

금융산업에서의 데이터 품질이 경제적인 성과에 주는 영향의 실증분석

An Empirical Analysis on the Effect of Data Quality on Economic Performance in the Financial Industry

이 상 호 (Sangho Lee)

선문대학교 IT경영학과 조교수 & 경희대학교

박 주 석 (Joo Seok Park)

경희대학교 경영대학 교수

김 재 경 (Jae Kyeong Kim)

경희대학교 경영대학 & 경영연구원 교수, 교신저자

요 약

본 연구는 한국 금융산업에서의 기업 데이터 품질이 경제적인 성과에 주는 영향을 2008년과 2009년 자료를 이용하여 실증적으로 분석하였다. 데이터 품질은 한국데이터베이스진흥원에서 측정한 데이터 품질관리 프로세스 지수와 데이터 품질기준을 이용하였으며, 기업의 재무 성과는 금융감독원의 금융 통계정보시스템을 이용하여 수집하였다. 분석 결과, 데이터 품질은 기업의 매출액, 영업이익, 부가가치에 통계적으로 유의한 영향을 주었다. 데이터 품질관리 프로세스 지수가 1만큼 증가하면, 부가가치는 약 2.3%가 증가한다. 또한 데이터 품질기준이 1만큼 증가하면, 부가가치는 약 72.6%가 증가한다. 기업의 경영자는 데이터 품질을 개선시키도록 노력해야 하고, 경제정책 입안자들은 기업들이 데이터 품질에 관심을 갖고 데이터 품질을 개선시키도록 하는 유인책과 정책 방향을 수립해야 한다.

키워드 : 데이터 품질, 데이터 품질기준, 데이터 품질관리 프로세스, 기업 성과, 금융산업

I. 서 론

미국의 소매업체는 연 매출액의 1%인 약 1조 원을 제품의 수요 예측과 창고관리를 위한 의사결정지원시스템에 사용하고 있다(DeHoratius and Raman, 2008). 정확하지 않은 재고 정보(데이터)는 이들 시스템의 가치를 훼손시키고 있다. 낮은 품질의 정보(데이터)를 이용한 의사결정지원시

스템의 지원을 받은 의사결정의 결과가 좋기는 어렵다. 정확한 정보에 의한 의사결정은 기업의 성과를 증가시킬 수 있지만, 정확하지 않은 정보에 의한 의사결정은 기업의 성과를 감소시킬 뿐만 아니라 기업의 흥망에도 영향을 준다. DeHoratius and Raman(2008)은 미국의 한 소매업체의 37개 지점에서 획득한 약 37만 개의 재고 레코드에서 약 65%의 부정확한 데이터를 발견하였다. 소매업체의 부정확한 데이터는 소매업체의 운영 성과에 부정적인 영향을 줄뿐만 아니라 채찍 효과(bullwhip effect)(Laudon and Laudon, 2009)때문에

† 본 연구는 2010년도 경희대학교 연구비(KHU-20100127)를 지원받았습니다.

공급망 파트너의 성과에 더 큰 부정적인 영향을 주게 된다.

한국데이터베이스진흥원(이하 진흥원)이 2010년도에 수행한 국내 기업과 공공기관의 데이터 품질 진단 결과에 따르면, 민간부문의 13개 데이터베이스의 오류율은 2%이고, 공공부문의 62개 데이터베이스의 오류율은 6%이었다(디지털타임스, 2010). 그리고 공공부문 데이터베이스의 27%는 데이터베이스 정비가 필요한 것으로 조사되었다.

이러한 데이터 오류가 성과에 부정적인 영향을 준다는 것은 연구자들과 실무자들은 모두 동의하고 있지만(예, 허희정, 김중우, 2008), 데이터 오류가 성과에 얼마나 부정적인 영향을 주는지를 분석한 연구는 제한적이다. 데이터 품질이 경제적인 성과에 어떠한 영향을 주는지를 실증적으로 연구하고 분석해야 데이터 품질의 개선이 기업 성과에 실질적인 영향을 주는 정도를 파악하고 기업이 데이터 품질의 개선에 어느 정도의 노력을 기울여야 하는지를 파악할 수 있다. 또한 이 영향 관계를 이용하여 금융산업의 경제적인 효과를 추정할 수 있고, 금융산업에서 발견된 영향 관계를 국가 수준에 동일하게 적용할 수 있다면 국가 수준의 경제적인 효과를 추정할 수 있다. 데이터 품질이 성과에 유의한 영향을 주지 못한다면, 데이터 품질을 개선하는 노력은 필요 없을 수 있다. 그러나 데이터 품질이 성과에 유의하게 영향을 준다면, 데이터 품질의 개선 노력(비용)이 성과 향상(효익)과 동일할 때까지 증가시켜야 한다. 본 연구의 결과에 기반하여 정책 입안자와 추진자들은 산업 수준 또는 국가 수준에서의 데이터 품질에 관련된 정책 방향이나 정책을 수립할 수 있다.

