Journal of the Korea Convergence Society
Vol. 10. No. 2, pp. 217-225, 2019

ISSN 2233-4890
https://doi.org/10.15207/JKCS.2019.10.2.217

연관규칙을 이용한 잠재성장모형의 개선방법론

조영빈[1], 전재훈[2], 최병우[1*]
[1]건국대학교 국제비즈니스학부 경영학전공 교수
[2]건국대학교 ICT융합공학부 의학공학전공 교수

A Methodology for Improving fitness of the Latent Growth Modeling using Association Rule Mining

Yeong Bin Cho[1], Jae-Hoon Jun[2], Byungwoo Choi[1*]
[1]Department of Business Administration, Division of International Business, Konkuk Univ.
[2]Department of Biomedical Engineering, College of Biomedical and Health Science, Konkuk University

요 약 대표적인 종단자료 분석방법인 잠재성장모형(Latent Growth Modeling)은 무조건적 모형과 조건적 모형으로 구분한다. 잠재성장모형의 무조건적 모형 성장궤적은 선형으로 가정하여 분석하는 경우가 많다. 본 연구는 선형 성장궤적으로 가정하여 모형 적합도가 미달하는 경우 연관규칙기법을 이용하여 모형 적합도를 제고하는 방법론을 제안한다. 방법론은 연관규칙 마이닝의 순차패턴(Sequential Pattern)을 사용한다. 이를 위하여 종단자료를 분위별로 나누고, 각 분위에 속한 종단자료의 기간 변화를 산출한 뒤 이를 순차 패턴 화하였다. SPSS AMOS를 이용하여 한국고용정보원의 2001년부터 6년간 조사한 청년 패널 자료로 효과성을 검증하였다. 기존 단순선형함수를 가정할 때와 비교하여 모형 적합도가 상승하는 것을 확인할 수 있었다.

주제어 : 종단 자료 분석, 비선형 잠재성장모형, 무조건적 모형, 조건적 모형, 연관규칙, 순차패턴

Abstract The Latent Growth Modeling(LGM) is known as the typical analysis method of longitudinal data and it could be classified into unconditional model and conditional model. It is common to assume that the growth trajectory of unconditional model of LGM is linear. In the case of quasi-linear, the methodology for improving the model fitness using Sequential Pattern of Association Rule Mining is suggested. To do this, we divide longitudinal data into quintiles and extract periodic changes of the longitudinal data in each quintiles and make sequential pattern based on this periodic changes. To evaluate the effectiveness, the LGM module in SPSS AMOS was used and the dataset of the Youth Panel from 2001 to 2006 of Korea Employment Information Service. Our methodology was able to increase the fitness of the model compared to the simple linear growth trajectory.

Key Words : Longitudinal Data Analysis, Nonlinear Latent Growth Modeling, Unconditional Model, Conditional Model, Association Rule, Sequential Pattern

1. 서론

수많은 데이터가 축적되고 확대되면서 빅데이터 시대가 오고 있다. 빅 데이터의 주류를 이루는 비정형 데이터 뿐만 아니라 정형화된 데이터도 계속 확대 저장되고 있다. 데이터가 축적되면서 정형 데이터와 비정형 데이터가 사용자 중심으로 통합되는 경향도 보이고 있다. 예를 들어 아마존이 개장한 무인상점의 과금 과정은 정형 데이터와 비정형데이터의 사용자 중심 데이터통합을 보여주고 있다[1]. 거기에다 기업이 보유하고 있는 POS 데이

*Corresponding Author : Byungwoo Choi (neocon@kku.ac.kr)
Received November 4, 2018 Revised January 16, 2019
Accepted February 20, 2019 Published February 28, 2019

터, 과금 데이터가 비정형 데이터와 사용자 중심으로 통합하는 데이터 통합작업이 이루어지면, 횡단 자료뿐만 아니라 종단 자료의 양도 급증할 것이다. 아직 경영학 분야에서 정기적인 조사를 통한 종단자료는 그리 많지 않으며 재무, 마케팅, 인적자원관리 분야를 중심으로 종단 연구가 이루어지고 있다[2-6]. 반면 교육, 의료, 노동 등의 분야에서는 정기적인 조사를 통하여 종단자료가 수집되어 공개되고 있다. 청소년 패널자료, 교육종단자료, 노동패널자료, 아동패널자료 등이 대표적이며. 해외에서는 교육학 분야를 중심으로 다양한 항목을 가진 종단자료가 수집되고 있다[7].

대표적인 종단자료 분석방법은 잠재성장모형(Latent Growth Modeling: LGM)이다. 잠재성장모형은 먼저 변수의 종단적인 변화추이를 분석한다. 다음으로는 무조건적 모형(Unconditional Model)은 종속변수와 관련된 모형으로서 모형 적합도가 높은 초기 값과 기울기를 찾고, 독립변인을 포함한 조건적 모형(Conditional Latent Curve Model)을 만들어 시간의 변화를 반영한 변수를 반영하여 종단 인과관계 검증을 하게 된다[8].

무조건적 모형을 추정할 때 통상적으로 기울기는 선형(linear)을 가정한다. 이는 LGM이 성장궤적 함수형태를 사전에 정하는 모형기반 방법이기 때문이다. 그렇지만 대부분의 경우 종속변수의 변화추이가 완벽하게 선형이 되기는 어렵다. 예를 들어 4기 종단자료의 경우 1기의 기울기를 0이라 한다면 2기의 기울기가 1, 3는 2, 4기가 3인 경우를 말한다. 다시 말하여 기간별 기울기는 동일한 차분 값을 가지기 어렵다. 특히 종단자료의 경우 패널의 성장궤적이 상이하기 때문에 측정시점별로 다른 기울기를 가지게 된다.

