

A study on patent evaluation model based on Bayesian approach of the structural equation model

Ho-young Woo^a · Jungae Kwak^b · Changwon Lim^{a,1}

^aDepartment of Applied Statistics, Chung-Ang University; ^bKorea Invention Promotion Association

(Received September 5, 2017; Revised October 5, 2017; Accepted November 6, 2017)

Abstract

Recently, the industrial paradigm shift to the fourth industry has already begun, and the importance of patents as intangible intellectual property in the fourth industry era is increasing day by day. Since the technical valuation of a patent is calculated according to the opinion of experts, it is costly and time consuming, and hence, the quality of the patent is judged based on subjective opinions of non-experts. Therefore, it is necessary to develop an objective and rational evaluation system for the qualitative level of patents. In this paper, we classify the valuation of patents into technicality, rights, and usability, and consider the quantitative and objective evaluation modeling of patents using Bayesian structural equation model. In particular, based on the data collected by the Korea Invention Promotion Association, we apply the Bayesian approach, which is capable of stable modeling even under small samples by using prior information, and the structural equation model, which is excellent for modeling and evaluating qualitative performance that is difficult to measure directly, to develop a patent evaluation model.

Keywords: evaluation of patent, structural equation models, Bayesian

1. 서론

지식기반 및 글로벌 경제 사회에서는 토지나 자본 등 유형 자산 보다는 특허, 상표, 디자인, 저작물 등과 같은 무형 지식재산이 국가나 기업 경쟁력의 핵심요소로 등장하고 있다. 그 가운데 특허는 무형자산의 대표적인 지적재산으로 개인과 기업 및 국가의 기술 수준과 혁신 역량을 측정하는데 있어 객관적인 척도로 중요하게 활용되고 있다 (Yoo와 Jung, 2010). 특허권은 새로운 기술적 발명에 대하여 일정 기간 동안 그 발명의 독점적 실시권을 갖고 배타적 지배권을 행사할 수 있기 때문에 기업의 수익 창출과 안정적인 사업 운영을 위해 매우 중요하다 (Kim 등, 2011a). 특허는 출원 후 심사를 거쳐 등록이 되면 출원일로부터 20년까지 독점배타적 지위를 유지할 수 있지만 20년 동안 연차 유지료를 납부해야 한다. 그렇기 때문에 보유 특허가 많은 기업은 특허 유지료에 부담을 갖게 되어 일부 기업은 특허를 20년이 되기 전에 포기하게 된다. 이때 포기 대상 특허를 선정하기 위해 특허에 대한 평가가 선행된다.

특허와 같이 무형자산에 해당되는 기술 가치 평가는 시장접근법, 수익접근법, 원가접근법으로 구분된다. 시장접근법은 대상기술과 동일 또는 유사한 기술이 활성화시장에서 거래된 가치에 근거하여 비교 분석을 통하여 상대적인 가치를 산정하는 방법이며, 수익접근법은 대상기술의 경제적 수명기간 동안 기술사업

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: clim@cau.ac.kr

화로 인해 발생될 경제적 이익을 적정 할인율을 적용하여 현재가치로 환산하는 방법이다. 마지막으로 원가접근법은 대상기술을 개발하는데 투입된 비용을 기초로 기술의 가치를 산정하거나 대체의 경제 원리에 기초를 두고 동일한 경제적 효익을 가지고 있는 기술을 개발하거나 구입하는 원가를 추정하여 가치를 산정하는 방법이다 (Ministry of Commerce, Industry and Energy, 2014). 하지만 기술가치평가는 평가대상기술의 기술성, 권리성, 활용성으로 구분하여 해당 전문가가 기술과 시장 관점의 자료 조사와 분석 후 주관적인 평가 의견을 제시하고, 취합된 전문가 의견에 따라 최종 가치를 산정하기 때문에 상당히 오랜 시간과 많은 비용이 수반된다. 그렇기 때문에 기업들이 포기대상 특허를 선별하기 위해서 전문가를 통한 가치평가를 수행하기보다는 특허 담당자의 주관적인 판단에 따라 포기 대상 특허를 선별하는 경우가 많다. 이에 기업들은 전문가의 주관적인 판단을 최대한 배제하고, 특허의 질적 수준을 판단 할 수 있는 객관적인 평가 방법을 필요로 하고 있다.

국내의 특허평가 서비스는 System to Measure, Analyze and Rate patent Technology 3 (SMART3); Korea Institute of Patent Information (KIPI)-Patent Evaluation & Grading 2.1 (K-PEG2.1); Science & Technology Information Analysis for R&D (STAR)-VALUE; PATENTPIA; FOCUST 등이 존재한다. 그 중 SMART3는 특허명세서에서 독립항과 종속항이 기술되는 패턴을 인식하여 청구항에 대한 정보를 추출하고, 특허 정보의 정제를 통하여 추출된 실시권, 무효심판, 패밀리 정보 등을 활용하여 전문가가 직접 평가한 정보를 학습셋으로 적용하여 다중회귀분석모델로 기술분야별 특허 등급 평가 서비스를 제공한다 (Korea Invention Promotion Association, 2012). 또한, K-PEG2.1은 텍스트마이닝 기법에 기반하여 생존 가능 확률이 높은 특허를 가치가 높은 특허로 평가하는 방식이고 (Korea Institute of Patent Information, 2013), STAR-VALUE는 특허 기술의 현금 흐름과 시장 정보를 이용하여 평가하는 시스템이다 (Korea Institute of Science and Technology Information, 2014).

한편, 설문 조사 기반의 평가시스템 모형 개발에는 이론을 객관화하는 통계적 방법인 구조방정식이 많이 사용되어 왔다. 설문조사로 측정된 자료를 토대로 구조방정식을 이용하여 Shin 등 (2009)은 신도시 보행자 서비스 질에 대한 평가지표 개발에 대한 연구를 진행하였으며, Kim 등 (2004)도 구조방정식을 통하여 중소기업들의 정보기술화 성과 정도를 평가하는 지표 개발에 대해 연구하였다. 또한, Kim 등 (2011a)의 공공연구기관의 기술사업화 프로세스와 성과분석에 대한 연구, Ham 등 (2004)의 게임만족도 평가시스템 모형 개발에 대한 연구 등 다양한 분야에서 구조방정식을 이용한 평가모델에 대한 연구가 있었다. 최근에는 사전정보를 모형에 반영 할 수 있는 특징을 내포하므로 전문가들의 사전 정보를 모형에 통합 할 수 있는 베이지안 접근법을 구조방정식과 함께 활용하여 평가 모형을 개발하는 연구도 있었다. 예를 들어, Kim 등 (2016)은 베이지안 구조방정식을 이용하여 기후요인 분석에 대해 연구하였고, Fong과 Ho (2015)은 직무 탈진의 평가 도구에 대한 연구를 수행하였다.

본 연구에서는 베이지안 방법의 특징과 구조방정식의 특성을 함께 내포하는 구조방정식의 베이지안 접근법을 활용하고자 한다. 특히 델파이 전문가 설문조사를 통해 주요 특허의 질적인 평가 관점에서 특허 정보의 중요도에 대한 전문가의 의견을 취합하여 이 정보를 토대로, 특허의 평가에 있어서 특허의 정성적 가치를 나타내는 전문가 의견을 베이지안 구조방정식을 통하여 일관되고 객관적인 평가를 하는 특허 평가 모델링을 제시해 보고자 한다. 2장에서는 연구방법인 전통적인 구조방정식과 구조방정식의 베이지안 접근법을 살펴보고 3장에서는 실제 전문가들의 설문 자료를 토대로 평가 모델링의 결과를 확인한다. 마지막 장에서는 본 연구의 결과를 요약해 보고자 한다.

