

연속적 이항 로지스틱 회귀모형을 이용한 R&D 투입 및 성과 관계에 대한 실증분석

박성민[†]

백석대학교 경상학부

Empirical Analysis on the Relationship between R&D Inputs and Performance Using Successive Binary Logistic Regression Models

Sungmin Park

Department of Business Administration, Baekseok University

The present study analyzes the relationship between research and development (R&D) inputs and performance of a national technology innovation R&D program using successive binary Logistic regression models based on a typical R&D logic model. In particular, this study focuses on to answer the following three main questions; (1) "To what extent, do the R&D inputs have an effect on the performance creation?"; (2) "Is an obvious relationship verified between the immediate predecessor and its successor performance?"; and (3) "Is there a difference in the performance creation between R&D government subsidy recipient types and between R&D collaboration types?" Methodologically, binary Logistic regression models are established successively considering the "Success-Failure" binary data characteristic regarding the performance creation. An empirical analysis is presented analyzing the sample $n = 2,178$ R&D projects completed. This study's major findings are as follows. First, the R&D inputs have a statistically significant relationship only with the short-term, technical output, "Patent Registration." Second, strong dependencies are identified between the immediate predecessor and its successor performance. Third, the success probability of the performance creation is statistically significantly different between the R&D types aforementioned. Specifically, compared with "Large Company", "Small and Medium-Sized Enterprise (SMS)" shows a greater success probability of "Sales" and "New Employment." Meanwhile, "R&D Collaboration" achieves a larger success probability of "Patent Registration" and "Sales."

Keywords: Binary Logistic Regression, Government Subsidy, R&D Collaboration, R&D Logic Model, Small and Medium-Sized Enterprise

1. 서론

공공부문 R&D 성과평가는 R&D 논리모형(Logic Model) 또는 논리사슬(Logic Chain) 등으로 지칭되는 체계적 흐름도에 근거하여 R&D 자원의 투입 이후 시간적 경과에 따라 단계적으로 창출되는 성과의 효율성, 효과성, 적절성 등을 다양한 연구방법론을 활용하여 가능한 계량적으로 분석한 후, 그 분석 결과

를 후속 R&D 사업 계획의 수립 및 전개 시에 활용하는 것이다 (Wholey, 1983; Bickman, 1987; Wholey, 1987; McLaughlin and Jordan, 1999; Ruegg and Feller, 2003; W. K. Kellogg Foundation, 2004). WKKF(2004)는 세 가지 서로 다른 기간 중 창출 가능한 성과에 대한 단계적 분류로서; (1) 단기적(1~3년) 산출, (2) 중기적(4~6년) 결과, (3) 대부분 사업 종료 이후 발생 가능한 사회·경제적 장기적(7~10년) 파급효과 등을 논의한 바 있다. 특히,

이 논문은 2014년도 백석대학교 대학연구비에 의하여 수행된 것임.

[†] 연락저자 : 박성민 교수, 330-704 충청남도 천안시 동남구 안서동 백석대학교 경상학부, Tel : 041-550-2497, Fax : 041-550-9172,

E-mail : smpark99@bu.ac.kr

2014년 1월 2일 접수; 2014년 2월 24일 수정본 접수; 2014년 3월 21일 게재 확정.

국가연구개발사업에 특화된 전형적인 R&D 논리모형으로는, Ruegg and Feller(2003)의 미국 상무부 산하 Advanced Technology Program(ATP) R&D 논리모형, McLaughlin and Jordan(1999)의 미국 에너지부 산하 Research and Technology Development and Deployment Program(RTDDP) R&D 논리모형 등이 개발되어, 정부지원 R&D 과제에의 성과를 심층적으로 분석한 바 있다.

한편, 대표적인 기술혁신 국가연구개발사업으로는, 미국 ATP, 대만 경제부 산하 Industrial Technology Development Program(ITDP) 및 대한민국 지식경제부 산하 지식경제기술혁신사업(Knowledge Economy Technology Innovation Program, KETIP) 등이 있다(Ruegg and Feller, 2003; Shipp *et al.*, 2005; Ruegg, 2006; Hsu and Hsueh, 2009; KEIT, 2010; KEIT, 2011; KEIT, 2013). 또한, 제 2장에서 자세히 논의되는 것처럼, 최근 국가연구개발사업 계획의 수립 및 전개 시 주관기관의 R&D 협력이 강조됨과 동시에, 주관기관 유형에 따른 성과 차이를 고려한 정부지원 R&D 과제의 세심한 사전 선별이 요구되는 실정이다(KEIT, 2010; KEIT, 2011; OMB and OSTP, 2012; OSTP, 2012; KEIT, 2013).

하지만, 앞선 기술혁신 국가연구개발사업들의 경우, 주관기관 및 협력 유형별로 성과 창출 정도에 얼마만한 차이가 실제로 존재하는지 등에 대한 선행연구는 현재까지 매우 제한적이며 특히, 대규모 분석 샘플을 이용한 실증분석 결과 실례는 더욱 찾기 어렵다. R&D 성과 분석을 시도했던 많은 선행연구의 공통적 한계점으로서, 투입 대비 성과 창출 시점 간 시차(time-lag)를 충분히 반영한 분석 샘플을 설계·수집하기가 현실적으로 힘든 점 등도 지적 가능하다(Wu *et al.*, 2006; Guan and Chen, 2010; Chen *et al.*, 2011). 또한, 성과 창출에 성공한 정부지원 R&D 과제의 희소성 등과 같은 태생적 이유로 분석 샘플 확보에 어려움이 있었던 또 다른 이유로 추정되며, 이와 함께 R&D 자원의 투입이 실질적인 영향력을 발휘할 수 있는 성과 창출

범위에 대한 연구 필요성도 제기된 바 있다(Shipp *et al.*, 2005; Ruegg, 2006; KEIT, 2011). 한편, 연구방법론 측면에서는 다양한 회귀모형들을 활용한 R&D 기술혁신성 및 생산성 관련 성과 분석을 논의한 선행연구들이 확인된다(Fritsch and Lukas, 2001; Belderbos *et al.*, 2004; Laursen and Salter, 2006; Berchicci, 2013; Robin and Schubert, 2013).

본 연구는 기술혁신 국가연구개발사업 내 정부지원 R&D 과제의 투입 대비 성과 간 관계를 일반적인 R&D 논리모형 상순서에 기초하여 분석한다. 구체적으로, 본 연구는 다음 세 가지 질문에 대한 답변을 도출하고자 한다; (1) “투입 대비 성과 창출의 영향력은 어느 범위까지 발휘 가능한가?”; (2) “직전 대비 후속 성과 간 관계는 확인되는가?”; (3) “주관기관 및 협력 유형 간 성과 창출 여부에 유의한 차이는 존재하는가?” 분석 자료는, 지식경제기술혁신사업 내 “중료년도” 기준 2007년부터 2011년까지, 총 5년 간 대기업 주관 311개 및 중소기업 주관 1,867개, 총 2,178개 정부지원 종료 R&D 과제를 분석 샘플 $n = 2,178$ 개로 선정하여 실증분석을 실시한다. 연구방법론으로서, 분석 샘플을 구성하는 각 관측 개체(즉, 정부지원 종료 R&D 과제)의 자료 특성 중 특히, 성과 창출 여부 관련 성공-실패 이진 자료 특성을 고려한 연속적 이항 로지스틱 (Logistic) 회귀모형을 기반으로 실증분석을 실시한다. 이후, 특별한 언급이 없는 한, “정부지원 종료 R&D 과제”는 “R&D 과제”로 지칭된다. 본 논문은, 제 1장 서론, 제 2장 이론배경 및 문헌고찰, 제 3장 연구모형, 제 4장 실증분석 및 제 5장 결론 등의 순서로 구성된다.

2. 이론배경 및 문헌고찰

<Figure 1>에서 실선 사각형으로 표시되어 연결된 부분들이

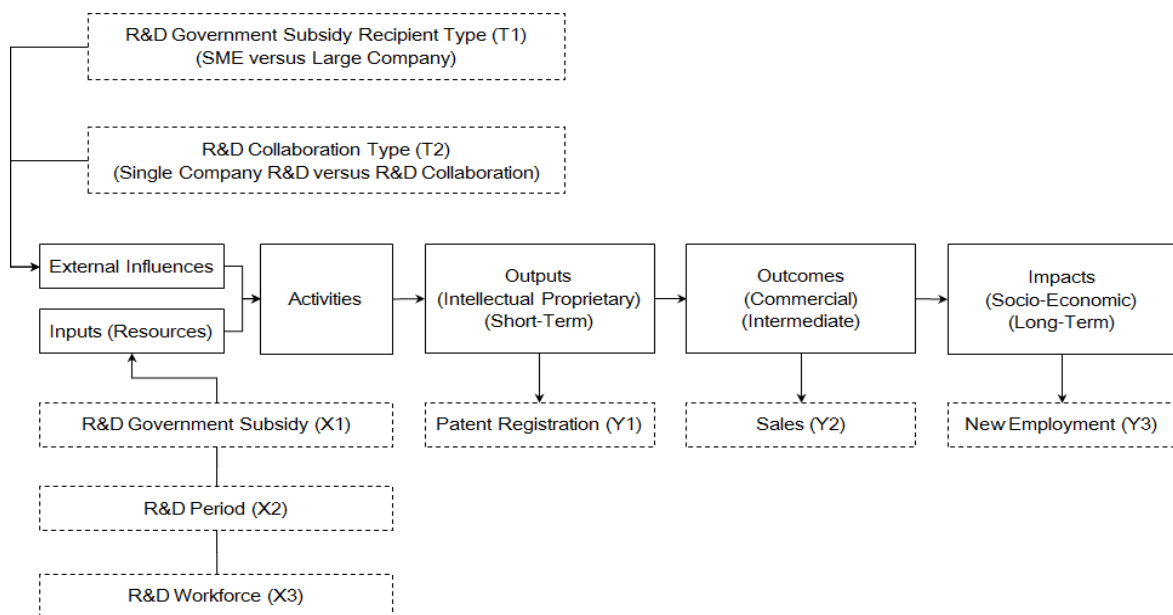


Figure 1. A R&D logic model with inputs, performance and external influences

공공부문 R&D 성과평가를 위한 전형적인 R&D 논리모형이다 (McLaughlin and Jordan, 1999; Ruegg and Feller, 2003; WKKF, 2004). Ruegg and Feller(2003), Hsu and Hsueh(2009)는 문헌고찰을 통해 R&D 과제 성과평가 시 고려해야 할 항목들을 세부적으로 제시하였다; (1) 투입 (Inputs) 항목은 정부지원금액, 수혜기관 대응투자금액, 인력, 성과조사기간; (2) 단기적, 기술적 산출(Outputs) 성과 항목은 논문, 특허 출원·등록·활용; (3) 중기적, 경제적 결과(Outcomes) 성과 항목은 사업화 이익; (4) 장기적, 사회·경제적 파급효과(Impacts) 항목은 신규고용 등이 확인된다. <Figure 1>에서 점선 사각형으로 표시된 내용들은 본 연구의 관심 대상으로서, R&D 투입, 성과 및 외부 영향요인들을 보여주는데, 이에 대한 자세한 설명은 제 3장, 제 4장에서 논의한다.

