

빅데이터를 활용한 가계부채 부실 조기경보지표 개발에 대한 연구

박성용¹⁾

가

.

.

1) 중앙대학교 경제학부 부교수

< 요약 >

I. 서론

II. 가계부채 결정요인에 대한 기존연구

1. 선행연구
2. 논문의 구성 및 차별점

III. 국내 가계부채의 추이 및 조기경보지수

1. 가계부채 추이 및 지수
2. 가계부채 조기경보지수

IV. 실증분석

1. 분석자료
2. 실증분석 모형: 인공신경망 모형
3. 실증분석 결과
4. 조기경보지수 예측

V. 결론

부록

참고문헌

〈 요약 〉

최근 저금리와 주택가격 상승에 따라 가계부채는 2016년 기준 약 1,270조원을 기록하며 지속적으로 상승하고 있다. 우리나라는 1997년 외환위기 이후부터 본격적으로 가계대출을 취급하기 시작하면서 2004년부터 본격적으로 가계대출은 꾸준히 증가하는 추세를 보여 왔다. 특히 2014년 이후 연속적으로 전년대비 10% 이상의 증가율을 보였으며, 2016년에는 11.6% 가량 증가하여 명목 GDP의 77%에 해당하는 수치를 나타냈다.

본 연구는 최근 중요한 이슈로 부각되고 있는 가계부채 부실에 대한 조기경보지표를 개발하는데 그 목적이 있다. 가계부채의 결정요인을 분석한 여러 연구들에서 제안한 거시금융자료와 인터넷 검색자료를 이용하여 조기경보지표를 예측하고 조기에 위험을 경보하는 모형을 제시하였다. 최근에 들어 가계부채의 국내총생산과 총처분가능소득 대비 증가율이 꾸준히 상승하여 그 위험수위에 다다르고 있는 바, 가계부채 위험에 대한 정확한 조기경보는 위험에 대한 대비와 향후 경제정책 설정에 아주 중요한 도움을 주는 것임에는 틀림없다.

가계부채의 결정요인을 분석하는 연구는 꾸준히 진행되어져 왔지만 가계부채 부실에 대한 조기경보모형은 아직 여러 연구에 소개되지 않은 실정이다. 가계부채의 시계열은 2002년 신용위기 시기를 제외하고는 약간의 변동을 보이면서 꾸준히 증가하는 추세를 가진다. 따라서 이렇게 꾸준히 증가하는 추세만으로는 가계부채 부실 위험 자체를 정의 내리기가 힘들며 가계부채 위험에 대한 정의가 없이 조기경보모형을 설정하기에 한계가 존재한다. 몇몇 국외 연구에서는 국내총생산 대비 가계부채의 비율이 50%나 85%가 넘어서면 위험수준이라고 제안하였지만 국내 가계부채 비율은 이미 그 수준을 넘어서 있어 그 임계치를 우리나라의 상황에 적용하기가 현실적이지 않다고 볼 수 있겠다. 따라서 가계부채 시계열 자료를 위기상황을 잘 정의하기 위한 특정한 변환이 필요하다.

본 연구에서는 가계부채 시계열을 적절히 변환한 지수를 사용하였다. 강종구(2017)의 연구는 가계부채의 저량 그리고 유량적인 성격에 따라 국내총생산 및 소비에 미치는 영향이 상이하다는 것을 보였다. 따라서 가계부채의 유량적인 성격과 저량적인 성격을 함께 고려한 지수를 작성한다면 가계부채 부실위험을 좀 더 효율적으로 측정할 수 있다. 본 연구에서는 적절한 가중치를 설정한 유량변수와 저량변수의 블록결합을 통하여 새로운 가계부채 부실지수를 작성하였다. 이렇게 작성된 지수로 가계부채 위험 조기경보모형을 설정하였다.

금융위기에 대한 조기경보모형은 지금까지 여러 연구들에서 여러 모형들이 제시되

어제 왔으며 가장 많이 쓰이는 모형은 신호접근법과 프로빗·로짓 모형 등이 있다. 이 모형들은 위기에 대한 변수로 이항확률변수를 사용하여 분석한다. 하지만 현실에서 위기 상황은 안정적인 상황에서 갑자기 위기상황으로 넘어가지 않으므로 이와 같은 모형설정은 한계가 존재한다. 따라서 본 연구에서는 가계부채의 위기를 5단계로 나누어 단계별 위기상황을 설정하였다.

기존의 조기경보모형은 종속변수와 설명변수의 복잡한 관계를 지나치게 단순화시킨 선형모형을 설정하여 분석하고 있다. 물론 가계부채 부실지수를 잘 설명할 수 있는 명확한 그리고 많은 수의 설명변수가 존재하고 취득이 가능하다면 부실 위험에 대한 선형모형의 예측력은 다른 복잡한 비선형모형의 그것에 비해 높을 수 있다. 하지만 현실에서는 그러한 많은 수의 명확한 설명변수들을 선정하는데 한계가 있다. 따라서 본 연구에서는 사용되는 모형의 함수를 선형적으로 설정할 필요가 없는 인공신경망 모형을 사용하였다. 많은 비선형 모형은 인공신경망 모형의 특별한 경우이다. 인공신경망 모형은 종속변수와 설명변수의 복잡한 관계를 잘 근사시킨다고 알려져 있다.

인공신경망 모형을 이용하여 추정된 결과를 가지고 본 연구에서는 2017년 2분기부터 2019년 2분기까지 가계부채 위험 단계를 예측을 하였다. 조건부 예측방법을 이용하여 분석한 결과 고려된 두 모형에서 공통적으로 예측기간에 단계 5에 이르는 가계부채 부실위험은 없는 것으로 나타났다. 또한 모형에 따르면 2017년 3분기까지는 보통단계인 단계 1-2를 유지하다가 2017년 4분기에 들어서 위기 단계 4까지 상승하는 것을 볼 수 있으며 따라서 2017년 4분기에 적극적으로 가계부채의 추이와 관련 변수들을 주의 깊게 살펴 볼 필요가 있다고 하겠다.

I. 서론

최근 저금리와 주택가격 상승에 힘입어 가계부채는 2016년 기준 약 1,270조원을 기록하며 지속적으로 상승하고 있다. 우리나라는 1997년 외환위기 이후부터 본격적으로 가계대출을 취급하기 시작하면서 2004년부터 가계대출은 본격적으로 꾸준히 증가하는 추세를 보여 왔다. 특히 2014년 이후 연속적으로 전년대비 10% 이상의 증가율을 보였으며, 2016년에는 11.6% 가량 증가하여 명목 GDP의 77%에 해당하는 수치를 나타냈다.