연구는 다음과 같이 서술된다. 제 II장은 데이터 품질과 성과와의 관계를 분석한 기존 연구에 대하여 조사하고, 제 III장 연구모형과 자료는 연구모형, 변수들, 자료원을 설명하고, 제 IV장 분석 결과는 연구모형을 이용하여 실증분석한 결과를 서술하며, 마지막으로 제 V장 결론은 연구의

요약, 연구의 기여점과 한계점, 미래연구의 방향 등을 제시한다.

II. 관련 연구 분석

2.1 정보 품질과 성과

정보 품질이 기업 성과에 어떤 영향을 주는가의 주제는 DeLone and McLean(1992)이 관련된 기존 연구를 이용하여 정보시스템 성공 모형을 개발한 이후로 실증연구자들의 많은 관심을 받았다. 약 10년 후, DeLone and McLean(2003)은 자신들의 모형을 이용하여 *Information Systems Research*, *MIS Quarterly*, *Journal of Management Information Systems*와 같은 주요 학술지에 발표된 연구들을 검토하여 갱신된 정보시스템 성공 모형을 발표하였다. 정보 품질이 정보시스템의 사용 의도와 사용자 만족에 영향을 주고, 정보시스템의 사용과 사용자 만족은 순 효익(net benefits)에 영향을 주는 모형이다.

이들 모형을 이용한 실증연구들은 정보 품질이 순효익에 유의한 영향을 주는 것을 보이고 있다(Petter and McLean, 2009). 그러나 이 모형들을 이용한 많은 연구들은 정보 품질 뿐만 아니라 성과 요소인 순효익의 측정도 자신들의 연구를 위한 리커트 5점(또는 7점) 척도를 이용한 하나의 설문지를 이용하여 주관적으로 측정하고 있어 산업이나 국가 수준의 정보 품질을 파악하기는 쉽지 않다.

2.2 데이터 품질과 성과

진흥원은 데이터 품질을 데이터 품질기준(data quality criteria) 관점과 데이터 품질관리 프로세스 관점에서 조사하고 있다(한국데이터베이스진흥센터,¹⁾ 2006). 데이터 품질기준은 데이터의 유

1) 현재는 한국데이터베이스진흥원으로 기관 이름을 변경하였다

효성과 활용성 기준으로 분류된다. 다시 데이터의 유효성은 데이터의 정확성과 일관성으로 세분되며, 데이터의 활용성은 데이터의 유용성, 접근성, 적시성, 보안성으로 세분된다. 그리고 진흥원은 데이터 품질관리 프로세스(data quality management process) 관점에서는 요구사항 관리, 데이터 구조 관리, 데이터 흐름 관리, 데이터베이스 관리, 데이터 활용 관리, 데이터 표준 관리, 데이터 오너쉽 관리, 사용자부 관리로 세분하여 조사한다.²⁾

이창수, 김선영(2010)은 2006년 12월부터 2007년에 3월까지 국내 152개 기업을 대상으로 조사된 데이터 품질 자료와 한국신용평가정보(주)에서 산출하고 있는 기업 신용도와와 상관관계를 분석하였다. 데이터 품질은 데이터 품질관리 프로세스를 세분화한 하위 프로세스를 측정단위로 이용하여 측정하였고, 기업 신용도는 100점 만점으로 환산된 신용 점수를 이용하였다. 데이터 오너쉽 관리를 제외한 다른 하위 프로세스들은 기업 신용도와 통계적으로 유의한 정의 상관관계를 보였다. 또한 기업의 규모가 클수록 데이터 품질은 우수하였으며, 신용 점수는 높았다. 이 연구는 데이터 품질이 성과에 주는 영향을 조사할 때 기업의 크기를 통제할 필요가 있음을 보여준다.

데이터 품질이 성과에 유의한 영향을 준다는 가설과 주장(예, Redman, 1995; 허의정, 김종우, 2008)을 반영하여 진흥원은 국내 산업이나 국가의 데이터 품질의 수준을 측정하고 있다. 그리고 기존 연구에서 이용된 기업 신용도는 성과의 예측 요소가 될 수 있으나, 기업 신용도 자체가 기업의 경제적인 성과는 아니다. 따라서 데이터 품질이

기업의 성과에 의미있는 영향을 준다는 기존 연구(예, Redman, 1995)의 가설과 주장을 뒷받침하기 위하여 기업의 데이터 품질이 기업의 경제적인 성과에 정량적이고 구체적으로 어느 정도 영향을 주는 지를 분석하는 실증적인 연구가 필요하다.