대한민국 정부는 국가연구개발사업을 통해 창출된 성과를 매년 정기적으로 국가과학기술지식정보서비스(National Science and Technology Information Service, NTIS)에 등록·관리하는데, 핵심 성과지표 6가지 즉, (1) 특허, (2) 논문, (3) 기술료, (4) 사업화, (5) 인력양성, (6) 학술·기술 연수성과 및 (7) 장비·설비 구축 정보, 총 7가지 주요 성과지표를 고려할 때, <Figure 1>은 특허 특허 → 사업화 → 인력양성 성과 사슬에 초점을 맞춘 구성이라고 판단되며, 이는 대학, 국립연구소 등과 같은 순수 교육·연구기관보다는 기술혁신 지향적 기업에 부합된 성과 사슬로 간주된다(KISTI, 2008; MKE, 2008; MST·OSTI, 2008). 일련의 지식경제기술혁신사업 성과활용현황 조사에 따르면, R&D 자원의 투입 대비 성과 간 관계 그리고, 주관기관, 협력, 과제규모 유형별 성과 창출 정도의 차이 비교 등이 주된 심층적 R&D 성과분석 주제인 것으로 파악된다(KEIT, 2010; KEIT, 2011; KEIT, 2013).

최근, 미국의 경우 국가연구개발사업 예산 편성 시 R&D 협력의 필요성을 강조하고 있다. 미국 예산관리국(Office of Management and Budget, OMB) 및 미국 대통령 직속기관인 과학기술정책국(Office of Science and Technology Policy, OSTP) 등은 2014년도 미국 국가연구개발사업의 예산 투입의 우선순위를 논의한 바 있는데 특히, 정보기술을 포함한 R&D 협력이 필요한 9가지 과학기술 분야를 공표하였다(OMB and OSTP, 2012). ATP도 정부지원금액 수혜기관이 “단일기업”(Single Company)인 경우, 정부지원금액 상한선을 최대 US\$ 2백만으로 책정하였고, 종업원 20인 미만 소기업을 우선 지원하도록 관리된 것으로 보고된다(Ruegg and Feller, 2003).

Chen *et al.*(2011)은 2005년부터 2007년까지 3년 간 중국 73개 정보통신 기업의 생산성 변화를 DEA 기반 맘퀴스트지수(Malmquist Index) 모형으로 분석하였는데, 특히 대기업과 중소기업 간 R&D 협력의 필요성을 주장하였다. Belderbos *et al.*(2004)는 1996년부터 1998년까지 3년 간 네덜란드 2,056개 기업을 대상으로 R&D 협력이 기업의 노동 및 기술혁신 생산성에 미치는 영향을 회귀모형으로 분석하였는데, 4 가지 협력 유형별로 추구하는 생산성이 상이함을 확인하였다. Fritsch and Lukas(2001)는

독일 1,800개 기업을 대상으로 실시한 설문조사 자료를 2단계 Logit-Poisson 회귀모형으로 분석하였는데, 협력 빈도는 기업 규모(즉, 종업원수), R&D 강도(즉, R&D 인력비율), 이 두 가지와 통계적으로 유의한 양(+)의 상관성이 있음을 확인하였다. Berchicci(2013)는 1992년부터 2004년까지 13년 간 이탈리아 2,537개 기업을 대상으로 한 설문조사 자료를 Tobit 회귀모형으로 분석하였는데, 기술혁신성 대리변수인 “신제품 매출액비율”은 기업이 활용하는 외부 R&D 활동 비율과 역-U자 형태의 비선형 관계가 있음을 확인하였다. 특히, R&D 역량이 작은 기업은 큰 기업에 비해 외부 R&D 활동을 높임으로써 기업혁신성을 함께 제고시킬 수 있다고 결론지었다. Robin and Schubert (2013)는 2004년부터 2008년까지 5년 간 프랑스 및 독일 기업을 대상으로 설문조사를 실시하고 그 자료를 역시 Tobit 회귀모형으로 분석하였는데, 기업이 수행한 공공부문 R&D 기관과의 협력이 기술혁신성 대리변수인 신제품 매출액비율을 증가시키는 경우가 있었음을 보고하였다.

3. 연구모형

본 연구에서 검토하고자 하는 각 관측 개체 (즉, R&D 과제)의 자료 특성을 나타내는 변수들은, <Figure 1>에 점선 사각형으로 표시된 총 8가지 R&D 투입, 성과 및 외부 영향요인 각 항목에 대응된다. 첫째, R&D 투입 항목에 대응되는 R&D 입력변수로는, X1(R&D Government Subsidy, 정부지원금액), X2(R&D Period, 연구개발기간), X3(R&D Workforce, 연구개발인력) 세 가지이다. 둘째, R&D 성과변수로는 Y1(Patent Registration, 특허등록), Y2(Sales, 매출액), Y3(New Employment, 신규고용) 세 가지이다. 마지막으로, R&D 입력변수와 성과변수 간 외부 영향요인변수는 T1(Company Type, 정부지원금액 수혜기관 유형 또는 주관기관 유형), T2(R&D Collaboration Type, 협력 유형) 두 가지로 본 연구모형에서는 한정하고자 한다. 이후 제 4장에서 설명되는 것처럼, 각 R&D 성과 항목별로 성과 창출에 성공한 관측 개체가 상대적으로 매우 희소할 뿐만 아니라, 그 값들의 극단적인 비대칭 분포를 감안하여, 세 가지 연속형 원자료 R&D 성과변수 Y1, Y2, Y3는 식 (1)부터 식 (3)과 같은 세 가지 이진 변수들로 변환되어 이항 로지스틱 회귀모형으로 분석된다. 예를 들어, i^{th} 관측 개체의 Y1 값, $Y1_i > 0$ 이면(즉, 특허등록 성과 창출 성공)이면, 이에 대응하는 특허등록 성과 창출 여부 관련 성공-실패 이진 변수 $B1_i = 1$ 로 변환됨을 의미한다. 한편, 본 연구에서 검토되는 8가지 변수들은 <Figure 1>에서 논의된 선행 연구에서 공통적으로 활용된 대표적인 항목들에 기초하여 선정된 것이며, 이와 더불어 NTIS 핵심 성과지표 중에서 본 실증 분석 대상 사업 특성과 가장 밀접한 관련성이 있다고 판단된 항목들을 추가로 반영하여 결정된 것이다(McLaughlin and Jordan, 1999; Ruegg and Feller, 2003; WKKF, 2004; Wu *et al.*, 2006; KISTI, 2008; MKE, 2008; MST·OSTI, 2008; Hsu and Hsueh,

2009; Guan and Chen, 2010).

$$B1_i = \begin{cases} 1 & \text{if } Y1_i > 0, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$B2_i = \begin{cases} 1 & \text{if } Y2_i > 0, \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$B3_i = \begin{cases} 1 & \text{if } Y3_i > 0, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

T1, T2 두 가지 이진 영향요인변수는 식 (4), 식 (5)와 같이 정의된다. 먼저, i^{th} 관측 개체의 주관기관 유형이 “대기업”(Large Company)이면 $T1_i = 1$, “중소기업”(Small and Medium-Sized Enterprise, SME)이면 $T1_i = 0$ 값을 갖는다. 이와 유사하게, i^{th} 관측 개체의 협력 유형이 “협력 R&D”(R&D Collaboration)이면 $T2_i = 1$, “단독 R&D”(Single Company R&D)이면 $T2_i = 0$ 값을 갖는다.

$$T1_i = \begin{cases} 1 & \text{if the } i^{th} \text{ observation by "Large Company",} \\ 0 & \text{otherwise (i.e. "SME")} \end{cases} \quad (4)$$

$$T2_i = \begin{cases} 1 & \text{if the } i^{th} \text{ observation by "R&D Collaboration",} \\ 0 & \text{otherwise (i.e. "Single Company R&D")} \end{cases} \quad (5)$$

식 (6)과 같이 정의된 이항 로지스틱 회귀모형 Model (1)은 반응변수(Response Variable) B1을 총 5가지 예측변수(Predictor) X1, X2, X3, T1, T2로 설명한다. 그러므로, Model (1)은 세 가지 R&D 투입 항목들과 함께 주관기관 및 협력 유형 등에 따라 특허등록 성과 창출 여부 관련 성공-실패 확률이 얼마나 민감하게 반응하는지를 분석한다. <Figure 1>의 R&D 논리모형의 성과 창출 순서에 따라 연속적으로 Model (2), Model (3)를 식 (7), 식 (8)과 같이 정의한다. Model (2)는 예측변수 B1과 반응변수 B2 즉, 매출액 성과 창출 여부 관련 성공-실패 간 관계를 추가로 분석한다. 마지막으로, Model (3)는 R&D 논리모형 순서 상 가장 마지막 단계에 위치한 신규고용 성과와 이전 단계에 위치한 특허등록, 매출액 성과 등을 함께 관계시킨 모형이다.

• Model (1) : $E(B1|X1, X2, X3, T1, T2)$
 $=f(X1, X2, X3, T1, T2)$

$$= \frac{\exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2)}{1 + \exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2)} \quad (6)$$

• Model (2) : $E(B2|X1, X2, X3, T1, T2, B1)$
 $=f(X1, X2, X3, T1, T2, B1)$

$$= \frac{\exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2 + \beta_{B1} B1)}{1 + \exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2 + \beta_{B1} B1)} \quad (7)$$

• Model (3) : $E(B3|X1, X2, X3, T1, T2, B1, B2)$
 $=f(X1, X2, X3, T1, T2, B1, B2)$

$$= \frac{\exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2 + \beta_{B1} B1 + \beta_{B2} B2)}{1 + \exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_{T1} T1 + \beta_{T2} T2 + \beta_{B1} B1 + \beta_{B2} B2)} \quad (8)$$

한편, 제 2장에서 자세히 서술된 관련 문헌고찰을 통한 이론적 배경에 기초하여, 제 1장에서 제시된 본 연구의 동기인 세 가지 질문에 대한 답변을 도출하기 위해, 본 연구모형을 통해 검정하고자 하는 가설을 항목별로 설정하면 아래와 같이 정리할 수 있다.