이러한 가계부채 증가의 원인은 크게 세 가지로 나누어 볼 수 있다. 첫 번째는 낮은 금리이다. 2017년 현재 한국은행의 기준금리는 2010년 이후 가장 낮은 1.25%를 유지하고 있으며, 이에 따라 유동성이 확대되고 가계의 차입 비용이 낮아지면서 가계부채 증가의 원인이 되었다. 두 번째는 부동산 시장의 호황이다. 낮은 금리와 함께 아파트 분양 물량이 증가하면서 주택담보대출의 수요가 크게 증가하며 가계부채 또한 증가하였다. 세 번째는 실질가구소득 증가율의 하락이다. 2014년 이후 증가된 가계대출에 비하여 실질가구소득 증가율은 꾸준히 감소하였고, 2016년에는 전년대비 실질소득이 0.4% 감소하였다. 늘어난 대출에 비하여 소득 하락이 또 다른 대출수요의 증가로 이어지게 된 것이다.

이와 같이 가계의 재무건전성은 거시경제 변수에 많은 영향을 받고 있으며, 경기가 수축하는 국면에서는 가계의 소득과 자산이 감소하여 부채상환 여건이 악화되고, 이에 따른 가계부채 증가는 수축국면을 장기화 시킬 수 있는 원인이 될 수 있다. 따라서 가계부채의 위험 신호를 정확하게 파악하고 추가적인 경기 침체의 충격에 앞서 대처할 수 있는 조기경보모형에 대한 중요성이 더욱 높아지고 있다.

가계부채의 결정요인에 대한 많은 연구들이 존재한다. 대부분의 연구들은 가계의 미시자료를 이용하여 결정요인을 분석하고 있다 (김현정 · 김우영(2009), 김현정 · 김우영 · 김기호(2009), 김우영 · 김현정(2009), 김영일 · 유주희(2013), 함준호 · 조현철 · 권영철(2015), 서상원 · 오권영(2016)). 하지만 가계부채의 부실위험성과 조기경보모형에 대한 연구는 거의 없다. 가계부채의 부실위험성 예측 및 평가를 한 연구는 서상원 · 오권영(2016)이 유일하다. 일반적으로 가계부채의 결정요인에 따라 가계부채의 위험도가 달라진다고 보고 있으므로 미시적 패널자료를 가지고 가계부채를 분석하는 것이 적절하다고 볼 수 있으나 이러한 미시적 패널자료는 거의 연도별 자료로 주어져 있기 때문에 이러한 자료를 가지고 조기경보모형을 설정하는 것은 큰 의미를 가지지 못한다고 볼 수 있겠다. 따라서 본 연구에서는 가계부채와 그 결정요인들의 총량 시계열 자료를 사용하였다.

조기경보모형이 주목받기 시작한 것은 1990년대 금융위기가 발생하면서 부터이다. 1997년 아시아 금융위기는 막대한 규모의 음의 경제적 파급효과를 발생시켰고, 이러한 위기에 대처하기 위한 각국의 방안으로 체계적인 조기경보시스템이 만들어지게 되는 계

기가 되었다 (Glostein et al., 2000). 이후 2008년 글로벌 금융위기는 조기경보시스템의 경제주체별 국가별 위기 전파를 대처하기 위한 모형 개발의 시작점이 되었다.

초기의 조기경보모형은 Kaminsky et al. (1998)가 제시한 비모수적 방법으로 거시경제 변수들의 비정상적인 움직임을 포착할 수 있는 모형과 Berg and Pattillo (1999)의 로짓-프로빗 모형을 이용한 모수적 추정방법을 기반으로 세계 각국의 중앙은행에서 많은 연구가 이루어졌다. 하지만 이러한 모형은 단순한 계산과 제한된 변수로 위기를 예측하기 때문에 정확성의 측면에서 좋은 성과를 내지 못하였다. Ghosh et al., (2009)은 조기경보 모형이 위기를 정확하게 포착하는 것은 불가능 하다는 것을 지적하며 위험 자체를 식별해 내는 것보다 위험요인에 대한 신호를 포착해야 위험을 사전에 방지할 수 있을 것이라고 주장하였다.

금융위기에 대한 조기경보모형은 위기의 시점 및 조기경보지수의 정의가 명확한 반면 가계부채 부실위험에 대한 조기경보모형은 그렇지 못한 면이 존재한다. 가계부채 시계열 자료에 의하면 국내총생산 대비 가계부채 비율 수준은 2002년 신용위기 기간을 제외하면 꾸준히 증가하는 추세를 가진다. 따라서 꾸준히 증가하는 추세를 지닌 시계열에 대해 위험수준은 정의할 수 있는 임계치를 객관적으로 설정하는 것은 그렇게 간단한 문제가 아니다. 몇몇 국외 연구(Cecchetti et al. (2011), Arcand et al.(2015))에서는 국내총생산 대비 가계부채의 비율이 50%나 84%가 넘어서면 가계부채가 경제성장을 저해하는 위험수준이라고 추정하였지만 국내 가계부채 비율은 이미 그 수준을 넘어서 있어 그 임계치를 적용하기가 현실적이지 않다고 볼 수 있겠다. 본 연구에서는 가계부채가 경제성장에 미치는 정의 영향과 부의 영향을 고려하여 적절히 변환한 가계부채 지수를 사용하였다. 가계부채의 저량 그리고 유량적인 성격에 따라 국내총생산 및 소비에 미치는 영향이 상이하므로 이와 같은 가계부채의 성격을 같이 고려한 지수를 작성하였다. 이러한 가계부채 지수는 적절한 가중치를 설정한 유량변수와 저량변수의 블록결합을 통하여 작성될 수 있으며 이 지수를 이용하여 가계부채 위험 조기경보모형을 설정하였다.

기존의 모형이 위기시점에 앞서 정확한 신호를 발생시키지 못하면서 광범위한 자료를 활용하기 위한 모형 개발로 자연스럽게 흐름이 바뀌게 되었다. 최근 금융위기 이후의 많은 정보를 포함하고 있는 자료가 이용 가능해지고 경제주체, 국가, 변수들 간의 복잡한 상관관계를 모형에 포함할 수 있게 되면서 빅데이터로부터 필요한 정보를 선별하여 정확하게 신호를 포착할 수 있는 새로운 형태의 모형이 주목을 받기 시작하였다. Choi and Varian(2012)을 시작으로 이탈리아(D'Amuri and Marcucci, 2017), 스페인(Artola and Galan, 2012), 칠레(Carriere-Swallow and Labbe, 2010), 터키(Chadwick and Sengül, 2012), 이스라엘(Suhoy, 2009) 등의 많은 중앙은행에서 빅데이터를 활용한 조기경보모형을 개발하여 활용하고 있다.

