

데이터기반의 신규 사업 매출추정방법 연구: 지능형 사업평가 시스템을 중심으로*

전승표

한국과학기술정보연구원
기술사업화분석센터
(spjun@kisti.re.kr)

성태응

한국과학기술정보연구원
사업기회분석실
(ts322@kisti.re.kr)

최산

과학기술연합대학원대학교
과학기술경영정책학과
(soullives83@kisti.re.kr)

사업타당성 분석이나 기업·기술가치평가 등 미래의 사업에 대한 진입이나 투자 타당성을 분석하기 위해서는 새로운 사업과 관련한 시장을 추정하고 그 안에서 확보 가능한 매출을 객관적으로 추정하는 과정이 필수 불가결하다. 이런 신규 매출이나 시장규모의 추정 방법은 다양한 방법으로 구분이 가능한데 크게 정량적인 방법과 정성적인 방법으로 구분할 수 있다. 그러나 두 가지 방법 모두 많은 자원과 시간을 필요로 한다. 그래서 우리는 신규 사업의 평가지원을 위한 데이터 기반의 지능형 매출 예측 시스템을 제안하고자 한다.

본 연구는 사업타당성 분석이나 기술가치평가를 위한 신규 사업의 매출 추정 시스템을 개발하는데, 알고리즘 기반으로 전통적인 정량 예측방법 중 하나인 유추방법에 주목했다. 동일한 국내 산업에서 최근 창업한 기업의 매출 실적을 국내 신규 사업의 매출액을 추정하는 유추 대상 변수로 활용할 수 있는지 검토한다. 여기서 유추 예측 대상은 최초 매출액과 초기 성장률이며, 주요 비교 차원은 산업분류, 창업시기 등이 고려된다. 특히 본 연구는 우리나라 창업 기업이 가지는 매출 성장률의 평균회귀 현상을 활용하는 지능형 정보 지원 시스템을 제안한다.

본 연구에서는 신규 매출 추정을 위해서 역사적 자료인 창업 매출 실적을 활용하는 방법이 적절한지 판단하기 위해서 잠재성장모형 등을 활용해 산업분류에 따른 신규 사업의 초기 매출액과 연도별 성장률이 산업분류별로 차이가 있는지 분석한다. 기존 기업의 창업 후 4년간 매출 성과의 중단자료를 잠재성장모형으로 분석하는데, 특정 산업분류에서 차이를 보여주는지 분석해 산업분류가 유추 예측에서 고려해야할 유의미한 변수인지 분석하는 것이다.

본 연구의 결과는 신속하고 객관적인 신규 사업 매출 추정을 가능하게 하는 지능형 정보시스템을 개발하게 해서 사업타당성 분석이나 기술가치평가 과정의 효율성을 개선시켜 줄 것으로 기대된다.

주제어 : 매출 예측 시스템, 유추 예측, 매출 추정, 기술가치평가, 사업타당성분석, 잠재성장모형

논문접수일 : 2017년 2월 20일 논문수정일 : 2017년 2월 20일 게재확정일 : 2017년 2월 27일

원고유형 : 일반논문 교신저자 : 전승표

1. 개요

기술이나 사업의 현재가치를 평가할 때는 일반적으로 10~15년까지의 추정기간을 이용하는 데, 흔히 현금흐름 추정을 두 시기로 분리하기도

한다. 두 기간 중에 첫 번째는 실제 변수(예를 들어 단위수량, 단위원가)들에 대해 정밀한 계산을 거친 완전한 대차대조표 및 손익계산서를 작성하는 구체적인 5~7년의 추정기간이 되고, 둘째는 매출성장, 마진 및 자본회전율과 같이 몇 가

* 본 논문에 2016년 한국기술혁신학회 추계학술대회에 발표한 학회발표 논문을 일부 포함하고 있음을 밝힌다.

지 중요한 변수들에 초점을 맞추는 잔여년도에 대한 단순화된 추정기간이 된다(Koller et al., 2010).

전체 기술, 사업, 기업의 가치평가는 미래의 수익을 기반으로 추정된 FCF(Free cash flow)를 기초로 한다. FCF 추정과정은 5~6단계로 나눌 수 있다. 1) 과거 재무자료 준비 및 분석, 2) 매출 추정, 3) 손익계산서 추정, 4) 대차대조표 추정, 5) ROIC (Return on invested capital) 및 FCF 계산의 단계로 나누어 볼 수 있다. 여기서 매출추정은 추정된 재무제표의 거의 모든 세부 항목과 직간접적으로 연동되어 있다. 따라서 FCF 추정에 가장 큰 영향을 주는 과정이라고 설명할 수 있는데, 추정 방법은 Top-down 방식(시장 기반)이나 Bottom-up 방식(고객 기반)을 선택하여 미래의 매출액을 예상할 수 있다. 추정된 성장은 과거 실적 및 경제 전반의 사례와 비추어 무리가 없어야 한다(Koller et al., 2010).

매출액을 추정하기 위해 활용하는 Top-down 방식은 전체 시장의 크기를 측정(또는 추정)하고 시장점유율을 결정하여 가격을 추정함으로써 매출액을 추정하는 방식이다. Bottom-up 접근법은 기존 고객, 고객회전을 및 신규 고객에 대한 잠재력으로부터 회사 자체적인 수요를 추정하는 방법이다. 일반적으로 가능하다면 두 방법을 모두 사용하여 매출 추정하는 방법이 권장된다(Koller et al., 2010). 이런 두 가지 방법은 모두 많은 노력과 시간이 투자되어야 한다. 따라서 본 연구는 신속하고 효율적으로 신규 사업의 평가를 지원하기 위한 데이터 기반의 지능형 수요예측 시스템을 제안하고자 한다.

앞서 설명된 바와 같이 기업, 사업 또는 기술의 평가를 위한 매출 추정은 전체 평가 결과에 끼치는 영향이 매우 커서 객관적 추정이 중요하

며, Top-down 또는 Bottom-up 방식을 활용하는데, 추정된 성장률은 과거 실적 및 경제 전반의 사례와 비추어 무리가 없어야 한다는 전제 조건이 있다. 우리는 이런 전제조건에 주목했고, Top-down 또는 Bottom-up 방식을 보완할 유추방식을 매출추정의 방식으로 제안하고자 한다. 본 연구에서는 동일한 산업에서 기존에 창업한 기업의 실적을 바탕으로 새로운 사업이나 기술에 의해 창출되는 신규 매출액을 추정하는 방법을 제안한다(Jun et al., 2016). 본 논문에서 제안된 데이터 기반의 매출 추정 알고리즘 관련 지능형 정보 시스템에 활용되면 다양한 의사결정 지원에 크게 기여할 것으로 기대된다.

본 연구에서는 창업 선례를 활용한 매출액 추정 알고리즘이 적절한지 살펴보기 위해서 제2장에서는 매출액 추정과 관련된 선행연구를 살펴 보며, 제3장에서는 제안된 모델(알고리즘)을 구체적으로 제시하며, 이 모델의 활용 가능성을 검토하기 위한 세 가지 문제를 설명하고 연구방법과 데이터 기반의 초기 매출액 통계값을 제시한다. 제4장에서는 제안된 모델의 활용을 위한 세 가지 문제에 대한 연구결과를 살펴본다. 제5장에서는 제안된 모델에 대한 특정 사례(산업분류) 개발 결과를 제시하고, 예측력을 검증한다. 제6장에서는 도출된 결과를 바탕으로 시사점과 제한점을 설명하게 된다.

2. 선행연구 분석

2.1 매출액과 매출성장률의 거동관련 연구

미래의 성과를 예측하는 최선의 방법은 이를 포괄적으로 수행하는, 즉 이익 예측뿐만 아니라

현금흐름 및 재무상태표 항목에 대한 예측까지 수행하는 것이다. 포괄적 접근법은 많은 예측을 수반하지만 대개의 경우 그 예측들은 전부 몇 개의 주요 동인의 거동(behavior)과 관련되어 있다. 주요 동인은 관련된 경영의 유형에 따라 달라지는데, 재무부문 외의 다른 경영활동이 없다고 가정하는 경우, 매출액 예측치와 이익률은 많은 경우 주요 동인이 된다. 대부분의 비용 항목 또한 이익률의 예상된 변동에 따라 매출액을 따르게 된다. 각 항목별 예측치를 매출액 예측치와 관련 시킴으로써 내적 불일치의 문제와 비현실적인 항목적 가정을 방지할 수 있다(Palepu & Healy, 2007).

모든 예측에는 적어도 암묵적으로 최초 “기준(benchmark)” 또는 출발점이 있게 마련이다. 직전기의 매출액을 참고하거나, 최근 추세를 조정 한 매출액을 활용하며, 일견 합리적인 것처럼 보이지만 유용하지 않은 것으로 밝혀진 것인데, 이는 과거 수년간의 평균성과이다(Palepu & Healy, 2007). 그런데 신규 사업이나 기술가치평가에서는 벤치마킹할 시작점을 찾는 것이 용이하지 않은 문제점이 있다. 직전기의 매출액이 존재하지 않기 때문이다.

경영전략분석, 회계분석 및 정밀 재무분석이 완성되는 시점에서의 예측치는 출발점과 크게 다를 수 있다. 그럼에도 불구하고 정밀분석의 착수에 도움이 되는 출발점을 가진다는 단순한 목적만으로도 어떤 주요 재무적 통계값의 “평균적” 형태를 아는 것이 유용하다. 이익과 같은 어떤 주요 통계값의 경우에는 단순히 그 수치의 과거 형태에 기초한 출발점은 기대 이상의 효과가 있다. 따라서 그 기준은 종종 좋은 출발점일 뿐만 아니라 정밀분석 후의 예측치에 근접한다. 그 기준에서 크게 이탈한 출발점은 기업의 상황이

분명 예외적인 경우에 한하여 정당화될 수 있다(Palepu & Healy, 2007). 앞서 언급된 바와 같이 추정된 실적은 과거 실적 및 경제 전반의 사례와 비추어 무리가 없어야 한다는 전제 조건이 있는 것이다(Koller et al., 2010).