2. 연구방법

2.1. 구조방정식

통계적인 접근 방법의 가장 중요한 이슈는 알고자하는 불확실한 현상들의 참된 원리를 충분히 근사시

키거나 또는 이를 확률적으로 표현하는 모형을 구성하는 것이다. 알고자하는 현상들에 대한 어떤 측정값의 집합을 반응변수 또는 확률변수 Y 라고 하면 탐구에 대한 미지의 현상은 $f(Y|\theta)$ 인 확률 함수로 표현 할 수 있으며 θ 는 확률변수의 특성치인 모수 집합이다. 특히 어떤 연구자가 여러 종류의 변수를 관측하고 이 변수들을 사용하여 구조(construct)를 정의한 다음 이 구조들이 어떻게 서로 연관되어 있는가를 연구한다고 가정해보자. 이 때 연구자는 다양한 이론적 모형을 구상할 수 있고 각 이론적 모형의 적합성에 대한 수량화된(quantitative) 가설 검정을 실시하고, 관측된 자료에 대해 가장 적합하다고 판단되는 모형을 찾는 과정을 구조방정식(structural equation modeling; SEM)이라 한다 (Lee, 2007).

구조방정식은 직접 관찰되거나 측정되지 않는 변수로 관측변수들의 잠재 요인인 잠재변수와 직접 측정되는 관측변수로 나눌 수 있다. 외생(잠재)변수(exogenous variable)는 독립 잠재변수이며 다른 잠재변수들에 영향을 주는 잠재변수를 의미한다. 내생(잠재)변수(endogenous variable)는 종속변수이며 직접 혹은 간접적으로 영향을 받게 되는 변수이다. 오차변수(error variable)는 크게 두 가지로 구분할 수 있으며, 관측 변수들에서 발생하는 오차를 나타내는 측정오차(measurement error)와 잠재변수에 발생하는 오차인 구조오차(structural error)로 나누어 생각 할 수 있다. 또한 SEM은 회귀분석(regression)과 경로분석(path analysis; PA) 그리고 확증적 요인분석(confirmatory factor analysis; CFA)의 세 가지 분석기법이 결합된 분석 방법이다. 만약 SEM에서 내생잠재변수가 없다면 CFA와 같다고 할 수 있다. 이 구조방정식을 식으로 나타내면 다음과 같다:

$$\begin{aligned} \eta &= B\eta + \Gamma\xi + \zeta, \\ y &= \Lambda_y\eta + \epsilon, \\ x &= \Lambda_x\xi + \delta. \end{aligned}$$

첫 번째 식에서 내생잠재변수는 $(q_1 \times 1)$ 벡터인 η 로 나타내며 B 는 다른 내생잠재변수와의 선형관계를 나타내는 회귀계수이다. 그리고 $(q_2 \times 1)$ 벡터인 ξ 는 외생잠재변수를 나타내며 Γ 는 외생잠재변수와 η 와의 선형관계를 나타내는 회귀계수이다. 또한 ζ 는 내생잠재변수에 대한 구조오차를 나타낸다. 이 식은 내생잠재변수와 외생잠재변수의 구조는 서로 독립이라는 점을 내포하고 있다. 두 번째 식에서 y 는 내생잠재변수 η 에 대한 관측변수이다. Λ_y 는 내생잠재변수에 대한 요인적재값을 나타내며 ϵ 은 내생관측변수 y 에 대한 측정오차를 나타낸다. 내생관측변수 y 의 개수는 q_1 개이다. 세 번째 식에서 x 는 외생잠재변수 ξ 에 대한 관측변수이다. Λ_x 는 외생잠재변수에 대한 요인적재값을 나타내며 δ 는 외생관측변수 x 에 대한 측정오차를 나타낸다. 외생관측변수 x 의 개수는 q_2 개이다. 위의 세 가지 수식을 토대로 관측변수들에 대한 분산 공분산 행렬을 계산하고 이 계산된 분산-공분산 행렬을 적합공분산행렬(implied covariance matrix)이라고 한다:

$$\Sigma(\theta) = \begin{pmatrix} \Lambda_y\Pi(\Gamma\Phi\Gamma^{-1} + \Psi)\Pi^T\Lambda_y^T + \Theta_\epsilon & \Lambda_y\Pi\Gamma\Phi\Lambda_x^T + \Theta_{\epsilon\delta} \\ \Lambda_x\Pi^T\Gamma^T\Phi^T\Lambda_x^T + \Theta_{\epsilon\delta} & \Lambda_x\Phi\Lambda_x^T + \Theta_\delta \end{pmatrix},$$

$$\Pi = (I - B)^{-1}, E(\xi\xi^T) = \Phi, E(\zeta\zeta^T) = \Psi, E(\epsilon\epsilon^T) = \Theta_\epsilon, E(\delta\delta^T) = \Theta_\delta, E(\epsilon\delta^T) = \Theta_{\epsilon\delta}.$$

만약 우리가 가정한 모형이 매우 타당하다면 실제로 데이터로부터 계산한 표본공분산행렬은 적합공분산행렬과 일치해야 할 것이다. 즉, 적합공분산행렬을 $\Sigma(\theta)$ 라고 하고 표본공분산행렬을 S 라고 한다면 잔여행렬인 $|\Sigma(\theta) - S|$ 를 최소화해야 할 것이다. 일반적으로는 잔여 행렬에 대한 함수를 합치함수(discrepancy function)라고 하며 이 합치함수를 최소화하여 모수를 추정하는 것이 구조방정식 모수추정의 큰 틀이다 (Park, 2000). 최대우도추정법(maximum likelihood estimation; MLE), 비가중최소제곱추정법(unweighted least square; ULS), 일반화최소제곱추정법(generalized least square; GLS)

등이 대표적인 추정방법이다. 하지만 분산-공분산 행렬에 기반한 추정 방법은 수치적으로 이루어지므로 표본수가 너무 작은 경우 추정이 적합하게 되지 않게 되므로 적절한 표본수를 확보한 후에 사용해야 한다 (Loehlin, 1992).

2.2. 구조방정식의 베이지안 접근방법

구조방정식에 있어서 베이지안 접근방법의 가장 큰 매력은 사전정보에 대한 적용이 유연하고 쉽다는 것이다. 대부분의 연구자들은 현재의 연구 이전의 연구들에 대한 정보 혹은 전문가들의 지식들을 알고 있을 것이다. 그러나 공분산 기반 구조방정식에서는 이전 연구나 전문가들의 사전정보들을 모형에 반영할 수 없다. 베이지안 기반의 구조방정식은 사전정보가 존재한다면 사전정보를 사전분포로 설정하여 사전정보를 모형에 반영하고 사전정보가 없다고 하더라도 무정보적 사전분포(noninformative prior)를 가정함으로써 논리적으로 타당한 구조방정식을 모형화 할 수 있다. 또한 많은 표본을 필요로 하지 않는다. 기존의 방법에서는 최대우도 추정을 가정하기 위해 많은 표본을 필요로 하였다. 하지만 베이지안 방법은 난수 발생을 통한 표집법에 기초하여 추론하므로 점근적 근사 이론의 영향을 덜 받으며 적은 표본 수에서도 추정치가 견고한(robust) 결과를 보여준다. 그리고 잠재변수들과 추정해야 하는 모수들은 Markov chain Monte Carlo (MCMC)에서 생성된 사후분포를 통해 직접적인 추정이 가능하다. 즉 분포를 알게 되므로 추정된 값들의 가정된 분포의 타당성도 직접적으로 알 수 있게 되고 가정한 사전분포를 수정하여 보다 타당한 사후분포를 유도 할 수 있다.