일반적인 사용되던 거시경제자료는 시계열 자료 사이의 간격이 길고 지표가 즉각적으로 발표되는 않기 때문에 상당기간 시차가 존재한다는 단점이 있었다. 인터넷 검색자료를 기반으로 하는 빅데이터 형태의 자료는 기존에 사용하던 일반적인 거시경제 자료와

함께 인터넷 검색자료를 함께 포함하고 있다. 이러한 웹 기반의 조기경보모형은 구글트랜드, 네이버트랜드, 뉴스키워드지표를 사용하여 거시경제지표가 포함하지 못했던 경제주체들의 움직임을 직접적으로 포착할 수 있고, 시장에 대한 많은 정보를 분석하여 정확하게 위험요인이 주는 신호를 구분해낼 수 있다(Beau, 2016).

빅데이터를 사용하여 위험을 식별해 내기 위해서는 새로운 방법을 사용해야 한다. 기존의 조기경보모형은 모형은 자기회귀모형 또는 자기회귀이동평균모형과 같은 단순한 모형을 사용하고 있다. 하지만 이러한 선형모형에 기반을 둔 모형은 변수들 간의 복잡한 상호관계를 근사시키는데 어려움이 존재한다. 따라서 본 연구에서는 변수들 사이의 복잡한 관계를 잘 근사시키는 것으로 알려져 있는 인공신경망(artificial neural networks: ANN)모형을 사용하고자 한다.

본 연구에서 사용한 인공신경망모형은 개별적인 문제나 자료를 분석하기보다 복잡한 문제를 해결하기 위한 모형으로서 빅데이터를 활용한 분석에 적합한 모형이다. 인간의 뇌 기능을 모방하여 문제를 효율적으로 해결하는 이러한 모형은 각각의 신경망이 다수의 정보를 분석하여 문제를 해결해 나아가고, 이러한 과정에서 학습을 하는 방식을 가지고 있다. 따라서 많은 변수가 주는 복합적인 신호를 분석할 수 있고 빅데이터를 활용하여 복수의 요인으로부터 오는 가계부채의 위험신호를 분석하기에 적합한 모형이라고 사료된다.

본 연구의 I장에서는 조기경보모형과 연구의 배경에 대하여 간략하게 소개하였다. 이어지는 보고서는 다음과 같이 구성되어 있다. II장에서는 가계부채 결정요인에 대한 기존연구사례들을 살펴보았다. III장에서는 국내 가계부채의 현황을 살펴보고, 부채증가가 조기경보지수 변화에 주는 영향에 대하여 살펴보았다. IV장에서는 인공신경망모형을 사용하여 가계부채의 위험 신호를 실증분석하고 빅데이터를 사용한 모형의 정확성을 파악하였다. 마지막으로 V장에서는 빅데이터와 인공신경망 모형을 사용한 결정요인 분석결과들을 정리하고 이를 기초로 정책적 시사점을 제시하고자 한다.

II. 주요 선행연구

그 동안 가계부채의 위험성을 예측하고 결정요인을 분석하기 위한 다양한 연구가 이루어져 왔다. 우선 가계부채의 부실위험성을 예측한 대표적인 연구는 서상원·오권영(2016)이 있다. 이 연구에서는 연도별 한국노동패널자료를 이용하여 가계부채 위험성에 대한 예측과 스트레스 테스트를 위한 거시경제금융 변수를 포함하는 모형을 구축하였다. 유사한 연구로 김우영·김현정(2009)과 김현정·손종칠·이동렬·임현준·나승호(2013)에서는 가계부채 결정요인과 지속가능성을 분석하기 위하여 한국노동패널자료와

함께 가계금융조사 자료를 이용하여 회귀분석을 실시하였다.

CB자료를 이용한 이동걸·전성인·정재욱·변동준(2014)에서는 자료의 한계를 보완하기 위하여 가구단위 CB자료를 활용하여 가구별 자산규모를 구축하였으며, 가계부채 연체결정모형을 추정하고 스트레스 테스트를 하였다. 또한 함준호·조현철·권영철(2015)은 금융권별, 지역별, 담보유무별, 신용등급 및 소득분위별 가계부채의 경기순응성 변화를 거시정보결합모형을 통하여 분석하였다.

한편 기존의 가계조사자료를 이용한 개별 가구의 분석과는 다르게 국가별 자료를 이용한 박대근·최우주(2015)는 12개 국가의 패널자료를 사용하여 주택가격과 대출심사기준 변화를 고려한 다국 패널모형으로 국가 전체의 가계부채비율 결정요인을 분석하였다. 이와 함께 정운영·정세은(2013)에서는 부동산에 관련된 요소들이 가계부채 증가에 주는 영향에 대하여 분석하였다.

최근의 국제 자료를 포함한 연구로 강종구(2017)는 우리나라와 주요 선진국 자료를 활용하여 자금순환표 기준 가계부채의 형태가 소비와 경제성장에 주는 영향에 대하여 GMM 방법을 사용하여 실증분석 하였다. 또한 성태윤·박기영·안현미(2017)는 가계부채 관리제도의 방향설정이 거시경제에 미치는 영향력을 국제 사례를 바탕으로 비교분석하였다.

기존의 조기경보모형은 일반적으로 Kaminsky and Reinhart(1999)와 Kaminsky et al. (1998)이 제안한 거시경제변수 및 금융 변수의 비이상적인 움직임을 포착하는 비모수적 신호분석 모형과, Berg and Pattillo (1999)가 제안한 로짓-프로빗모형을 사용하여 위기발생 확률을 계산하는 모수적 모형으로 나눌 수 있다. 한 국가 안에서 개별 분야에 대한 조기경보모형은 비모수적 신호분석이 비교적 정확한 결과를 나타내면서 경제부문별 분석에는 신호분석모형이 일반적으로 사용되고 있다 (Davis and Karim, 2008).