이상에서 살펴본 바와 매출액 자체의 추정은 벤치마킹 외에 기존의 Top-down이나 Bottom-up 추정을 대신할만한 대안은 찾기 쉽지 않다. 그러나 매출액 추정을 초기 매출액과 기간별 성장률로 구분해본다면, 다른 대안을 찾아 볼 수 있다. 일반적으로 매출 성장은 “평균회귀(mean reverting)”의 경향이 있기 때문이다.

Palepu와 Healy(2007)에 따르면 평균 이상 또는 그 이하의 매출성장을 유지하는 기업은 3~10년 내에 시간의 경과에 따라 “정상(Normal)”수준으로 수렴하는 경향이 있다. 그들은 Thomson financial worldscope 자료에 수록된 1989~2005년 모든 유럽 상장기업의 매출성장률이 평균회귀 경향을 보여주는 것을 밝혔다. 산업과 기업이 성숙기로 접어 들어감에 따라 수요의 포화와 산업내 경쟁으로 인하여 매출성장이 수그러드는 형태를 나타낸다는 것이다. 그러므로 어떤 기업이 현재 급속하게 성장하고 있어도 그러한 고도성장이 무한히 지속될 것으로 보는 것은 비현실적이라고 주장한다. 물론 기업의 성장률이 얼마나 빨리 평균회귀 경향을 보일지는 산업의 특성과 산업 내에서 경쟁적 지위에 따라 달라진다(Palepu & Healy, 2007). 이런 경향은 국내도 비슷해서 Kim(2015)에 따르면 상장기업의 매출성장률에 따라 5개 그룹을 구분한 후 매출성장률을 추적했는데, 2~3년 후에 모두 유사해 지는 경향을 확인했다. 이런 실증은 일반적인 기업에서 기업의 규모는 생존 기간이나 성장률과 무관하다는 지브랏의 법칙(Gibrat's law of

proportionate effect)과도 일맥상통한다(Daunfeldt & Halvarsson, 2015).

이상에서 확인한 매출성장률의 평균회귀 현상은 신규기업에 대한 것은 아니며 기존 기업들을 성장률에 따라 구분한 후 추적 조사한 결과다(Palepu & Healy, 2007; Welch, 2012). 물론 신규 사업에 대한 매출성장률을 다루는 연구도 있었지만 역시 역사적 실적을 바탕으로 하는 경우가 많았다(Lundholm et al., 2010; Park & Kim, 2012). 본 연구에서는 이런 평균회귀 현상이 창업기업만을 대상으로 했을 때도 나타나는지 확인하게 된다. 본 연구가 매출액이나 매출성장률의 추정에 활용하는 방법은 유추 예측이지만, 매출성장률에서 이런 평균회귀 현상이 확인된다면, 매출성장률 추정 알고리즘을 벤치마킹 단계와 평균 수렴을 따르는 단계로 나누어 추정하는 것이 적절하기 때문이다.

2.2 시장규모와 점유율관련 연구

시장 규모나 성장률 자료도 일반적인 자료의 구분과 같이 조사방법에 따라서 1차와 2차 자료로 구분할 수 있다. 기존의 시장 규모 조사는 주로 전문 기관이 이미 조사해 공개한 2차 자료를 사용하는 경우와 연구 문제 해결을 위해 정보를 새롭게 수집하는 1차 자료로 구분된다(Yoo et al., 2015). 1차 자료로 수집된 시장 규모 추정은 조사 목적에 적합하다는 장점이 있지만, 수집 과정에 많은 노력이 필요하고, 수집 비용과 시간이 많이 소요되는 단점이 있다(Shin, 2010). 이런 1차 조사는 특정 조건(예 10인 이상)에 해당되는 사례를 전수 조사하는 경우(Jun & Park, 2016)도 있지만, 1차 자료 활용이 가지는 시간과 비용한계를 극복하기 위해 샘플링 조사가 많이 사용되

는데, 주로 샘플링한 기업의 매출액 정보나 설문 조사를 통한 예상을 바탕으로 전체시장을 추정한다(Jun et al., 2014). 특히 신규 시장이나 직접 거래가 어려운 제품 또는 서비스의 시장 규모를 추정할 때 1차 정보를 활용한다(Kim & Kim, 2008).

반면 주로 활용되는 2차 자료 기반의 시장 규모 추정은 수집과정이 쉽고, 수집 비용과 시간이 짧은 장점이 있다. 그러나 연구 목적에 맞는 정확한 정보 수집이 어렵기 때문에 목적에 맞게 정보를 가공하는 과정에서 가정이 다수 포함될 수밖에 없는 한계가 있다(Shin, 2010). 2차 자료의 정보원으로는 주로 시장 조사 전문기관이 제공하는 유료 정보나, 통계청과 같은 공공기관이 제공하는 공개 자료가 많이 활용된다(Yoo et al., 2015). 그러나 이들 자료는 연구 목적에 맞는 정확한 조사 통계자료가 존재하지 않아, 여러 가정을 통해 다수의 기관 자료를 종합해서 사용하거나, 기존 조사에서 기준이 된 분류를 재가공하여 활용하기도 한다(Yang et al., 2010). Yoo et al.(2015)는 수집 과정, 비용, 시간에서 비교 우위가 있는 2차 자료를 활용함에 있어서 자주 활용되는 통계청과 공공기관(관세청)의 자료 활용 방법을 합리적으로 연계할 수 있는 방법을 제공함으로써, 2차 자료를 활용한 시장 규모 추정 방법의 장점을 크게 높이려고 했다.

조사방법에 따라서 1차와 2차로 구분할 수도 있지만, 조사된 정보를 활용한 시장예측은 역시 정량적인 방법과 정성적인 방법들이 존재한다. 이 중에서 대표적인 정량적인 시장 예측 방법이 성장모형이나(Park et al., 2015) 데이터 마이닝과 같은 방법이다(Kim & Ahn, 2011). 그러나 역시 신규시장에서는 이런 정량적 방법의 적용이 거의 불가능하거나 전문가가 모수나 유사시장을

추정해야 하는 문제점이 존재한다(Lee et al., 2015).

본 연구에서는 초기 매출액과 매출성장률을 구분하여 신규 사업 추정기간의 매출을 예측하는 방법을 제시하는데, 초기 매출액은 앞서 설명한 Top-down이나 Bottom-up 방식의 추정을 기본적으로 추천한다. 그러나 본 연구의 제3장부터는 추정에 소요되는 시장과 비용을 고려해 유추방법을 활용한 초기 매출액 추정도 가능하도록 일부 통계값을 시범적으로 제시하며, 예측 결과 비교에도 초기 매출액은 유사기업의 통계값을 활용하게 된다.

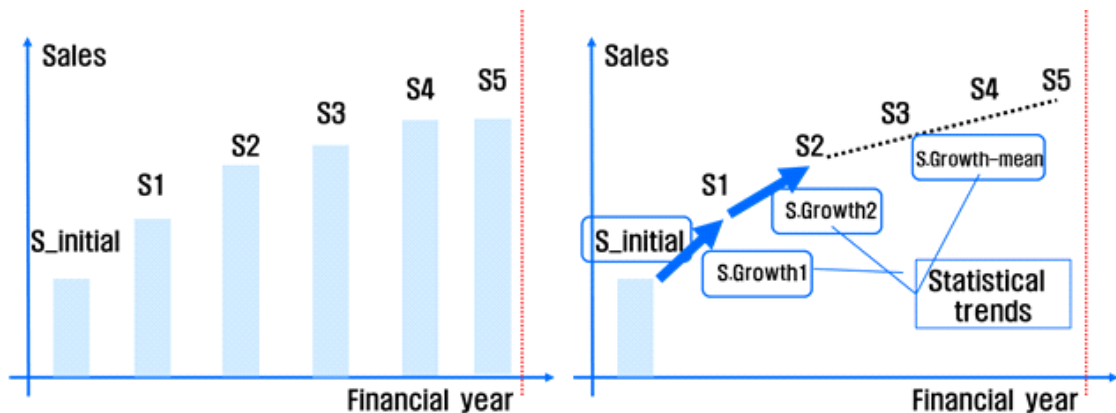
3. 연구 방법

3.1 신규 사업 매출 추정 모델

본 연구에서 신규 사업, 창업 기업 그리고 기술 사업화 등의 타당성 분석을 위한 매출 실적 추정 모형을 제안하는데, 기존의 Top-down 또는 Bottom-up 방식을 보완할 유추방식을 매출추정

의 방식으로 제안한다. 특히 사업화 후 매 기수의 매출액을 개별 추정하는 방법대신 본 연구는 초기 매출액과 매출성장률을 구분해 추정하는 알고리즘을 시스템화 하는 방법을 제안한다. 이것은 매출액 추정을 일일이 진행하면 추정하는 연도(기수)만큼 많은 가정이 필요하게 되는데 (<Figure 1> 참조), 초기 매출액과 각 기수의 매출액 증가율(이하 매출성장률)을 추정할 경우 매출성장률의 평균회귀 현상 때문에 추정하는 연도 후반기에는 그 만큼 불필요한 가정을 줄일 수 있기 때문이다.