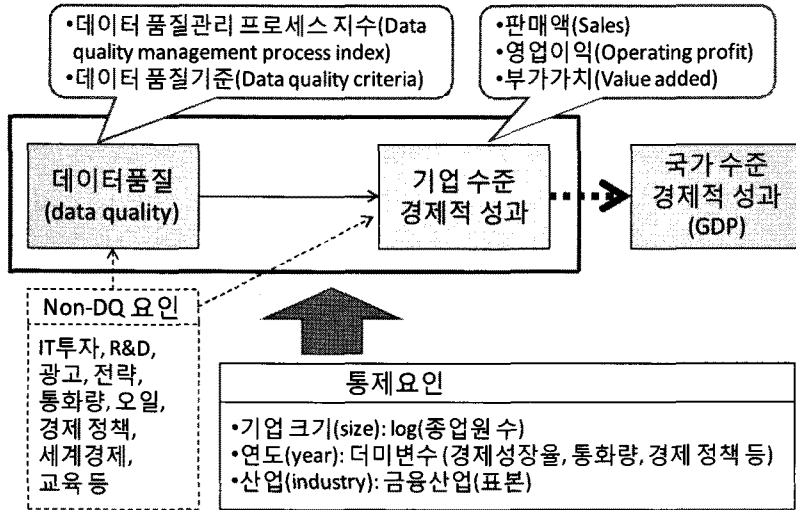
III. 연구모형과 자료

3.1 연구모형

본 연구는 최근까지 연구가 제한적이었던 한국에서 데이터 품질이 기업의 경제적인 성과에 주는 영향을 분석한 실증연구로, <그림 1>과 같은 연구모형을 이용하여 분석한다.

기업 수준의 경제적인 성과는 데이터 품질 뿐만 아니라 IT 투자, 연구개발(R&D: research and development), 광고, 전략, 통화량, 경제정책, 교육 등의 다양한 요인으로부터 영향을 받는다(Lee and Kim, 2006). 그러나 각 요인들은 각 학문 분야(경제학, 마케팅, 전략, 교육학 등)에서 집중적으로 조명하며 분석되고 있으며, 본 연구에서 조명할 부분은 기존 연구(예, Redman, 1995)에서 주장하고 있는 기업의 데이터 품질이 기업 성과에 주는 영향이다. 특히, 기업의 IT 투자는 기업 성과에 영향을 줄 수 있어 통제가 필요하나, 한국의 금융산업의 기업 수준 IT 투자금액 자료의 획득에 제약이 있어 본 연구에서는 고려하지 않는다. 데이터 품질과 관련이 없으면서 기업의 경제적인 성과에 영향을 줄 수 있는 요인중의 일부인 기업의 크기, 경제성장률, 통화량 등은 기존 연구(예, 이상호 등, 2009)와 같이 기업의 크기와 연도를 통제변수로 이용하여 통제하였다. 산업의 정보(데이터)강도 효과(Kim et al., 2009)는 분석 대상 표본을 금융산업으로 한정하여 통제할 수 있었다. 그리고 기업 수준의 경제적 성과를 분석하여 산업이나 국가 수준의 경제적 성과를 추정한다.

2) 본 연구에서는 데이터(정보) 품질과 성과와의 관계에 관련된 기존 연구를 분석한다. 데이터 품질과 데이터 품질관리 프로세스에 대한 기존 연구의 분석은 허희정, 김종우(2008, pp. 92-95)를 참조할 수 있고, 한국데이터베이스진흥원에서 사용하고 있는 설문지의 구체적인 설문 항목과 설문 변수는 이창수, 김선영(2010, pp. 204-206)을 참조할 수 있다.



〈그림 1〉 연구모형

데이터 품질이 기업 성과에 주는 영향은 다음 수식을 이용하여 분석한다.

$$FP_{it} = \alpha + \beta_1 \cdot DQ_{it} + \beta_2 \cdot size_{it} + \beta_3 \cdot D + \epsilon_{it} \quad (1)$$

여기서, FP_{it} 는 기업 성과(즉, 매출액, 영업이익, 부가가치의 로그값), DQ_{it} 는 데이터 품질(즉, 데이터 품질관리 프로세스 지수, 데이터 품질기준), $size_{it}$ 는 기업의 크기(종업원 수의 로그값), D 는 더미변수로 2008년 자료이면 $D = 0$, 2009년 자료이면 $D = 1$, α 는 상수항, β_1 은 데이터 품질이 기업 성과에 영향을 주는 정도를 나타내는 계수, β_2 는 통제변수인 기업의 크기가 기업 성과에 영향을 주는 정도를 나타내는 계수, β_3 는 통제변수인 연도의 효과를 나타내는 계수, 그리고 ϵ_{it} 는 에러항이다.