- **가설 1** : 투입 대비 성과 창출의 양 (+)의 상관성을 취하는 영향력은 존재하나, 그 영향력은 기술적 산출에서 경제적 결과 등으로 진행될수록 감소할 것이다.
- **가설 2** : 직전 성과 대비 후속 성과 간 양 (+)의 상관성을 취하는 관계가 존재할 것이다.
- **가설 3** : 주관기관 및 협력 유형 간 성과 창출 여부에 유의한 차이가 존재할 것이다.
- **가설 3-1** : 주관기관 유형 간(즉, “중소기업” 대비 “대기업”) 성과 창출 여부에 유의한 차이가 존재할 것이다.
- **가설 3-2** : 협력 유형 간(즉, “단독 R&D” 대비 “협력 R&D”) 성과 창출 여부에 유의한 차이가 존재할 것이다.

4. 실증분석

4.1 분석자료

본 실증분석은 지식경제기술혁신사업 “종료년도” 기준 2007년부터 2011년까지 5년 간 총 3,033개 R&D 과제 중, 대기업이 주관한 358개 및 중소기업이 주관한 1,899개, 총 2,257개 R&D 과제를 초기 분석 대상 샘플로 선정한다. 비록, 종료년도가 약간 상이하기는 하지만 “계속과제”가 아닌 “종료과제”를 분석 대상으로 추출했다는 점에서 R&D 투입 대비 성과 창출의 시차가 충분히 반영된 분석 자료의 설제라고 판단된다.

단, 초기 분석 대상 샘플 총 2,257개 R&D 과제의 원자료를 검토한 결과, X3 및 Y1, Y2, Y3 모두에서 극단적인 이상점(Outlier)들이 확인되어, 이 4가지 변수들에 대해서는 각각 99% 백분위수를 초과하는 값을 갖는 관측 개체들은 본 실증분석 샘플에서는 제외하였다. 구체적인 99% 백분위수는 X3, Y1, Y2, Y3 순서대로 각각 82.20, 15.60, 89.51, 103.00이며, 각 변수의 단위는 <Table 1>에 제시되어 있다. 이와 같은 경계 값을 적용하여 총 79개 관측 개체를 이상점으로 판정하여 제거한 후, 최종적으로 정리된 본 실증분석 대상 샘플 크기는 $n = 2,178$ 이다. 한편, X1의 경우 분석 샘플 정비를 위해 적용된 환율은 1,000원/US\$ 1이다. Y1은 지식경제부 산하 한국산업기술진흥원의 “지식경제 R&D 성과지표 설계 가이드라인” 지침 중에 포함된 가중성과지표 정의식(즉, 국내 특허등록 : 국외 특허등록 = 0.6 : 1.0)이 적용된 국내 및 국외 특허등록 성과의 가중합

계이다 (MKE · KIAT 2012).

<Table 1>은 본 실증분석 샘플 $n = 2,178$ 의 주요 기술통계량을 보여준다. 먼저, R&D 입력변수 및 성과변수 6가지(즉, X1, X2, X3, Y1, Y2, Y3)의 왜도(Skewness), 첨도(Kurtosis) 등을 보면 심한 우변사향 분포를 취함을 알 수 있다. 특히, 이 6가지 변수의 히스토그램을 보여주는 <Figure 2>에서도 확인되는 것처럼, 세 가지 성과변수 Y1, Y2, Y3의 경우에는 0 초과 값을 취하는 성과 창출 관측 개체 수가 기술적 성과변수 Y1(897개, $897/2,178 \times 100 = 41\%$)을 시작으로, 경제적 성과변수 Y2(840개, 39%), 사회·경제적 파급효과 성과변수 Y3(678, 31%)와 같이, <Figure 2>의 R&D 논리모형 순서 상 시작 점에서 끝 점으로 접근할수록 점점 더 감소하는 현상이 확인된다. 이와 같은 비율은 순서대로 세 가지 이진변수 B1, B2, B3의 평균(Mean) 0.41, 0.39, 0.31과도 일치된다. 이와 같은 현상은 세 가지 성과변수 Y1, Y2, Y3 모두 중위수(Median)가 0 값을 취하고 있음과도 일치되며, 이로써 성과 창출 관측 개체의 최소적 측면이 재차 확인된다. 영향요인변수 T1, T2 관련 각각 평균이 0.14, 0.58이므로 분석 샘플 $n = 2,178$ 에서; (1) “대기업” 14%, “중소기업” 86%; (2) “협력 R&D” 58%, “단독 R&D” 42% 등으로 조사된다. 본 실증분석 시 정의된 “협력 R&D” 유형은 해당 관측 개체가, (1) “산학”, (2) “산연”, (3) “산학연”, 이 세 가지 유형 중 어느 한 가지에라도 분류됨을 의미한다. <Table 1>은 이외에도 표준편차(StDev), 변동계수(CoefVar), 최대값(Max) 등을 보여주며, “종료과제”로 구성된 샘플이므로 X2 최대값은 10.17로 확인된다. <Table 2>는 분석 샘플 $n = 2,178$ 을 주관기관 유형(즉, T1 = 0, 1), 협력 유형(즉, T2 = 0, 1) 및 세 가지 이진 성과변수 유형(즉, B1 = 0, 1; B2 = 0, 1; B3 = 0, 1) 각각에 대한 관측 개체 구성을 자세히 정리한 것이다. 한편, 본 연구에서는 R&D 입력변수 간 상관분석을 매개로 로지스틱 회귀분석을 위한 완전모형을 축소 모형으로 변환시킬 필요성을 언급하고자, 이는 제 4.4절에서 자세히 논의한다.

4.2 빈도분석

<Table 2>는 분석 샘플 $n = 2,178$ 을 대상으로 성과 창출 관측

개체 수가 R&D 논리모형 상 시간적 순서에 따라 최소해지는 현상이 경우마다 상이함을 보여준다. 앞서 설명된 것처럼 B1 = 1 조건을 만족하는 관측 개체 수 897개 중에서 B1 = 1 & B2 = 1 조건을 만족하는 관측 개체 수는 429개로, 그 비율이 $429/897 \times 100 = 48\%$ 인 반면, B1 = 0 조건을 만족하는 관측 개체 수 1,281개 중에서 B1 = 0 & B2 = 1 조건을 만족하는 관측 개체 수는 411개로, 그 비율이 $411/1,281 \times 100 = 32\%$ 와 같이 상대적으로 더 축소되어 결국, Y2 성과 창출 여부가 R&D 논리모형 상의 직전 단계인 Y1 성과 창출 여부에 의존될 가능성이 있음을 보여준다. 마찬가지로, B1 = 1 & B2 = 1 조건을 만족하는 관측 개체 수 429개 중에서 B1 = 1 & B2 = 1 & B3 = 1 조건을 만족하는 관측 개체 수는 327개로, 그 비율이 $327/429 \times 100 = 76\%$ 인 반면, B1 = 1 & B2 = 0 조건을 만족하는 관측 개체 수 468개 중에서 B1 = 1 & B2 = 0 & B3 = 1 조건을 만족하는 관측 개체 수는 30개로 그 비율이 $30/468 \times 100 = 6\%$ 와 같이 계산되므로 상대적으로 더 축소되므로 결국, Y3 성과 창출 여부가 Y2 성과 창출 여부에 매우 의존적임이 판명된다.

한편, 영향요인변수 T1, T2에 따라서도, 예를 들어 Y1 성과 창출 관측 개체 수 비율이 매우 다를 수 있음을 확인할 수 있다. B1 = 1 조건을 만족하는 관측 개체 수 897개 중에서 T2 = 1 조건을 만족하는(즉, “협력 R&D”) 관측 개체 수는 637개로, 그 비율이 $637/897 \times 100 = 71\%$ 인 반면, T2 = 0 조건을 만족하는 관측 개체 수는 260개로, 그 비율이 $260/897 \times 100 = 29\%$ 이므로 결국, Y1 성과 창출 여부는 영향요인변수 T2에 큰 차이를 보인다. 또한, T1 = 1 조건을 만족하는(즉, “대기업”) 관측 개체 수는 175개로, 그 비율이 $175/897 \times 100 = 20\%$ 인 반면, T1 = 0 조건을 만족하는 관측 개체 수는 722개로, 그 비율이 $722/897 \times 100 = 80\%$ 이므로 역시, “중소기업”의 Y1 성과 창출 성공 확률이 더 클 가능성이 있다. 즉, 이후 보다 자세히 설명될 로지스틱 회귀분석과 같은 본 분석에 앞서, 이를 위한 예비분석의 일환으로서 상기 빈도분석의 필요성이 있음을 알 수 있다. 즉, 이와 같은 빈도분석을 통해 로지스틱 회귀분석 실시를 위한 방향과 체계 그리고 본 분석 후의 주요 시사점을 미리 가늠할 수 있다고 판단된다.