구조방정식은 표본수가 너무 적으면 추정에 있어서 문제가 발생하게 된다. 따라서 적정 표본수에 대한 논의가 있어 왔으며 적정 표본수에 대한 이론은 크게 두가지로 분류되는데, 하나는 절대적인 표본 수 크기에 대한 이론이며 다른 하나는 표본 수와 추정모수 비율에 대한 이론이다. 절대적인 표본수는 Hutcheson과 Sofroniou (1999)가 제안한 표본수 150개 이상의 기준과 Kline (2011)이 주장한 100개에서 200개 사이의 기준 등 있으며, Bentler와 Chou (1987), Kline (2011) 등이 주장한 표본 수와 추정모수의 변수 비율이 각각 5:1, 10:1 이상이라는 기준이 존재한다. 하지만 분석을 진행하는 자료의 표본수는 30개이므로 일반적인 구조모형의 적합에 부적절하다고 할 수 있다. 반면에, 베이지안 접근 방법은 공분산 구조 접근방법과는 달리 MCMC 방법을 통하여 잠재변수와 모수들의 사후분포를 생성하므로 표본 수에 상관없이 정확한 추정이 가능하다 (Lee, 2007).

구조방정식에서 유의확률(p -value)로 가설검정을 하는 것은 매우 위험한 일이다. 왜냐하면 표본의 수가 증가 할수록 구조모형의 적합성을 판단하는 가설검정에서 이론을 반영하는 모형과 실제 관측된 모형이 같다는 귀무가설을 기각할 빈도가 높아지기 때문이다. 즉, 표본수가 많으면 가정한 모형의 설명력이 아무리 높더라도 검정통계량의 값이 너무 크게 산출 되므로 유의확률값이 너무 작게 나오게 된다. 이렇게 되면 가설의 모형은 타당함에도 불구하고 모형을 선택 할 수 없게 된다. 따라서 전통적인 구조방정식에서는 다양한 적합도 지수들로 모형의 적합도를 판단하고 있다. 실제로 다양한 적합도 지수에서 0.95 이상의 수치가 나오에도 불구하고 유의확률 값은 0.01 미만으로 산출 되는 경우가 많다. 또한 가설 검정에서 유의성 검정의 유의확률은 귀무가설에 반하는 증거의 척도라는 점이다. 즉, 모형이 타당한지 또는 모형이 적합한지를 지지하는 증거의 척도가 아니다. 유의성 검정의 결론은 귀무가설을 기각하는데에만 사용해야하지 귀무가설이 얼마나 타당한지를 말하는 증거로 사용해서는 안된다 (Lee, 2007). 마지막으로 이러한 유의확률이나 기타 적합도 지수들은 서로 다른 이론을 반영하는 구조 모형(예를 들어, 비-지분(non-nested)모형)과의 비교를 위해 사용할 수 없다. 그러나 구조 모형을 토대로 이론을 검증함에 있어서 대부분의 경우는 특정 이론이 다른 이론을 내포하는 모형 보다는 서로 다른 형태의 비-지분 모형들에 대한 검정을 필요로 한다. 하지만 베이지안 구조방정식에서는 기존의 구조방정식이 가지고 있는 모형에 대한 검정의 문제를 가지고 있지 않다. 베이지안 방법에서는 귀무가설과 대립가설을

각각 하나의 모형으로 간주하고 각 모형에 대한 비교를 하는 것으로 생각한다. 또한 단순한 가설 검증 뿐만 아니라 지분되어 있지 않은 모형들과의 비교에서도 논리적인 검증 방법을 제시한다. 바로 베이즈 인자(Bayes factor)이다. 이러한 베이즈 인자 중에서는 베이저안 정보 기준(Bayesian information criterion; BIC)과 같이 널리 알려진 지표들이 있으며 이러한 베이저안 지표들은 사후확률(posterior probability)에 대한 MCMC를 토대로 계산되어진다.

구조방정식의 베이저안 접근은 다음과 같이 나타낼 수 있다:

$$\begin{aligned} \mathbf{y}_i &= \mu + \Lambda\omega_i + \epsilon_i, \\ \eta_i &= \Pi\eta_i + \Gamma\xi_i + \delta_i \end{aligned}$$

\mathbf{y}_i 는 내생관측변수 y 와 외생관측변수 x 로 이루어진 $p \times 1$ 벡터이고, $\omega_i = (\eta_i^T, \xi_i^T)^T$ 는 $\xi_i(q_2 \times 1)$ 와 $\eta_i(q_1 \times 1)$ 로 이루어진 확률벡터이다($q_1 + q_2 = p$). 공분산 기반 구조방정식에서는 외생잠변수와 모형오차, 측정오차에 대한 다변량 정규분포를 가정한다. 이러한 가정을 토대로 베이저안 구조방정식은 잠재변수와 구조모형의 오차, 회귀계수 등 모든 변수를 확률변수로 취급한다. 외생잠변수는 $\xi_i \sim N(0, \Phi)$ 이고 구조모형의 오차는 $\delta_i \sim N(0, \Psi_\delta)$, 측정모형의 오차는 $\epsilon_i \sim N(0, \Psi_\epsilon)$ 이고 절편은 $\mu \sim N(\mu_0, \Sigma_0)$ 이다. 베이저안 구조방정식의 적합공분산 행렬 Σ_ω 은 다음과 같이 나타낼 수 있다:

$$\Sigma_\omega = \begin{pmatrix} \Pi^{-1}(\Gamma\Phi\Gamma^T + \Psi)\Pi^{-T} & \Pi^{-1}\Gamma\Phi \\ \Phi\Gamma^T\Pi^{-T} & \Phi \end{pmatrix}.$$

θ 를 $\Lambda, \Psi_\epsilon, \Psi_\delta, \Pi, \Gamma, \Phi$ 를 가진 미지의 모수들의 벡터라 하고 모형을 조건부분포로 나타내면 다음과 같다:

$$\begin{aligned} (\mathbf{y}_i|\omega_i, \theta) &\sim N(\mu + \lambda\omega_i, \Psi_\epsilon), \\ (\omega_i|\mathbf{y}_i, \theta) &\sim N(\Sigma^{*-1}\Lambda^T\Psi_\epsilon^{-1}\mathbf{y}_i, \Sigma^{*-1}), \end{aligned}$$

여기에서, $\Sigma^* = \Sigma_\omega^{-1} + \Lambda^T\Psi_\epsilon^{-1}\Lambda$ 이다. Ψ_ϵ^{-1} 의 k 번째 대각원소를 $\psi_{\epsilon k}^{-1}$, Ψ_δ 의 k 번째 대각원소를 $\psi_{\delta k}^{-1}$ 라고 하면 $\psi_{\epsilon k}^{-1}$ 과 $\psi_{\delta k}^{-1}$ 의 사전분포는 각각 $\psi_{\epsilon k}^{-1} \sim G(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$, $\psi_{\delta k}^{-1} \sim G(\alpha_{0\delta k}, \beta_{0\delta k})$ 로 감마분포(gamma distribution)이다. 관측변수의 절편(intercept) μ 의 분포는 정규분포를 따르며 $\mu \sim N(\mu_0, \Sigma_0)$ 이다. 외생 잠재변수의 분산-공분산 행렬의 역행렬은 차원이 q_2 인 위샤트 분포(Wishart distribution)이며 $\Phi^{-1} \sim W_{q_2}(R_0^{-1}, \rho_0)$ 이다. 또한 구조모형과 측정모형의 오차항에 대한 사전분포가 주어졌을 때 각 모형의 계수들의 조건부분포는 다음과 같이 나타낼 수 있다: 측정모형 계수의 k 번째 행을 Λ_k^T 라고 하고 구조모형의 계수 $\Lambda_\omega = (\Pi, \Gamma)$ 의 k 번째 행을 $\Lambda_{\omega k}^T$ 라고 하면 각각 $(\Lambda_k|\psi_{\epsilon k}) \sim N(\Lambda_{0k}, \psi_{\epsilon k}H_{0yk})$, $(\Lambda_{\omega k}|\psi_{\delta k}) \sim N(\Lambda_{0\omega k}, \psi_{\delta k}H_{0\omega k})$ 로 나타낼 수 있다.