본 연구는 잠재성장모형 성장궤적의 기울기를 추정하는 데 있어서 선형성장궤적을 가정하지 않고, 종단자료의 특성에 따라 기울기를 추정하는 방법을 제안한다. 이를 위하여 연관 규칙 마이닝 중 순차패턴(Sequential Pattern)을 이용하여 모형 적합도를 제고하는 방법론을 제시한다. 비선형 잠재성장모형의 성장궤적의 기간별 기울기를 추정하기 위하여, 종단자료를 4, 5, 10분위로 각각 나누고, 각 분위에 속한 종단자료의 기간 변화를 산출하였다. 다음으로 종단 자료별 기간 변화 자료를 연관규칙 기법을 이용하여 순차 패턴 화하였다. 기존 연구[9]는 추정한 기울기가 선형 성장궤적보다 모형 적합도를 증진시키기는 했으나, 모형 적합도가 너무 낮아서 실제 연구에 제안한 방법론을 적용하기는 어려웠다. 본 연구에서는 대규모 자료를 기반으로 좀 더 일반화된 방법론을 제안하였다. 또한 잠재성장모형의 조건적 모형도 구축하여 분석 완결성을 높였다.

본 연구에서 제시한 방법론은 한국고용정보원의 2001년부터 2006년까지의 청년 패널 데이터를 사용하여 방법론의 효과성을 검증하였다. 사용한 데이터의 샘플 수는 2,172개 이었다. 제시한 방법론을 적용한 결과, 잠재성장모형의 모형 적합도가 상승함을 확인할 수 있었고 모형 적합도 값도 일반적인 분석에 사용할 수 있는 수준이 되었다.

2. 관련 연구

2.1 종단자료와 종단연구(Longitudinal Data and Study)

종단자료(Longitudinal Data)란 연구 기간 동안 여러 번 수집한 자료를 지칭한다. 반면 횡단자료(Cross-sectional Data)는 전체 연구 기간 동안 한번 만 수집한 자료를 말한다. 따라서 연구 기간 동안 여러 번 횡단자료가 수집된다면 종단자료라 할 수 있다.

종단연구(Longitudinal Study)는 종단자료를 기반으로 하는 연구방법이다. 그렇지만 횡단연구를 연속적으로 하면 종단연구와 동일한 결과를 얻을 수도 있다. 종단 연구 설계가 아니더라도 종단자료를 수집할 수 있다. 그렇지만 수집한 횡단자료는 결측 치와 이상치를 조심스럽게 전처리(Pre-processing)하여야 한다. 종단자료를 얻을 수 있는 종단 연구 설계는 다음 7가지로 정리할 수 있다[11]. ① 동시적 횡단연구(Simultaneous Cross- sectional Study): 동일 횡단연구를 여러 연구대상에게 동시에 실시하는 연구방법을 지칭한다. 예를 들어 동일한 설문지로 여러 연령층의 연구대상자에게 동시에 조사하는 경우가 해당한다. 이렇게 수집한 자료로 시간흐름에 따른 변화를 분석하기는 어렵다. 왜냐하면 연령이 핵심 변수이지만 연령의 영향은 복합적이기 때문이다. ② 경향 분석(Trend Study): 반복적인 횡단연구(Repeated Cross-sectional Study) 라고 생각할 수 있다. 동일 모집단의 상이한 샘플에 대하여 각각 두 번이상의 횡단분석을 할 경우를 지칭한다. 개인차원(individual level)의 변화보다는 집단차원(aggregate level)의 변화에 관심이 있는 경우에

적합한 연구방법이다. ③ 시계열 연구(Time-series Analysis): 동일집단 연구대상에 대한 반복적인 측정으로 정의할 수 있다. 개입연구(Intervention Study)나 패널 연구(Panel Study)는 시계열연구의 변형으로 볼 수 있다. 통상적으로 작은 연구범위와 매우 많은 샘플과 적은 변수 수로 구성되어 있다. ④ 개입연구(Intervention Study): 사전통제-사후통제 연구(pretest-posttest control study)라 할 수 있으며, 실험 군과 대조군으로 구분하여 참여자를 배정하고 특정한 개입을 가한 후 두 그룹사이의 차이를 분석하는 연구방법이다. 실험군과 대조군에 대하여 복수의 측정을 하는 경우에 해당한다. ⑤ 패널 연구(Panel Study): 참여자 중 특정 집단을 대상으로 동일한 설문지를 사용하여 시간 흐름에 따라 반복적으로 조사하는 방법이다. 개인 수준에서 시간흐름에 따른 변화에 관한 정보를 획득할 수 있다. 동일 집단에 같은 설문을 반복하기 때문에 다른 집단에 설문을 하는 반복 횡단 연구에 비해서는 비용이 적게 소요된다. ⑥ 회상 연구(Retrospective Study): 회상 설문을 기반으로 한 연구이며, 대부분의 범죄조사과정에서 사용된다. 그렇지만 회상한 자료의 신뢰성과 정확도를 검증해야하는데, 일반적으로 응답의 50%는 부정확한 것으로 알려져 있다. ⑦ 코호트 연구(Cohort Study): 패널 연구와 유사한 연구방법으로, 동일 기간, 동일 사건을 경험한 개인들의 집단을 대상으로 한다. 특정 연도에 태어난 개인을 대상으로 하는 탄생 코호트(Birth Cohort)가 대표적인 예이다. 일반적으로 종단 연구에 사용되는 연구 방법은 패널 연구나 코호트 연구이며, 대부분의 종단 자료도 패널이나 코호트를 기반으로 하고 있다. 종단 연구는 자료가 매우 많고, 수집 비용도 많이 든다는 단점이 있다[12]. 또한 대부분의 종단 연구는 시간흐름에 따른 종속 변수의 변화에 대한 추정을 먼저 하고 독립 변수를 추정한다. 시간흐름에 의한 변수들의 변화량과 개인의 변화(Within-individual Change)를 추정한다[10].