초기매출액과 매출성장률을 구분해서 신규 사업의 매출을 추정하는 방법에도 여러 가지 방법이 가능하다. 먼저 초기매출액 추정 방법은 앞서 선행연구에서 언급된 방법들로 Top-down이나 Bottom-up 방식을 활용할 수 있다(Koller et al., 2010), 다음으로 초기 매출액을 추정하는 방법은 벤치마킹하는 방법으로 유사한 제품이나 산업의 창업기업 사례를 참조하는 것이다(Palepu & Healy, 2007). 본 연구에서는 정확도는 경우 따라 조금 떨어지지만 간편하고 객관적이며 신속한



<Figure 1> Comparison of the conventional (left) and new (right) methods of sales estimation

추정 알고리즘을 제안하고자하기 때문에 초기 매출액 추정은 벤치마킹(유추 예측)을 활용한다. 시스템에서는 산업별로 최근 10년 안에 창업한 기업의 첫째 매출액 또는 첫 번째 매출액의 분포를 제시하여 활용할 수 있도록 지원한다.¹⁾

매출성장률 추정 방법도 여러 가지 방법이 가능한데, 본 연구에서는 기본적으로 창업기업의 매출성장률도 평균회귀 현상이 관찰된다면, 특정 기간부터는 그 산업 또는 전체 제조업의 추정 평균 성장률로 매출성장률이 대체되는 것으로 가정한다. 따라서 성장 초기 일정 기간 동안의 매출성장률 추정방법에 대한 고려가 필요한데, 본 연구에서는 3가지 방법을 추천한다. 첫 번째는 매출성장률이 평균에 수렴한다는 점에서 착안해서 평균 대비 특정 배수에서 시작해서 점차 감소하는 방법으로 예측하는 방법으로, 여기서 배수는 창업기업의 사례를 분석해서 분포를 구하는데 연도별 경향을 분석해 평균과 차이 즉 배수를 구하면 된다. 두 번째는 유사 기업이나 특정 산업의 창업 후 특정 기간까지의 매출성장률을 유추 예측에 활용하는 방법이다. 역시 창업기업의 창업 후 매출 실적을 분석하는데, 절대적인 값인 연도대신 창업 후 시기를 기준으로 추산한 상대적인 시기의 매출성장률 분포를 구하게 된다. 마지막 매출성장률 추정 방법은 두 번째 방법과 유사하지만, 창업 후 매출성장률을 유추하지 않고 첫 매출 후 매출성장률을 추정하는 방법이다.

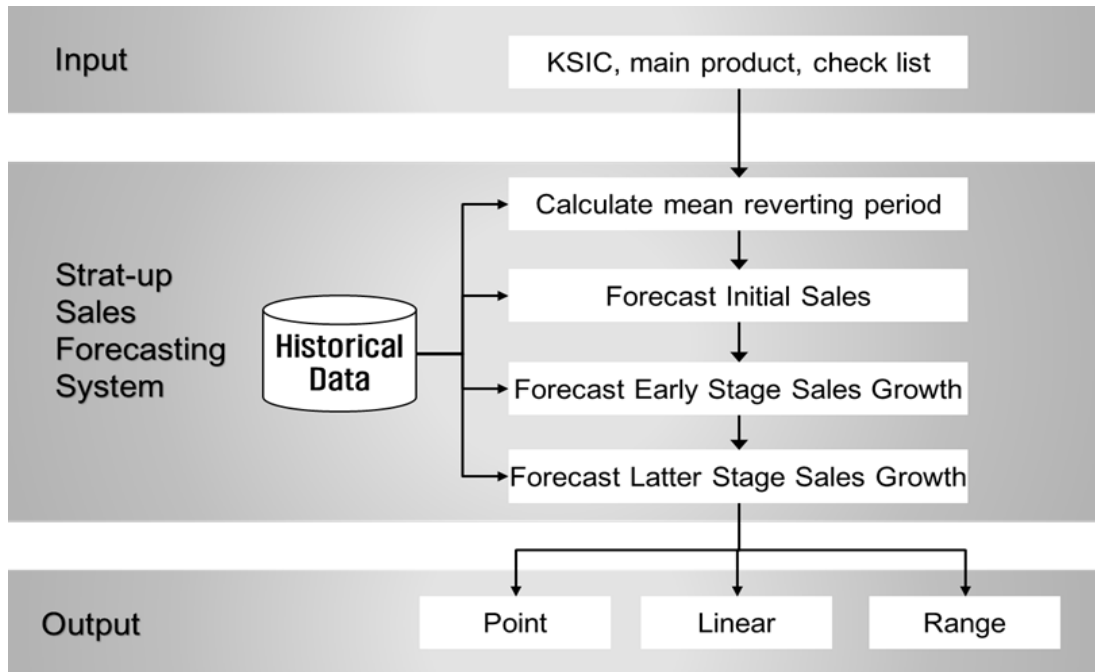
이상의 세 가지 방법은 각각 다른 조건에서 활용할 수 있는데, 먼저 특정 제품군이나 산업의 평균 성장률만 객관적 예측이 가능한 경우 첫 번째 매출성장률 추정방식을 추천한다. 신규 사업

의 창업 후 매출 지연 시기와 첫 매출액을 비교적 정확히 추정할 수 있는 경우는 세 번째 추정 방법이 좋은 매출성장률 추정 대안이 될 수 있다. 마지막으로 창업시기와 주요제품 또는 산업 분류 정도의 정보만 있는 경우는 두 번째 매출성장률 추정 방식의 활용을 추천한다. 본 연구에서는 세 번째 즉 첫 매출 후 매출성장률을 바탕으로 추정하는 방법을 중심으로 설명하게 된다. 이 상에서 설명된 창업기업 매출 추정 시스템의 구조는 <Figure 2>와 같은데, 입력 조건(정보)에 따라서 매출 추정 모델을 선택하고, 평균회귀 기간, 초기 매출액, 회귀기간 전후의 매출성장률을 차례로 추정해서 특정시기, 선행 또는 구간으로 신생기업의 매출액을 추정하게 된다.

3.2 연구 문제

본 연구가 신규 사업(또는 기술)의 매출액을 추정하기 위해서 <Figure 2>에서 제시한 매출추정 시스템을 실무적으로 활용하기 위해서는 몇 가지 이슈에 대한 연구가 필요하다. 첫째, 창업기업의 매출성장률도 평균 수렴하는가를 검토할 필요가 있다. 결과에 따라서 매출성장률에서 평균회귀가 관찰된다면 본 연구가 제시한 예측 모델(<Figure 1>)이 기존 보다 효율적이라고 설명할 수 있고, 매출 성장률은 두 단계로 나누어 볼 수 있다는 결론을 도출할 수 있기 때문이다. 두 번째, 연도별 창업의 시기는 성장률에 차이를 가져오는가를 검토할 필요가 있다. 창업시기에 따라서 거시적 경제상황에 서로 다른 영향을 받을 수 있기 때문인데, 만약 창업시기에 따라 매출성장률 차이가 크게 발생된다면, 거시경제상황의

1) 창업 첫째 매출액이 부재한 경우가 다수 존재하기 때문에 첫 매출액과 설립 후 첫째 매출액은 구분해야 한다. 본 연구에서는 첫 매출액을 첫째 매출액이 아니라 첫 번째 매출액으로 가정한다.



〈Figure 2〉 Start-up Sales Forecasting System

변수에 대한 정규화 과정(normalization)이 필요해지기 때문이다. 세 번째, 산업별로 매출 성장률이 다른지 검토할 필요가 있다. 만약 산업별로 매출성장률이 다르다면 당연히 산업 분류별로 성장률을 각각 예측해야하기 때문이다. 이상의 연구 문제를 정리하면 다음과 같으며, 제4장에서 문제에 대한 연구결과를 제시한다.

- Q1. 창업 기업의 매출성장률도 평균 수렴하는가?
- Q2. 연도별 창업의 시기는 성장률에 차이를 가져오는가?
- Q3. 매출성장률은 산업별로 차이가 있는가?

이상의 분석 결과에 따라 우리는 새로운 신규 사업 매출추정 모델을 제안할 수 있다. 본 연구

에서는 두 가지 예측모델을 제안하는데 Range model과 Linear model이다. Range model은 매출 성장률을 기존 창업기업의 매출성장률 자료 분석을 통해 산출된 사분위수를 바탕으로 제시하는데, 추후 추정 대상 기술이나 기업의 경쟁력을 고려해 선택적으로 성장률을 적용할 수 있게 만든 모델이다. 반면에 Linear model은 중위수를 중심으로 단일 값을 제시하는 모델로 전체적이고 대표적인 경향을 제시하는 목적으로 개발되었다. 마지막으로 본 연구가 제시하는 문제는 제안된 예측 모델은 신뢰성이 있고 기존 예측 방법보다 효율적인가에 대한 것이다. 제5장에서는 특정 사례(산업)에 대해서 두 가지 예측모델을 실제 구축하고 결과를 제시하며, 제안된 모델의 예측력을 검증하게 된다.

3.3 연구 방법 및 연구 사례

3.3.1 연구 사례와 초기 매출액

본 연구는 최근 10년 즉 2005년 이후 한국에서 설립된 기업의 2005년부터 2014년말까지 매출실적 707,704건을 대상으로 분석했는데, 기업으로는 157,790개 기업의 실적에 해당된다. 산업 분류는 KSIC(korean standard industrial classification) 대분류를 기준으로 1부터 63까지 분석했으며, 분류 46이 28.8%로 가장 많았고, 47은 8.6%, 42는 8.2%, 29는 6.6%로 나타났다. 이들 기업에 대한 초기매출액의 분포를 분석하면 다음 <Table 1>과 같은데 평균의 차이가 산업별로 매우 큰 것을 알 수 있으며, 7천억 원 이상에서 1억 5천만 원까지 산업별로 큰 차이를 보이고 있었다.²⁾ 여기서 주목할 것은 중앙값과 평균의 큰 차이로, 침도와 왜도 분석 결과도 매우 높게 나타나서 초기 매출액의 분포가 정규분포를 따르지 않는 것을 확인할 수 있었다. 따라서 본 연구에서 제안하는 지능형 정보 지원 시스템에서는 분포를 고려해 초기 매출액을 활용할 때는 평균보다 중위수를 추천한다.