박원암(2015)은 신호분석모형으로 1997년 아시아 외환위기와 2008년 글로벌 금융위기를 예측하였으며, 우리나라 외환위기의 경우 비모수적 신호분석이 더 적합하다고 분석하였다. 하지만 우리나라 금융산업의 경우 모수적모형인 로짓-프로빗모형이 일반적인 조기경보모형으로서 사용하고 있으며 이렇듯 자료의 형태와 시장의 구조에 따라 모형의 정확도가 다르게 나타나고 있다.

이러한 기존의 연구들은 가계부채의 결정요인을 분석하기 위하여 다양한 자료를 활용하였다는 장점이 있지만 많은 자료를 포함하거나 비선형 관계를 포착할 수 없고, 가계부채의 추세 및 변동성을 예측하고 위험성을 알려주지 못한다는 단점을 보이고 있다. 가계부채 조기경보시스템을 개발하는데 있어 기존의 연구들과 같이 미시적 자료를 사용하면 가계대출 위기를 식별하는데 도움은 되겠지만 시계열 자료가 연도별로 주어진 패널자료를 조기경보시스템 개발에 사용하는 것에는 시간의 간격이 길다는 분명한 한계가 있다. 따라서 본 연구에서는 새로운 총량자료를 이용하여 조기경보시스템 모형을 설정함으로써 더욱 정확한 예측력을 가지는 모형을 제시하고자 한다.

시계열 자료로서 구할 수 있는 가계부채 자료는 일반적으로 분기별 자료이다. 가계대

출 월별자료를 사용할 수 있기는 하지만 가계부채를 정의하는 것에 문제가 있을 수 있으며, 다른 자료와 함께 분석하는 것에도 많은 제약이 따른다. 이러한 자료의 한계를 극복하기 위하여 최근 인터넷 검색자료를 이용하여 경제변수 예측을 실시하는 연구가 활발히 이루어지고 있다.

구글트렌드를 선형회귀모형에 활용한 Choi and Varian(2012)을 시작으로 Koop and Onorante(2013)에서는 월별 주요 거시경제변수의 단기예측(nowcasting)을 위한 동태적 모형에 구글트렌드 자료를 활용하였으며 기존의 자료를 활용한 결과보다 더 높은 예측력을 보였다. 특히 구글트렌드에 따른 각각의 거시경제변수에 대한 포함확률(inclusion probability)을 적용시켰다는 차이점을 보였다.

Li(2016)는 일별, 월별 구글트렌드 자료를 거시경제 변수를 예측하는데 사용하였다. 이 연구에서는 특히 빅데이터 형태의 자료를 사용함으로써 단기간의 미래에 대한 예측력을 강화하였다. 또한 D'Amuri and Marcucci(2017)에서는 구글트렌드 자료를 활용하여 미국의 실업률을 예측한 결과 기존 자료를 사용했을 때보다 높은 예측력을 보였다.

국내에서도 구글과 같이 네이버트렌드를 통하여 검색자료를 활용할 수 있으며 Kim, Koop and Chung(2016)에서는 이러한 자료를 사용하여 국내거시경제변수를 예측하기위한 방법을 제안하였다. 이 연구에서는 국내 인터넷 사용자들의 특징이 반영된 자료를 사용하여 기존의 연구보다 더 높은 예측력을 보이는 분석을 하였다.

이완수·박양수(2016), Tobback et al.(2016)는 뉴스자료를 기반으로 하는 지수를 사용하여 경제 주체별 거시경제변수에 대한 반응을 경기 국면과 정부 구조에 따라 예측하여 검색자료를 기반으로 하는 분석의 정확성을 확인하였다. 한국언론진흥재단은 신문기사를 기반으로 하는 빅카인즈(BIG KINDS : Korean Integrated News Database)라는 통계서비스를 제공하고 있으며, 이궁희(2017)에서는 빅카인즈를 활용한 경기지수로 거시경제 변수를 예측한 결과가 기존의 지수보다 더 정확한 예측치를 보이는 것으로 나타났다.

최근 많은 연구의 흐름이 빅데이터를 활용하기 위한 방향으로 흘러가면서 예측모형에 대한 연구에서도 방대한 자료를 활용할 수 있는 모형이 주목받고 있다. 특히 인공지능망모형은 뉴런이나 세포가 정보를 처리하는 것과 같이 각각의 결과에 따라 더욱 정확한 분석이 가능하며, 기존의 수학적으로 증명된 모형이 최근 기술의 발전으로 많은 분야에 적용되기 시작하였다. Lawrence et al. (1985)에서는 인공지능망모형이 분석을 하는 과정에서 학습을 함으로써 많은 분야에서 높은 성과를 나타낼 수 있다고 증명하였다. Abbas (2015, 2017)는 복잡한 시차가 포함된 시계열의 비선형관계를 분석할 수 있는 신경망모형이 기존의 모형보다 더욱 정확한 결과를 보인다는 것을 확인하였다.

이와 같이 인터넷 검색자료를 모형에 포함한 연구는 모형의 예측력이 대체로 통계적으로 유의하게 상승되었다는 공통된 특징을 보이고 있다. 인터넷 검색자료의 장점은 기존의 거시경제변수들이 측정하지 못하는 경제 주체의 활동을 변수로 나타낼 수 있으며, 기존의 변수들이 포함하지 못하는 정보를 포함하고 있다는 점이다. 또한 기존의 자료에 비하여 통계자료가 발표되는데 까지 걸리는 시간이 매우 짧아 충격에 대한 반응을 시차

없이 확인할 수 있다.