3.2 자료

본 연구에서 사용되는 데이터 품질관리 프로세스 지수와 데이터 품질기준은 진흥원에서 2008년과 2009년에 조사한 품질 자료를 이용하였다. 이 자료는 한국의 기업 수준에서 실증연구를 위

한 금융산업의 데이터 품질과 관련된 2차 자료(secondary data)의 거의 유일한 자료원으로 추정된다. 그리고 금융산업의 기업 성과와 관련된 자료는 기업들이 금융감독원에 제출 및 발표하는 연간 보고서를 이용하여 구축한 금융감독원의 금융통계정보시스템(2010)을 이용하여 획득하였다.

진흥원에서 2008년과 2009년에 조사한 금융산업의 데이터 품질 자료에는 64개의 기업이 포함되어 있다. 그러나 그중 41개의 기업만이 금융통계정보시스템에 포함되어 있어, 표본 기업의 성과정보를 수집할 수 있었다. 본 연구의 표본 기업들은 금융통계정보시스템에 분류된 금융산업 중 은행업, 생명보험업, 손해보험업, 투자매매중개업 I(증권), 리스업, 할부금융업, 상호저축은행업에 분포되어 있다.

IV. 분석 결과

<표 1>은 기초통계량이다. 기업의 연평균 매출액은 7조 1,400억 원, 최소 매출액은 57억 원, 그리고 최대 매출액은 50조 5천억 원이다. 연평균 영업이익은 1,980억 원, 최소 영업이익은 -6,460억 원(손실), 최대 영업이익은 1조 1,900억 원이다.

<표 1> 기초통계량

변수	설명	최소값	최대값	산술평균	표준편차
Sales	Sales, 매출액	57	505,000	71,400	135,000
OP	Operating profit, 영업이익	-6,460	11,900	1,980	3,650
Profit	After-tax profit, 세후이익(당기순이익)	-1,190	8,920	1,560	2,290
VA	Value added, 부가가치	-3,440	35,500	6,240	8,290
Worker	The number of worker, 종업원 수	16	10,800	2,150	2,530
DQI	Data quality index, 데이터 품질관리 프로세스 지수	43.1	99.6	85.1	14.7
DQC	Data quality criteria, 데이터 품질기준	0.0	4.0	1.2	0.77

참고사항: Sales, OP, Profit, and VA의 단위는 억 원, 표본 수는 41.

연평균 당기순이익은 1,560억 원, 최소 당기순이익은 -1,190억 원(손실), 최대 당기순 이익은 8,920억 원이다. 연평균 부가가치는 6,240억 원, 최소 부가가치는 -3,440억 원, 그리고 최대 부가가치는 3조 5,500천억 원이다. 평균 종업원수는 2,150명, 최소 종업원수는 16명, 최대 종업원수는 10,800명이다. 금융산업 기업의 평균 데이터 품질관리 프로세스 지수는 85.1(100점 만점), 최소 데이터 품질관리 프로세스 지수는 43.1, 최대 데이터 품질관리 프로세스 지수는 99.6이다. 평균 데이터 품질기준은 1.2(5.0만점), 최소 데이터 품질기준은 0.0, 최대 데이터 품질기준은 4.0이다.

종속변수로 사용되는 매출액, 영업이익, 부가가치와 데이터 품질관리 프로세스 지수나 데이터 품질기준의 산포도를 보면, 매출액, 영업이익, 부가가치는 로그값이 이용되어야 함을 알 수

있다. <표 2>는 피어슨 상관관계 분석 결과이다. 기업의 성과를 나타내는 매출액의 로그값, 영업이익의 로그값, 부가가치의 로그값은 1% 유의수준에서 0.85이상의 높은 정의 상관관계를 보이고 있다. 이는 본 연구에서 기업의 성과변수의 선정이 적절하게 일관된다는 것을 의미한다. 데이터 품질관리 프로세스 지수와 데이터 품질기준은 1% 유의수준에서 0.7이상의 높은 정의 상관관계를 보이고 있다. 데이터 품질관리 프로세스 지수는 성과변수들과 0.5내외, 데이터 품질기준은 성과변수들과 0.6내외의 정의 상관관계를 보이고 있다. 통제변수로 사용되는 기업의 크기(종업원 수의 로그값)(size)는 성과변수들과 1% 유의수준에서 0.7이상의 높은 정의 상관관계를 보이고 있어, 기업의 크기가 클수록 성과가 좋을 것이라는 것을 의미할 수 있다. 기업의 크기는 데이터 품

<표 2> 피어슨 상관관계 분석

	logSales	logOP	logVA	DQI	DQC	logWorker(size)
logSales	1.000					
logOP	0.854***	1.000				
logVA	0.888***	0.924***	1.000			
DQI	0.472***	0.486***	0.578***	1.000		
DQC	0.592***	0.560***	0.625***	0.722***	1.000	
logWorker(size)	0.868***	0.718***	0.837***	0.516***	0.542***	1.000

*** $p < 0.01$.

질관리 프로세스 지수와 데이터 품질기준과 0.5 정도의 정의 상관관계를 보이고 있어, 기업의 크기가 클수록 데이터 품질이 좋을 것이라는 것을 의미할 수 있다. 이 결과는 이창수, 김선영(2010)의 발견과 일치한다.