본 연구는 매출성장률의 변화에 산업분류가 영향을 주는지 분석하게 되는데, 매출성장률을 추정하기 위한 방법으로 유사기업의 첫 매출 후 매출성장률을 유추하는 방법을 본 연구가 선택했기 때문에 산업분류의 영향 분석에서는 첫 매출 후 4년간 매출액이 계속 존재하는 70,099개 기업을 분석 대상으로 한정했다. 또한 <Table 1>에서 나타난 바와 같이 매출액이나 매출성장률의 분포는 정규분포를 따르지 않는다. 따라서 본

<Table 1> Comparison of initial sales and sales delays by KSIC classification

KSIC	Initial sales(Million ₩)				Lag time until initial sales(Years)			
	Q1	Q2	Q3	mean	Q1	Q2	Q3	mean
1	66	303	1,057	1,170	0.00	1.00	2.00	1.05
2	89	228	422	338	0.00	0.00	1.00	0.88
3	347	840	1,467	1,359	0.00	1.00	2.00	1.15
5	313	892	12,311	5,228	0.00	1.00	2.50	1.20
6	315	692	2,649	1,339	0.00	2.00	3.75	1.88
7	227	640	1,491	1,359	0.00	1.00	2.00	1.28
8	32	791	1,989	3,024	0.25	1.00	1.75	1.00
10	112	432	1,353	2,401	0.00	0.00	1.00	0.79
11	26	127	417	5,013	0.00	1.00	2.00	1.11
13	267	655	1,581	3,208	0.00	0.00	1.00	0.64
14	243	577	1,570	2,463	0.00	0.00	1.00	0.63
15	210	549	1,346	1,534	0.00	0.00	1.00	0.66
16	193	475	1,095	1,262	0.00	0.00	1.00	0.65
17	219	541	1,286	1,321	0.00	0.00	1.00	0.63
18	199	410	951	872	0.00	0.00	1.00	0.84
19	119	556	2,779	781,676	0.00	0.50	1.75	0.92
20	104	363	1,088	10,798	0.00	0.00	1.00	0.69
21	53	167	552	2,155	0.00	0.00	1.00	0.61
22	195	489	1,239	2,361	0.00	0.00	1.00	0.61
23	136	480	1,348	1,773	0.00	0.00	1.00	0.73
24	278	775	2,027	4,879	0.00	0.00	1.00	0.59
25	187	439	1,066	1,150	0.00	0.00	1.00	0.63
26	128	387	1,026	5,049	0.00	0.00	1.00	0.54
27	90	260	707	1,544	0.00	0.00	1.00	0.54
28	117	330	900	1,590	0.00	0.00	1.00	0.56
29	168	401	996	1,103	0.00	0.00	1.00	0.66
30	226	624	1,843	2,240	0.00	0.00	1.00	0.66
31	319	925	2,188	2,027	0.00	0.00	1.00	0.58
32	175	433	1,020	879	0.00	0.00	1.00	0.79
33	121	370	889	935	0.00	0.00	1.00	0.83

연구는 매출성장률을 중심으로 첫 매출 후 각 기수에서 매출성장률이 상하위 10%를 벗어난 이상값이 포함된 케이스를 제외한 31,984건을 최종 분석대상으로 선택했다.³⁾

- 2) 지면 관계로 분류 1~33까지만 초기 매출액과 초기 매출 지연기간을 제시한다.
- 3) 상한값(10%)을 제외한 첫 매출성장률의 왜도의 값만 2를 상회(3.599)했으며, 대부분 1이하 값으로 안정적인 분포를 나타냈다. 따라서 매출성장률은 정규분포를 가정해 분석한다.

또한 본 연구가 매출성장률 모델 수립에 첫 매출 후 4년간 매출액이 계속 존재하는 70,099개 기업을 대상으로 좁혔기 때문에 2012년 이후 최근까지 설립된 기업은 자연스럽게 모델 수립에서 제외된다. 이렇게 제외된 2012년 이후 설립 기업의 매출과 성장률은 제5장에서 모델 검정을 위해 활용하게 된다.

3.3.2 연구 방법

먼저 창업시기에 따른 평균회귀 현상을 통계적으로 살펴보기 위해서 반복측정 분산분석을 활용했다. 이 방법은 동일한 실험개체에 3회 이상의 측정된 자료에 대해 분산분석을 적용한 것이다. 반복측정은 실험에서 충분한 수의 실험개체를 확보하기 어려운 경우나 실험개체 간의 차이가 커서 전체 결과에 영향을 주는 것을 방지하기 위해 실험개체를 블럭으로 처리한다(Lee & Kang, 2013). 한 개체를 반복적으로 측정함으로써 처음의 처리가 이후 실험에 영향을 줄 수 있는 태생적 단점이 있지만, 의·약학분야에서 치료 효과의 추이를 파악하거나 사회과학분야의 시간에 따른 변화 연구에서는 오히려 이런 단점이 크게 문제가 되지 않을 수 있다. 다음에서 설명할 잠재성장모형이 변화 추이를 가진 모형에 대한 전체 적합도에 치중한 분석이라면 반복측정 분산분석은 변화의 추이보다는 각 단계별로 차이여부를 중점적으로 해석한다(Lee & Kang, 2013). 이런 반복측정 분산분석은 구형성에 대한 가정이 만족되어야 하는 제약이 있다(Girden, 1992). 모집단의 분포 가정이거나 구형성에 대한 제약이 없는 방법으로 대응 k-표본의 비모수적 분석방법인 프리드만 검정이 있다(Song et al., 2003).

다음으로 산업분류 집단간 동질성 분석을 위해서 일원분산분석(ANOVA)과 잠재성장모형(Latent Growth Model, LGM)을 활용했다.⁴⁾ 먼저 일원분산분석을 위해 산업분류집단을 줄였는데, 50개 이상의 집단에 대한 분산분석은 수행되지 않고 의미도 없기 때문이다. 따라서 케이스가 충분하면서 유사할 수 있는 대분류 22~31만을 대상으로 분석했다. 대상은 정제된 즉 첫 매출 후 4년간 매출액이 지속적으로 발생되고, 이상값이 제거된(상하위 10%) 자료를 대상으로 했는데, 보다 보수적인 분산분석을 위해 이상값이 제거된 자료를 선택했다. 앞서 설명된바와 같이 정제된 자료에서 매출성장률은 왜도와 첨도가 안정적이었고, 케이스도 매우 많기 때문에 정규분포에 가깝다고 설명할 수 있으나, 확인차원에서 비모수검정인 Kruskal-Wallis 검정도 같이 수행했다.

잠재성장모형(LGM)은 시간에 따라 측정된 종단적 자료를 분석하는 방법으로 세 번 이상 측정된 종단자료에 대해 집단 또는 개인에 대한 변화량을 확인하는 연구방법이다(Duncan et al., 2006). 잠재성장모형 분석은 기본적으로 두 단계를 거쳐 분석한다. 첫 단계를 비조건적 모델분석, 다음 단계를 조건적 모델이라 한다(Kline, 1998). 또한 Preacher et al.(2008)과 Kim(2009)은 잠재성장모형을 6가지 유형으로 나누었으며, Kaplan(2009)은 좀 더 다양한 모형을 제시했다.

일반적으로 선형 모형 6개의 유형별로 적합도를 비교하여 가장 적합한 모형을 결정하거나, 무변화, 선형, 2차 모형에서 모형을 비교하여 가장 적합한 모형을 선정하는데, 본 연구에서는 6가지 선형 모형에 대한 적합성 분석결과를 바탕으로 적정모형을 선택했다.⁵⁾

4) 집단간 동질성 분석은 SPSS(ver 20.0) 프로그램과 AMOS 프로그램을 사용하여 분석하였다.

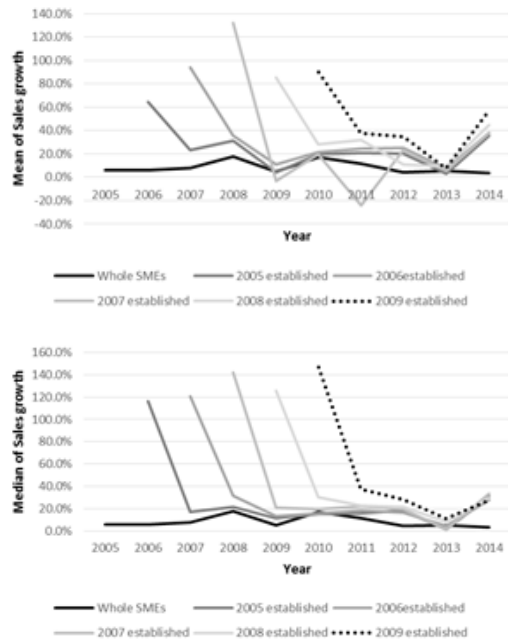
4. 연구 결과

4.1 창업 매출성장의 형태

<Figure 1과 2>의 매출성장률 기반 매출추정 알고리즘의 활용 가능성을 검토하기 위한 창업 기업의 매출성장률 검토 결과가 <Figure 3>에 제시되어 있다. 최근 10년(2005~2014년) 동안 국내에서 설립된 157,790개 기업의 설립 후 매출성장률을 살펴보면 <Figure 3>과 같은데,⁶⁾ <Figure 3>에는 2005년부터 2009년까지 설립된 기업의 매출성장률에 대한 평균 또는 중위수가 각 연도의 한국 중소기업의 매출성장률과 비교되어 있다. <Figure 3>에 따르면, 설립년도별 첫 번째 매출성장률은 설립년도에 따라 다르지만, 다음 연도부터 그 차이가 줄어들며, 3~4년차에는 거의 같아지고 심지어 중소기업 평균 매출성장률에 회귀되는 경향을 확인할 수 있다. 여기서 주목할 것은 기존 문헌에 확인된 상장기업 평균회귀는 전체는 당연히 평균을 구성하지만, 일시적으로 다른 추세를 보인 일부 기업들의 평균회귀 현상에 가까운 설명이라면, <Figure 3>에서 나타난 신규 기업의 매출성장률 추세는 값의 차이가 있을 뿐 1~3년간은 공통적으로 평균보다 크게 높은 성장률을 보인다점도 평균회귀 추세만큼 주목할 필요가 있는 발견이다. 물론 산업에 따라서 젊은 기업의 성장률이 높다는 Jovanovic(1982)의 주장과 일치하는 결과라고도 할 수 있다. 또한 <Figure 3>에서 보면 앞서 설명된바와 같이 매출액이나 성장률의 분포는 정규분포를 따르지 않기 때문에 역시 중위수에서 평균회귀 경향이 잘

관찰되는 것도 확인할 수 있다.