2.2 잠재성장모형(Latent Growth Modeling)

잠재성장모형은 1987년 McArdle과 Esptein[13]가 처음으로 제안한 종단자료 분석방법이다. 또한 잠재성장모형은 구조방정식 모형 (Structural Equa- tional Modeling : SEM)의 특수 형으로 볼 수도 있다[2, 14, 15]. 구조방정식모형과 같이 잠재성장모형도 잠재변인과 관측변인은 중요하다[14]. 잠재변인과 관측변인은 잠재성장모형에서도 동일하게 적용된다. 잠재성장모형의 무조건적 모형에서 잠재변인은 기울기(Slope)와 초기 값(Initial Status)에 해당하고, 관측변인은 매회에 측정된 종단자료들로 정의한다.

잠재성장모형은 다음과 같이 표현할 수 있다.

$$y_{it} = \beta_{0i} + \beta_{1i} x_t + \varepsilon_{it} \quad (1)$$

수식 (1)에서 y_{it}는 i 번째 패널 자료 중 t 주기에 측정된 종단 값을 의미하며, β_{0i}는 패널 i의 초기 값을 의미한다. β_{1i}는 i 번째 패널의 기울기 값을 나타내며, x_t는 t 주기의 기울기 계수를 나타낸다[16]. 수식 (1)은 2개의 하위 계층으로 분리되어 개별효과를 추정하는데 사용된다.

수식 (1)을 두 개의 공식으로 분리하면 다음과 같다.

$$\beta_{0i} = \beta_0 + \zeta_{0i} \quad (2)$$

$$\beta_{1i} = \beta_1 + \zeta_{1i} \quad (3)$$

수식 (2)에서 β_0는 전체 평균의 초기 값이고 ζ_{0i}는 전체 평균과 i 번째 패널 간의 차이가 된다. 마찬가지로 수식 (3)은 수식 (1)의 개별 패널의 기울기 값을 전체 평균의 기울기(β_1)와 전체평균과 개인의 차이(ζ_{1i})점수로 나타낸다. 수식은 전체 평균의 기울기와 개별 패널의 기울기는 격차가 발생할 수 있으며, 이러한 격차는 개별 패널의 특수성으로 설명할 수 있음을 나타낸다. 따라서 전체 평균의 기울기와 함께, 집단의 개별 패널의 기울기에 대한 분산이 유의한지 통계적으로 밝혀 전체 변화와 개별 패널의 변화를 동시에 추정할 수 있게 한다[16].

잠재성장모형은 비조건적 모형의 성장궤적 형태에 따라 선형과 비선형 잠재성장모형(Linear and Non-linear LGM)으로 분류할 수 있다[2]. 무조건적 모형의 성장 궤적을 추정에 사용할 수 있는 방법은 다항(Polynomial) 함수를 이용하는 것으로 독립변수인 시간 변수의 2차 항또는 3차 항, 지수함수 등을 모형에 포함하는 방식으로 비선형 성장궤적을 추정할 수 있다[2,17,18]. 기울기 추정에 가장 많이 사용하는 방법은 종속변수의 개략 추세에 따라 기울기를 추정하는 방법이다. 이는 시간의 흐름에 따라 변화하는 종단적 성장궤적을 찾는 일이다[12]. 최근 종단 자료 분석을 할 수 있는 통계 프로그램(예, AMOS, LISREL프로그램)이 있으며, 이러한 프로그램은 복잡한 통계적 수식보다는 도형을 도식화하여 모형을 설계하고, 결과를 쉽게 그리고 정확하게 해석할 수 있도록 되어 있

다[14]. 본 연구에서는 SPSS AMOS의 잠재성장모형 모듈을 바탕으로 잠재성장모형의 성장궤적을 추정하는 방법을 제안한다.

2.3 연관(순차패턴) 마이닝 (Association[Sequential Pattern] Mining)

연관 규칙 마이닝 기법은 사건 간 상호 연관성을 추정하는 기법이다. 연관규칙 마이닝은 'X사건이 일어나면 Y사건도 일어난다.' 는 결과를 산출하고 다음과 같이 표현한다.

$$\text{If X, Then Y, } (X \rightarrow Y) \quad (4)$$

연관규칙 추출에 사용하는 척도는 지지도(Support), 신뢰도(Confidence), 지지도×신뢰도 등이다. 연관규칙 알고리즘 중 가장 먼저 개발되었고, 빈번하게 사용되고 있는 것은 Apriori 알고리즘이다[19].

순차패턴 마이닝(Sequential Pattern)은 시간의 흐름에 따라 사건 간 연관성을 찾는 기법이다. 다시 말해서 연관규칙기법에 시간을 추가하여 시계열에 따른 패턴을 찾아 상호 연관성을 탐색하는 기법이라 할 수 있다. 기본적으로는 연관 규칙 마이닝과 유사하지만 거래 데이터베이스로부터 고객 선호 항목들의 순차 패턴을 추출하고 순차(Sequence)적인 데이터베이스를 생성한다. 거래 순서에 따라 각 거래의 상품 항목들의 연관성을 탐색하고 거래의 순차패턴을 파악하여 선행 상품을 검색한다. 즉 고객 구매 패턴을 분석하여 미래 구매 가능 상품 예측에 사용할 수 있다[20-22]. 순차패턴은 본 연구에서 잠재성장모형의 성장궤적추정에 사용한다.