연구자가 가진 사전정보를 사전분포가 가진 초모수(hyperparameter)에 부여 할 수 있으며 사전정보에 대한 신뢰성이 높다면 초모수를 사전정보의 값과 최대한 가깝게 설정하기 위해 위치모수(location parameter)에 해당하는 초모수들의 값을 사전정보에 가깝게 부여 하고 척도모수(scale parameter)에 대한 초모수는 작게 설정해야 한다. 반면에, 사전정보의 신뢰성이 떨어지거나 알 수 없는 경우에는 위치모수는 0에 가깝게 척도모수에 대한 초모수는 최대한 크게 설정하여 사전분포를 무정보적 사전분포 가정 할 수 있다 (Congdon, 2001). 이렇게 구조모형을 계층적인 확률모형으로 표현할 수 있으며 깃스표집(Gibbs sampling) 방법을 통해 잠재변수와 모수들을 생성하여 각각의 사후분포를 토대로 추론을 진행할 수 있다.

2.3. 베이저안 구조방정식의 모형 적합도 지수

베이저안 구조방정식에도 모형을 평가하기 위한 다양한 적합도 지수가 개발 되어 있으며 이 적합도 지

수는 모형의 로그우도함수의 사후예측도(posterior predictive checks)를 제공한다. 이러한 사후예측도에는 BIC, 편차정보기준(deviance information criterion; DIC), (라플라스-근사) 로그 베이즈 인자((Laplace-approximated) log Bayes factor), 널리적용되는정보기준(widely applicable information criterion; WAIC) 등이 있다 (Merkle와 Rosseel, 2015). 잠재변수에 관측값들이 주어졌을 때의 모수의 사후분포와 잠재변수의 분포 하에서 생성된 우도함수를 사용하여 모수와 잠재변수를 반복적으로 추정한다. 잠재변수를 사전확률변수로 생각한다면 관측값들이 주어졌을 때 모수의 사후분포를 얻기 위해서는 잠재변수들을 적분해서 제거할 수 있다. 특히 이 사후분포는 모형을 평가하는 통계량으로 주로 계산된다. 구조방정식은 다변량정규분포를 가정하기 때문에 깃스 표집을 사용하여 쉽게 계산할 수 있다.

BIC는 Schwarz (1978)가 제안한 기준으로 사전분포에 의존하지 않는 로그 베이즈 인자의 간단한 근사라고 할 수 있다. 베이즈 인자는 $B_{10} = p(Y|M_1)/p(Y|M_0)$ 로 나타내며 각 모형의 사전정보와 관측값으로 계산된 사후 확률 간의 비(ratio)를 의미한다. 값이 클수록 첫 번째 모형을 더 강하게 지지한다는 의미이다 (Kass와 Raftery, 1995). 모형 k 에 대한 BIC는 다음 식과 같이 계산된다:

$$BIC_k = -2 \log p(Y|\theta_k^{ML}, M_k) + d_k^{\theta} \log n$$

여기에서, θ_k^{ML} 은 모형 k 의 모수 θ_k 에 대한 최우추정량이며 M_k 는 모형 k 를 의미한다. d_k^{θ} 는 모형 k 의 모수 θ 의 차원이며 n 은 자료의 크기이다. 또한 $BIC_{10} = 2[\log p(Y|\theta_1^{ML}, M_1) - \log p(Y|\theta_0^{ML}, M_0)] - (d_1^{\theta} - d_0^{\theta}) \log n$ 으로 나타낼 수 있으며 BIC_{10} 은 $2 \log B_k$ 에 근사한다. 따라서 BIC_{10} 가 0에 가까울수록 비교하고자하는 두 모형의 차이가 없음을 의미한다.

DIC는 사후분포 모수의 평균들에 대한 우도함수를 기반으로 산출되는 값인데 이 값에 모형의 효율적인 모수의 개수를 더해서 편차정보기준을 산출한다 (Spiegelhalter 등, 2002). 즉, DIC는 사후분포로부터 생성된 표본들의 로그우도함수들의 평균과 사후분포 모수의 평균들에 대한 우도함수의 차이를 나타내며, 이는 다음 식과 같이 계산된다:

$$DIC = -2 \log p(\theta^E|Y) + 2\mathbf{efp}_{DIC},$$

여기에서, θ^E 은 사후분포에서 생성된 모수들의 평균값들이고, \mathbf{efp}_{DIC} 은 모형의 효율적인 모수들의 개수를 의미한다. 이 기준은 절대값이 0에 가까우면 가까울수록 좋은 모형을 나타내지만 효율적인 모수의 개수가 증가하면 기하급수적으로 편차정보기준의 값이 커지므로 보편적으로 사용하는 적합도 지수는 아니다.

마지막으로, WAIC는 DIC보다 더 어려운 계산을 수반하는 반면에 DIC보다 더 장점을 많이 가진 적합도 지수이다. 최근에 복잡한 계산을 극복할 수 있는 방법들이 제시되었다 (Vehtari 등, 2015). WAIC는 점근적으로 편차정보기준과 동등한 적합도 측정법이며 관측된 자료로부터 계산한 로그우도함수와 효율적 모수의 개수를 사용하여 적합도를 계산한다. 특히, 예측의 정교성을 측정하는 모형의 로그점별예측 밀도함수의 기댓값(expected log pointwise predictive density)을 사용하여 계산한다. WAIC의 계산식은 다음과 같다:

$$WAIC = -2\text{lppd} + 2\mathbf{efp}_{WAIC},$$

여기에서, lppd 는 획득한 자료의 로그점별예측밀도함수이며, \mathbf{efp}_{WAIC} 은 효율적 모수의 개수이다. 이 기준은 절대값이 0에 가까울수록 좋은 모형을 나타낸다.

3. 실증 분석

특허에 대하여 정성적인 평가 기준을 산정하기 위하여 전문가들을 대상으로 델파이(Delphi) 기법을 활

Table 3.1. Classification code and evaluation items by technology field and evaluation index (Invention Promotion Society, 2017)