이런 평균회귀 경향을 통계적으로도 분석할 수 있는데, 첫 매출액 후 매출액이 4년 이상 존재하고 이상값(상하위 10% 제외)이 정제된 31,984건에 대해서 반복측정 분산분석(일반선형모형 반복측정 ANOVA)으로 분석했다. 먼저 첫 매출 후 3년간 성장률 분석 결과에 따르면 Mauchly의 구형성 검정결과가 귀무가설을 기각하는 것으로 나타나서($p < 0.000$)⁷⁾ 구형성 가정을 충족하지 못했다. 이 경우 개체 내 효과검정은 보다 보수적인 검증으로 결과를 판단할 필요가 있는데, 자유도가 훨씬 크고 보수적인 하한값이나 Greenhouse-



<Figure 3> Comparison of the annual sales growth rates (Upper: mean, Lower: median)

5) 2차 모형도 검토되었으나, 유의미한 결과가 도출되지 않았다.

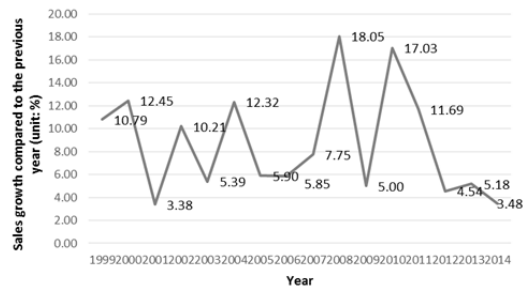
6) 창업 기업의 실적 자료는 한국기업데이터(KED) 자료를 바탕으로 재가공 후 활용했다.

7) 구형성 가정은 반복측정 디자인의 경우 요구되는 가정으로 일원분산분석의 경우 분산의 동일성 가정에 해당된다.

Geisser 검정의 결과에 따르면 반복측정 분산분석 결과는 첫 4년간 3번의 성장률은 서로 다른 것으로 결론내릴 수 있었다($F=16733.7, p<0.000$). 반면 6~9년차까지 3년간의 반복측정 분산분석 측정 결과는 Mauchly의 구형성 검정결과가 가정을 충족시켰고($p=0.855$), 구형성 가정을 만족한 개체 내 효과검정 결과는 마지막 4년간 3번의 성장률은 서로 다른 것으로 결론내릴 수 없었다($F=0.721, p=0.486$). 첫 매출 후 6년차부터는 각각 평균 3.29%, 3.75%, 2.64%로 매출이 성장했고, 통계적으로 거의 차이가 없는 것이다.⁸⁾

4.2 창업 시기에 따른 매출성장률 차이

<Figure 4>에는 <Figure 1과 2>의 매출성장률 기반의 매출추정 방법의 활용 가능성을 검토하기 위한 두 번째 이슈인 창업시기에 따른 매출 성장률 차이 가능성의 검토 결과가 제시되었다. <Figure 4>에는 창업시기에 따른 거시적 환경 차이를 검토하기 위한 1999년 이후 2014년까지 한국 제조업 중 중소기업의 매출성장률 추세가 제시되어 있다. 중소기업에 한정된 것은 창업기업을 분석하는 본 연구의 특성상 대기업은 창업기업이 거의 없기 때문이다. 동기간 동안 국내 중소기업은 최소 3.48%에서 18.05%까지 폭넓은 변화를 보이고 있었으며, 최소값과 최대값이 모두 최근 10년(2005~2014년)안에 나타났다. 따라서 본 연구가 창업기업의 매출액을 분석할 때 모두 같은 조건의 매출액인지 주의 깊게 다루어야 할 필요가 생긴다. 매출성장률이 유난히 높았던 2008년의 매출액과 성장률이, 가장 낮았던 2014년의 매출액과 같은 조건은 아니기 때문이다.



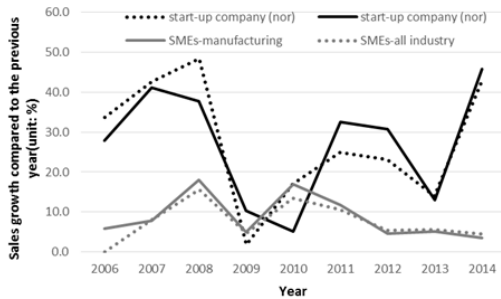
<Figure 4> Trend in the sales growth rate of SMEs and start-up companies in Korea, compared to the previous year (Data source: Bank of Korea's statistical system on business management and economics)

이런 문제를 고려하기 위해서 본 연구에서 창업기업의 전년 대비 매출 성장에서 체계적 성장 부분(전체 중소기업 매출성장률)을 제외한 매출성장률의 추이 분석했고 그 결과는 <Figure 5>와 같다. <Figure 5>에 보면 기존 성장률(창업기업)에 비해서 조정된 성장률(창업기업(nor))의 성장률은 다소 감소하는 것을 관찰할 수 있는데, 경향 자체는 큰 차이를 보이지 않았다. 따라서 경제적 상황 고려해서 정규화(normalization)할 수는 있지만, <Figure 5>에서 나타난바와 같이 경향에 큰 차이를 보이지 않았고, 정규화해서 체계적 성장을 제외하는 모형을 제시하는 경우 예측 모델에서 다시 경제적 상황을 예상해야 하는(체계적 성장 예측) 문제점도 있다. 또한 본 연구가 제시하고자 하는 객관적이고 신속한 모델 제시라는 측면에서도 비효율적이라고 판단해서 본 연구에서는 체계적 성장에 대한 정규화는 진행하지 않았다.

본 연구가 분석한 KSIC 분류에는 서비스 분

8) 비모수분석인 Friedman 분석에서도 반복측정 분산분석과 동일한 결과를 확인할 수 있었다.

류도 일부 포함되어 있기 때문에 <Figure 5>에는 중소기업 전산업에 대한 매출성장률도 제시되어 있는데, 제조업의 경향과 거의 차이가 없었기 때문에 본 연구와 시스템에서 산업의 평균은 중소기업제조업의 연평균성장률로 대푯값을 활용하게 된다.



(Figure 5) Comparison of the sales growth rate of SMEs and start-up companies in Korea, compared to the previous year

4.3 산업별 창업 매출성장 형태 차이

본 연구가 제시한 매출성장률 기반의 신규 사업 매출추정 방법의 활용 가능성을 검토하기 위한 마지막 질문인 산업별 매출성장률 차이 가능성 검토를 위해서 분산분석과 잠재성장모형을 활용한 결과가 다음에 제시되어 있다.

4.3.1 분산분석의 결과

산업분류(KSIC)가 매출성장에서 유의미한 차이를 주는지 간단히 분석하기 위해서는 일원 분산분석을 활용할 수 있다. 그런데 본 연구가 활용한 KSIC가 코드가 너무 많기 때문에(53개 집단), 정제된 31,984건에서 최소한 200건 이상의

케이스가 존재하는 일련의 산업분류 코드를 선택했다. 대분류 22~31까지 10종의 분류를 선택했고, 해당 케이스는 9,128건에 해당된다. 이 사례에 대한 분석한 결과에 따르면 분산의 동질성 검정이 만족되지 못했다($p=0.000\sim 0.001$). 따라서 Welch 검정결과로 평균의 동질성을 분석했는데, 첫 3개 기간에서 산업분류는 유의미한 차이가 있는 것으로 나타났다($p=0.000\sim 0.003$). 분산의 동질성 검정이 위반되었기 때문에 사후검정은 Games-Howell 검정 결과를 살펴보았는데, 먼저 창업 후 첫 번째 매출성장률에서는 분류 26이 다른 산업과 다수 유사해서 분류 22, 25, 28, 29, 30과 동일한 집단으로 나타났으며($p<0.05$), 분류 23, 27과 31은 다른 모든 산업과 다른 것으로 나타났다($p>0.05$). 두 번째 성장률에서는 오히려 분류 31이 분류 22, 26, 28 그리고 29와 동일한 집단으로 나타났으며($p<0.05$), 분류 23, 24, 27과 30이 타 산업과 다른 것으로 나타났다($p>0.05$). 마지막 세 번째 매출성장률은 단지 분류 28과 31 그리고 분류 29와 31만 동일 집단으로 판단되었으며($p=0.025$, $p=0.040$), 다른 산업분류 간에는 모두 통계적으로 유의미한 차이가 있었다($p>0.05$).⁹⁾ 즉 일원 분산분석결과로 보면, 매출성장률은 산업별로 차이가 있다고 주장할 수 있는 것이다.

4.3.2 잠재성장모형 분석의 결과

앞서 살펴본 반복측정 분산분석이나 일원 분산분석은 시계열 자료를 분석하는데 한계를 가진다. 이런 한계를 극복하기 위해서 잠재성장모형을 활용해서 비교 분석했다. 비조건적 모델 분석을 통하여 창업 기업의 첫 매출 후 4년간 매출

9) 비모수분석인 Kruskal-Wallis 검정에서도 유사한 결과를 확인할 수 있었다.

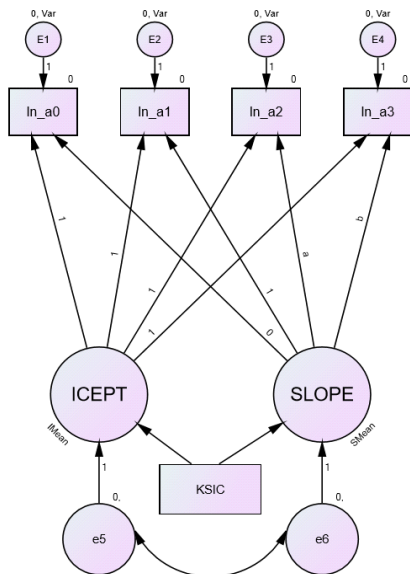
액 변화(성장)의 초기치(intercept)와 변화율(slope), 2차항(quad)을 구한 후 2단계 조건적 모델 분석단계에서 1단계에서 얻어진 잠재요인(latent factor)으로서의 초기치, 변화율, 이차항을 다양한 요인에 연결시켜 초기치와 변화율에 영향을 미치는 효과를 찾아내는 것이다(Bollen & Curren, 2006). 본 논문에서는 1단계의 비조건적 모델을 설정하는데 초점을 두고 있다.