회귀분석을 위한 기본 가정들(예를 들어, 분석 자료에 이상치(outlier) 없음, 설명(독립)변수의 수와 표본 수의 적정 비율, 자료의 선형성과 정규성, 잔차의 동분산성, 잔차간의 비자기상관성, 독립변수간의 비다중공선성 등)(Hair et al., 2005; Lee and Kim, 2006)이 충족됨을 확인하였는데, 이는 기존 연구(Lee and Kim, 2006, p. 53)의 방법들을 이용하였다. 그리고 연구에서 독립변수의 수는 3이고, 표본 수는 30이상으로 적정 비율을 유지하고 있다(<표 3>과 <표 4> 참조).

<표 3>은 데이터 품질관리 프로세스 지수를 이용한 회귀분석 결과이다. 금융산업 기업의 데이터 품질관리 프로세스 지수는 매출액의 로그값에 영향을 주고 있다(모형 1). 모형의 설명력은 27.7%

이고, 모형은 1% 수준에서 유의하다. 그러나 기업의 크기를 통제하면, 데이터 품질관리 프로세스 지수는 매출액의 로그값에 영향을 주지 못한다(모형 2). 데이터 품질관리 프로세스 지수와 영업이익의 로그값에 대한 분석도 매출액의 경우와 유사하다(모형 3과 모형 4).

기업의 데이터 품질관리 프로세스 지수는 기업이 산출하는 부가가치의 로그값에 유의한 영향을 주고 있다. 모형 5는 기업의 크기를 통제하지 않은 상태이며, 이 때 데이터 품질관리 프로세스 지수는 부가가치의 로그값에 0.03의 영향을 준다. 모형의 설명력은 35.7%이다. 그러나 기업의 크기는 기업 성과에 영향을 주기 때문에 기업의 크기를 통제할 필요가 있다(Lee and Kim, 2006). 기업의 크기는 통계적으로 유의하였지만, 유의한 연도 효과는 없었다. 기업의 크기와 연도별 차이 효과를 통제한 상태에서 데이터 품질관리 프로세스 지수는 부가가치의 로그값의 변동의 73.2%를 설명할 수 있으며, 모형은 1% 수준

<표 3> 데이터 품질관리 프로세스 지수의 회귀분석 결과

	<i>logSales</i>		<i>logOP</i>		<i>logVA</i>	
	모형 1	모형 2	모형 3	모형 4	모형 5	모형 6
상수	3.88*** (0.721)	2.72*** (0.448)	2.96*** (0.692)	2.29*** (0.570)	2.90*** (0.603)	2.02*** (0.414)
<i>DQI</i>	0.029*** (0.008)	0.002 (0.006)	0.025*** (0.008)	0.007 (0.008)	0.030* (0.007)	0.010* (0.005)
<i>D</i>	-0.409* (0.241)	0.003 (0.151)	-0.140 (0.242)	0.119 (0.201)	-0.233 (0.201)	0.081 (0.139)
<i>size</i>		1.091*** (0.129)		0.683*** (0.158)		0.833*** (0.117)
관측 수	41	41	33	33	40	40
R^2	0.277	0.754	0.245	0.541	0.357	0.732
Adjusted R^2	0.239	0.734	0.194	0.494	0.322	0.710
<i>F</i> -통계량	7.29***	37.7***	4.86**	11.4***	10.3***	32.8***
더빈-왓슨 <i>d</i>	2.28	1.87	2.20	1.96	2.48	1.78

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

참고사항: 2008년 자료이면 $D = 0$, 2009년 자료이면 $D = 1$. 괄호안의 값은 표준편차.

<표 4> 데이터 품질기준의 회귀분석 결과

	<i>logSales</i>		<i>logOP</i>		<i>logVA</i>	
	모형 7	모형 8	모형 9	모형 10	모형 11	모형 12
상수	5.47*** (0.245)	2.95*** (0.348)	4.33*** (0.258)	2.75*** (0.434)	4.64*** (0.222)	2.62*** (0.323)
<i>DQC</i>	0.668*** (0.148)	0.199* (0.108)	0.562*** (0.152)	0.244† (0.144)	0.628*** (0.131)	0.237** (0.103)
<i>D</i>	-0.300 (0.224)	-0.006 (0.142)	-0.052 (0.230)	0.131 (0.190)	-0.111 (0.196)	0.116 (0.133)
<i>size</i>		0.991*** (0.123)		0.627*** (0.151)		0.805*** (0.114)
관측 수	41	41	33	33	40	40
<i>R</i> ²	0.379	0.774	0.314	0.570	0.396	0.746
Adjusted <i>R</i> ²	0.347	0.755	0.269	0.525	0.363	0.724
<i>F</i> -통계량	11.6***	42.2***	6.88***	12.8***	12.1***	35.2***
더빈왓슨 <i>d</i>	2.23	1.75	2.34	2.02	2.43	1.76

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; † $p = 0.102$.