Table 1. Descriptive statistics of the sample($n = 2,178$)

Variable	Name	Unit	Mean	StDev	CoefVar	Median	Max	Skewness	Kurtosis
R&D Government Subsidy	X1	(US\$ 10 ⁶)	1.13	1.47	129.91	0.50	14.94	2.54	9.31
R&D Period	X2	(Years)	2.44	1.85	75.87	2.00	10.17	1.31	1.96
R&D Workforce	X3	(Man-years)	9.33	12.94	138.71	5.00	80.00	2.14	5.78
Patent Registration	Y1		0.85	1.82	213.82	0.00	15.60	3.81	18.40
Sales	Y2	(US\$ 10 ⁶)	1.43	6.20	434.05	0.00	87.67	7.65	72.26
New Employment	Y3	(Man-years)	2.91	8.97	308.27	0.00	103.00	6.01	46.33
Company Type	T1		0.14						
R&D Collaboration Type	T2		0.58						
Binary Conversion of Y1	B1		0.41						
Binary Conversion of Y2	B2		0.39						
Binary Conversion of Y3	B3		0.31						

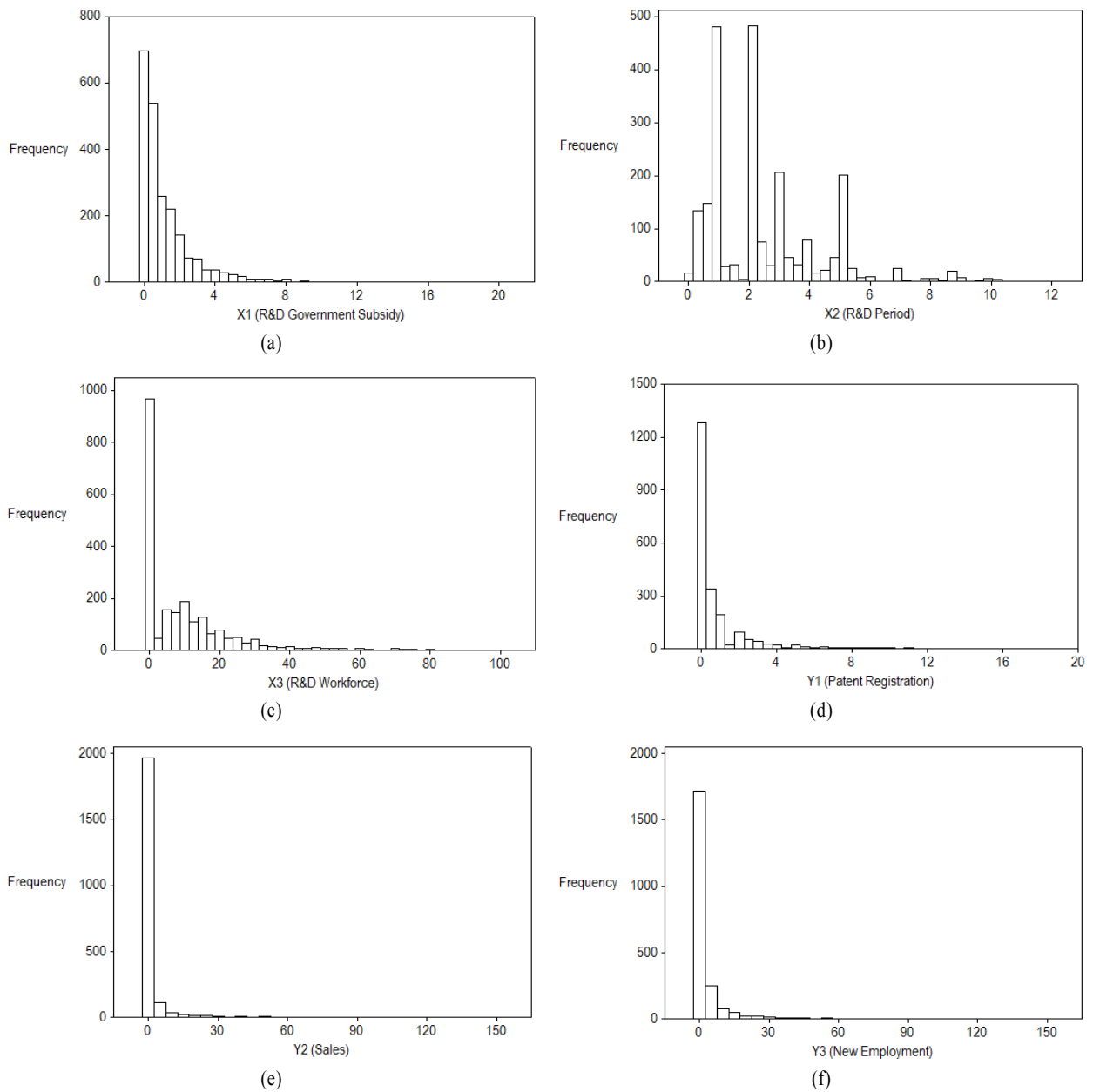


Figure 2. Histograms of R&D inputs and performance variables($n = 2,178$)

Table 2. The number of observations in the sample based on the performance creation

		B1 = 0					B1 = 1					Total
		B2 = 0		B2 = 1		Subtotal	B2 = 0		B2 = 1		Subtotal	
		B3 = 0	B3 = 1	B3 = 0	B3 = 1		B3 = 0	B3 = 1	B3 = 0	B3 = 1		
T1 = 0	T2 = 0	398	26	78	118	620	98	13	22	73	206	826
	T2 = 1	315	26	55	129	525	225	13	63	215	516	
Subtotal		713	52	133	247	1,145	323	26	85	288	722	1,867
T1 = 1	T2 = 0	27	1	4	8	40	37		7	10	54	94
	T2 = 1	74	3	9	10	96	78	4	10	29	121	
Subtotal		101	4	13	18	136	115	4	17	39	175	311
Total		814	56	146	265	1,281	438	30	102	327	897	2,178

4.3 로지스틱 회귀분석 - 완전모형

<Table 3>은 로지스틱 회귀분석 완전모형 Model (1)의 주요 분석 결과를 보여준다. “완전모형”(Full Model) 용어는 제 4.5절 “축소모형”(Reduced Model)과 구별하기 위해 사용된다. 반응 변수 B1 = 1 값이 성공(Success) 즉, “기준사건”(Case)으로 정의이며, 2-수준 요인(Factor) 예측변수 T1, T2 “기준수준”(Reference Level)은 각각 T1 = 0, T2 = 0 값으로 정의된다. Model (1)의 주요 분석 결과를 설명하면 다음과 같다.

첫째, X1, X2, X3 모두 양(+)의 추정계수를 갖고 P-value가 0.05 미만이므로, 반응변수 B1에 통계적으로 유의한 예측변수들로 판단된다. 추정계수 승산비(Odds Ratio) 즉, $\exp(\hat{\beta})$ 를 X1 경우를 예로 들어 설명하면, $\exp(\hat{\beta}_1) = 1.20$ 으로 X1 값 1 단위 증가에 따라 B1 = 1 값 즉, 성공 확률 승산비가 1.20배 증가하는 것으로 분석된다. 이와 같은 해석을 통해 결국, X1, X2, X3 모두 1 단위 증가에 따라 특허등록 성과 창출 성공 확률이 높아지는 것으로 해석된다. 특히, B1대비 X3보다는 X1, X2의 통계

Table 3. Binary logistic regression analysis with Model (1)

Model Description								
Model								
B1 versus X1, X2, X3, T1, T2								
Link Function								
Logistic Distribution								
Response Variable		Value	Count					
B1		1	897	(Success)				
		0	1,281	(Failure)				
		Total	2,178					
Factor		Level	Value					
T1		2	0, 1					
T2		2	0, 1					
Binary Logistic Regression Analysis								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-1.665	0.096	-17.42	0.000			
X1		0.180	0.050	3.63	0.000***	1.20	1.09	1.32
X2		0.328	0.037	8.91	0.000***	1.39	1.29	1.49
X3		0.011	0.005	2.36	0.018**	1.01	1.00	1.02
T1	1	-0.372	0.153	-2.42	0.015**	0.69	0.51	0.93
T2	1	0.400	0.104	3.86	0.000***	1.49	1.22	1.83
Log likelihood = -1,294.117, $\chi^2 = 363.057$, DF = 5, P-value = 0.000***								
Goodness-of-fit Statistics								
Method		χ^2	DF	P				
Pearson		1,843.28	1,662	0.001				
Deviance		2,156.23	1,662	0.000				
Hosmer-Lemeshow		90.31	8	0.000				
Measures of Association								
Somers' D		0.50						
Goodman-Kruskal γ		0.50						
Kendall's τ_A		0.24						
Pair Type		Count	Percentage					
Concordant Pairs		858,104	74.7					
Discordant Pairs		283,361	24.7					
Tied Pairs		7,592	0.7					
Total		1,149,057	100.0					

*, **, *** indicate statistical significance at the significance level $\alpha = 10\%$, 5% , 1% respectively.

적 유의성 및 승산비가 상대적으로 더 큰 것이 확인된다. 한편, 성공 확률 승산비 변화는 원래 성공 확률이 0.5 근처에 가까울 수록 더 큰 것으로 알려져 있다(Hosmer and Lemeshow, 2000; Montgomery *et al.*, 2001; Minitab, 2005; IBM SPSS, 2009).

둘째, T1은 음(-)의 추정계수 $\hat{\beta}_{T1} = -0.372$, P-value = 0.015**이므로, 반응변수 B1에 대해 통계적으로 유의한 예측변수로 판단된다. 한편, *, **, ***는 순서대로 유의수준 $\alpha = 10\%$, 5%, 1%에서 통계적으로 유의함을 표시한다. 음(-)의 추정계수를 취하므로 승산비 $\exp(\hat{\beta}_{T1}) = 0.69$ 로 그 값이 1보다 작게 계산되어 결국, T1 = 0기준 수준에서 T1 = 1로 1단위 증가할 때, B1 = 1 값 즉, 성공 확률 승산비는 오히려 0.69배 감소하는 것으로 해석된다. 따라서, “중소기업”(즉, T1 = 0)과 비교할 때 “대기업”(즉, T1 = 1)이 주관한 관측 개체일 때 특허등록 성과 창출 성공 확률이 저하됨을 의미한다. 한편, T2는 양(+)의 추정계수 $\hat{\beta}_{T2} = 0.400$, P-value = 0.000***이므로, 반응변수 B1에 통계적으로 유의한 예측변수로 판단된다. 양(+)의 추정계수를 취하므로 승산비 $\exp(\hat{\beta}_{T2}) = 1.49$ 로 그 값이 1보다 크게 계산되어 결국, T2 = 0기준 수준에서 T2 = 1로 1단위 증가할 때, B1 = 1 값 즉, 성공 확률 승산비는 1.49배 증가하는 것으로 해석된다. 특히, T2는 Model (1) 총 5가지 예측변수 중 통계적 유의성 및 승산비 절대값이 가장 큰 예측변수로 확인된다. 즉, “단독 R&D”(즉, T2 = 0) 대비 “협력 R&D”(즉, T2 = 1)일 경우 특허등록 성과 창출 성공 확률이 제고됨을 의미한다.