잠재성장모형은 Preacher et al.(2008)나 Kim(2009)에서 보면 6가지 유형으로 나누었는데, <Figure 3>을 바탕으로 성장의 추세 추정이 가능하기 때문에 본 연구는 그 중에서 (0.1.a.b)모형을 선택해서 분석했으며, 매출액의 비정규분포를 고려해서 자연로그로 변환된 매출을 변수로 활용했다. 또한 공변량으로 산업분류(KSIC)를 고려했는데 더미변수를 활용했다(<Figure 6> 참조). <Table 2>에는 모델의 적합도가 분석되어

있는데, 더미변수로 선정된 분류는 23, 24, 27 산업으로 앞선 분산분석에서 다른 산업과 특히 다르게 나타난 분류다. 결과를 보면 세 가지 분류에서 적합도 지수 차이는 거의 없었으며, 일부 지수에 부족한 부분도 있었지만(예 RMSEA), 다른 전반적인 모델 적합도는 우수한 것으로 나타났다.

<Table 2> The results of goodness-of-fit index analysis for the non-structural basic model

Index Models	χ^2	CMIN/df	NFI	TLI	CFI	RMSEA
Dummy (KSIC=23)	2388	298.5	0.966	0.958	0.966	0.181
Dummy (KSIC=24)	2389	298.6	0.966	0.958	0.966	0.181
Dummy (KSIC=27)	2388	298.6	0.966	0.958	0.966	0.181
Criteria			>0.9	>0.9	>0.9	<0.1



<Figure 6> The latent growth model reflecting covariance

분석결과를 보면 세 가지 산업분류 모두 ‘KSIC→ICEPT’는 유의미한 영향을 주었으며 ($p < 0.05$, C.R. 값 $> \pm 1.965$), ‘KSIC→SLOPE’는 KSIC가 분류 24인 경우만 유의미한 결과가 나타났다. 산업분류 24에 대한 구체적인 모형의 주요 추정값은 <Table 3>과 같은데, 절편의 비표준화 계수와 기울기가 모두 0보다 큰 것을 보면 대분류 24는 다른 산업보다 초기값과 성장률이 높은 것을 확인할 수 있다. 시계열 관계를 고려한 잠재성장모형에서도 산업분류에 따라 창업기업의 초기 매출액은 물론 매출변화(매출성장률)에서도 통계적으로 유의미하게 달라질 수 있음을 확인할 수 있었다.

(Table 3) The results of non-standardized coefficient and relation analysis by covariance (KSIC) (Regression weight)

Statistics Relationships	Estimate	S.E.	C.R.	P
KSIC→ICEPT	0.686	0.060	11.5	>0.001
KSIC→SLOPE	0.072	0.025	2.9	0.004
SLOPE→LN (1 st year later)	1.000	N/A	N/A	N/A
SLOPE(a)→LN (2 nd year later)	1.238	0.004	347.4	>0.001
SLOPE(b)→LN (3 rd year later)	1.371	0.004	356.5	>0.001
D KSIC 24 (Dummy)	0.055	0.002	23.0	>0.001

이상의 세 가지 문제에 대한 결과를 정리하면, 첫째 국내 사례에서 창업기업은 매출 성장률이 기존 기업의 매출성장률 평균보다 높지만 평균 회귀 현상이 관찰되었다. 따라서 일정시기 이후 매출 성장률은 산업 평균을 활용할 수 있다는 점을 확인했으며, <Figure 1과 2>에 제시된 본 연구의 매출성장률 기반의 신규 사업 매출추정 방법이 효과적일 수 있음을 확인했다. 두 번째로 창업시기에 따라 거시경제적 영향이 다를 수 있음을 확인했지만, 정규화의 효과가 크지 않고 오히려 예측 효율성을 낮출 수 있어서 정규화는 고려하지 않는 것으로 결정했다. 마지막으로 산업별로 초기 매출성장률의 경향이 다른 것을 분산분석과 잠재성장모형을 통해 확인했다. 따라서 <Figure 1과 2>의 매출성장률 기반의 신규 사업 매출추정 방법은 산업별로 다르게 추정되어야 한다는 결론을 도출했다.

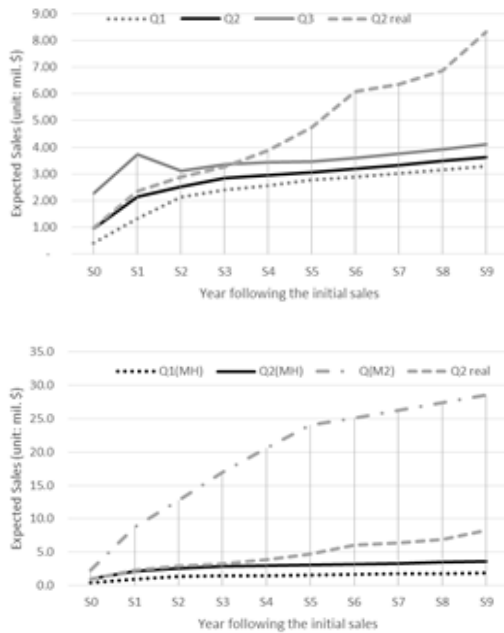
5. 토의

5.1 산업별 매출 추정 모델 개발

본 연구에서 제시된 산업별 초기 매출액 분포와 산업별 첫 매출 후 성장률을 활용한다면, <Figure 2>의 시스템과 같이 신규 사업의 매출 분포 범위를 추정해 볼 수 있다. 앞선 잠재성장모형 분석에서 타산업과 다른 특징이 확인된 대분류 24산업을 사례로 분석하면, 초기 매출액은 4.85억(Q1, 사분위수 25%), 11.68억원(Q2, 중위수), 27.45억원(Q3, 사분위수 75%)과 첫 매출 후 매출성장률을 활용해서 매출액을 추정할 수 있다.¹⁰⁾ 대분류 24 산업에 속한 신생기업의 최근 10년간 첫 매출 후 성장률은 119.6%, 18.3%, 12.5%, 3.9%, 3.4%로 변화하는 것으로 나타났다(Q2 기준). 이런 추세를 활용하면 매출액을 추정할 수 있는데, 앞선 분석에서 신규 사업의 매출 성장률도 특정 기간(6년차)부터 큰 차이가 없고 평균에 회귀하는 경향이 확인된 만큼 매출성장률은 6년차부터는 일정한 값으로 추정한다(중소기업 최근 3년간 평균 매출성장률 4.4%). 먼저 Range model에서 고려되어야 하는 것은 매출성장률도 범위로 적용해야 한다는 것인데, 전기의 매출액을 중위수로 동일하게 적용할 것인가 아니면 매출액도 매출성장률과 같은 사분위수를 선택할 것이냐에 따라서 추정된 매출액의 범위는 큰 차이를 나타내기 때문이다. 같은 매출액 범위를 활용해도, 성장률도 범위가 있기 때문에 평균에 수렴하는 경향(성장률)의 가정에 따라서 극단적으로 빠르게 수렴하는 경향과 극단적으로 느리게 수렴하는 경향으로 나타날 수 있다. 이를 비교한 Range model의 예측 결과가 <Figure 7>에

10) 이 값이 <Table 1>과 다른 것은 이상값을 제거한 자료를 바탕으로 산출되었기 때문이다.

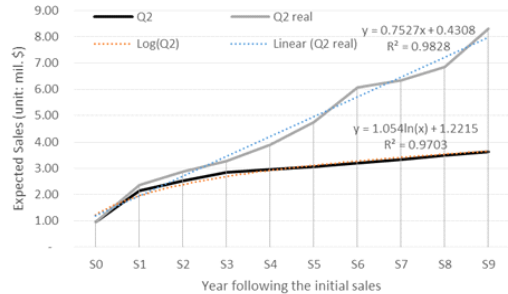
제시되어 있다.



<Figure 7> Sales estimation results based on the range model (KSIC=24, Upper: median reverting assumption, Lower: non-reverting assumption)

산업별 매출 추정을 위한 선형 추정식(Linear model)도 개발이 가능한데 <Figure 8>과 같이 중위수 성장률과 중위수 초기 매출액을 기반으로 제시할 수 있다. 예측된 결과는 매출성장률이 급격히 감소해서 평균에 수렴하는 형태가 되기 때문에 근사식은 로그함수로 표현될 수 있다. <Figure 8>에서 보면 실제 달성한 매출액의 통계량(Q2 real)의 중위수와 Linear model의 예측결과와의 차이가 크게 나타나고 있다. 기간별 통계적 중위수의 경향은 오히려 선형함수에 가까워 우리의 추정을 혼란스럽게 하거나 Linear model의 예측력을 의심하게 한다. 여기서 주의할 것은 산

업의 중위수 매출액의 선형적 증가는 특정 기업의 지속적 성과라기보다는 전반적인 산업의 성장추세로 보아야 한다는 것이다. 특정 기업의 매출액은 앞선 선행 연구 분석에서도 제시된바와 같이 특정한 경향이 없다. 따라서 <Figure 8>의 결과만으로 선형예측 모형의 우수성을 논하기는 무리가 있고, 오히려 Jovanovic(1982)의 주장이나 지브라의 법칙을 고려한다면 동일 기업의 매출 경향은 본 연구가 제시한 로그함수 형태의 Linear model 추정결과가 현실적이다.