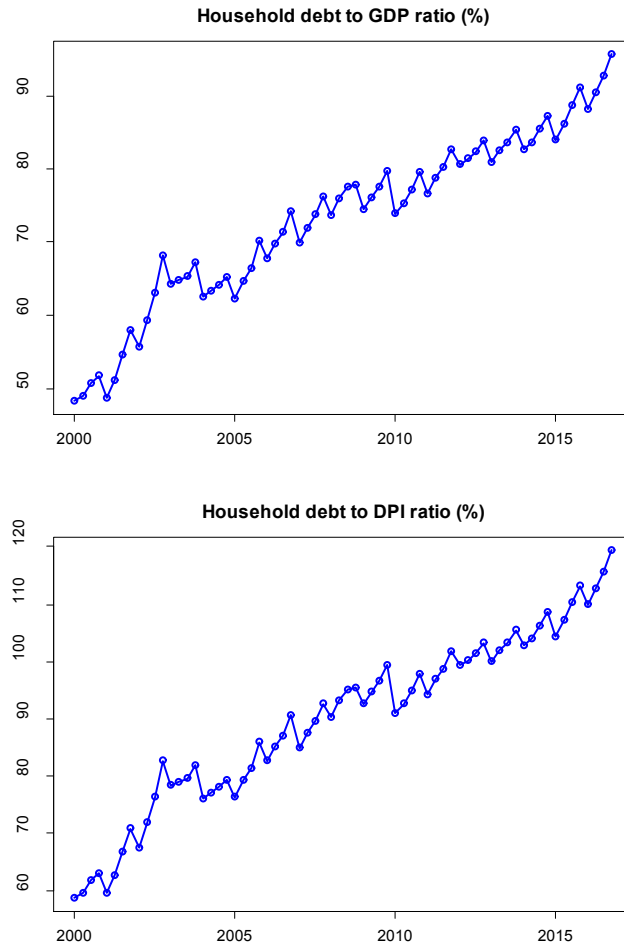
인터넷 검색자료는 사람들의 경험이나 인식을 직접적으로 빠르게 반영하고 있으며 그렇기 때문에 가계부채 위험증가에 따라 즉각 반응할 수 있는 조기경보 모형을 구축하는데 활용하기에 적합한 자료가 될 수 있다. 이러한 빅데이터를 활용할 수 있는 인공지능망모형은 여러 변수가 주는 신호를 복합적으로 분석하여 위험신호를 감지할 수 있다는 장점이 있다. 따라서 본 연구는 일반적으로 고려되는 거시경제변수와 인터넷 검색자료를 함께 사용한 인공지능망모형을 사용함으로써 더욱 정확한 가계부채 조기경보지표 모형을 고안하는 것을 목적으로 한다.

Ⅲ. 국내 가계부채의 추이 및 조기경보지수

1. 가계부채의 추이 및 지수

국내총생산(GDP)나 총처분가능소득(DPI) 대비 가계부채 (자금순환표 기준)는 2000년대에 들어서 시간에 따라 계속하여 증가하는 추세를 보인다. <그림 1>은 GDP와 DPI대비 가계부채의 추이를 보여주고 있다. <그림 1>에서 볼 수 있듯이 꾸준히 증가하는 GDP나 DPI대비 가계부채비율은 국내 거시경제에 많은 부정적인 불확실성을 제공한다는 명제는 명확해 보이나 어느 시점과 어느 수준이 진정한 위기시점인지 식별하기 힘들다. 예를 들어, Cecchetti et al.(2011)는 OECD 18개국의 자료를 이용하여 GDP대비 가계부채비율이 경제성장을 저해하는 임계치가 약 84%라고 추정하였으나 이 추정된 임계치는 통계적으로 유의하지 않았다. 반면 Arcand et al.(2015)는 1960년부터 2010년까지의 120개국 자료를 사용하여 GDP대비 가계부채비율이 경제성장을 저해하는 임계치가 50%라고 추정하였다. 이러한 추정된 임계치를 가지고 국내 상황을 비교해볼 때 국내 GDP대비 가계부채비율은 이미 2001년도에 50%를 넘어섰으며 2015년에는 85%를 넘어서고 있다. 따라서 이렇게 계속 증가하는 가계부채비율의 추세에서는 어떤 임계치가 진정한 국내 가계부채 위기상황을 나타내는가는 명확히 정의되기 힘들다.

<그림 1> 국내총생산(GDP) 및 총처분가능소득(DPI) 대비 가계부채 비율



주 : 상단에 위치한 그림은 2000년 1분기부터 2017년 1분기까지 국내총생산(GDP) 대비 가계부채(자금순환표 기준)의 비율을 나타내고 하단에 위치는 그림은 총처분가능소득(DPI) 대비 가계부채(자금순환표 기준)의 비율을 나타낸다.

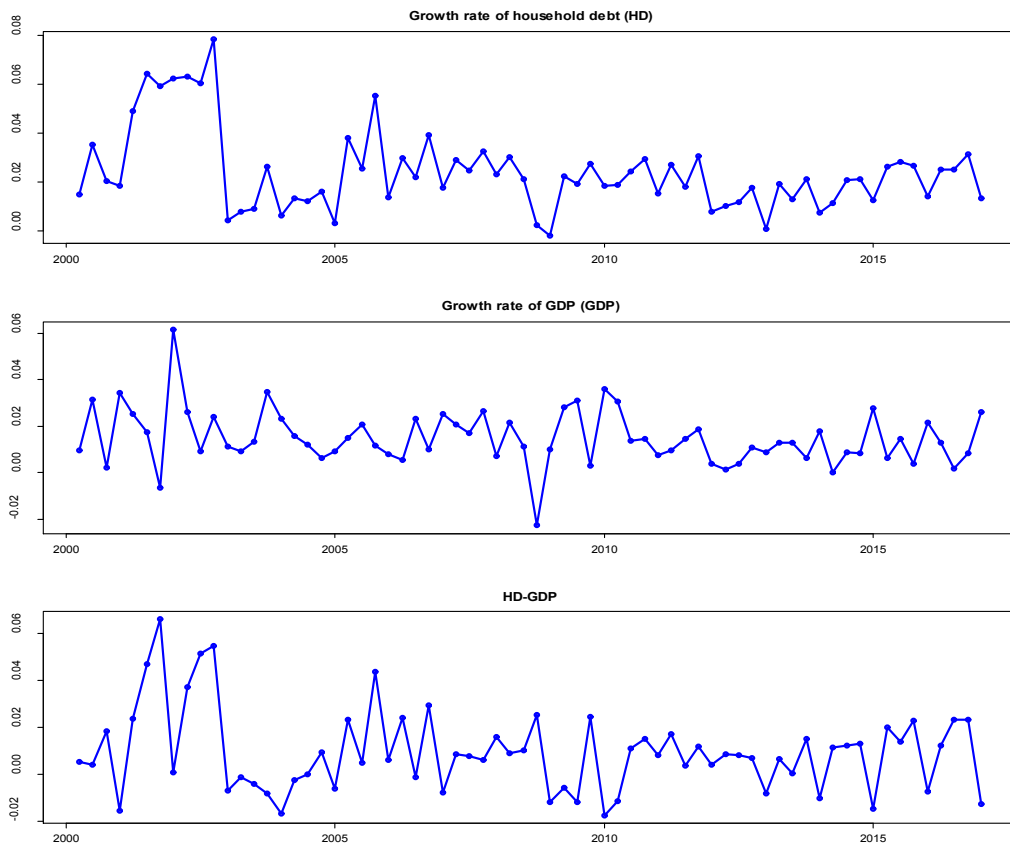
가계부채의 증가가 항상 국내 경제성장에 바람직하지 못한 영향을 미치는 것은 아니다. 소비목적이나 자산구입목적의 가계부채가 증가하게 되면 총소비 증가에 영향을 미치게 되면 이는 경제성장을 촉진시키는 관점도 존재한다. 따라서 가계부채 증대에 따른 국내 경제의 위기시점 및 수준을 식별하는데 가계부채비율의 원시계열을 사용하는 것은 바람직한 방법이 아닐 수 있다.