셋째, Model (1) 총 5가지 예측변수 계수 전체에 대한 유의성 검정을 위해 귀무가설 식 (9), 대립가설 식 (10)을 수립한다. 식 (9), 식 (10)에 대한 검정은 Model (1) 대비 상수항(Intercept)만을 갖는 “귀무가설모형” 사이 우도비 검정(Likelihood Ratio Test)으로 실시한다. 이 두 모형 간 우도비 검정은 Model (1) 이탈도 (Deviance) $D_{Fitted(1)}$ 대비 “귀무가설 모형” 이탈도 D_{Null} 사이 차이로 계산된다. $\chi^2 = D_{Fitted(1)} - D_{Null} = 363.057$, P-value = 0.000***이므로, 식 (3)을 기각하고 식 (4)를 채택할 수 있다.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_{T1} = \beta_{T2} = 0 \quad (9)$$

$$H_1 : \beta_i \neq 0 \text{ for at least one } i \ (i = 1, 2, 3, T1, T2) \quad (10)$$

넷째, χ^2 적합도 검정 통계량을 (1) Pearson, (2) Deviance, (3) Hosmer-Lemeshow, 세 가지 방법으로 계산하여 검토하면 모두 P-value 0.05 미만이므로, Model (1)의 분석 샘플에 대한 적합도는 낮은 것으로 판단된다. <Figure 3.(a)>는 Model (1)의 분석 샘플 중 적합도가 낮은 관측 개체를 보여주는 Delta Chi-sq($\Delta\chi^2$) 대비 예측 확률(Probability) 그림이다. <Figure 3.(a)>에서 가장 극단적인 부적합 관측 개체의 경우를 예를 들어 설명하면, 예측 확률 = 0.967, $\Delta\chi^2 = 29.801$ 인 경우가 확인된다. 이 관측 개체의 원자료를 확인하면 X1 = 10.28, X2 = 8.83, X3 = 25.00으로 상당히 큰 R&D 입력변수 값을 취하는 반면, R&D 성과변수 Y1, Y2, Y3 모두 0 값을 갖는 것으로 확인된다. 즉, 분석 샘플 중에는 상대적으로 큰 R&D 자원의 투입을 받았음에도 불구하고 성과 창출에 실패한 관측 개체가 다수 존재하며, 이 같은 현상이 상대적으로 낮은 적합도의 한 가지 원인으로 추측된다. <Figure 3.(a)>에서 예측 확률 0.5 값 근처를 초과할 때 $\Delta\chi^2$ 값이 기하급수적으로 증가하는 궤적 상에 존재하는 다수 관측 개체들이 바로 상대적으로 큰 R&D 자원의 투입을 갖는 관측 개체들로서, 예측 확률은 높게 계산되었지만 실제로는 B1 = 0 값 즉, 특허등록 성과 창출 실패 관측 개체들인 것으로 해석된다. 단, <Figure 3.(b)>에서 확인되는 것처럼, 앞서 일례로 설명된 극단적 부적합 관측 개체가 갖는 예측변수 계수 추정 영향력 Leverage = 0.004로 매우 작아, 이 관측 개체로 인한 계수 추정에 대한 문제는 심각하지 않은 것으로 판단된다.

다섯째, Model (1) 예측 능력을 나타내는 연관성 측도(Measures of Association) 세 가지, (1) Somer's D = 0.50, (2) Goodman-Kruskal $\gamma = 0.50$, (3) Kendall's $\tau_A = 0.24$ 처럼 계산되었다. 연관성 측도들은 이후 설명될 Model (2), Model (3) 대비 예측 능력에 대한 상대적 비교에 활용될 수 있는데, Model (1), Model (2)와 비교할 때 Model (3)가 세 가지 연관성 측도 모두 가장 큰 값

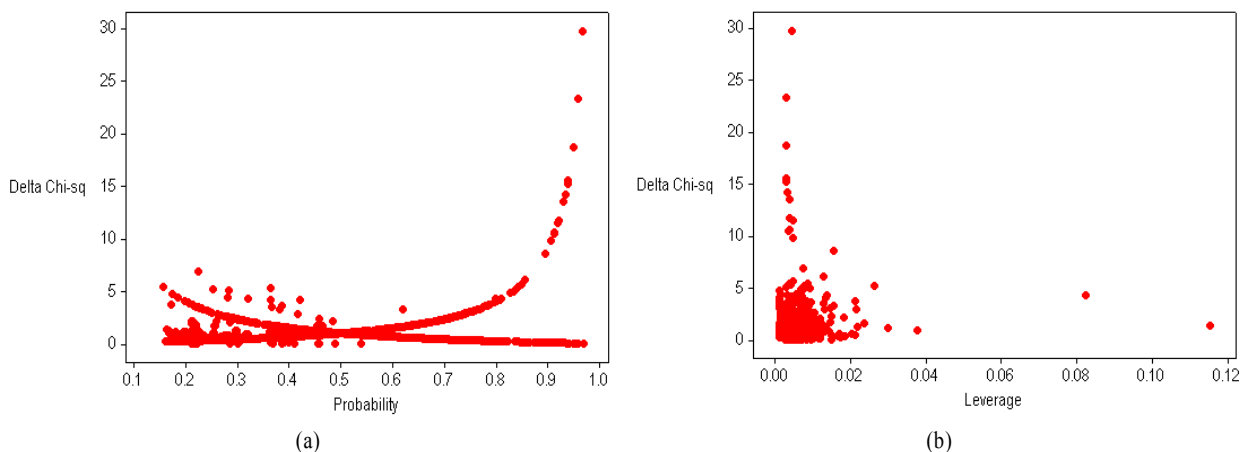


Figure 3. Diagnostic plots of Model (1); (a) Delta Chi-sq versus Probability, (b) Delta Chi-sq versus Leverage

을 취하므로 상대적으로 예측 능력이 우수한 모형으로 판단된다. 총 1,149,057(= 897×1,281)개 (성공, 실패) 관측 개체 짝(Pair) 중 성공 관측 개체의 성공 예측 확률이 실패 관측 개체의 성공 예측 확률보다 높은 “일치 짝”(Concordant Pair) 경우 858,104개 (74.7%)로 그 반대인 “불일치 짝”(Discordant Pair) 경우 283,361개(24.7%)와 비교할 때 상대적으로 높으므로, Model (1)의 반응변수 B1에 대한 예측 능력은 양호하다고 판단된다.

<Table 4>는 Model (2) 관련 주요 분석 결과이다. Model (2)는 Model (1)에 포함된 예측변수 5가지와 함께 Model (1) 반응변수 B1을 예측변수로 갖는, 총 6가지 예측변수로 반응변수 B2를 설명한다. 반응변수 B2 = 1 값이 성공 즉, 기준사건으로 정의되며, 2-수준 요인 예측변수로 지정된 T1, T2 및 B1의 기준수준은 각각 T1 = 0, T2 = 0, B1 = 0 값으로 정의된다. 단, <Table 4> 및 이후 설명될 <Table 5>에 정리된 분석 결과의 해석은 특

Table 4. Binary logistic regression analysis with Model (2)

Model Description								
Model								
B2 versus X1, X2, X3, T1, T2, B1								
Link Function								
Logistic Distribution								
Response Variable		Value	Count					
B2		1	840	(Success)				
		0	1,338	(Failure)				
		Total	2,178					
Factor		Level	Value					
T1		2	0, 1					
T2		2	0, 1					
B1		2	0, 1					
Binary Logistic Regression Analysis								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-0.684	0.084	-8.12	0.000			
X1		0.023	0.047	0.49	0.624	1.02	0.93	1.12
X2		-0.108	0.034	-3.15	0.002***	0.90	0.84	0.96
X3		0.014	0.005	3.00	0.003***	1.01	1.00	1.02
T1	1	-0.762	0.157	-4.85	0.000***	0.47	0.34	0.64
T2	1	0.195	0.100	1.95	0.052*	1.22	1.00	1.48
B1	1	0.733	0.100	7.34	0.000***	2.08	1.71	2.53
Log likelihood = -1,399.840, $\chi^2 = 104.787$, DF = 6, P-value = 0.000***								
Goodness-of-fit Statistics								
Method		χ^2	DF	P				
Pearson		1,778.84	1,724	0.175				
Deviance		2,274.15	1,724	0.000				
Hosmer-Lemeshow		18.86	8	0.016				
Measures of Association								
Somers' D		0.25						
Goodman-Kruskal γ		0.25						
Kendall's τ_A		0.12						
Pair Type		Count	Percentage					
Concordant Pairs		689,851	61.4					
Discordant Pairs		413,815	36.8					
Tied Pairs		20,254	1.8					
Total		1,123,920	100.0					

징적인 내용에만 한정하여 아래와 같이 논의한다.

첫째, X1, X2, X3 중 양(+)의 추정계수 $\hat{\beta}_3 = 0.014$, P-value = 0.003*** X3만 반응변수 B2에 통계적으로 유의한 예측변수로 판단된다. 하지만, 추정계수 승산비 $\exp(\hat{\beta}_3)$ 에 대한 95% 신뢰 구간(Confidence Interval, CI) 하한(Lower Bound, LB) = 1.00, 상한(Upper Bound, UB) = 1.02이므로, X3 값 1단위 증가에 따라 B2 = 1 값 즉, 성공 확률 승산비 증가는 매우 작다. 즉, X1, X2, X3 관련 (1) 추정계수가 기대부호와는 다르거나(즉, X2 경우), (2) 통계적으로 유의하지 않거나(즉, X1 경우), (3) 추정계수가 기대부호를 갖고 통계적으로는 유의하지만 그 승산비 증가는 미미한 것으로 종합된다. 그러므로, R&D 자원의 투입이 시간적으로 인접한 단기적, 기술적 성과 Y1에는 유의한 영향을 줄 수 있지만, 이후 시간적 경과에 따라 떨어진 중기적, 경제적 성과 Y2에는 Y1만큼 유의한 영향을 발휘하기란 쉽지 않다고 해석될 수 있다.

둘째, T1은 음(-)의 추정계수 $\hat{\beta}_{T1} = -0.762$, P-value = 0.000** 이므로, 반응변수 B2에 대해 통계적으로 유의한 예측변수로 판단된다. 승산비 $\exp(\hat{\beta}_{T1}) = 0.47$ 로 그 값이 1보다 작게 계산되어, T1 = 0 기준수준에서 T1 = 1로 1단위 증가할 때, B2 = 1 값 즉, 성공 확률 승산비는 0.47배 감소한다. 한편, <Table 3>, Model (1) 반응변수 B1에 대한 T1 승산비 $\exp(\hat{\beta}_{T1}) = 0.69$ 와 비교할 때, “중소기업”(즉, T1 = 0) 대비 “대기업”(즉, T1 = 1) 주관 관측 개체의 매출액 성과 창출 성공 확률의 저하 정도는 특허 등록 성과 창출 성공 확률의 저하 정도보다 더 클 수 있을 것으로 해석된다. T2는, 양(+)의 추정계수 $\hat{\beta}_{T2} = 0.195$, P-value = 0.052* 이므로, 역시 반응변수 B2에 통계적으로 유의한 예측변수로 판단된다. 승산비 $\exp(\hat{\beta}_{T2}) = 1.22$ 이므로, T2 = 0 기준수준에서 T2 = 1로 1단위 증가할 때, B2 = 1 값 즉, 성공 확률 승산비는 1.22배 증가한다. <Table 3>, Model (1) 반응변수 B1에 대한 T2 승산비 $\exp(\hat{\beta}_{T2}) = 1.49$ 와 비교할 때, “단독 R&D”(즉, T2 = 0)에서 “협력 R&D”(즉, T2 = 1)로 바뀌면 관측 개체의 매출액 성과 창출 성공 확률의 개선 정도는 특허등록 성과 창출 성공 확률의 개선 정도보다는 상대적으로 약간 낮은 것으로 해석된다.

