 것으로 간주된다. 극단적인 경우 완벽한 외생적인 변수는 예측오차 분산의 100%가 모든 시간대에 걸쳐서 그 변수에 의해 설명될 것이다. 예측오차 분해를 위해, 1개월에 해당하는 4주간의 창(window)을 허용하였다.

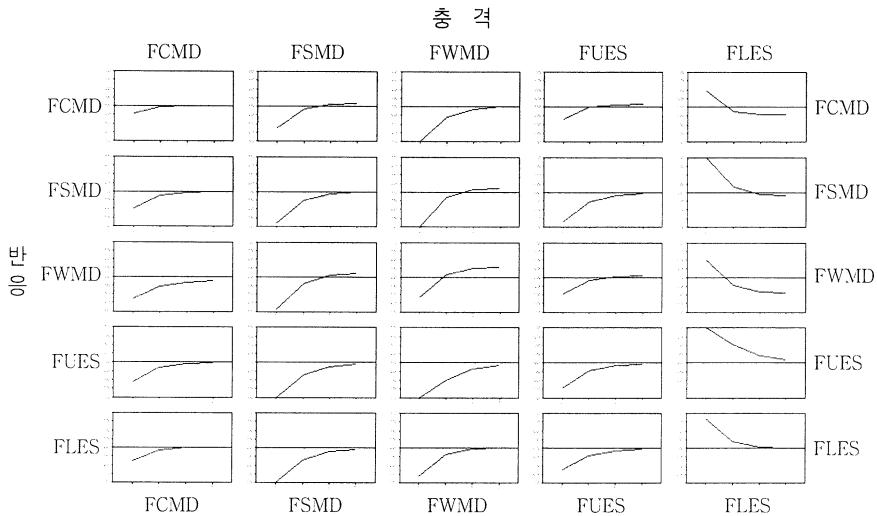
〈표 5〉 예측오차 분산분해

시 차		FCMD	FSMD	FWMD	FUES	FLES
FCMD	1	2.558	20.58	59.438	6.231	11.193
	2	2.427	19.793	60.616	5.879	11.284
	3	2.362	19.329	59.11	5.953	13.246
	4	2.304	19.102	57.633	6.064	14.897
FSMD	1	5.777	18.091	28.236	16.76	31.136
	2	5.834	18.637	27.480	17.629	30.420
	3	5.812	18.616	27.499	17.705	30.368
	4	5.790	18.546	27.601	17.658	30.405
FWMD	1	19.83	40.536	16.406	11.096	12.133
	2	21.715	38.784	15.359	10.619	13.523
	3	20.404	34.739	16.372	9.509	18.976
	4	18.716	31.488	17.505	8.612	23.679
FUES	1	4.534	17.546	32.713	7.499	37.707
	2	4.428	17.589	32.705	7.458	37.820
	3	4.393	17.580	32.708	7.439	37.880
	4	4.382	17.569	32.708	7.431	37.910
FLES	1	4.222	40.135	20.637	12.250	22.756
	2	4.090	41.012	20.135	12.564	22.198
	3	4.068	41.159	20.042	12.626	22.105
	4	4.071	41.174	20.028	12.637	22.090

〈표 5〉는 유의수준 10%에서 얻은 DAG 인과관계를 이용하여 Bernanke 순서화를 한 후 각 베이시스의 오차가 어떻게 분해되는지를 보여준다. 예를 들어, 〈표 5〉를 보면 FCMD 오차분해는 1의 시차를 두고 2.558%, 20.58%, 59.438%, 6.231%, 11.193%만큼이 각기 FCMD, FSMD, FWMD, FUES, FLES에 의해 설명된다. 아울러 타 지역의 베이시스 오차를 설명하는 데 있어서 SMD와 WMD의 영향이 평균적으로 높음을 알 수 있다. CMD의 베이시스

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

〈그림 4〉 충격반응 분석



오차에 대한 자기변수 설명력은 매우 낮아 타 지역의 베이시스에 민감한 정보의 흡수원임을 알 수 있는데, 이는 위의 DAG 결과와 합치하는 것으로 판단된다. 즉, CMD 베이시스 변동의 대부분이 자기변수 외의 타 변수로 설명되기 때문에 FCMD는 모형에서 가장 내생성이 높은 변수라고 판단된다. CMD의 높은 내생성은 CMD 지역에 Baltimore를 중심으로 한 대두 수출항이 위치해 있기 때문인 것으로 판단된다.