가계부채부실의 정도를 나타내기 위해 간단히 고려될 수 있는 지표는 가계부채 전분기대비 증가율에서 GDP전분기대비 증가율을 차감한 시계열을 생각해 볼 수 있다. 이는 일반적으로 부채관리는 소득보다 부채의 증가속도가 낮아야 한다는 것에 기인한다. 이 시계열은 다음과 같이 정의될 수 있다.

$$Index_{1,t} = \Delta \log(HD_t) - \Delta \log(GDP_t), \quad (1)$$

여기서 HD_t 는 t 시점의 가계부채를 의미한다. <그림 2>는 가계부채와 GDP증가율 및 $Index_1$ 을 나타내고 있다. GDP증가율의 변동성보다 가계부채의 변동성이 더 크게 나타나고 있는 것으로 보이며 $Index_{1,t}$ 는 2002년 신용위기에 약 6%로 큰 값을 지니고 있으며 최근으로 올수록 그 증가율이 크지 않으며 0을 기준으로 -1%와 2.4%사이에 변동하고 있는 것을 볼 수 있다. 이는 <그림 1>에서 가계부채의 수준은 GDP대비 꾸준히 증가하고 있는 것에 반해 <그림 2>의 $Index_1$ 은 최근 증가하지 않는 추세를 보이는데 이는 가계부채가 계속 증가하지만 공격적인 총량이 정해져있다면 그 증가율은 감소하는 경향을 보이기 때문이다. 따라서 <그림 2>에 나타나있는 $Index_{1,t}$ 만으로 가계부채 부실위험을 측정하는 것에는 무리가 있다고 할 수 있겠다.

<그림 2> 가계부채 증가율 및 $Index_{1,t}$



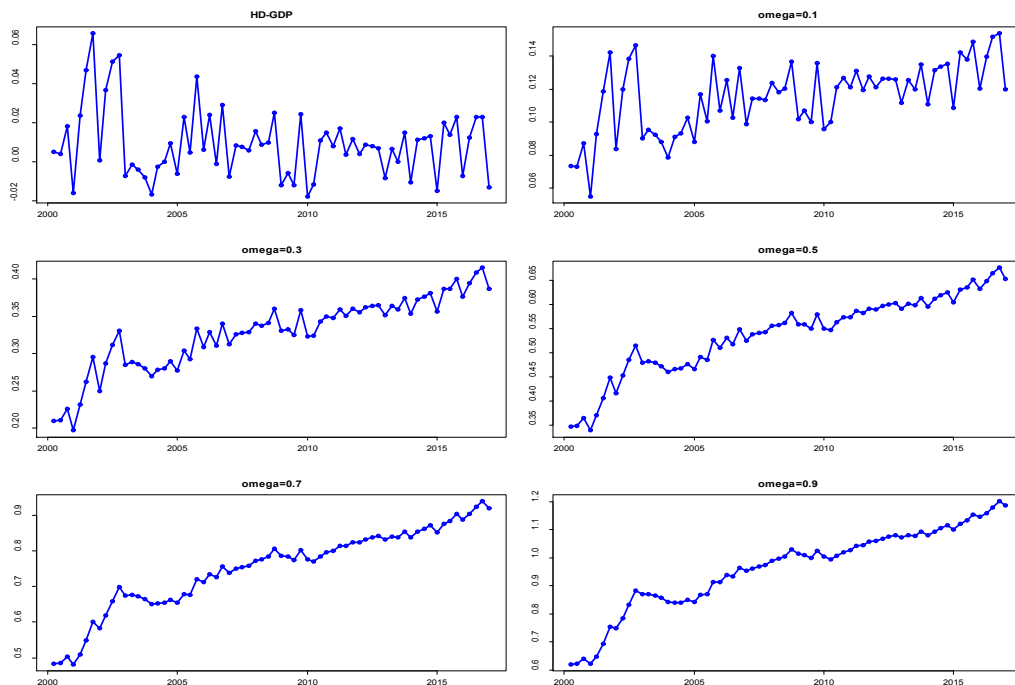
주 : 가장 위에 위치한 그림은 2000년 1분기부터 2017년 1분기까지 가계부채의 전분기대비 증감율을 나타내고 그 아래 그림은 GDP의 전분기대비 증감율을 나타내며 이는 각각 $\log(HD_t) - \log(HD_{t-1})$ 와 $\log(GDP_t) - \log(GDP_{t-1})$ 로 계산되었다. 마지막 그림은 가계부채와 GDP의 전분기대비 증감율의 차이를 나타내고 이는 식 (1)나 표현되어 있는 $Index_{1,t}$ 이다.

최근 강종구 (2017)의 연구에 따르면 가계부채는 유량효과와 저장효과를 지니고 있으며 유량효과는 경제성장을 촉진하는 영향을 미치고 저장효과는 경제성장을 저해하는 효과를 지닌다고 주장하였으며 실증분석을 통하여 그 영향을 통계적으로 검정하였다. 강종구 (2017)는 유량변수로는 가계부채의 증가율을 사용하였으며 저장변수로는 가계부채/잠재GDP 비율을 사용하였다. 따라서 가계부채의 위험을 적절하게 표현하기 위해서는 유량과 저량을 함께 고려하는 것이 필요하다고 하겠다. 이러한 유량과 저장효과를 함께 표현하기 위해 본 연구에서는 가계부채의 1차차분변수와 수준변수간의 블록결합(convex combination)을 이용하였다. 이러한 블록결합을 이용한 가계부채 부실위험을 나타내는 시계열은 다음과 같이 표현될 수 있다.

$$Index_{2,t}(\omega) = \omega Z_t + (1 - \omega)\Delta Z_t, \quad \omega \in [0, 1], \quad (2)$$

여기서 $Z_t = \log(HD_t) - \log(GDP_t)$ 로 주어지고 ΔZ_t 는 $Index_{1,t}$ 이다. 따라서 식 (2)의 $Index_{2,t}$ 는 저장변수인 Z_t 와 유량변수인 ΔZ_t 를 함께 고려한 시계열이다. 식 (2)에서 $Index_{2,t}$ 는 ω 에 대한 함수로 주어지기 때문에 적절한 시계열을 구성하기 위해서는 블록결합의 가중치인 ω 를 적절하게 선택하는 것이 중요하다고 하겠다. <그림 3>에서는 각기 다른 ω 값에 따른 $Index_{2,t}(\omega)$ 를 나타내고 있다.