<Figure 8> Example of a sales estimation formula developed based on the linear model (KSIC=24)

5.2 모델의 검증

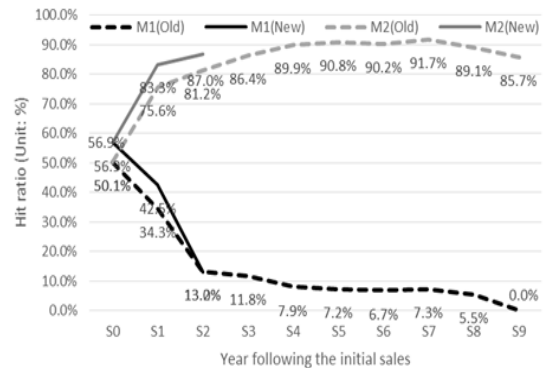
Range model의 예측력은 기존에 모델을 개발에 하는데 활용한 케이스와 활용하지 않은 새로운 케이스에 대해서 첫 매출과 그 후 경과기간마다 <Figure 7>의 범위가 얼마나 잘 맞추는지 검증한다. 이것은 전기 매출액에 대한 정보가 주어지지 않은 상황에서 매출 후 경과기간만으로 모델이 정확히 예측했는지 분석하는 가혹한 방법으로 전체 케이스 중에서 몇 케이스나 범위안에 드는지 정확도 즉 정분류비율(hit ratio)로

<Figure 7> 예측 결과의 적절성을 다루게 되며, 절대적 값으로만 검정하게 된다.

<Figure 7>에서 제시된 모델이 제안한 추정 매출액의 범위가 얼마나 예측력을 가졌는지 검토하기 위해서 우리는 두 가지 자료를 활용하며, <Figure 7>이 제시한 두 가지 범위 모두의 범주 예측력을 비교한다. 먼저 검정에 활용되는 첫 번째 자료는 본 연구에서 <Figure 7>을 도출하기 위해서 이미 활용된 자료(이하 Old DB)다. Old DB는 2005년부터 2011년까지 창업한 기업 중에서 첫 매출 후 4년간 지속적인 매출이 발생했고, 성장률이 상하위 10%를 벗어나지 않았던 501개 기업의 매출 실적이 대상이 된다(이하 KSIC=24). 다음 자료는 2012년~2014년에 창업해서 본 연구의 성장률 분석에서 제외되었던 즉 성장률 모델 구축에 활용된 적 없었던 자료(이하 New DB)를 활용했는데, 지속적인 매출에 대한 제약이나 상하위 10%와 같은 이상값에 대한 제약이 없는 모든 보유 자료를 대상으로 했는데 153개의 매출 실적이 대상이 된다.¹¹⁾ 이런 두 가지 자료군에 대해 범위의 정확성(분류정확도)을 분석한 결과가 <Figure 9>에 제시되어 있다. 여기서 M1은 <Figure 7 상>의 중위수 수렴 가정에 대한 범위를 의미하며, M2는 <Figure 7 하>의 중위수 수렴 가정 없이 동일 사분위 성장률이 적용된 범위를 의미한다.

<Figure 9>에서 제시된 모델 정확도 비교 결과에 따르면, 첫 매출액 추정은 정분류 비율은 50~57% 수준으로 보통으로 판단된다.¹²⁾ 좀 더 보수적인 M1(중위수 회귀)의 범위는 Old DB에 대해서 첫 매출액 추정 이후 2년차까지 정확도가

크게 낮아져 13% 수준까지 낮아지는 것으로 나타났다. 이후 감소세는 낮아졌지만, 7% 수준을 유지하며 매우 저조한 성능을 보였다. 성장률 구축에 활용되지 않은 New DB에서도 비슷한 성능이 예측되었다. 매우 낙관적인 가정인 M2(사분위 유지)의 범위는 Old DB에 대해서 첫 매출액 이후 지속적으로 정확도가 높아져서 5년차에 90%를 상회하는 우수한 정확도를 보였으며, 평균회귀하는 구간에서도 지속적 90%를 상회하다 8년차부터 다소 정확도가 감소하는 것으로 나타났다. New DB에서도 훌륭한 정확도를 보였는데 1년과 2년차 모두 83%에 이르는 높은 정확도를 보였는데, 새로운 자료는 이상값의 제약이 없었다는 것을 감안하면, 제시된 모델의 활용가능성을 확인할 수 있었다.



<Figure 9> Comparison of the accuracy (hit ratio) of the sales estimate model (KSIC=24)

11) 창업 후 매출 실적이 없는 기업은 대상에서 제외되었다.

12) 일반적으로 50%는 hit ratio로 낮은 수준이지만, 초기 매출액의 범위를 25~75%로 적용한 것을 감안하면 50% 수준은 당연하며, 오히려 새로운 자료에서 좀 더 높았다는데(56.9%) 의미가 있다.

6. 결론 및 연구의 제한점

본 연구는 최근 10년내 국내에서 창업한 15만여 개 신규기업만의 매출을 분석해서, 국내 창업기업은 매출성장률이 창업후 초기 몇 연간은 산업 평균대비 크게 높지만, 평균수렴 현상을 보이는 것을 밝혔다. 또한 창업시기가 매출성장률에 영향을 주지만, 그 정도가 높지 않음을 보였고, 매출성장률이 산업분류별로 다를 수 있음을 보였다. 이런 현상과 관련 산업의 창업기업의 실적을 활용해서 매출성장률 기반의 신규 사업 매출액 추정 모델 두 가지를 제안했다. 또한 제안된 Linear model을 포함한 Range model의 예측력이 받아들일만한 수준임을 밝혔다.

본 연구에서 제안된 방법은 객관적이고 신속하게 산업별 신규 사업의 매출액을 추정하게 하며, 다른 방법(Top-down, Bottom-up)에 의해 추정된 매출액이 관련 산업의 일반적인 경우에서 벗어나는지 판단할 수 있는 참조 정보를 제공할 것으로 기대된다. 특히 본 연구의 결과는 실무적으로 신규 사업 진입을 위한 타당성 검토나 기술가치평가에서 유용한 참조 정보로 활용할 수 있을 것이다. 기존의 Top-down 방법을 활용할 때는 시장규모나 시장 점유율에 대한 범위를 정하는데도 참고 할 수 있고, Bottom-up 방법을 활용할 때는 성장률 평균 수렴기간 정보를 고려해서 추정기간을 설정할 수도 있을 것이다. 본 연구가 제시한 두 가지 모델은 신속하고 객관적인 신규 사업 매출을 가능하게 해서 사업성타당성 분석이나 기술가치평가 과정의 효율성을 개선시켜 줄 것으로 기대한다.

학문적으로 볼 때 상장기업과 같은 안정적인 기업뿐만 아니라 일반 중소기업 중에서도 창업기업에서 평균 수렴현상이 발견된다는 것을 다

시 확인했다. 특히 대규모 자료를 활용해서 창업기업의 매출성장률에서 일반 상장기업과는 다른 경향의 평균 수렴현상이 발견됨을 보였고, 산업별로 차이가 있다는 것을 통계적으로 보였는데 본 연구는 의미가 있다고 할 수 있다. 실무적으로 특정기업의 매출 추정 방법에 유용한 Linear model의 활용 가능성이 높다면, 정책적 측면에서 불특정 기업들의 매출 추정 방법에 활용 가능한 Range model의 활용 가능성이 높다고 설명할 수 있다. 특정 산업군이나 기업군들의 창업활동과 실적을 분석할 때는 표준적 비교 자료를 제공해 비교할 수 있게 해준다는 측면에서 정책적 활용 가능성도 있다는 것이다.

본 연구는 신규 사업(기술사업화 등)의 매출 추정을 위한 산업별 매출액 추정 모델 개발 방법을 제안했는데, Range model에서는 초기 매출액과 매출성장률을 범위로 제시했다. 이것은 산업의 특성은 물론 기업의 산업 내에서 경쟁적 지위에 따라 평균회귀 속도가 달라진다는 주장(Palepu & Healy, 2007)을 고려한 것으로 기업의 경쟁력에 따라 산업내에서 초기 매출액과 매출성장률이 달라질 수 있기 때문에 이런 부분의 반영을 염두에 두고 범위의 활용을 제시한 것이다. 따라서 향후에는 적절한 분포 활용을 위해서 산업내 기업의 경쟁력을 순위로 판단할 수 있는 정량적 또는 정성적 방법(체크리스트 등)의 개발이 병행된다면 본 연구가 제시한 매출 추정모델이 좀 더 정교하게 활용될 수 있을 것이다. 또한 표준산업분류의 사업적 대표성과 동질성에 대한 한계를 고려하면, 기업의 유사도 지표를 개발해서 유사한 기업만 뽑아서 매출성장률을 산출하는 방법도 좋은 대안이 될 수 있을 것이다. 본 연구가 제시한 Linear model은 창업기업의 매출정보가 다소 존재하거나 분석 목적상 단일 예측값

이 필요한 경우 좋은 예측 모델이 될 수 있다.

본 연구에서는 초기 매출 지연 현상에 대한 분포 정보만 제시되어 있고 구체적인 분석은 누락되어 있는데, 본 연구가 제시한 매출 추정 방법을 실무에 적용하기 위해서는 창업 후 첫 매출까지 지연되는 기간도 고려되어야 한다는 한계가 있다. 지연 기간을 산정할 때는 <Table 1>에서 제시된 통계값을 활용할 수도 있지만, 분석 대상에 대한 객관적 정보를 바탕으로 예측할 것을 권장한다.