에서 유의하다(모형 6). 기업의 크기와 연도별 차이 효과를 통제한 상태에서 데이터 품질관리 프로세스 지수는 부가가치의 로그값에 0.01의 영향을 준다. 이를 백분율로 환산하면, 약 2.3%가 된다. 2008년과 2009년도 표본기업의 평균 부가가치는 약 6,240억 원이고, 평균 데이터 품질관리 프로세스 지수는 85.1이다(<표 1>). 금융산업 기업의 데이터 품질관리 프로세스 지수가 86.1로 1만큼 증가하면, 부가가치는 약 6,380억 원으로 약 140억 원(약 2.3%)이 증가할 수 있다.

2008년과 2009년의 금융산업 전체의 연평균 부가가치는 약 65.7조 원이다(한국은행, 2010). 금융산업의 데이터 품질관리 프로세스 지수가 1만큼 증가하면, 금융산업의 연평균 부가가치는 약 67.2조 원으로 약 1.5조 원이 증가한다. 또한 2008년과 2009년 한국의 연평균 부가가치는 약 939조 원이며, 연평균 국내총생산(GDP: gross domestic product)은 약 1,044.8조 원이다(한국은행, 2010). 한국의 데이터 품질관리 프로세스 지수가 1만큼 증가하면, 한국의 부가가치는 약 960.8조 원으로

21.8조 원만큼 증가하고, 국내총생산은 1,069.1조 원으로 24.3조 원만큼 증가한다.3)

<표 4>는 데이터 품질기준을 이용한 회귀분석 결과이다. 금융산업 기업의 데이터 품질기준은 매출액의 로그값에 영향을 주고 있다(모형 7). 모형의 설명력은 37.9%이고, 모형은 1% 수준에서 유의하다. 기업의 크기를 통제하면, 모형의 설명력은 77.4%로 증가한다. 데이터 품질기준은 매출액의 로그값에 10% 유의수준에서 0.199의 영향을 준다(모형 8). 통계적인 유의수준이 조금 낮기($p = 0.102$)는 하지만, 영업이익의 로그 값에 주는 데이터 품질기준의 영향도 매출액의 경우와 유사하다(모형 9와 모형 10).

데이터 품질기준은 부가가치의 로그값에 5% 수준 이상에서 유의한 영향을 주고 있다. 모형 11은 기업의 크기를 통제하지 않은 상태이며, 이 때

3) 산업 특성에 따라 데이터 품질의 영향 관계가 상이할 수 있으며, 본 연구에서는 금융산업만을 이용하여 발견된 데이터 품질의 영향이 한국의 국가적인 차원에서 동일하게 작용한다고 가정한다.

데이터 품질기준은 부가가치의 로그값에 0.628의 영향을 준다. 모형 12는 기업의 크기와 연도별 차이효과를 통제한 모형이며, 이 때의 데이터 품질기준은 부가가치의 로그값의 변동을 74.6%를 설명할 수 있고 모형은 1% 수준에서 유의하다. 기업의 크기와 연도별 차이 효과를 통제한 상태에서 데이터 품질기준은 부가가치의 로그값에 5% 유의수준에서 0.237의 영향을 준다. 즉, 데이터 품질기준이 1만큼 증가하면, 부가가치의 로그값은 0.237만큼 증가한다. 이를 백분율로 환산하면, 약 72.6%가 된다. 2008년과 2009년 표본기업의 평균 부가가치는 약 6,240억 원이고, 평균 데이터 품질기준은 1.2이다(<표 1>). 금융산업 기업의 데이터 품질기준이 2.2로 1만큼 증가하면, 부가가치는 약 1조 770억 원으로 약 4,530억 원이 증가한다. 산업 관점에서 보면, 2008년과 2009년의 금융산업 전체의 연평균 부가가치는 약 65.7조 원이다(한국은행, 2010). 금융산업의 데이터 품질기준이 1만큼 증가하면, 금융산업의 연평균 부가가치는 약 113.4조 원으로 약 47.7조 원이 증가한다. 또한 국가적인 관점에서 2008년과 2009년 한국의 연평균 부가가치는 약 939조 원이며, 연평균 국내총생산은 약 1,044.8조 원이다(한국은행, 2010). 한국의 데이터 품질기준이 1만큼 증가하면, 한국의 부가가치는 약 1,620.5조 원으로 681.5조 원만큼 증가하고, 국내총생산은 1,803.1조 원으로 758.3조 원만큼 증가한다.