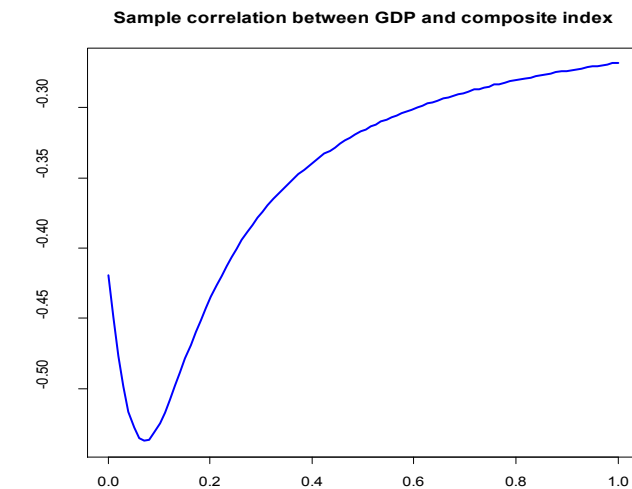
<그림 3> ω 에 따른 $Index_{2,t}$



주 : 각 그림들은 ω 의 값 (0.1, 0.3, 0.5, 0.7, 0.9)에 따라 변화에 따른 $Index_{2,t}(\omega)$ 의 그래프이다.

<그림 3>에서 ω 가 0.1일 때와 0.9일 때를 비교해보면 ω 가 0.1일 때는 $Index_{2,t}$ 의 증가율이 빠르지 않지만 ω 가 0.9일 때는 아주 빠른 속도로 $Index_{2,t}$ 가 증가하는 것을 알 수 있다. ω 가 0.3만 되더라도 ΔZ_t 의 성격은 Z_t 에 의해 상쇄되는 것을 알 수 있다. 본 연구에서는 가계부채의 부실위험이 경제성장에 미치는 영향을 최소화하기 위한 조기경보지표를 개발하는 것에 연구의 목적이 있으므로 ω 를 선택할 때 $Index_{2,t}(\omega)$ 가 $\Delta \log(GDP_t)$ 와 가장 큰 상관관계를 가지게 하는 ω 를 선택하였다. 이렇게 ω 를 선택함으로써 GDP증가율에 가장 크게 연동되는 저량 및 유량의 가중평균을 나타내는 가계부채 지수를 작성할 수 있다. 이러한 $\omega \in [0,1]$ 를 선택하기 위해서 $[0,1]$ 구간에 균등한 거리로 1000개 값을 선택하고 각 ω 에 대하여 $Index_{2,t}(\omega)$ 와 $\Delta \log(GDP_t)$ 의 표본상관계수를 계산하였다. 이렇게 계산된 표본상관계수의 절대값이 가장 큰 ω 값을 선택하였다. <그림 4>는 이와 같은 방법으로 계산과 각 ω 에 대한 $Index_{2,t}(\omega)$ 와 $\Delta \log(GDP_t)$ 의 표본상관계수를 나타내고 있다.

<그림 4> ω 에 따른 $Index_{2,t}(\omega)$ 와 $\Delta \log(GDP_t)$ 의 표본상관계수

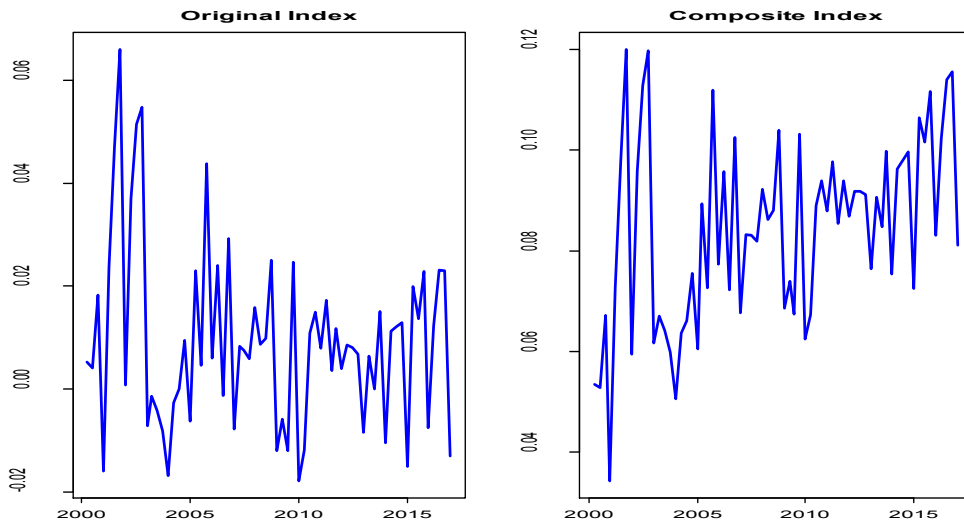


<그림 4>에 따르면 표본상관계수는 0에서부터는 급격히 감소하다 0.07에서 저점을 가지고 그 이후로 계속하여 증가하는 추세를 지니고 있으며 표본상관계수의 절댓값으로 비교해보았을 때 $Index_{2,t}(\omega)$ 와 $\Delta \log(GDP_t)$ 의 표본상관계수가 0.07에서 가장 큰 값을 지니는 것을 알 수 있다. 따라서 식 (2)의 $Index_{2,t}$ 는 $\omega = 0.07$ 을 선택하여 작성하였다.¹⁾

1) ω 를 선택하기 위해 표본상관계수와 회귀모형을 이용하였는데 회귀모형에서의 결과도 표본상관계수를 사용한 결과와 아주 비슷하였다.