셋째, R&D 논리모형 상 직전 성과 창출 여부 관련 성공-실패 예측변수 B1 대비 반응변수 B2와의 관계를 살펴보면, B1은 양(+)의 추정계수 $\hat{\beta}_{B1} = 0.733$, P-value = 0.000** 이므로, 반응변수 B2에 대해 통계적으로 유의한 예측변수로 판단된다. 승산비 $\exp(\hat{\beta}_{B1}) = 2.08$ 로 결국, B1 = 0 기준수준에서 B1 = 1로 1단위 증가할 때, B2 = 1 값 즉, 성공 확률 승산비는 2.08배 증가하는 것으로 분석된다. 특히, B1은 Model (2)에 포함된 총 6가지 예측변수 중 통계적 유의성 및 승산비 절대값이 가장 큰 예측변수임을 알 수 있다. 그러므로, 특허등록 성과 창출에 성공한(즉, B1 = 1) 관측 개체의 매출액 성과 창출 성공 확률이 크게 제고될 수 있음이 확인된다.

<Table 5>는 Model (3) 관련 주요 분석 결과이다. Model (3)는 Model (2)에 포함된 예측변수 6가지와 함께 Model (2) 반응변수 B2를 예측변수로 갖는, 총 7가지 예측변수로 반응변수 B3를 설명한다. 반응변수 B3 = 1 값이 성공 즉, 기준사건으로 정의되며, 2-수준 요인 예측변수로 지정된 T1, T2, B1 및 B2의 기준수준은 각각 T1 = 0, T2 = 0, B1 = 0, B2 = 0 값으로 정의된다.

첫째, X1, X2, X3 중 양(+)의 추정계수 $\hat{\beta}_3 = 0.015$, P-value = 0.018** X3만이 반응변수 B3에 통계적으로 유의한 예측변수이다. 하지만, 추정계수 승산비 $\exp(\hat{\beta}_3)$ 에 대한 95% CI LB = 1.00 이므로, X3 값 1단위 증가에 따라 B3 = 1 값 즉, 성공 확률 승산비 증가는 매우 작다. 또한, Model (2)와 비교할 때 Model (3)에 포함된 X1, X2, X3의 P-value가 모두 증가되어 통계적 유의성은 더욱 낮아졌고, 추정계수가 기대부호와는 다른 예측변수가 X1, X2 두 가지로 늘어난 점 등이 확인된다. 따라서 역시, R&D 자원 투입은 시간적 경과에 따라 가장 멀리 떨어진 장기적, 사회·경제적 과급효과 성과 Y3에는 Y2, Y1만큼은 유의한 영향을 발휘하기가 쉽지 않을 수 있다고 해석 가능하다.

둘째, T1은 음(-)의 추정계수 $\hat{\beta}_{T1} = -0.594$, P-value = 0.007*** 이므로, 반응변수 B3에 대해 통계적으로 유의한 예측변수로 판단된다. 승산비 $\exp(\hat{\beta}_{T1}) = 0.55$ 로 그 값이 1보다 작게 계산되어, T1 = 0 기준수준에서 T1 = 1로 1단위 증가할 때, B3 = 1 값 즉, 성공 확률 승산비는 0.55배 감소한다. 즉, “중소기업”(즉, T1 = 0) 대비 “대기업”(즉, T1 = 1) 주관 관측 개체의 신규고용 성과 창출 성공 확률은 저하된다. 반면, T2는 양(+)의 추정계수 $\hat{\beta}_{T2} = 0.152$, P-value = 0.277이므로, 반응변수 B3에 통계적으로 유의한 예측변수는 아닌 것으로 판단된다. 즉, 협력 유형은 관측 개체의 신규고용 성과 창출 성공 확률에 영향을 주지 않는 것으로 해석된다.

셋째, R&D 논리모형 상 직전 성과 창출 여부 관련 성공-실패 예측변수 B2, B1 대비 반응변수 B3와의 관계를 살펴보면, B2는 양(+)의 추정계수 $\hat{\beta}_{B2} = 3.473$, P-value = 0.000***이므로, 역시 반응변수 B3에 대해 통계적으로 유의한 예측변수로 판단된다. 승산비 $\exp(\hat{\beta}_{B2}) = 32.22$ 로 결국, B2 = 0 기준수준에서 B2 = 1로 1단위 증가할 때, B3 = 1 값 즉, 성공 확률 승산비는 32.22배 증가한다. 마찬가지로, B1은 양(+)의 추정계수 $\hat{\beta}_{B1} = 0.392$, P-value = 0.005***이므로, 역시 반응변수 B3에 대해 통계적으로 유의하다. 승산비 $\exp(\hat{\beta}_{B1}) = 1.48$ 로 결국, B1 = 0 기준수준에서 B1 = 1로 1단위 증가할 때, B3 = 1 값 즉, 성공 확률 승산비는 1.48배 증가한다. 특히, B2, B1은 Model (3)에 포함된 총 7가지 예측변수 중 통계적 유의성 및 승산비 절대값이 가장 큰 예측변수들로 확인된다. 더불어, R&D 논리모형 상 반응변수 B3 직전에 위치한 예측변수 B2 영향력이 B1 대비 상대적으로 더 큰 것도 확인된다. 그러므로, 관측 개체의 매출액 성과 창출 성공(즉, B2 = 1)이 후속 신규고용 성과 창출 성공(즉, B3 = 1)에 결정적 인자로 작용될 수 있음을 알 수 있다.

결국, <Table 3>부터 <Table 5>까지의 완전모형에 기초한 로지스틱 회귀분석 결과를 제 3장에서 설정한 가설과 결부시켜 종합하면 다음과 같이 요약된다. 첫째, 가설 1은 채택된다. 구체적으로, 세 가지 R&D 입력변수 X1, X2, X3 모두가 통계적으로 유의한 양(+)의 상관성을 보인 경우는 Model (1) 즉, 특허등록 성과 창출과 관련된 경우 단 한 가지로 확인된다. 반면, 매

출액, 신규고용과 관련된 나머지 두 가지 경우에는 일관된 기대부호를 취하면서 동시에 통계적으로 유의한 결과는 확인할 수 없었다. 이는 이후 설명될 제 4.5절의 축소모형에 기초한 로지스틱 회귀분석 결과에서도 재차 확인되는 사항이다. 둘째, 가설 2는 채택된다. 즉, 직전 성과 대비 후속 성과 간 강한 양(+)의 상관성이 확인된다. 특히, 매출액 성과 창출에 가장 큰

Table 5. Binary logistic regression analysis with Model (3)

Model Description								
Model								
B3 versus X1, X2, X3, T1, T2, B1, B2								
Link Function								
Logistic Distribution								
Response Variable		Value	Count					
B3		1	678	(Success)				
		0	1,500	(Failure)				
		Total	2,178					
Factor		Level	Value					
T1		2	0, 1					
T2		2	0, 1					
B1		2	0, 1					
B2		2	0, 1					
Binary Logistic Regression Analysis								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-2.832	0.153	-18.55	0.000			
X1		-0.022	0.066	-0.33	0.745	0.98	0.86	1.11
X2		-0.048	0.049	-0.97	0.333	0.95	0.87	1.05
X3		0.015	0.006	2.37	0.018**	1.02	1.00	1.03
T1	1	-0.594	0.221	-2.69	0.007***	0.55	0.36	0.85
T2	1	0.152	0.140	1.09	0.277	1.16	0.88	1.53
B1	1	0.392	0.139	2.83	0.005***	1.48	1.13	1.94
B2	1	3.473	0.136	25.60	0.000***	32.22	24.70	42.03
Log likelihood = -816.857, $\chi^2 = 1,067.583$, DF = 7, P-value = 0.000***								
Goodness-of-fit Statistics								
Method		χ^2	DF	P				
Pearson		1,837.30	1,804	0.287				
Deviance		1,423.15	1,804	1.000				
Hosmer-Lemeshow		25.06	8	0.002				
Measures of Association								
Somers' D		0.75						
Goodman-Kruskal γ		0.76						
Kendall's τ_A		0.32						
Pair Type		Count	Percentage					
Concordant Pairs		887,178	87.2					
Discordant Pairs		122,127	12.0					
Tied Pairs		7,695	0.8					
Total		1,017,000	100.0					

양(+)¹의 관계 예측변수는 B1(즉, 특허등록 성과 창출 여부), 신규고용 성과 창출에 가장 큰 양(+)¹의 관계 예측변수는 B2(즉, 매출액 성과 창출 여부)로 확인된다. 가설 2에 대한 이와 같은 판정은, 이후 설명될 제 4.5절 축소 모형에 기초한 로지스틱 회귀분석 결과에서도 동일하게 도출된다. 셋째, 가설 3은 부분적으로 채택된다. 먼저, 가설 3-1의 경우, “대기업” 대비 “중소기업”의 특허등록, 매출액 및 신규고용 성과 창출 성공 확률이 더 큰 것으로 확인된다. 한편, 가설 3-2의 경우, “단독 R&D” 대비 “협력 R&D”의 특허등록 및 매출액, 이 두 가지 경우에서만 성과 창출 성공 확률이 더 큰 것으로 판명된다.

4.4 R&D 입력변수 상관분석

앞서 설명된 Model (2) 반응변수 B2, Model (3) 반응변수 B3 각각에 대해 세 가지 R&D 입력변수 X1, X2, X3 계수 추정기 기대와는 달리 일관적이지 않다. 이와 같은 현상의 원인 중 한 가지로 세 가지 R&D 입력변수 간 다중공선성(Multicollinearity) 문제를 지적할 수 있다. 일반적인 회귀모형 분석 결과에서는 분산확대인자(Variance Inflation Factor, VIF) 등으로 예측변수 간 다중공선성이 쉽게 확인된다. 본 실증분석에서는 분석 샘플의 비모수적 특성을 고려하여 <Table 6>과 같이 X1, X2, X3 간 모수적, 비모수적 상관분석을 실시한다. 모수적 상관계수 Pearson’s *r* 및 비모수적 상관계수 Kendall’s τ_B , Spearman’s ρ_s , 총 세 가지 값이 정리된다.