반면, FLES의 자기변수 설명력은 4주간에 걸쳐 평균 22%로서 외생성이 상대적으로 높은 것으로 나타났는데, 이는 정보채널로 LES가 역할한다는 DAG의 결과보다 훨씬 강한 것이다. LES가 상대적으로 높은 수준의 외생성을 보이는 것은 LES가 대두의 최다 생산지역으로서 타 지역의 가격형성에 미치는 영향이 크기 때문인 것으로 판단된다.

〈그림 4〉의 충격반응 분석 결과는 위의 예측오차 분산분해 결과를 재확인해 준다. 그림의 첫 번째 열은 CMD 베이시스의 확률충격 한 단위 변화가 각 지역의 베이시스에 미치는 영향을 나타낸다. 모형의 외생적 충격에 대하여 대부분

의 변수가 매우 빠른 속도로 조정되어 2주 내에 안정세에 접어들었다. 그럼에서 한 가지 특기할 사항은 LES를 제외한 지역의 베이시스 확률충격에 대해 타 지역은 (-)의 크기로 반응하나, LES 베이시스 확률충격에 대해서는 (+)로 반응 한다는 것이다. 이는 지역 내 대두의 최다생산지역인 LES 베이시스에 발생하는 확률충격이 타 지역의 현물가격을 상승시키는 역할을 하기 때문인 것으로 판단 되는데, 이에 관한 세부적인 이해를 위해서는 각 시장간의 미시적 구조(micro-structure)에 대한 추가 연구가 필요할 것이다.

#### IV. 결 론

다면량 시계열 분석은 그 유용성에도 불구하고 모형구조의 식별방법이 유일 하지 않은 관계로 연구자의 주관적 판단 또는 사전에 전제되어 있는 이론적 모형에 의존해야 한다는 제약이 있다. 이 과정에서 흔히 지적되는 변수의 임의나 열 문제를 해결하기 위한 한 방법으로 최근에 주목받고 있는 그래프이론을 소개하였다.

미국 메릴랜드의 대두 베이시스 자료를 이용한 VAR-DAG 분석 결과, LES는 타 지역의 베이시스 형성에 선도역할을 하는 외생성이 가장 큰 반면, CMD는 타 지역으로부터의 베이시스에 민감하게 영향을 받는 내생성이 가장 큰 지역인 것으로 나타났다. 이는 동시적 인과성을 보여주는 DAG에서 LES는 정보의 채널로, CMD는 정보의 흡수원이라는 결과와 일치하는데, 이러한 결과는 LES는 대두의 최다생산지역이며 CMD는 대두의 수출항이라는 데에 기인하는 것으로 판단된다.

본 연구에서 소개된 방법론은 VAR 모형을 이용하는 대부분의 시계열 분석에 응용될 수 있으며, 최근에는 농수산물 도소매 시장가격, 에너지의 원별 및 산지별 가격, 금융시계열 분석 등 다양한 분야에 적용된 연구결과들이 나오고

## VAR와 그래프이론을 이용한 시계열의 인과성 분석

있다. 물론 시계열간에 공적분이 존재할 경우 VAR 대신 오차수정모형을 사용하면 되고, 그 외의 전개과정은 본 연구에서 소개된 것과 동일하다. 또한 구조 VAR 모형에도 그래프이론을 적용하는 등 최근 다양한 방향으로 연구가 진행되고 있어 이에 대한 지속적인 관심과 주목이 요구된다.