3. 제안방법론

3.1 적용자료

본 연구의 종단자료는 한국고용정보원의 2001년부터 2006년까지 조사한 청년 패널 데이터이다. 청년 패널 조사는 15~29세 청년층의 직업선택 및 노동시장 이동 등을 조사한 종단 자료이다. 비록 최근 데이터는 아니지만 방법론의 효과성을 검증하기 위한 목적으로 사용하였다. 데이터는 분석목적에 맞게 발췌하여 종단자료를 구성하였다. SPSS AMOS 25와 STATA, SPSS Modeler를 사용하였다.

Fig. 1에서와 같이 샘플 수는 362명이며, 6년 동안 종단조사를 실시하였기 때문에 총 샘플은 2,172개이다. 2001년부터 2006년까지 6년간 임금과 임금에 영향을 미치는 독립변수 2개를 선정하였다. 독립변수는 성별과 대학졸업여부이다. 결측치와 이상치는 사전에 제거하거나 적당한 값으로 대체하는 전처리를 실시하였다.

pid: 1104, 2804, ..., 955400	n = 362
year: 2001, 2002, ..., 2006	T = 6
Delta(year) = 1 unit Span(year) = 6 periods (pid*year uniquely identifies each observation)	

Fig. 1. Longitudinal data characteristics

종속변수인 임금은 월 평균임금이며, 본 연구에서는 2005년 불변가격기준의 로그 치환치(lr_wage)를 사용하였다. Figure. 2.는 임금의 종단 추이를 나타낸다. 전체적으로는 우 상향 선형이지만, 미세하게 비선형임을 알 수 있다. 기울기는 2001년부터 2003년까지 완만한 상승세를 보이다가 2004년에 기울기가 줄어들었고, 2005년, 2006년에는 상승폭이 커지는 것을 발견할 수 있다. 매년 임금수준의 수직선은 95% 신뢰구간의 상한치와 하한치를 나타낸다. 표준편차는 연도별로 유사한 것을 알 수 있다.

3.2 무조건적 모형의 모형 적합성

Fig. 3은 SPSS AMOS 잠재성장모형 중 무조건적 모형 화면을 나타낸다. SPSS AMOS에서 무조건적 모형의 초기 값과 기울기는 사용자가 계수 값을 입력하는 방식이다. 사용자가 자료특성에 따라 적절한 초기 값과 기울기를 추정하면 SPSS AMOS는 기울기와 초기 값 분산의 통계적 유의성을 산출해준다.

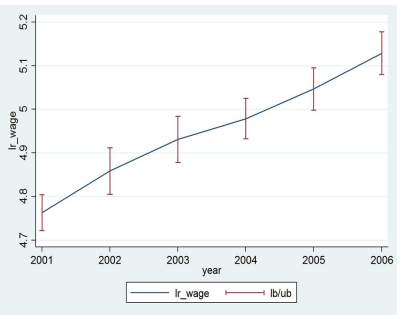

Fig. 2. longitudinal trajectory of 'lr_wage' variable

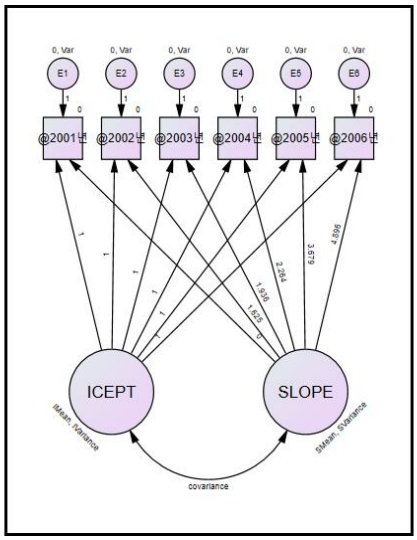

Fig. 3. Unconditional model of LGM in SPSS AMOS

Fig. 3의 사각형은 관측 변인을 의미하며 원은 잠재변인을 나타낸다. 초기 값과 기울기는 잠재변인이기 때문에 원으로 표시되었고, 측정 값은 관측변인이기 때문에 사각형으로 표시되었다. SPSS AMOS에서는 화살표로 표현된 경로계수에 적절한 값을 입력하도록 되어있다. 기울기(SLOPE) 경로계수 설정은 선형함수를 가정하여 0, 1, 2, 3, 4, 5, 6를 입력하였다. 모형 적합도는 Table 1.에 제시하였다. 잠재성장모형에서 모형 적합도는 CFI(Comparative Fit Index), NFI(Normed Fit Index), IFI (Incremental Fit Index), TLI (Tucker-Lewis Index)와 같은 적합도 지수와 상대적 적합도 지수인 χ^2 값을 사용한다. 절대적 적합도 지수의 최소 값은 0이고 최대 값은 1이다 . 통상적으로 0.9이상이면 적합도가 우수한 것으로 알려져 있다. 적합도의 크기가 클수록 좋은 것으로 알려져 있다. χ^2 적합도 지수는 작으면 작을수록 적합도가 우수하다고 평가할 수 있다[12]. 산출된 절대적 적합도 지수는 0.9를 근소하게 상회하고 있고 대부분 하회하고 있다. 따라서 선형모형으로 가정하여 조건적 모형 추정을 진행하기에는 부족한 것으로 드러났다.