<Table 6>은 X1, X2, X3 간 매우 강한 상관성을 보여준다. 그러므로, X2, X3가 X1에 의해 결정 가능하다는 전제 하에, X1에 대한 의존성이 높은 X2, X3 두 가지 입력변수들이 제거한 축소 모형으로 재분석하여 앞서 설명된 회귀분석 결과와의 일관성을 추가적으로 확인할 필요성이 있다. 한편, 세 가지 R&D 입력변수 X1, X2, X3와 관련하여, 만약 추정계수 부호가 양(+)¹의 기

대부호만을 갖는 축소모형으로 제시될 경우, 이를 제공받는 실무자는 축소모형을 활용하여 특정 R&D 성과에 대해 그 R&D 성과 창출 성공 확률을 계산하고 예측하고자 하는 데에 더 큰 관심을 보일 수 있으리라 판단된다.

4.5 로지스틱 회귀분석 - 축소모형

<Table 7>은 앞선 세 가지 완전모형 각각 대해서 X2, X3 두 가지 예측변수들을 제거하고 재분석한 결과이다. <Table 7.(a)>는 반응변수 B1 대비 예측변수 X1, T1, T2 세 가지에 대한 회귀분석 결과이다. 추정계수 부호가 Model (1) 분석 결과와 일치하면서 통계적으로 유의한 예측변수는 X1, T2, 두 가지가 확인된다. <Table 7.(b)>는 반응변수 B2 대비 예측변수 X1, T1, T2, B1 4가지에 대한 회귀분석 결과이다. 추정계수 부호가 Model (2) 분석 결과와 일치하면서 통계적으로 유의한 예측변수는 T1, T2, B1 세 가지가 확인된다. 마지막으로, <Table 7.(c)>는 반응변수 B3 대비 예측변수 X1, T1, T2, B1, B2 5가지에 대한 회귀분석 결과이다. 추정계수 부호가 Model (3) 분석 결과와 일치하면서 통계적으로 유의한 예측변수는 T1, B1, B2 세 가지가 확인된다. <Table 7> 회귀분석 결과에서도, X1은 반응변수 B1, B2, B3에 대해서 모두 양(+)¹의 추정계수 부호를 갖지만, 통계적 유의성이 확인된 반응변수는 오직 B1뿐이므로, R&D 자원 투입의 영향력은 시간적 경과에 따라 약화될 수 있음을 재확인된다.

한편, <Table 7>의 축소 모형에 기초한 로지스틱 회귀분석 결과로써 제 3장에 제시된 가설 검정을 실시하면, 이는 앞선 제 4.3절 완전모형의 해석과 관련된 부분과 일치함이 확인된다. 단, 한 가지 차이가 발생하는 부분이 가설 3-1인데, 축소 모형의 경우에서는 “대기업” 대비 “중소기업”의 매출액 및 신규고용, 이 두 가지 경우에서만 성과 창출 성공 확률이 더 큰 것으로 해석

Table 6. Correlation coefficients of R&D input variables

		X1 (R&D Government Subsidy)	X2 (R&D Period)
X2 (R&D Period)	Pearson’s <i>r</i>	0.617	
	(P-value)	(0.000 ^{***})	
	Kendall’s τ_B	0.693	
	(P-value)	(0.000 ^{***})	
X3 (R&D Workforce)	Spearman’s ρ_s	0.844	
	(P-value)	(0.000 ^{***})	
	Pearson’s <i>r</i>	0.596	0.434
	(P-value)	(0.000 ^{***})	(0.000 ^{***})
	Kendall’s τ_B	0.411	0.397
	(P-value)	(0.000 ^{***})	(0.000 ^{***})
	Spearman’s ρ_s	0.549	0.521
	(P-value)	(0.000 ^{***})	(0.000 ^{***})

Table 7. Binary logistic regression analyses with reduced models

(a)								
Model								
B1 versus X1, T1, T2								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-1.242	0.081	-15.42	0.000			
X1		0.458	0.044	10.41	0.000***	1.58	1.45	1.72
T1	1	-0.094	0.148	-0.64	0.525	0.91	0.68	1.22
T2	1	0.647	0.098	6.57	0.000***	1.91	1.57	2.32
Log likelihood = -1,340.988, $\chi^2 = 269.316$, DF = 3, P-value = 0.000***								
(b)								
Model								
B2 versus X1, T1, T2, B1								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-0.773	0.077	-10.04	0.000			
X1		0.020	0.037	0.55	0.579	1.02	0.95	1.10
T1	1	-0.750	0.151	-4.96	0.000***	0.47	0.35	0.64
T2	1	0.160	0.096	1.67	0.096*	1.17	0.97	1.42
B1	1	0.679	0.097	7.04	0.000***	1.97	1.63	2.38
Log likelihood = -1,409.137, $\chi^2 = 86.195$, DF = 4, P-value = 0.000***								
(c)								
Model								
B3 versus X1, T1, T2, B1, B2								
Variable	Factor Level	Coefficient	SE	Z	P-value	Odds Ratio	Odds Ratio 95% CI	
							LB	UB
Intercept		-2.867	0.142	-20.18	0.000			
X1		0.026	0.052	0.50	0.618	1.03	0.93	1.14
T1	1	-0.525	0.213	-2.46	0.014**	0.59	0.39	0.90
T2	1	0.161	0.135	1.19	0.234	1.17	0.90	1.53
B1	1	0.374	0.135	2.78	0.005***	1.45	1.12	1.89
B2	1	3.485	0.135	25.74	0.000***	32.62	25.02	42.53
Log likelihood = -820.073, $\chi^2 = 1,061.152$, DF = 5, P-value = 0.000***								

된다.

5. 결론

국가연구개발사업으로 창출된 성과를 분석하고, 그 결과를 후속 사업 계획 수립과 전개 및 이와 관련된 예산 결정과 집행 등에 연계시키는 과정을 통해, R&D 성과평가 본연의 책무성(Accountability)을 제고할 수 있다고 판단된다. 하지만, R&D 자원 투입과 이후 창출된 성과 사이 시차에 대한 고려, 그리고 특히, 기술혁신 국가연구개발사업의 경우에는 성과 창출 실패

R&D 과제들이 다수 발생하는 등의 현실적인 어려움으로 인해, 현재까지 이와 관련된 실증분석에 기초한 충분한 선행연구들은 매우 제한적이었다.

본 연구는 대한민국 정부의 대표적인 기술혁신 국가연구개발사업 중 한 가지인 지식경제기술혁신사업을 대상으로 2007년부터 2011년까지 5년 간 대기업 및 중소기업이 주관한 정부 지원 종료 R&D 과제로 구성된 분석 샘플 $n=2,178$ 개를 분석하였다. 본 실증분석 결과의 주요 내용을 요약하면 다음과 같다.

첫째, R&D 입력변수 X1(즉, 정부지원금액)은 특허등록 성과 창출에 통계적으로 유의한 양(+)의 관계를 보였다. 하지만, R&D 논리모형 상 시간적 경과에 따라 그 강도가 약화되어 매

출액, 신규고용 성과 창출에는 양(+)의 관계는 보였지만 통계적 유의성은 확인되지 않았다. 둘째, R&D 논리모형 상 직전 대비 후속 성과 창출 성공-실패 여부 사이에는 강한 의존 관계가 존재하는 것으로 분석되었다. 즉, 매출액 성과 창출에 가장 큰 양(+)의 관계 예측변수는 B1, 신규고용 성과 창출에 가장 큰 양(+)의 관계 예측변수는 B2로 확인되었다. 셋째, 주관기관 유형 영향요인변수 T1의 경우, “대기업” 대비 “중소기업” 매출액 및 신규고용 성과 창출 성공 확률이 더 큰 것으로, 협력 유형 영향요인변수 T2의 경우, “단독 R&D” 대비 “협력 R&D” 특허등록 및 매출액 성과 창출 성공 확률이 더 큰 것으로 판명되었다. 넷째, 세가지 R&D 입력변수 간 강한 상관성에 확인되어 회귀모형 계수 추정 시 세심한 주의가 요구될 수 있다고 판단된다.

특히, R&D 성과관리 측면에서의 정책적 함의를 도출하면 아래와 같다. 첫째, 주관기관 및 협력 유형별로 기술적, 경제적 성과 창출 여부에 통계적으로 유의한 차이가 존재함으로, 후속 사업 계획 수립과 전개 시 보다 세심한 R&D 과제 선별이 요구된다. 둘째, 직전 대비 후속 성과 창출 성공-실패 여부 사이의 강한 의존 관계에 기초할 때, R&D 논리모형 상 순차적 성과 창출을 유도함이 R&D 과제의 성과 창출 수명을 연장할 수 있다고 판단된다. 셋째, 투입 대비 성과 간 관계가 R&D 논리모형 순서 상 증기적, 경제적 결과 이후 크게 감소될 수 있으므로, 투입 대비 창출된 성과의 크기를 측정하는 효율성 지수의 발굴과 활용 시 투입 영향력이 실존하는 단기적, 기술적 성과에 주로 한정하는 방안도 고려될 필요가 있다.

한편, 본 연구모형에서 검토된 8가지 변수 중, 세 가지 R&D 입력변수들을 제외한 나머지 5가지 성과 및 영향요인 관련 항목들은 모두 이진 변수들로 변환되었으므로, 이 항목들이 취하는 값을 좀 더 다항화함으로써 순서형(Ordinary) 로지스틱 회귀모형 등으로 분석을 시도할 수 있다. 더불어, 대학, 국공립 연구소 등과 같은 순수 교육·연구기관도 향후 실증분석에서는 포함시켜 본 연구의 범위를 확장시킬 수 있다. 단, 이 경우에는 논문, 기술이전 등과 같이 새로운 성과 항목들이 추가로 연구모형에 포함되어 설계되어야 할 것으로 판단된다. 또한, 최근 국가연구개발사업과 관련하여 정부지원금액 수혜기관 유형별로 추구하는 R&D 성과에 유의한 차이가 있음이 실제로 확인되고 있으므로, 이와 같은 현상을 정확히 반영할 수 있는 정부지원금액 수혜기관 유형별 맞춤형 R&D 성과평가 체계의 수립에 유용한 결과를 제공할 수 있는 이론적, 실증적 주제들도 향후 연구에서 고려될 수 있다.

참고문헌

Belderbos, R., Carree, M., and Lokshin, B. (2004), Cooperative R&D and Firm Performance, *Research Policy*, **33**(10), 1477-1492.
 Berchicci, L. (2013), Towards an Open R&D System : Internal R&D Investment, External Knowledge Acquisition and Innovative Performance, *Research Policy*, **42**(1), 117-127.