### ◎ 참고 문 헌 ◎

1. 김명직 · 장국현, 『금융시계열 분석』, 경문사, 2002.
2. 윤병삼, “베이시스 추정치를 활용한 곡물 저장전략의 유효성에 대한 검증”, 「농업 경제연구」, 제43권 제2호, 2002.
3. Awokuse, T. and D. Bessler, “Vector Autoregressions, Policy Analysis, and Directed Acyclic Graphs: An Application to the U.S. Economy,” *Journal of Applied Economics*, Vol. 11, 2003, pp. 1~24.
4. Bernanke, B., “Alternative Explanations of the Money-Income Correlation,” *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25, 1986, pp. 49~100.
5. Bessler, D. and J. Yang, “The Structure of Interdependence of International Stock Markets,” *Journal of International Money and Finance*, Vol. 22, 2003, pp. 261~287.
6. Cooley, T. and M. Dwyer, “Business Cycle Analysis without Much Theory: A Look at Structural VARs,” *Journal of Econometrics*, Vol. 83, 1998, pp. 57~88.
7. Cooley, T. and S. LeRoy, “Atheoretical Macroeconomics: A Critique,” *Journal of Monetary Economics*, Vol. 16, 1985, pp. 283~308.
8. Haigh, M. and D. Bessler, “Causality and Price Discovery: An Application of Directed Acyclic Graphs,” *Journal of Business*, 2003.
9. Haigh, M., Park, H. and D. Bessler, “Revisited to Sound Matters: An Investigation of Sound Effect on Treasury 30 Bonds using Directed Acyclic Graph,” University of Maryland, 2003, mimeo.
10. Jeon, S., “The Proper Monetary Policy Stance During the Currency Crisis in East Asia:

- Tight or Expansionary?,” University of California at Davis, 2002, mimeo.
11. Lütkepohl, H. and J. Breitung, “Impulse Response Analysis of Vector Autoregressive Process,” in C. Heij, J. Schumacher, B. Hanzon and C. Praagman ed., *System Dynamics in Economic and Financial Models*, John Wiley & Sons, 1997.
  12. Pearl, J., “Causal Diagrams for Empirical Research,” *Biometrika*, Vol. 82, 1995, pp. 669~710.
  13. Pearl, J., *Causality: Models, Reasoning and Inference*, Cambridge Press, 2000.
  14. Pesaran, M. H. and Y. Shin, “Generalized Impulse Response Analysis in Linear Multiplicative Models,” *Economic Letters*, Vol. 58, 1997, pp. 17~29.
  15. Scheines, R., Spirtes, P., Glymour, C. and C. Meek, *TETRAD II: Tools for Causal Modeling*, 1994.
  16. Sims, C., “Are Forecasting Models Usable for Policy Analysis?,” *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 10, 1986, pp. 2~15.
  17. Spirtes, P., Glymour, C. and R. Scheines, *Causation, Prediction and Search*, Cambridge, MA: MIT Press, 2000.
  18. Swanson, N. and C. Granger, “Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions,” *Journal of the American Statistical Association*, Vol. 92, 1997, pp. 357~367.

---

Abstracts

---

---

Time-Series Causality Analysis using VAR and Graph Theory:  
The Case of U.S. Soybean Markets

---

Hojeong Park · Won-Cheol Yun

The purpose of this paper is to introduce time-series causality analysis by combining time-series technique with graph theory. Vector autoregressive (VAR) models can provide reasonable interpretation only when the contemporaneous variables stand in a well-defined causal order. We show that how graph theory can be applied to search for the causal structure in VAR analysis. Using Maryland crop cash prices and CBOT futures price data, we estimate a VAR model with directed acyclic graph analysis. This expands our understanding the degree of interconnectivity between the employed time-series variables.

Keywords : Graph Theory, VAR, Basis

## Abstracts

---

# VAR와 그래프이론을 이용한 시계열의 인과성 분석 — 미국 대두 가격 사례분석 —

---

박호정 · 윤원철

VAR(벡터자기회귀)에서 모형의 식별가정에 관한 주된 비판은 변수의 나열순서에 따라 결과가 달라진다는 것이다. 본 논문은 Swanson and Granger (1997) 이후 시계열 분석에 활발히 적용되기 시작한 그래프이론이 이와 같은 임의식별 문제를 해결함으로써, 자원가격의 가격발현과정을 이해하는데 유용한 수단임을 보여준다. 모형의 이론적 방법론을 소개한 후, 미국 대두의 지역 베이시스를 이용한 실증추정 결과를 제시한다.

주제어 : 그래프이론, 벡터자기회귀, 베이시스