Table 1. Model fitness of simple linear growth trajectory

	model fit-simple linear growth trajectory				
fit index	χ^2	CFI	NFI	IFI	TLI
value	101.9	.868	.838	.867	.906

3.3 제안방법론

연관 규칙 기법은 이전 t 기간 동안 $(X_{t-1}, X_{t-2}, \cdots\cdots, X_t)$의 순차 패턴을 갖게 되면 t+1기에 Y 값을 갖는다고 해석한다. 따라서 연관 규칙의 순차 패턴은 종단자료의 성장 궤적을 추정하는데 사용할 수 있다.

그렇지만 추출된 규칙 중 타당성이 높지 않은 순차 패턴을 사용하면 패턴의 신뢰성이 떨어질 것이다. 왜냐하면 그러한 순차 패턴은 특정 패널만의 패턴을 반영하고 있기 때문이다. 연관규칙 기법은 이를 위하여 지지도와 신뢰도를 사용하여 규칙의 신뢰성을 조절한다. 지지도와 신뢰도는 연구자가 결정할 수 있다[19].

SPSS Modeler의 연관규칙 모듈을 사용하여 연관 규칙을 도출하였다. lr_wage변수를 분위 수(fractile)로 전환하였다. 사용한 분위 수는 각각 4분위, 5분위, 10분위였다. 분위 수 전환은 SPSS Modeler의 기능을 사용하였다. 따라서 lr_wage변수는 각각 4, 5, 10개의 범주를 가진 3개의 변수로 치환되었고, 치환된 변수를 이용하여 연도별 순차패턴을 추출하였다.

Fig. 4 - Fig. 10은 산출된 순차패턴을 나타낸다. 지지도와 신뢰도는 지지도 10%이상, 신뢰도는 50%이상을 사용하였고, 실제 성장궤적 추정에는 지지도×신뢰도가 0.1이상인 패턴만을 사용하였다. 이는 지지도와 신뢰도를 각각 사용하는 경우보다 좀 더 엄격한 기준이다. Fig. 4의 4분위 2순차(2-sequence)중 첫 번째 패턴은 2001년에 1분위였던 패널이 2002년에도 1분위에 있는 경우를 나타내며, 이 패턴의 지지도는 20.99%, 신뢰도는 64.47%이고, 지지도×신뢰도는 0.1353임을 나타낸다. 또한 2001년과 2002년 사이에 산출된 패턴은 2001년 1분위에서 2002년 1분위로의 패턴이 1개, 2001년 4분위에서 2002년 4분위로의 패턴이 1개로 총 2개였다. 2순차 패턴으로 산출된 패턴의 수는 총 15개였다. 그림에 제시한 바와 같이 4분위에서 2, 3, 4, 5, 6순차가 산출되었고, 5분위에서는 2, 3, 4, 5 순차가 산출되었다. 산출된 순차패턴은 연도별 해당 분위 변화량에 대한 가중평균으로 사용한다.

식(5)와 같이 가중평균을 이용하여 순차별 기울기의 가중치를 산출하였다.

$$\sum_{i=1}^{T}\sum_{j=1}^{n} \triangle_j \times \{1 \pm (S_{ij} \times C_{ij})\} \qquad (5)$$

여기서 T: 분위 수, n: 연도, Δ_j: (상위 년도 분위 평균

- 하위 년도 분위 평균), S_{ij}: i번째 분위의 j번째 순차패턴의 지지도, C_{ij}: i번째 분위의 j번째 순차패턴의 신뢰도

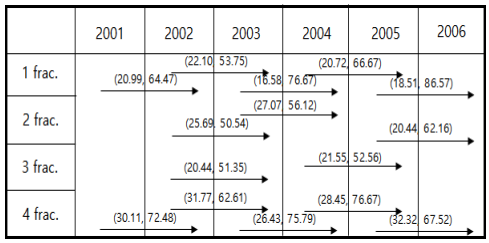

Fig. 4. 4-fractile 2-sequences association (sequential pattern) rules

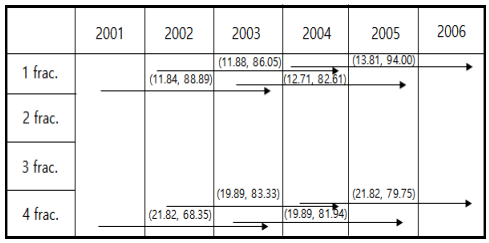

Fig. 5. 4-fractile 3-sequences association (sequential pattern) rules

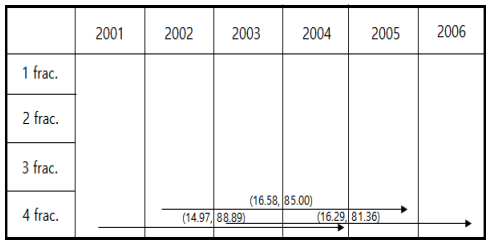

Fig. 6. 4-fractile 4-sequences association (sequential pattern) rules

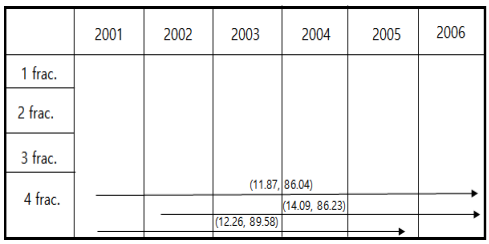

Fig. 7. 4-fractile 5-sequences and 6-sequences association (sequential pattern) rules