기술분야별 구분코드			1차 요인	평가항목
전기전자IT 화학바이오				
A	B			
평가지표별 구분코드				
권리성	기술성	활용성		
101	201	301	출원정보	독립항수가 많으면 해당하는 평가지표에 높게 영향을 준다.
102	202	302		독립청구항 길이가 짧으면 해당하는 평가지표에 높게 영향을 준다.
103	203	303		종속항 수가 많으면 해당하는 평가지표에 대한 영향도가 높아진다.
104	204	304		종속항의 평균길이 길이가 길수록 해당하는 평가지표에 높게 영향을 준다.
105	205	305		청구항계열 수가 많으면 해당하는 평가지표에 높게 영향을 준다.
106	206	306		도면수가 많으면 해당하는 평가지표에 높게 영향을 준다.
107	207	307		발명의 설명의 길이가 길수록 해당하는 평가지표에 높게 영향을 준다.
108	208	308		분할출원·우선권주장 특허가 많을수록 해당하는 평가지표에 높게 영향을 준다.
109	209	309		해외패밀리가 많을수록 해당하는 평가지표에 높게 영향을 준다.
110	210	310		IPC가 많이 부여될수록 해당하는 평가지표에 높게 영향을 준다.
111	211	311	조기공개를 했으면 해당하는 평가지표에 높게 영향을 준다.	
112	212	312	우선심사청구를 했으면 해당하는 평가지표에 높게 영향을 준다.	
113	213	313	재심사청구를 했으면 해당하는 평가지표에 높게 영향을 준다.	
114	214	314	의견서제출수가 많을수록 해당하는 평가지표에 높게 영향을 준다.	
115	215	315	정보제공이 많을수록 해당하는 평가지표에 높게 영향을 준다.	
116	216	316	총 피인용을 많이 받을수록 해당하는 평가지표에 높게 영향을 준다.	
117	217	317	피인용과 출원일 차이가 길수록 해당하는 평가지표에 높게 영향을 준다.	
118	218	318	선행문헌 중 논문·외국특허수가 많을수록 해당하는 평가지표에 높게 영향을 준다.	
119	219	319	피인용문헌 중 논문·외국특허수가 많을수록 해당하는 평가지표에 높게 영향을 준다.	
120	220	320	등록정보	연차등록을 오래 할수록 해당하는 평가지표에 높게 영향을 준다.
121	221	321	특허정보	발명자수가 많을수록 해당하는 평가지표에 높게 영향을 준다.
122	222	322		존속기간 연장등록이 될수록 해당하는 평가지표에 높게 영향을 준다.
123	223	323		실시권자수가 많을수록 해당하는 평가지표에 높게 영향을 준다.
124	224	324		권리자변동수가 많을수록 해당하는 평가지표에 높게 영향을 준다.
125	225	325		금융기관 질권설정이 많을수록 해당하는 평가지표에 높게 영향을 준다.
126	226	326		무효심판이 많을수록 해당하는 평가지표에 높게 영향을 준다.
127	227	327	거절결정불복심판이 많을수록 해당하는 평가지표에 높게 영향을 준다.	
128	228	328	심판정보	적극적 권리범위확인심판 인용이 많을수록 해당하는 평가지표에 높게 영향을 준다.
129	229	329		적극적 권리범위확인 심판 기각·취하·각하가 많을수록 해당하는 평가지표에 높게 영향을 준다.
130	230	330		소극적 권리범위확인심판 기각이 많을수록 해당하는 평가지표에 높게 영향을 준다.
131	231	331		소극적 권리범위확인심판 인용·취하·각하가 많을수록 해당하는 평가지표에 높게 영향을 준다.
132	232	332		정정심판이 많을수록 해당하는 평가지표에 높게 영향을 준다.

용하였다. 델파이 기법은 예측하고자 하는 문제에 대하여 전문가들에게 반복 질문하여 의견을 종합하는 하나의 연구방법이다 (Kang, 2008). 이를 토대로 Table 3.1에서 보여지듯이 전문가들이 중요하다고 생각되는 32가지의 평가 항목들에 대하여 상대적 중요도를 확인하고자 하였으며 우선적으로 전자전기IT, 화학바이오 두 가지 기술분야별로 구분하고 권리성, 기술성, 활용성의 3가지 잠재변수를 가정하였다. 32가지 평가 항목들은 델파이 기법을 통하여 5점 척도로 측정된 자료이다. 자료의 갯수는 분야별로 35개이며 이 자료를 토대로 특히 평가 모델링을 위하여 베이지안 구조방정식을 사용하였다.

3.1. 연구모형

특히 평가 지표 산정을 위한 모형은 출원정보, 심사정보, 등록정보, 특허정보, 심판정보 등 5개의 1차 잠재요인에 큰 틀의 평가 지표인 권리성, 기술성, 활용성이라는 잠재요인이 영향을 준다고 하는 2차 잠재요인 모형과 2차 잠재요인이 존재하지 않고 평가 지표의 잠재요인만 존재하는 단일 잠재요인 모형의 두 가지 모형에 대하여 베이지안 구조방정식에 적합하였다. 또한 전자전기IT, 화학바이오의 2가지 산업별로 평가 항목의 중요도가 다를 수 있기 때문에 기술분야별로도 각각의 평가 지표에 대하여 베이지안 구조방정식에 적합시켜 보았다. 적합된 모형의 베이지안 적합도 지수와 추정된 계수 값을 토대로 평가 항목별 점수를 산정해 보겠다. 2차 요인모형은 내생관측변수는 존재하나 외생관측변수는 존재하지 않는 구조모형이다. 2차 요인모형은 다음과 같이 나타낼 수 있다:

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\eta}_i &= \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \\ (\mathbf{y}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) &\sim N(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_\epsilon), \\ (\boldsymbol{\omega}_i|\mathbf{y}_i, \boldsymbol{\theta}) &\sim N\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i, \boldsymbol{\Sigma}^{*-1}\right). \end{aligned}$$

한편, 단일 잠재요인 모형은 다음과 같다:

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i, \\ (\boldsymbol{\xi}_i|\mathbf{y}_i, \boldsymbol{\theta}) &\sim N\left(\left(\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i, \left(\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}\right)^{-1}\right). \end{aligned}$$

2차 잠재요인 모형은 1차 잠재변수들에 대한 경로계수 중에서 등록정보의 경로계수를 0.02로 지정하였다. 이는 등록정보의 평가 항목이 하나만 존재하여 다른 평가 항목에 비하여 상대적으로 점수가 크게 배정 될 수 있기 때문이다. 또한 내생잠재변수들의 관측변수 중 첫 번째 평가 항목에 대하여 경로계수를 1로 설정하고 분석하였으며 단일 잠재요인 모형은 잠재변수의 척도를 부여하기 위해 첫 번째 관측변수의 경로계수를 1로 고정하고 모형에 적합하였다. 본 연구에서 사용한 사전분포는 Table 3.2와 Table 3.3에서 확인할 수 있다.

blavaan 패키지는 Muthen과 Asparouhov (2012)가 제안한 방법을 따르고 있으며 이 방법은 추정하고자 하는 공분산 행렬의 주변 분포가 역위샤프트 분포를 따르도록 하여 부적절 사전분포를 사용할 수 있게 한 방법이다 (Merkle와 Rosseel, 2015). 측정모형의 절편에 대한 사전분포는 $U(1, 5)$ 로 하였으며 이는 평가 항목에 대한 설문 점수가 1점에서 5점 사이에서 정의되기 때문이다. 또한 경로계수의 크기에 따라 평가지표를 배분 하므로 추정된 경로계수값이 0보다 작으면 평가지표 배분에 어려움이 있기에 경로계수에 대한 사전분포로 $U(0, 2)$ 를 사용하였다. 그러나 균등분포를 적용한 경우 수렴하지 않는 모형인 전자전기IT 분야의 권리성 단일모형과 화학바이오 분야의 권리성과 기술성 단일모형은 경로계수에 대한 사전분포로 $N(0, 0.01)$ 를 사용하였다.