V. 결론

본 연구는 한국 금융산업의 기업 데이터 품질 자료와 재무성과 자료를 이용하여 데이터 품질이 기업 성과에 주는 영향을 분석하고, 데이터 품질이 한국의 금융산업과 국가의 경제적 성과에 주는 영향을 추정하였다. 분석 결과, 데이터 품질은 기업의 매출액, 영업이익, 부가가치에 유의한 영향을 주었다. 데이터 품질관리 프로세스 지수가 1만큼 증가하면, 부가가치는 약 2.3%가

증가한다. 또한 데이터 품질기준이 1만큼 증가하면, 부가가치는 약 72.6%가 증가한다.

본 연구는 데이터 품질이 기업 성과에 주는 영향을 실증적으로 분석한 상대적으로 초기의 연구이며, 데이터 품질이 금융산업과 국가의 경제적 성과에 어느 정도의 영향을 주고 있는지를 정량적으로 추정하였다. 이는 개념적인 주장이 아니라 데이터 품질이 기업 성과에 영향을 주는 것을 정량적인 수치로 증명한 연구이다. 또한 연구를 위한 단일 설문에서 데이터 품질과 기업 성과를 조사하는 1차 자료가 아니라, 진흥원에서 분석한 데이터 품질과 기업의 실제적인 성과 자료인 2차 자료(secondary data)를 이용하기에, 1차 자료의 한계성을 피할 수 있다.

본 연구는 한계점을 갖고 있다. 분석 대상 산업인 금융산업은 정보(데이터)집약적 산업이다. 정보집약적 산업에서 IT 투자가 기업 성과에 더 많은 영향을 주는 것(Lee and Kim, 2006)과 마찬가지로 데이터 품질이 기업 성과에 더 많은 영향을 줄 수 있다. 따라서 연구 결과를 이용하여 국가 수준에서 데이터 품질의 경제적인 효과를 추정하는 것은 과대 추정될 수 있다. 그리고 IT 투자도 기업 성과에 유의한 영향을 주지만(Lee and Kim, 2006; Lee et al., 2011), IT 투자 자료의 제약(표본 수의 부족)으로 고려할 수 없었다. 또한 2008년과 2009년의 자료를 이용하기는 하였지만, 자료의 제약으로 종단적 연구를 수행할 수 없었다.

추후 연구에서는 자료의 제약이 해결될 수 있다면, 기업의 IT 투자 자료도 모형에 포함시켜 분석해야 할 필요가 있으며, 분석 자료를 전체 산업으로 확장하여 국가 수준에서 데이터 품질의 경제적인 효과를 추정할 필요가 있다. 그리고 데이터 품질과 기업 성과간의 시간 지연을 고려한 종단적 연구를 수행하여 데이터 품질의 기업 성과와의 연관(association) 관계가 아닌 인과(causality) 관계를 이해하는 것이 필요하다.

참고 문헌

- 금융감독원, 금융통계정보시스템, <http://fisis.fss.or.kr/>, 2010.
- 디지털타임스, [사설] 공공데이터 품질이 국가경쟁력 좌우한다, 2010, http://www.dt.co.kr/contents.html?article_no = 2010080302012351676001.
- 이상호, J. Y. Xiang, 김재경, 중국과 한국 기업의 정보기술 투자와 기업 성과의 관계에 대한 실증 연구”, *Information Systems Review*, 제 11권, 제3호, 2009, pp. 1-21.
- 이창수, 김선영, “데이터 품질관리와 기업신용도 상관성 분석에 관한 탐색적 연구”, *인터넷전자상거래연구*, 제10권, 제1호, 2010, pp. 197-218.
- 한국데이터베이스진흥센터, 데이터 품질관리 성숙 모형(Ver 1.0), 2006, http://www.dbq.or.kr/pds/pds_read.php?DBS_IDX = 123&page = 1&field = &keyword = &DBM_IDX = 3.
- 한국은행, 계간국민계정, 2010, <http://www.bok.or.kr/contents/total/ko/boardView.action?boardBean.sdt = &boardBean.edt = &boardBean.searchColumn = title&boardBean.search Value = &menuNavId = 589&boardBean.menuid = 589&boardBean.brdid = 72378&boardBean.rnum = 4&boardBean.cPage = 1>.
- 허희정, 김종우, “데이터 품질 향상을 위한 데이터 관리 프로세스 개선 사례 연구: 데이터 표준과 요구사항 관리 중심으로”, *Information Systems Review*, 제10권, 제1호, 2008, pp. 91-113.
- DeHoratius, N. and A. Raman, “Inventory record inaccuracy: An empirical analysis”, *Management Science*, Vol.54, No.4, 2008, pp. 627-641.
- DeLone, W. H. and E. R. McLean, “Information systems success: The quest for the dependent variable”, *Information System Research*, Vol.3, No.1, 1992, pp. 60-95.
- DeLone, W. H. and E. R. McLean, “The DeLone and McLean model of information systems success: A ten-year update”, *Journal of Management Information Systems*, Vol.19, No.4, 2003, pp. 9-30.
- Hair, J. Jr., R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis*, 5th ed., Prentice-Hall International, Upper Saddle River, NJ, 2005.
- Kim, J. K., J. Y. Xiang, and S. Lee, “The impact of IT investment on firm performance in China: An empirical investigation of the Chinese electronics industry”, *Technological Forecasting and Social Change*, Vol.76, No.5, 2009, pp. 678-687.
- Laudon, K. C. and J. P. Laudon, *Management Information Systems: Managing the Digital Firm*, 10th ed., Pearson Prentice Hall, Upper Saddle River, NJ, 2009.
- Lee, S. and S. H. Kim, “A lag effect of IT investment on firm performance”, *Information Resources Management Journal*, Vol.19, No.1, 2006, pp. 43-69.
- Lee, S., J. Y. Xiang, and J. K. Kim, “Information technology and productivity: Empirical evidence from the Chinese electronics industry”, *Information and Management*, Vol.48, No.2-3, 2011, pp. 79-87.
- Petter, S. and E. R. McLean, “A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level”, *Information and Management*, Vol.46, No.3, 2009, pp. 159-166.
- Redman, T. C., “Improve data quality for competitive advantage”, *Sloan Management Review*, Vol.36, No.2, 1995, pp. 99-107.