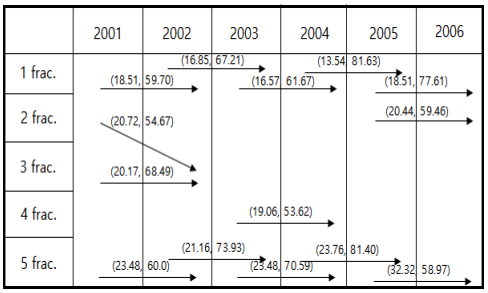

Fig. 8. 5-fractile 2-sequences association (sequential pattern) rules

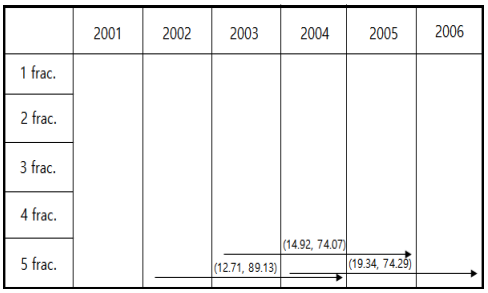

Fig. 9. 5-fractile 3-sequences association (sequential pattern) rules

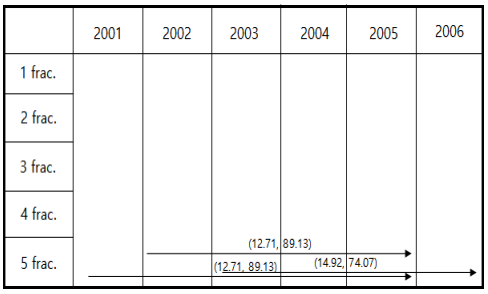

Fig. 10. 5-fractile 4-sequences association (sequential pattern) rules

순차패턴을 적용하지 않을 때는 성장궤도의 계수가 0, 1, 2와 같이 단순 선형함수가 되지만 식(5)를 적용하게 되면 순차패턴에 따라 가중치를 적용할 수 있다. 수식의 의미는 i번째 분위의 j번째 순차패턴이 일정 지지도×신뢰도를 상회할 경우 해당 순차패턴에 대응하는 연도의 분위 평균에 지지도×신뢰도만큼 가중치를 적용한다는 것이다. 지지도×신뢰도 값은 최소 0이고 최고 1이므로 해당 순차패턴의 최고 가중치는 2이고 최저 가중치는 0이 된다. 가중(+)하는 경우는 순차패턴이 연도별로 동일 분위이거나 상위 분위일 경우이고, 감중(-)하는 경우는 순차패턴이 연도별로 하위 분위일 경우가 된다. Fig. 9에서와

같이 5분위 2순차에서 감중(-) 경우가 있었다. 이렇게 순차패턴을 적용하면 선형 성장궤도가 아닌 비선형 성장궤도를 산출할 수 있다. 다시 말해 연도별 기울기가 달라지면서 종단 궤적에 더 근접하는 결과를 얻을 수 있다.

추출된 8개의 패턴 모형에 대한 적합도와 단순선형모형의 적합 도를 Table 2에 제시하였다. 분석 결과 8개 패턴 모형 모두 단순선형모형보다 적합도가 좋은 것으로 나타났으며, 가장 적합도가 높은 순차패턴은 4분위 3순차패턴으로 나타났다. 특히 절대적 모형적합도가 0.9에 근접하였다. 4분위 3순차 패턴 모형에 대한 기울기와 초기 값의 평균과 분산에 대한 추정치와 표준오차는 Table 3에 제시하였다. 초기 값과 기울기 값의 분산은 통계적으로 유의하였다($p<0.01$).

먼저 초기 값의 추정된 평균은 4.797로 나타났다. 초기 값이 4.797라는 의미는 환산하면 월 평균 급여가 1,204,217원이며 분산 환산치는 10,790원이 된다.

Table 2. Model fitness of derived sequential patterns vs simple linear pattern

		model fitness					
fit index		χ^2	CFI	NFI	IFI	TLI	
simple linear		101.9	.868	.838	.867	.906	BM
4-frac.	2seq.	94.58	.878	.848	.877	.913	
	3seq.	91.30	.884	.857	.883	.917	**
	4seq.	94.20	.879	.848	.878	.914	
	5seq.	94.40	.879	.848	.877	.913	
	6seq.	94.05	.879	.848	.878	.914	
5-frac.	2seq.	92.44	.882	.851	.881	.916	
	3seq.	92.56	.882	.851	.881	.916	
	4seq.	91.97	.852	.852	.882	.916	

Table 3.Rresult of initial status and slope

		4-fractile 3-sequence nonlinear model			
		mean		variance	
		estimate	SD	estimate	SD
lr_wage	icept	4.797*	0.02	0.076*	0.10
	slope	0.068*	0.05	0.000*	0.001

*: $p<0.01$

3.4 조건적 모형 분석

무조건적 모형 분석에서 초기 값과 기울기의 분산은 통계적으로 유의하였다. 다음으로는 개인 임금상승의 차이를 설명할 수 있는 독립변인이 무엇인지를 찾아내야 한다. 무조건적 모형과는 달리 조건적 모형에서는 독립변인들과 잠재변인(기울기와 초기 값)과의 관계를 검증하는 것이다. Fig. 12. 에는 조건적 모형을 제시하고 있다. 조건적 모형분석을 위하여 사용한 독립변인은 성별(gender)과 대학졸업여부(college)이다. 두 개 변인 모두 0와 1만 갖는 이산형 변수 이며, 시간에 따라 변화하지 않는 변수로 정의하였다. 잠재성장모형에서 독립변인은 잠재변인이 아닌 관측변인이기 때문에 사각형으로 모형에 포함하였다. 또한 기존 무조건적 모형에서의 기울기와 초기 값과의 연결은 화살표로 표시하였다. 새로운 독립변인이 추가된 후 모형 적합도를 Table 3.에 제시하였다. 조건적 모형의 χ^2값은 107.18로 무조건적 모형의 91.30보다 악화되었다. 이는 추가된 독립변인이 모형 적합도를 저하시키는 요인이라는 것을 나타낸다. 이러한 현상은 다른 모형 적합도 지수에도 공통적으로 나타났다.