Table 3.2. Prior distribution of electronic, electronics & IT field and PSRF index

평가지표	평가항목	요인명	2차요인모형				단일요인모형			
			1차요인		2차요인		사전분포	PSRF		
			사전분포	PSRF	사전분포	PSRF				
권리성	A101	출원정보	$U(0, 2)$	1.001						
	A102						$U(0, 2)$	1.001	$N(0, 0.01)$	1.000
	A108						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
	A109						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
	A115	심사정보	$U(0, 2)$	1.001			$N(0, 0.01)$	1.000		
	A116						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
	A119						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
	A120								$N(0, 0.01)$	1.000
	A122	특허권정보	$U(0, 2)$	1.000			$N(0, 0.01)$	1.000		
	A123						$U(0, 2)$	1.000	$N(0, 0.01)$	1.001
	A124						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
	A126								$N(0, 0.01)$	1.000
	A128	심판정보	$U(0, 2)$	1.001			$U(0, 2)$	1.001		
	A128						$U(0, 2)$	1.001	$N(0, 0.01)$	1.000
A130						$U(0, 2)$	1.001	$N(0, 0.01)$	1.001	
A132						$U(0, 2)$	1.000	$N(0, 0.01)$	1.000	
기술성	A201	출원정보	$U(0, 2)$	1.001						
	A203						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A205						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A208						$U(0, 2)$	1.001	$U(0, 2)$	1.001
	A209	심사정보	$U(0, 2)$	1.001			$U(0, 2)$	1.000		
	A215								$U(0, 2)$	1.000
	A217						$U(0, 2)$	1.002	$U(0, 2)$	1.001
	A218						$U(0, 2)$	1.001	$U(0, 2)$	1.000
	A219	등록정보	$U(0, 2)$	1.000			$U(0, 2)$	1.001		
	A220								$U(0, 2)$	1.000
	A222								$U(0, 2)$	1.001
	A223						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A224	심판정보	$U(0, 2)$	1.001			$U(0, 2)$	1.000		
	A226								$U(0, 2)$	1.001
A228						$U(0, 2)$	1.002	$U(0, 2)$	1.000	
A130						$U(0, 2)$	1.002	$U(0, 2)$	1.000	
활용성	A311	심사정보	$U(0, 2)$	1.002						
	A312						$U(0, 2)$	1.001	$U(0, 2)$	1.000
	A315						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A316						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A320	특허권정보	$U(0, 2)$	1.002			$U(0, 2)$	1.000		
	A322								$U(0, 2)$	1.001
	A323						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A324						$U(0, 2)$	1.000	$U(0, 2)$	1.001
	A325	심판정보	$U(0, 2)$	1.001			$U(0, 2)$	1.001		
	A326								$U(0, 2)$	1.001
	A328						$U(0, 2)$	1.000	$U(0, 2)$	1.000
	A329						$U(0, 2)$	1.004	$U(0, 2)$	1.000
A330		$U(0, 2)$	1.001			$U(0, 2)$	1.002			
A330						$U(0, 2)$	1.000	$U(0, 2)$	1.002	
A331						$U(0, 2)$	1.004	$U(0, 2)$	1.000	
A332						$U(0, 2)$	1.001	$U(0, 2)$	1.000	

PSRF = potential scale reduction factor.

Table 3.3. Prior distribution of chemistry & biology field and PSRF index

평가지표	평가항목	요인명	2차요인모형				단일요인모형	
			1차요인		2차요인		사전분포	PSRF
			사전분포	PSRF	사전분포	PSRF		
권리성	B101	출원정보	$U(0, 2)$	1.001	$U(0, 2)$	1.001	$N(0, 0.01)$	1.003
	B103							
	B105							
	B108							
	B109							
	B116	심사정보	$U(0, 2)$	1.001	$U(0, 2)$	1.000	$N(0, 0.01)$	1.001
	B120	등록정보						
	B122	특허권정보	$U(0, 2)$	1.002	$U(0, 2)$	1.000	$N(0, 0.01)$	1.001
	B123							
	B126							
	B128	심판정보	$U(0, 2)$	1.001	$U(0, 2)$	1.002	$N(0, 0.01)$	1.002
	B130							
	기술성	B201	출원정보	$U(0, 2)$	1.000	$U(0, 2)$	1.000	$N(0, 0.01)$
B208								
B209								
B215		심사정보	$U(0, 2)$	1.000	$U(0, 2)$	1.001	$N(0, 0.01)$	1.001
B216								
B220		등록정보	$U(0, 2)$	1.001	$U(0, 2)$	1.000	$N(0, 0.01)$	1.000
B222		특허권정보						
B226								
B228		심판정보	$U(0, 2)$	1.001	$U(0, 2)$	1.000	$N(0, 0.01)$	1.002
B130								
활용성	B305	출원정보	$U(0, 2)$	1.001	$U(0, 2)$	1.001	$U(0, 2)$	1.000
	B309							
	B312	심사정보	$U(0, 2)$	1.001	$U(0, 2)$	1.007	$U(0, 2)$	1.000
	B315							
	B320							
	B322	특허권정보	$U(0, 2)$	1.003	$U(0, 2)$	1.002	$U(0, 2)$	1.000
	B323							
	B325							
	B326							
	B328	심판정보	$U(0, 2)$	1.002	$U(0, 2)$	1.001	$U(0, 2)$	1.000
B330								
B332								
B332								

PSRF = potential scale reduction factor.

MCMC의 수렴성 검토를 위하여 Gelman과 Rubin (1992)이 제시한 potential scale reduction factor (PSRF)를 사용하였다. PSRF는 각 체인의 분산의 평균과 각 체인에서 생성된 모수의 평균들에 대한 분산으로 표현되는 수렴성 판단 지표이다. 생성된 모수의 PSRF가 모두 1에 가깝고 1.1보다 값이 작다면 모형은 수렴했다고 판단 할 수 있다 (Gelman 등, 2004). 각 모수에 대한 PSRF 지수는 Table 3.2와 Table 3.3에서 확인할 수 있다.

분석에 활용한 통계 소프트웨어 및 관련함수는 R version 3.3.0에서 blavaan 패키지 (Merkle와 Rosseel, 2015)를 사용하였으며 blavaan 패키지의 blavaan 함수를 사용하였다. burn-out 이후 표집 할 표본

Table 3.4. Goodness-of-fit index by model

기술분야별 구분	평가지표	모형	Num of para.	ln L	BIC	DIC	WAIC
전자전기	권리성	단일잠재요인모형	45	-374.891	908.469	827.694	835.832
		2차잠재요인모형	49	-377.798	928.388	824.739	829.658
	기술성	단일잠재요인모형	48	-293.418	756.101	664.514	684.123
		2차잠재요인모형	52	-286.934	757.240	642.977	655.555
	활용성	단일잠재요인모형	45	-396.209	951.105	868.460	895.593
		2차잠재요인모형	48	-388.719	946.703	843.658	862.226
화학바이오	권리성	단일잠재요인모형	36	-384.934	896.816	832.022	857.594
		2차잠재요인모형	40	-394.085	929.224	844.591	864.314
	기술성	단일잠재요인모형	33	-346.544	809.458	748.880	790.393
		2차잠재요인모형	37	-333.051	796.577	715.165	743.601
	활용성	단일잠재요인모형	35	-413.262	949.946	880.501	899.835
		2차잠재요인모형	40	-353.586	848.227	760.268	779.037

BIC = Bayesian information criterion; DIC = deviance information criterion; WAIC = widely applicable information criterion.

의 수를 10,000개로 하였으며 burn-in 되는 표본의 수는 2,000개로 하였다. 또한, blavaan 패키지는 JAGS(Just Another Gibbs Sampling) 기반으로 Gibbs sampling을 수행하므로 MCMC 초기 단계에서 MCMC의 수렴을 위해 사용되는 적응적(adaptive) 표집은 1,000개로 설정 하였다. 적응적 표집은 성공적인 burn-in을 보장해준다 (Plummer, 2015). 그리고 수렴성 진단을 위하여 사용한 사슬(chain)의 수는 3개로 하였다. 추정된 계수의 유의성은 로그베이지언자를 토대로 확인하였으며 로그베이지언자가 1보다 작은 경우 추정된 계수가 유의하지 않다고 판단하고 모형에서 제외하였다.