An Empirical Analysis on the Effect of Data Quality on Economic Performance in the Financial Industry

Sangho Lee* · Joo Seok Park** · Jae Kyeong Kim***

Abstract

This study empirically investigated the effect of firm-level data quality on economic performance in the Korean financial industry during 2008~2009. The data quality was measured by data quality management process index and data quality criteria by Korea Database Agency, and financial firm performance data was acquired from Financial Statistics Information System of the Financial Supervisory Service. The result showed that the data quality has statistically significant impacts on financial firm performance such as sales, operating profit, and value added. If the data quality management process index increases by one, the value added can increase by 2.3 percent. Moreover, the data quality criteria increase by one, the value added can increase by 72.6 percent.

Keywords: *Data Quality, Data Quality Criteria, Data Quality Management Process, Firm Performance, Financial Industry*

* Department of IT Management, Sun Moon University and Kyung Hee University

** School of Management, Kyung Hee University

*** School of Management and Management Research Institute, Kyung Hee University

○ 저 자 소 개 ○



이 상 호 (slee@sunmoon.ac.kr, sangholee@kaist.ac.kr)

선문대학교 IT경영학과 조교수이며, 국방과학연구소 선임연구원, 한국국방연구원 선임연구원, 경희대학교 경영대학 연구교수를 역임하였다. 성균관대학교 경영학과에서 학사 및 석사를, 한국과학기술원(KAIST) 경영대학에서 경영공학박사(MIS)를 취득하였다. 주요 관심분야는 정보기술 투자 성과 측정, S/W개발 프로젝트관리, S/W 프로세스 개선, 인과성 분석 등이다.



박 주 석 (jspark@khu.ac.kr)

서울대학교 산업공학과에서 학사학위를 받았고, KAIST 산업공학과에서 석사학위를 받았으며, University of California, Berkeley 경영대학에서 박사학위를 받았다. 현재 경희대학교 경영대학 경영학과 교수로 재직 중이며 ISO SPICE 심사위원이다. 한국경영학회 부회장, 한국경영정보학회 부회장, 한국EA학회 부회장을 역임하였으며, 공기업 경영평가위원을 맡고 있다. 주요 관심 분야는 정보화전략, 모델링, 데이터베이스, CRM 등이다.



김 재 경 (jaek@khu.ac.kr)

서울대학교에서 산업공학 학사, 한국과학기술원(KAIST)에서 경영정보시스템 전공으로 석사 및 박사학위를 취득하였으며, 현재 경희대학교 경영대학 교수(e-비즈니스 전공)로 재직하고 있다. 한국지능정보시스템학회 편집위원장 및 회장, 한국경영정보학회 부회장을 역임하였으며, 현재 BK21사업팀 단장, 저탄소녹색성장국민포럼산하 그린IT분과위원, 그리고 *Information Technology and Management(SCI)*의 AE(Associate Editor)를 역임중이다. 주요 연구분야로는 비즈니스 인텔리전스, 네트워크 경영, 그린 비즈니스/IT 등이다.

논문접수일 : 2010년 12월 15일
1차 수정일 : 2011년 02월 25일

게재확정일 : 2011년 03월 08일