이는 독립변인 성별(gender)과 대학졸업여부(college)가 개인 임금상승의 격차를 설명할 수 있는 독립변인이 아니라는 의미이다. 다시 말하여 남녀 간, 학력 간 임금수준 격차는 존재하지만 개인차원에서는 남녀, 학력에 상관없이 임금상승이 일어나고 있다고 설명할 수 있다. 일단 임금노동자가 되면 성별과 학력보다는 다른 변인에 의하여 개인별 임금격차가 벌어지고 있었다.

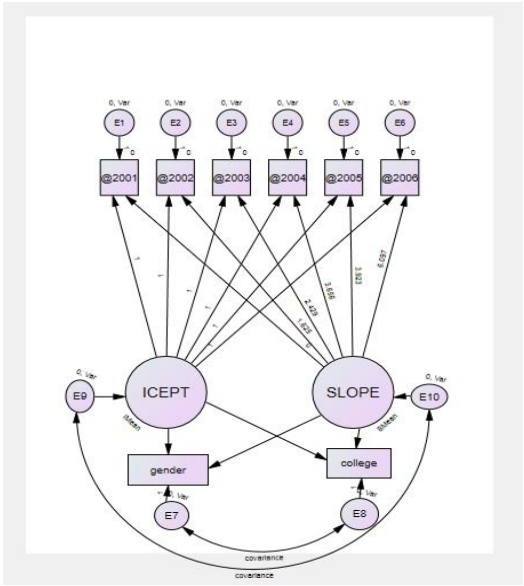

Fig. 12. Conditional Model of LGM

Table 4. Model fitness of Conditional Model

	model fitness(conditional model) -4 frac. 3seq.				
fit index	χ^2	CFI	NFI	IFI	TLI
value	107.18	.893	.854	.893	.907

4. 결론 및 시사점

아마존의 무인 상점사례와 같이 데이터 측적이 계속된다면[1], 마케팅이나 IS분야에서 종단자료 축적을 가속화할 것이다.

본 연구의 의의는 단순 선형 성장궤적으로 모형 적합도가 미달할 경우 모형 적합도를 제고할 수 있는 방법론을 제시하였다는 데 있다. 기존 연구[13-16]에서는 성장궤적 기울기의 함수형태를 가정하고 기울기 계수를 함수에 따라 추정하는 방법을 사용하는 것이 일반적이었다. 그래서 기존 연구에서도 선형성 검증을 통하여 무조건 모형의 성장궤적을 추정하였다[6]. 그렇지만 단순 선형 성장궤적의 모형 적합도가 미달할 경우, 이를 조정할 방법이 없었다. 이런 경우 모형 적합도를 제고할 수 있는 방법을 제시하였다.

두 번째 도출된 순차패턴의 가중치 계산과정에서 가중(+)과 감중(-)되는 경우를 구분하여 가중치 부여의 논리적인 일관성을 제고하였다. 이는 성장궤적이 증가하는 경우 뿐만 아니라 감소하는 경우에도 적용할 수 있는 방법을 제시하였다는 의의가 있다. 세 번째는 이전 연구[9]의 절대적 모형 적합도는 0.37~0.58인데 비하여 본 연구는 0.87~0.92로 상승하여 방법론의 타당성을 높였다는 점이다. 모형 적합도가 상승하게 된 것은 샘플 수증가, 종속변수 통제, 방법론 적용 정교화가 동시에 영향을 미친 것으로 판단된다.

본 연구의 한계는 조건적 모형 분석과정에서 사용한 독립변인이 너무 적고, 시간 고정 형(time-invariant) 변수로만 한정하였다는 점에 있다. 다양한 독립변인을 사용하여 방법론의 효과성을 검증해볼 필요가 있다. 또한 기울기는 분야에 따라 다르게 적용될 수 있음을 고려해야 한다는 것이다. 예를 들어 최저임금의 급격한 상승 등은 종속변인의 시간적 변화로는 설명할 수 없어 선형성장궤적으로 설명할 수 없는데, 방법론에는 이를 반영할 수 없다. 또한 연관규칙의 지지도와 신뢰도 적용을 사용자가 임의로 정한 면이 있다. 최적 수준에 대한 추가연구가 필요하다.

추가 연구로는 선형성(linearity) 검증에 대한 방법론 개선이 필요하다. 본 연구에서는 선형 궤적을 가정하고 모형 적합도에 따라 가정의 타당성을 검증하였지만 선형성을 통계적으로 검증하면 방법론의 효율성을 제고할 수 있다. 또한 선형 성장궤적이 아닌 2차 함수, 지수 함수에도 제안한 방법론이 사용될 수 있는지 추가 연구가 필요하다. 이를 위해서 다양한 종단자료 확보 및 적용이 필요할 것이다.

REFERENCES

[1] Hankyoreh(2018.1.22.) http://www.hani.co.kr/arti/international/america/828794.html

[2] E. J. Lee & C. H. Cho. (2013). A Longitudinal Study on the Effects of Franchise's Factors and Performance - Disclosure Agreement, Korean J. of Business Administration, 26(8), 2185-2209.

[3] E. J. Lee & C. H. Cho. (2014). A Longitudinal Study on the Service Quality in Korean Service Industry: Focusing on KS-SQI, Journal of Korea Service Management Society, 15(2), 23 - 47.
DOI : https://doi.org/10.15706/jksms.2014.15.2.002

[4] H. J. Lim & J. S. Cho. (2012). The Effect of Ownership Concentration on Firm Performance : Static and Dynamic Panel Data Analysis, Korean J. of Business Administration, 25(8), 3265-3291.