3.2. 분석 결과

분석 결과 Table 3.4에서 확인 할 수 있듯이, 2차 잠재요인 모형의 적합도 지수가 단일 잠재요인 모형의 적합도 지수보다 전반적으로 높았다. 하지만 일부 적합도 지수에서는 단일 잠재요인 모형의 적합도 지수가 더 높아서 2차 잠재요인모형 뿐만 아니라 단일 잠재요인모형도 고려하여 문항별 평가 점수를 산출하는 것이 바람직 할 것이다. 분야별로 살펴 보면 전자전기 분야에서는 기술성, 권리성, 활용성의 순서로 모형 적합도가 높으므로 기술성, 권리성이 상대적으로 활용성에 부여하는 평가 점수 보다 높아야 할 것이다. 또한 화학바이오 분야에서도 기술성, 권리성, 활용성의 순서로 모형 적합도가 높게 나타났으므로 기술성, 권리성, 활용성 순으로 평가 점수를 부여하는 것이 바람직 할 것이다.

추정된 잠재변수의 계수의 크기에 따라 각 평가 항목들에 배점을 할당하였다. 특정 평가 항목의 추정된 계수가 높다면 낮게 추정된 계수의 항목 보다 더 크게 배점하였으며 반대의 경우에는 적게 배점하였다. 2차 잠재요인 모형은 평가항목들이 유사한 평가 문항으로 묶어서 점수를 배정하고 해석하기 용이하였지만, 평가 배점이 2차 잠재요인에 할당된 배점을 통하여 세부 항목들의 점수가 배정되므로 특정 2차 잠재요인에 배점이 작게 되면 세부 항목들의 계수가 크더라도 작게 배점을 해야 하는 단점이 있었으므로 단일 잠재요인 모형의 평가 점수와 비교하여 특허 전문가들의 의견을 반영하여 적절하게 평가 항목에 점수를 배정해야 할 것이다.

Table 3.5에서 볼 수 있듯이, 전자전기IT 분야에서 권리성 지표는 모든 모형에서 ‘정보제공이 많을수록 해당 하는 권리성에 높게 영향을 준다’의 평가항목이 가장 높은 점수가 산정되었으며 2차 잠재요인 모형에서 등록정보에 해당하는 ‘연차 등록을 오래 할수록 권리성에 높게 영향을 준다’의 항목이 가장 낮은 배점을 할당 받았다. 이는 등록정보라는 2차 잠재요인이 다른 요인들에 비하여 잠재변수의 계수가 작

Table 3.5. Score by evaluation item for electronic, electrics & IT field

평가 지표	평가항목	1차요인	1차 요인별 점수	2차요인모형 평가항목별점수	단일요인모형 평가항목별점수		
권리성	A101	출원정보	20	6	4		
	A102			5	3		
	A108			5	4		
	A109			4	2		
	A115	심사정보	29	12	11		
	A116			9	10		
	A119			8	8		
	A120	등록정보	1	1	5		
	A122	특허권정보	26	10	8		
	A123			7	8		
	A124			9	7		
	A126			7	10		
	A128	심판정보	24	7	9		
	A130			8	10		
A132	2			2			
A201	기술성				6	6	
A203		6	6				
A205		출원정보	21		4	5	
A208		3	7				
A209		2	3				
A215		9	8				
A217		심사정보	23		5	6	
A218					5	5	
A219					4	3	
A220		등록정보	1		1	6	
A222	특허권정보	34	20	11			
A223			4	3			
A224			10	7			
A226			7	9			
A228	심판정보	21	7	8			
A130	7	8					
A311	활용성			17	6		
A312				심사정보	42	14	7
A315						8	6
A316						3	2
A320						등록정보	1
A322				7	8		
A323				특허권정보	24	4	6
A324						4	5
A325	9	10					
A326	5	6					
A328	2	4					
A329	심판정보	33	9	9			
A330			4	8			
A331			8	8			
A332			5	8			

Table 3.6. Score by evaluation item for chemistry & biology field

평가 지표	평가항목	1차요인	1차 요인별 점수	2차요인모형 평가항목별점수	단일요인모형 평가항목별점수
권리성	B101			5	5
	B103			5	9
	B105	출원정보	23	5	9
	B108			5	9
	B109			3	8
	B116	심사정보	25	25	7
	B120	등록정보	1	1	6
	B122	특허권정보	30	14	4
	B123			16	10
	B126			7	12
	B128	심판정보	21	7	11
	B130			7	10
	기술성	B201			12
B208		출원정보	26	4	7
B209				9	7
B215				7	12
B216		심사정보	25	9	15
B220		등록정보	1	1	6
B222		특허권정보	35	35	13
B226				4	8
B228		심판정보	13	4	6
B130				5	8
활용성	B305	출원정보	21	12	7
	B309			9	7
	B312	심사정보	28	16	10
	B315			12	11
	B320	등록정보	1	1	8
	B322			11	8
	B323	특허권정보	23	9	7
	B325			3	3
	B326			9	11
	B328	심판정보	27	5	10
	B330			4	6
	B332			9	12

게 추정되었기 때문이다. 하지만 단일 잠재요인모형은 2차 잠재요인을 포함하지 않기 때문에 ‘연차 등록을 오래 할수록 권리성에 높게 영향을 준다’의 평가 항목이 가장 작은 점수를 배점 받지 않았다. 전자전기IT 분야의 기술성 지표에서는 ‘존속기간 연장등록이 될수록 해당하는 평가지표에 높게 영향을 준다’의 항목이 각 모형에서 가장 높게 산출되었다. 활용성 지표에서는 각 모형마다 가장 높게 산출된 평가 항목이 달랐으며 2차 잠재요인 모형의 경우 ‘조기공개를 했으면 해당하는 평가지표에 높게 영향을 준다’의 평가 항목이, 단일 잠재요인 모형의 경우 ‘금융기관 질권 설정이 많을수록 해당하는 평가지표에 높게 영향을 준다’의 평가 항목이 높은 배점을 받았다. 따라서 모형의 적합도 지수와 항목 별 배점의 실제적 타당성을 전반적으로 고려하여 2차 잠재요인 모형 또는 단일 잠재요인 모형의 평가 점수를 선택

하는 것이 바람직하겠다.

Table 3.6에서 확인해보면, 화학바이오 분야에서 권리성 지표에 대한 모형은 심사정보에 해당하는 평가 항목이 하나만 유의하게 산출되어 2차 요인 모형의 경우 ‘총피인용을 많이 받을수록 해당하는 평가지표에 높게 영향을 준다’의 항목이 가장 높은 점수를 받았다. 단일요인 모형은 ‘무효심판이 많을수록 해당하는 평가지표에 높게 영향을 준다’의 평가 항목이 가장 높게 산정되었다. 기술성 지표에서도 특허권 정보에 속하는 평가 항목이 하나만 산출 되어 2차 요인 모형은 ‘존속 기간 연장등록이 될수록 해당하는 평가지표에 높게 영향을 준다’의 항목이, 단일잠재요인 모형은 ‘총피인용을 많이 받을수록 해당하는 평가지표에 높게 영향을 준다’의 항목이 각각 가장 높은 점수로 산출 되었다. 단일잠재모형의 경우 권리성과 기술성 모두에서 ‘총피인용을 많이 받을수록 해당하는 평가지표에 높게 영향을 준다’의 항목이 가장 높은 점수를 배정받았는데 이는 연구에 있어서 총피인용수를 중요하게 여기는 것과 비슷한 맥락이라고 볼 수 있다. 활용성 지표에서는 2차 요인 모형은 심사정보에 해당하는 평가 항목인 ‘우선 심사청구를 했으면 해당하는 평가지표에 높게 영향을 준다’가, 단일 모형에서는 심판정보에 해당하는 항목인 ‘정정심판이 많을수록 해당하는 평가지표에 높게 영향을 준다’가 각각 가장 높은 점수를 배정받았다. 그러므로 화학바이오 분야에서도 실무에서 고려하는 사항을 반영하여 2차 잠재 요인 모형 또는 단일 잠재 요인 모형 중 보다 타당한 모형의 배점을 사용하는 것이 합리적일 것이다.