참고문헌(References)

- Bollen, K. A. and Curren, P. J., *Latent curve models a structural equation perspective*, Wiley- Interscience, Hobken, NJ, 2006.
- Daunfeldt, S. O. and Halvarsson, D., “Are high-growth firms one-hit wonders? Evidence from Sweden,” *Small Business Economics*, Vol.44, No.2(2015), 361-383.
- Duncan, T. E., Duncan, S. C. and Strycker, L. A., *An introduction to latent variable growth curve modeling : Concepts, issues, and applications*, Lawrence Erlbaum Associates, Mahwah, NJ, 2006.
- Girden, E. R., *ANOVA: Repeated measures*, Sage University paper series on quantitative applications in the social sciences, 07-084, SAGE, Newbury Park, CA, 1992.
- Jun, S. P., Sung T, E, and Park H. W., “Forecasting by analogy using the web search traffic,” *Technological Forecasting and Social Change*, Vol.115, (2016), 37-51.
- Jun, S. P., & Park, D. H., “Consumer information search behavior and purchasing decisions: Empirical evidence from Korea,” *Technological Forecasting and Social Change*, Vol.107, (2016), 97-111.
- Jun, S. P., Park, D. H., & Yeom, J., “The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference,” *Technological Forecasting and Social Change*, Vol.86, (2014), 237-253.
- Jovanovic, B., “Selection and the Evolution of Industry.” *Econometrica*, Vol.50, No.3 (1982), 649-670.
- Kaplan, D. W., *Structural equation modeling: Foundations and extensions*, SAGE, CA, 2009.
- Kim, K. S., *Analysis of latent growth modeling and structural equation model*, Hannarae, Seoul, 2009.
- Kim, K. J. and Ahn, H. C., “Optimization of Support Vector Machines for Financial Forecasting,” *Journal of Intelligence and Information Systems*, Vol.17, No.4(2011), 241-254.
- Kim, K. J., *K-IFRS Financial Statement Analysis and Valuation*, 5th ed., Changmin, Seoul, 2015.
- Kim, M. C. and Kim, D. S., “An Estimation of Market Potential in Ubiquitous Health industry - Focused on Communication Companies,” *International Journal of Information and Communication Engineering*, Vol.12, No.12(2008), 2131-2135.
- Kline, R. B., *Principles and practices of structural equation modeling*, Gilford press, NY, 1998.
- Koller, T., Goedhart, M., and Wessels, D.,

- Valuation: measuring and managing the value of companies 5th edition*, John Wiley and Sons, 2010.
- Lee, H. J. and Kang, S. B., "Analysis of latent growth model using repeated measures ANOVA in the data from KYPS," *Journal of the Korean Data & Information Science Society*, Vol.24, No.6(2013), 1409-1419.
- Lee, D., Jung, Y., Jung, J. and Park, D., "An Expert System for the Estimation of the Growth Curve Parameters of New Markets," *Journal of Intelligence and Information Systems*, Vol.21, No.4(2015), 17-35.
- Lundholm, R., McVay, S. and Randall, T., *Forecasting sales: A model and some evidence from the retail industry*, Unpublished working paper, University of British Columbia and University of Washington, 2010.
- Shin, M. C., *Basic Statistics for Business & Economics*, Changmin, Seoul, 2010.
- Song, M. S., Park, C. S. and Lee, J. J., *Nonparametric statistics*, Freedom academy, Gyeonggi-do, 2003.
- Palepu, K. G. and Healy, P. M., *Business Analysis and Valuation: IFRS Edition*, Cengage Learning EMEA, 2007.
- Park, D. H., Chung, J., Chung, Y. J. and Lee, D., "Development of Market Growth Pattern Map Based on Growth Model and Self-Organizing Map Algorithm: Focusing on ICT Products," *Journal of Intelligence and Information Systems*, Vol.20, No.4(2014), 1-23.
- Park, H. S. and Kim, J. H., "The Relationship between Firm Size and Firm Growth in the Korean Non-manufacturing Industry," *Korean Business Education Research*, Vol.26, No.4(2012), 419-431.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C. and Briggs, N. E. *Latent growth curve modeling, Review of statistical models for analyzing repeated measures data*, SAGE, LA, 2008.
- Welch, J., "Combining Mean-Reversion of Sales Growth and Low Valuation Multiples into Single Investment Strategy," *World Review of Business Research*, Vol.2, No.2(2012), 32-44.
- Yang, H. B., Moon, J. H. and Jung, M. A., "Estimation of Market Size and Value Added by Embedded SW Industry Cluster," *The Journal Of Korean Institute Of Communications And Information Sciences*, Vol.35, No.8(2010), 1211-1216.
- Yoo, H. S., Seo, J. H., Jun, S. P., and Seo, J., "A Study on an Estimation Method of Domestic Market Size by Using the Standard Statistical Classifications," *Journal of Korean Technology Innovation Society*, Vol.18, No.3(2015), 387-415.

Abstract

A Data-based Sales Forecasting Support System for New Businesses

Seung-Pyo Jun* · Tae-Eung Sung** · San Choi***

Analysis of future business or investment opportunities, such as business feasibility analysis and company or technology valuation, necessitate objective estimation on the relevant market and expected sales. While there are various ways to classify the estimation methods of these new sales or market size, they can be broadly divided into top-down and bottom-up approaches by benchmark references. Both methods, however, require a lot of resources and time. Therefore, we propose a data-based intelligent demand forecasting system to support evaluation of new business.

This study focuses on analogical forecasting, one of the traditional quantitative forecasting methods, to develop sales forecasting intelligence systems for new businesses. Instead of simply estimating sales for a few years, we hereby propose a method of estimating the sales of new businesses by using the initial sales and the sales growth rate of similar companies. To demonstrate the appropriateness of this method, it is examined whether the sales performance of recently established companies in the same industry category in Korea can be utilized as a reference variable for the analogical forecasting.

In this study, we examined whether the phenomenon of “mean reversion” was observed in the sales of start-up companies in order to identify errors in estimating sales of new businesses based on industry sales growth rate and whether the differences in business environment resulting from the different timing of business launch affects growth rate. We also conducted analyses of variance (ANOVA) and latent growth model (LGM) to identify differences in sales growth rates by industry category. Based on the results, we proposed industry-specific range and linear forecasting models.

This study analyzed the sales of only 150,000 start-up companies in Korea in the last 10 years, and identified that the average growth rate of start-ups in Korea is higher than the industry average in the first

* Corresponding Author: Seung-Pyo Jun

Technology Commercialization Analysis Center, Korea Institute of Science and Technology Information
66, Hoegi-ro, Dongdaemun-gu, Seoul 130-741, Republic of Korea

Tel: +82-2-3299-6095, Fax: +82-2-3299-6041, E-mail: spjun@kisti.re.kr;spjunn@msn.com

** Technology Commercialization Analysis Center, Korea Institute of Science and Technology Information

*** University of Science & Technology (UST)

few years, but it shortly shows the phenomenon of mean-reversion. In addition, although the start-up founding juncture affects the sales growth rate, it is not high significantly and the sales growth rate can be different according to the industry classification. Utilizing both this phenomenon and the performance of start-up companies in relevant industries, we have proposed two models of new business sales based on the sales growth rate.

The method proposed in this study makes it possible to objectively and quickly estimate the sales of new business by industry, and it is expected to provide reference information to judge whether sales estimated by other methods (top-down/bottom-up approach) pass the bounds from ordinary cases in relevant industry. In particular, the results of this study can be practically used as useful reference information for business feasibility analysis or technical valuation for entering new business. When using the existing top-down method, it can be used to set the range of market size or market share. As well, when using the bottom-up method, the estimation period may be set in accordance of the mean reverting period information for the growth rate. The two models proposed in this study will enable rapid and objective sales estimation of new businesses, and are expected to improve the efficiency of business feasibility analysis and technology valuation process by developing intelligent information system.

In academic perspectives, it is a very important discovery that the phenomenon of ‘mean reversion’ is found among start-up companies out of general small-and-medium enterprises (SMEs) as well as stable companies such as listed companies. In particular, there exists the significance of this study in that over the large-scale data the mean reverting phenomenon of the start-up firms' sales growth rate is different from that of the listed companies, and that there is a difference in each industry. If a linear model, which is useful for estimating the sales of a specific company, is highly likely to be utilized in practical aspects, it can be explained that the range model, which can be used for the estimation method of the sales of the unspecified firms, is highly likely to be used in political aspects. It implies that when analyzing the business activities and performance of a specific industry group or enterprise group there is political usability in that the range model enables to provide references and compare them by data based start-up sales forecasting system.

Key Words : Sales Forecasting Systems, Analogical Forecasting, Sales Estimation, Technology Valuation, Business Feasibility Analysis, Latent Growth Models(LGM)

Received : February 20, 2017 Revised : February 20, 2017 Accepted : February 27, 2017

Publication Type : Regular Paper Corresponding Author : Seung-Pyo Jun

저 자 소개



전승표

KAIST에서 경영학으로 석사학위를 취득하고, 고려대학교에서 과학관리학 전공으로 이학박사를 취득했다. 현재 한국과학기술정보연구원 기술사업화분석센터에 책임 연구원 및 센터장으로 재직 중이며, 과학기술연합대학원대학교 과학기술경영정책학과 부교수로 재직중이다. Technological forecasting and social change, Scientometrics, Energy policy, Internet research 등 해외학술지와 한국기술혁신학회지, 지능정보연구 등 국내학술지에 주저자로 다수의 논문을 게재했다. 주요 관심분야는 빅데이터를 활용한 수요 예측, 유망 기술 탐색, 기술가치평가, 산업시장분석 등을 위한 지능형 정보 시스템 연구이다.



성태응

텍사스오스틴 주립대에서 전자공학으로 석사학위를 취득하고, 코넬대학교에서 동 전공으로 공학박사를 취득했다. 현재 한국과학기술정보연구원 책임연구원로 재직 중이며, 경희대학교 테크노경영대학원 기술경영학과와 과학기술연합대학원대학교 과학기술경영정책학과 겸임교수로 재직중이다. 관심분야는 기술가치평가, 산업시장분석, 벤처창업 및 데이터 사이언스 기반의 정보시스템 연구이다.



최산

University of Georgia 에서 경제학으로 학사학위를 취득하고, 과학기술연합대학원대학교에서 과학기술경영정책 전공으로 석박통합과정 진행 중이다. 한국과학기술정보연구원 사업화기획분석실에 학생 연구원으로 재직 중이다. 주요 관심분야는 기술경영, 과학기술정책, 기술가치평가, 개발경제 연구이다.