Bickman, L. (1987), The Functions of Program Theory, *Special Issue : Using Program Theory in Evaluation, New Directions for Program Evaluation*, **1987**(33), 5-18.
 Chen, X., Wang, X., and Wu, D. D. (2011), Analysing Firm Performance in Chinese IT Industry : DEA Malmquist Productivity Measure, *International Journal of Information Technology and Management*, **10**(1), 3-23.
 Fritsch, M. and Lukas, R. (2001), Who Cooperates on R&D?, *Research Policy*, **30**(2), 297-312.
 Guan, J. and Chen, K. (2010), Modeling Macro-R&D Production Frontier Performance : An Application to Chinese Province-level R&D, *Scientometrics*, **82**(1), 165-173.
 Hosmer, D. W. and Lemeshow, S. G. (2000), *Applied Logistic Regression*, 2nd ed., New York NY : John Wiley and Sons.
 Hsu, F. M. and Hsueh, C. C. (2009), Measuring Relative Efficiency of Government-Sponsored R&D Projects : A Three-Stage Approach, *Evaluation and Program Planning*, **32**(2), 178-186.
 IBM SPSS (2009), *PASW Statistics Release 18*, Armonk NY : IBM Corp.
 Korea Evaluation Institute of Industrial Technology(KEIT) (2010), *2010 Performance Investigation and Analysis of Knowledge Economy Technology Innovation Program*, Korea : KEIT.
 Korea Evaluation Institute of Industrial Technology(KEIT) (2011), *2011 Performance Investigation and Analysis of Knowledge Economy Technology Innovation Program*, Korea : KEIT.
 Korea Evaluation Institute of Industrial Technology(KEIT) (2013), *2012 Performance Investigation and Analysis of Knowledge Economy Technology Innovation Program*, Korea : KEIT.
 Korea Institute of Science and Technology Information(KISTI) (2008), Law-Enforcement Ordinance-Enforcement Regulations, Available Online at : <http://www.ntis.go.kr/ThMain.do> (Accessed 1 August 2013).
 Laursen, K. and Salter, A. (2006), Open for Innovation : The Role of Openness in Explaining Innovation Performance among U.K. Manufacturing Firms, *Strategic Management Journal*, **27**(2), 131-150.
 McLaughlin, J. A. and Jordan, G. B. (1999), Logic Models : A Tool for Telling Your Program's Performance Story, *Evaluation and Program Planning*, **22**(1), 65-72.
 Ministry of Knowledge Economy(MKE) (2008), *2007 Electric Power Industry R&D Programs*, Korea : MKE.
 Ministry of Knowledge Economy and Korea Institute for Advancement of Technology(KIAT)(MKE · KIAT) (2012), *2012 Guideline of Knowledge Economy R&D Performance Index Design*, Korea : MKE · KIAT.
 Ministry of Science and Technology and Office of Science and Technology Innovation(MST · OSTI) (2008), *2008 Internal Evaluation Manual of National R&D Programs*, Korea : MST-OSTI.
 Minitab^R (2005), *Minitab^R Release 14.20 StatGuide*, State College PA : Minitab Inc.
 Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001), *Introduction to Linear Regression Analysis*, 3rd ed., New York NY : John Wiley and Sons.
 Office of Management and Budget(OMB) and Office of Science and Technology Policy(OSTP) (2012), *Memorandum for the Heads of Executive Departments and Agencies, Subject : Science and Technology Priorities for the FY 2014 Budget*, Available Online at : <http://www.whitehouse.gov/omb/>, (Accessed 1 August 2013).
 Office of Science and Technology Policy(OSTP) (2012), *Innovation for America's Economy, America's Energy, and American Skills : Science, Technology, Innovation, and STEM Education in the 2013 Budget*, Available Online at : <http://www.ostp.gov/>, (Accessed 1 August 2013).

- Robin, S. and Schubert, T. (2013), Cooperation with Public Research Institutions and Success in Innovation : Evidence from France and Germany, *Research Policy*, **42**(1), 149-166.
- Ruegg, R. (2006), *Bridging from Project Case Study to Portfolio Analysis in a Public R&D Program : A Framework for Evaluation and Introduction to a Composite Performance Rating System*, Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology(NIST), U.S. Department of Commerce, Gaithersburg, MD.
- Ruegg, R. and Feller, I. (2003), *A Toolkit for Evaluating Public R&D Investment : Models, Methods and Findings from ATP's First Decade*, Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology(NIST), U.S. Department of Commerce, Gaithersburg, MD.
- Shipp, S., Chang, C., and Wisniewski, L. (2005), *Evaluation Best Practices and Results : The Advanced Technology Program*, Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology(NIST), U.S. Department of Commerce, Gaithersburg, MD.
- W. K. Kellogg Foundation(WKCF) (2004), *W. K. Kellogg Foundation Logic Development Guide*, Battle Creek MI.
- Wholey, J. S. (1983), *Evaluation and Effective Public Management*, Boston MA : Little Brown.
- Wholey, J. S. (1987), Evaluability Assessment : Developing Program Theory, *Special Issue : Using Program Theory in Evaluation, New Directions for Program Evaluation*, **1987**(33), 77-92.
- Wu, W., Tsai, H., Cheng, K. and Lai, M. (2006), Assessment of Intellectual Capital Management in Taiwanese IC Design Companies : Using DEA and the Malmquist Productivity Index, *R&D Management*, **36**(5), 531-545.

<부 록>

<Figure A.1>은 본 연구의 실증분석과 관련된 전체적인 흐름을 요약한 절차화된 그림이다. 총 일곱 단계로 구분하여 연속적인 로지스틱 회귀모형을 이용한 R&D 투입 및 성과 관계를 분석하는 전 과정을 정리하고 있다. Stage I, II는 제 1장, 제 2장에 대응되는 내용이며, Stage III은 제 3장 연구모형 부분에서 설명된 모형 수립, 변수 정의 및 가설 설정 등과 관련된 내용으로 구성됨을 알 수 있다. Stage IV, V, VI은 제 4장 실증분석 결

과를 어떻게 해석할지와 관련된 자세한 절차와 이 때 사용되는 통계량 및 방법론 등을 요약하여 보여준다. 한편, Stage VI, 축소모형과 관련된 세부 절차는 Stage IV, 완전모형과 관련된 동일한 절차를 거쳐 진행하되, 직전 단계인 예측변수 간 상관 분석 결과를 반영하여 축소모형이 수립됨을 나타낸다. Stage VII은 제 5장에 대응되는 내용이 실시됨을 언급한 것이다.

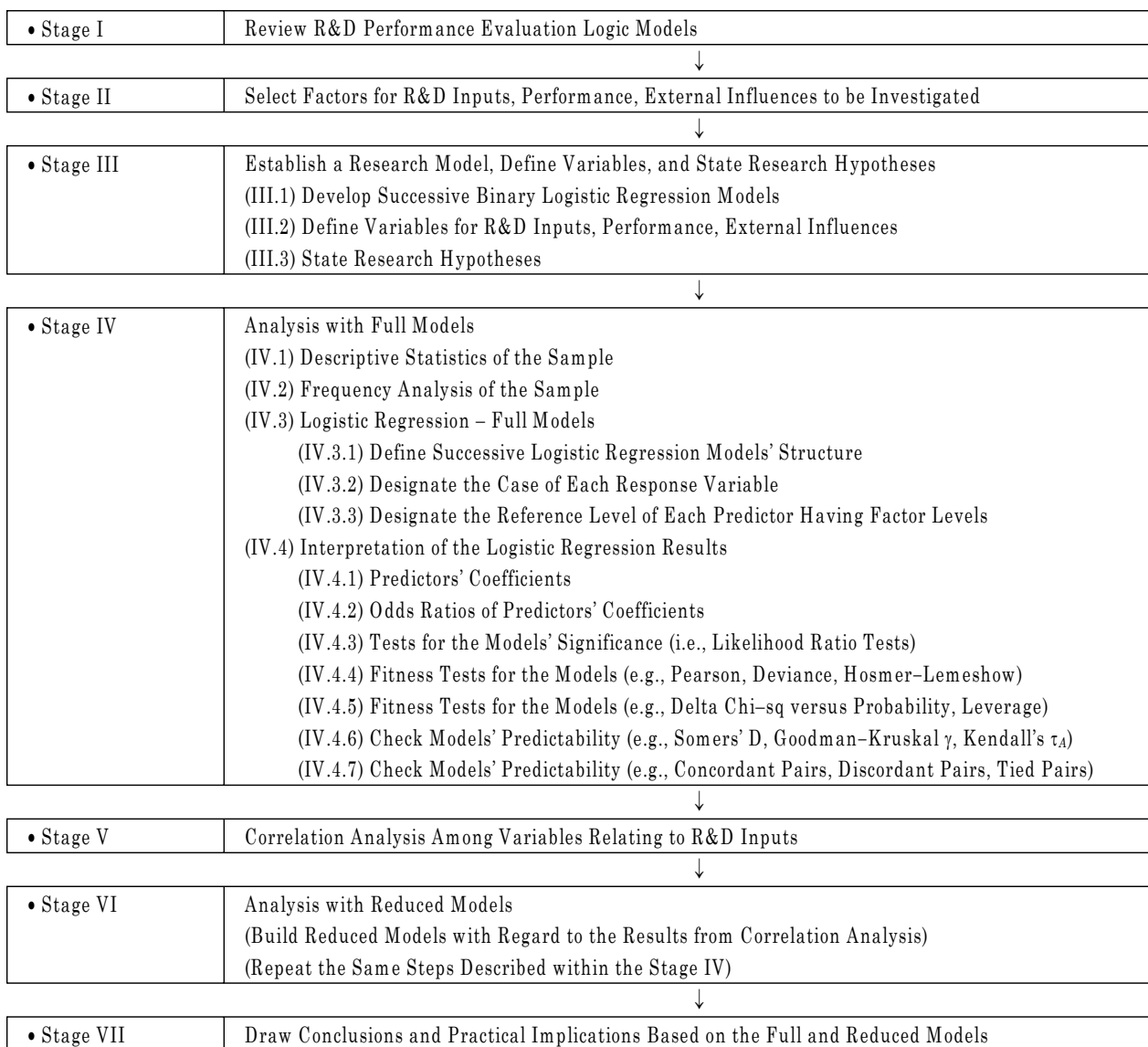


Figure A.1. A schematic diagram for the empirical analysis procedure of the present study