4. 결론

정성적인 자료에 대하여 정량적으로 표현하고자하는 것은 통계적 추론의 목적이라고 할 수 있다. 특허의 가치는 많은 요인들에 의해서 설명 될 수 있으며 단순히 정량적으로 설명되는 것에는 한계가 존재할 것이다. 그럼에도 불구하고 측정될 수 없는 정성적인 요인들에 대하여 구조적인 관계를 확인하여 전문가들의 중요도를 반영할 수 있는 구조방정식은 합리적인 접근이라고 할 수 있다.

본 논문에서는 특허에 대한 평가 항목을 기술분야별로 구분하고 세 가지의 평가 지표 별로 구분하여 평가 항목 중에서 전문가들이 중요하다고 여겨지는 항목들에 대하여 더 높은 점수를 반영 할 수 있게 하였다. 5가지 1차 잠재 요인을 포함하는 모형과 1차 잠재요인이 없는 단일 잠재요인 모형을 각각 적합시켜 베이시안 적합도를 확인하고 추정된 계수를 토대로 평가 점수를 산정하였다. 적합도 지수 상에서는 2차 잠재요인 모형이 전반적으로 높았으나 2차 잠재요인 모형은 평가 항목을 하나만 포함하는 1차 잠재요인이 존재하기 때문에 그 평가 항목이 단일 잠재요인 모형에 비하여 상대적으로 작게 배정된다. 또한 단일 잠재요인 모형은 평가 항목들 중에서 두드러지는 항목이 없기 때문에 적합도 지수 뿐만 아니라 배정된 점수들에 대한 실무적인 관점이 필요하다.

특허에 대한 평가는 국가 발전을 위한 중요한 요소이며 연구자들 개개인을 위해서도 매우 중요한 요소이다. 그러므로 특허에 대한 평가를 좀 더 합리적이고 논리적으로 진행하기 위하여 평가 항목들에 대한 정량화에 대한 연구는 앞으로도 지속되어야 할 것이다.

References

- Bentler, P. M. and Chou, C. (1987). Practical issues in structural modeling, *Sociological Methods and Research*, **16**, 78–117.
- Congdon, P. (2001). *Bayesian Statistical Modelling*, Wiley, New York.
- Fong, T. C. and Ho, R. T. (2015). Dimensionality of the 9-item Utrecht Work Engagement Scale revisited: a Bayesian structural equation modeling approach, *Journal of Occupational Health*, **57**, 353–358.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed), CRC Press, Florida.

- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, **7**, 457–511.
- Ham, H.-B., Lee, Y.-S., and An C.-H. (2004). An assessment system model for game satisfaction degree to establish game development strategy, *Journal of Korea Multimedia Society*, **7**, 1630–1638.
- Hutcheson, G. D. and Sofroniou, N. (1999). *The Multivariate Social Scientist: An Introduction to Generalized Linear Models*, Sage Publications.
- Kang, Y. J. (2008). Understanding and application of Delphi technique, *Occasional Task Report*, 1–17.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.
- Kim, B.-K., Cho, H.-J., and Og, J.-Y. (2011a). A study on the technology commercialization process and performance of public research institutes in Korea using the structural equation model, *Journal of Korea Technology Innovation Society*, **14**, 552–577.
- Kim, C.-H., Park, S.-P., and Koh, Y.-H. (2011b). Intellectual property strategy, *Korea Invention Promotion Association*, 15–19.
- Kim, J. G., Moon, T. H., and Sohn, S. Y. (2004). Structural equation model for information technologization success index (ITSI) of small and medium enterprise, *IE Interfaces*, **17**, 338–348.
- Kim, M., Jeon, M., Sung, K.-I., and Kim, Y.-J. (2016). Bayesian structural equation modeling for analysis of climate effect on whole crop barley yield, *Korean Journal of Applied Statistics*, **29**, 331–344.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed), The Guilford Press, New York.
- Korea Institute of Patent Information (2013). *Qualitative Evaluation of Patent Utilizing K-PEG Evaluation System*, Korea Patent Information Center, Patent Information Promotion Center, Seoul, 1–8.
- Korea Institute of Science and Technology Information (2014). *Science & Technology Information Analysis for R&D-Value System*, Korea Institute of Science and Technology Information Technology Valuation LAB, Daejeon.
- Korea Invention Promotion Association (2012). System to Measure, Analyze and Rate patent Technology 3. URL https://smart.kipa.org/kipa2/valmodel/smart_process.do?flag=VALMODEL_INTRO
- Korea Invention Promotion Association (2017). Introduction of Patent Analysis Evaluation System Evaluation Model, URL <http://smart.kipa.org/>.
- Lee, S.-Y. (2007). *Structural Equation Modeling: A Bayesian Approach*, John Wiley & Sons, New York.
- Loehlin, J. C. (1992). *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis* (2nd ed), Lawrence Erlbaum Associates, New Jersey.
- Merkle, E. C. and Rosseel, Y. (2015). blavaan: Bayesian structural equation models via parameter expansion, *arXiv preprint arXiv:1511.05604*, Retrieved URL <https://arxiv.org/abs/1511.05604>.
- Ministry of Commerce, Industry and Energy (2014). *Technical Value Assessment Practical Guide*, 27–33.
- Muthen, B. and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory, *Psychological Methods*, **17**, 313–335.
- Park, K. B. (2000). *Multivariate Analysis*, Hakjisa, Seoul.
- Plummer, M. (2015). *JAGS Version 4.0.0 User Manual*.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **2**, 461–464.
- Shin, H.-M., Kim, T.-H., Jeong, K.-S., and Won, J.-M. (2009). A development of service quality evaluation models for pedestrians in newtown using structural equation modeling - a case of Bundang New Town -. *Seoul Studies*, **10**, 183–196.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.
- Vehtari, A., Gelman, A., and Gabry, J. (2015). Loo: Ecient leave-one-out cross-validation and WAIC for Bayesian models. URL <https://github.com/jgabry/loo>
- Yoo, J. B. and Jung, Y. M. (2010). Analysis of factors affecting patent citation, *Journal of the Korean Society for Information Management*, **27**, 103–118.

구조방정식 모형의 베이저안 접근법 기반의 특허평가 모델링에 대한 연구

우호영^a · 광정애^b · 임창원^{a,1}

^a중앙대학교 응용통계학과, ^b한국발명진흥회

(2017년 9월 5일 접수, 2017년 10월 5일 수정, 2017년 11월 6일 채택)

요약

최근 4차 산업으로의 산업 패러다임의 변화가 이미 시작되었으며, 이러한 4차 산업 시대에 무형 지식재산인 특허의 중요성은 날로 증대되고 있다. 특허의 기술가치평가는 전문가의 의견에 따라서 산정되기 때문에 많은 비용과 시간이 소모되므로 비전문가들의 주관적인 의견에 기인하여 특허의 질적 수준을 판단하게 된다. 따라서 특허의 질적 수준에 대한 객관적이고 합리적인 평가 체계 개발이 필요하다. 본 논문에서는 특허의 가치평가를 기술성, 권리성, 활용성으로 구분하고 베이저안 구조방정식을 사용하여 특허의 정량화되고 객관적인 평가 모델링에 대해 고려했다. 특히, 한국발명진흥회에서 수집한 자료를 토대로, 직접적으로 측정되기 어려운 질적 성과들을 모형화하고 평가하는데 탁월한 구조방정식과 사전 정보를 활용함으로써 작은 표본 하에서도 안정적인 모형화가 가능한 베이저안 접근법을 함께 적용하여 특허 평가 모형을 개발하였다.

주요어: 특허평가, 베이저안, 구조방정식

¹교신저자: (06974) 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: clim@cau.ac.kr