[5] J. H. Kim. (2018). A longitudinal study of the relationships between commitment type HRM, work team autonomy and innovation performance. The Korean Journal of Human Resource Development Quarterly, 20(2), 1 - 24.
DOI : https://doi.org/10.18211/kjhrdq.2018.20.2.001

[6] Y. B. Cho, S. K. Lee & K. H. Ro. (2015). A Methodology for Analyzing the Longitudinal Data using SOM Technique, Korean J. of Business Administration, 28(1), 93-102.

[7] J. S. Lee & S. Y. Kim. (2017). An Exploration of Nonlinear Latent Growth Model Using Exponential Function: As an Alternative to Quadratic LGM, J. of Educational Evaluation, 30(4), 791-816.

[8] K. S. Kim. (2009). AMOS and LISREL, Han Academy.

[9] Y. B. Cho. (2018). A Data Based Methodology for Estimating the Unconditional Model of the Latent Growth Modeling, J. Digital Convergence, 16(6), 85-93.

[10] S. W. Menard. (2002). Longitudinal research (2nd. ed.). London: Sage Publications Inc.

[11] Toon Taris. (1999). A Primer in Longitudinal Data Analysis, SAGE Publications Inc.
DOI : http://dx.doi.org/10.4135/9781849208512.n1

[12] S. S. Yeo & S. H. Park. (2012). An Appliation of Latent Growth Modeling: Use of Curriculum-Based Measurement as longitudinal Data. Asian J. of

Education, 13(4), 247-273.
DOI :https://doi.org/10.15753/aje.2012.13.4.011

[13] K. L. McArdle & D. B. Epstein. (1987). Latent Growth curves within development structural equation models. Child Development, 58, 110-133.
DOI : https://doi.org/10.2307/1130295

[14] B. M. Byrne. (2016). Structural Equation Modeling With AMOS Basic Concepts, Applications, and Programming, Third Edition. New York; Routledge,
DOI : https://doi.org/10.4324/9781315757421

[15] R. B. Kline. (2004). Principles and practice of structural equation modeling. New York: Guilford.

[16] K. A. Bollen & P. J. Curran. (2006). Latent curve models: a structural equation perspective. Hoboken, NJ: Wiley-Interscience.

[17] Annie Britton, Yoav Ben-Shlomo, Michaela Benzeval, Diana Kuh & Steven Bell. (2015). Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. BMC Medicine, 13(1), 47.
DOI : https://doi.org/10.1186/s12916-015-0273-z

[18] James A. Cranford, Patrick E. Shrout, Masumi Iida, Eshkol Rafaeli, Tiffany Yip & Niall Bolger. (2006). A Procedure for Evaluating Sensitivity to Within-Person Change: Can Mood Measures in Diary Studies Detect Change Reliably? Personality and Social Psychology Bulletin, 32(7), 917-929.
DOI :https://doi.org/10.1177/0146167206287721

[19] R Agrawal, T. Imielinski & A. Swami. (1993). Mining association rules between sets of items in large databases, Proceedings of the ACM SIGMOD Conference on Management of Data, 207-216.
DOI : https://doi.org/10.1145/170036.170072

[20] B. W. Jin, Y. S. Cho & K. H. Ryu. (2010). Personalized e-Commerce Recommendation System using RFM method and Association Rules. J. of the Korea Society of Computer and Information, 15(12), 227-235.
DOI : https://doi.org/10.9708/jksci.2010.15.12.227

[21] J. C. Kim, H. I Jung, H. Yoo & K. Y. Chung. (2018). Sequence Mining based Manufacturing Process using Decision Model in Cognitive Factory. Journal of the Korea Convergence Society, 9(3), 53-59.

[22] Y. J. Shin & M. S. Yim. (2012). A Study of the Relationship Analysis between Mobile Application by Using An Association Rules. Journal of the Korea Convergence Society, 3(2), 19-25.

[23] C. G Park & K. E Lee. (2014). A linearity test statistic in a simple linear regression. Journal of the Korean Data and Information Science Society, 25(2), 305-315.
DOI : 10.7465/jkdi.2014.25.2.305

조 영 빈(Cho, Yeong Bin) [정회원]

- 1985년 고려대학교 산업공학과 (공학사)
- 1988년 한국과학기술원 산업공학과(공학석사)
- 2005년 한국과학기술원 경영대학 경영공학(경영정보학 박사)
- 2006년 3월 ~ 현재 : 건국대학교 글로컬캠퍼스 국제비즈니스학부 경영학과 교수
- 관심분야 : CRM, 데이터마이닝, 온라인고객
- E-Mail : ybcho111@kku.ac.kr

전 재 훈(Jun Yae-Hoon) [정회원]

- 1986년 고려대학교 화학공학과 졸업.
- 1993년 MS, Chemical Eng., Texas A&M University, USA
- 2001년 Ph.D., Biomedical Eng., Texas A&M University, USA
- 2001년 ~ 2004년 Research Associate, Biomedical Eng., VCU(MCV campus)., USA.
- 2004년 ~ 현재 건국대학교 의학공학부
- E-Mail: jjun81@kku.ac.kr

최 병 우(Choi, Byungwoo) [정회원]

- 1986년 서울대학교 수학교육과 (이학사)
- 1998년 울산대학교 경영학과 (경영학 석사)
- 2002년 인하대학교 경영학과 (경영학 박사)
- 2004년 3월 ~ 현재 : 건국대학교 글로컬캠퍼스 국제비즈니스학부 경영학과 교수
- 관심분야 : 인사조직관리, OCB
- E-Mail : neocon@kku.ac.kr