<그림 5>는 식 (1)과 (2)을 통하여 작성된 $Index_{1,t}$ 와 $Index_{2,t}$ 를 나타낸다. 기본적으로 $Index_{1,t}$ 와 $Index_{2,t}$ 의 차이는 전자는 증가추세가 없는 반면 후자는 증가추세가 뚜렷히 존재한다는 것에 있다. 그리고 두 시계열의 국지적인 변동은 아주 비슷하게 나타나 있다. 이는 저량변수인 Z_t 의 증가하는 추세와 유량변수인 ΔZ_t 의 국지적 변동이 같이 $Index_{2,t}$ 에 나타나 있음을 의미한다. 그리고 2002년의 신용위기 시점에서 $Index_{2,t}$ 의 값은 아주 높게 나타나 있으며 2016년을 지나면서 2002년 신용위기 때의 값에 가까워지고 있음을 알 수 있다. 이는 $Index_{2,t}$ 를 사용하였을 때 유량효과만을 나타내는 $Index_{1,t}$ 에서 식별하지 못하는 가계부채 위험을 $Index_{2,t}$ 가 식별하고 있다고 볼 수 있다.

<그림 5> $Index_{1,t}$ 와 $Index_{2,t}$



주 : 좌측의 그림은 $Index_{1,t}$ 를 나타내고 우측의 그림은 $Index_{2,t}$ 를 나타낸다.

2. 가계부채 조기경보(EWS)지수

일반적으로 신호접근방법이나 프로빗·로짓 모형을 사용한 조기경보모형은 가계부채 부실 조기경보지수를 구축하기 할 때 위기 여부를 가리기 위한 0 아니면 1인 이항변수를 고려한다 (박원암(2015)). 이러한 방법으로 EWS지수를 정의하게 되면 어느 시점에서 갑자기 위기가 발생한다는 방식으로 경보를 하게 되는데 일반적으로 위기가 갑자기 발생할 수도 있지만 예측기간 내에서 서서히 나타나는 것이 일반적이다. 따라서 본 연구에서는 이항변수를 사용하여 EWS지수를 정의하기 보다는 Song(2010)에서 고려한 것과 같은 위기단계를 가장 낮은 위기단계인 단계 1부터 가장 위기단계인 단계 5까지 5단계

로 나누어 정의하였다.

$$\begin{aligned}
 \text{단계 1:} & \quad Index_{2,t} \leq \gamma_t + c_1\sigma \\
 \text{단계 2:} & \quad \gamma_t + c_1\sigma \leq Index_{2,t} \leq \gamma_t + c_2\sigma \\
 \text{단계 3:} & \quad \gamma_t + c_2\sigma \leq Index_{2,t} \leq \gamma_t + c_3\sigma \\
 \text{단계 4:} & \quad \gamma_t + c_3\sigma \leq Index_{2,t} \leq \gamma_t + c_4\sigma \\
 \text{단계 5:} & \quad \gamma_t + c_4\sigma \leq Index_{2,t} \quad .
 \end{aligned}$$

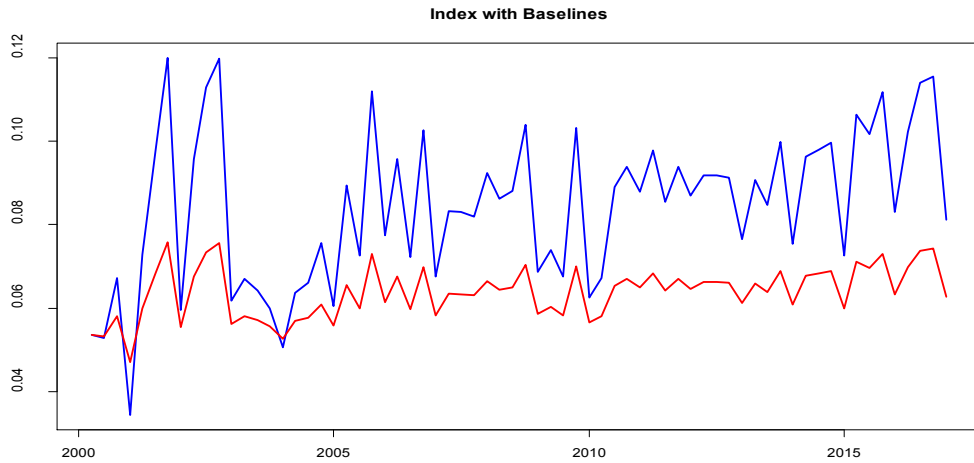
여기서 σ 는 $Index_{2,t}$ 의 비조건부 표준오차를 나타내며²⁾ γ_t 는 위기의 기준선을 나타낸다. 위기의 기준선은 가계부채 상승으로 인한 경제적 충격이 0인 선을 나타내며 이는 다음의 식으로 고려하였다.

$$\gamma_t = \gamma_{t-1} + (Index_{2,t} - Index_{2,t-1})/\delta, \quad (3)$$

여기서 δ 는 기준선의 변동성의 크기를 결정하는 모수이다. 일반적인 조기경보모형에서는 위기의 기준선은 어떠한 상수로 고정되어져 있다. 하지만 <그림 5>에서 볼 수 있듯이 $Index_{2,t}$ 는 근래에 들어 시간에 따라 증가하는 추세를 보이고 있으며 또한 큰 변동성을 가지고 있음을 알 수 있다. 이러한 시계열을 고려할 때는 상수로 고정되어 있는 기준선을 사용하게 되면 조기경보(early warning system: EWS) 지수가 필요하지 않게 자주 급변할 수 있는데 식 (3)과 같이 시변하는 기준선을 고려하게 되면 EWS 지수가 급변하는 것을 방지할 수 있다.³⁾ 식 (3)과 같이 $Index_{2,t}$ 가 증가하게 되면 γ_t 도 역시 증가하게 되며 그 증가폭은 δ 에 따라 결정된다. δ 가 너무 작은 수를 지니면 γ_t 가 급변하게 되면서 단계 5에 도달하기 아주 힘들어지게 되고 반대로 δ 가 너무 큰 수를 지니면 γ_t 의 변동이 적어지고 단계 5에 쉽게 도달하게 된다. 본 연구에서는 δ 값을 3으로 설정하였다. 각 위험단계를 결정하는 또 하나의 중요한 모수는 c_i , $i=1,2,3$ 이다. 본 연구에서는 c_i 를 $(Index_{2,t} - \gamma_t)/\sigma$ 의 분위수 (quantile)로 결정하였다. 각 c_1 , c_2 그리고 c_3 는 $(Index_{2,t} - \gamma_t)/\sigma$ 의 0.6, 0.8 그리고 0.9 분위수이며 c_4 는 2002년의 신용위기를 고려하여 2.3으로 결정하였다.

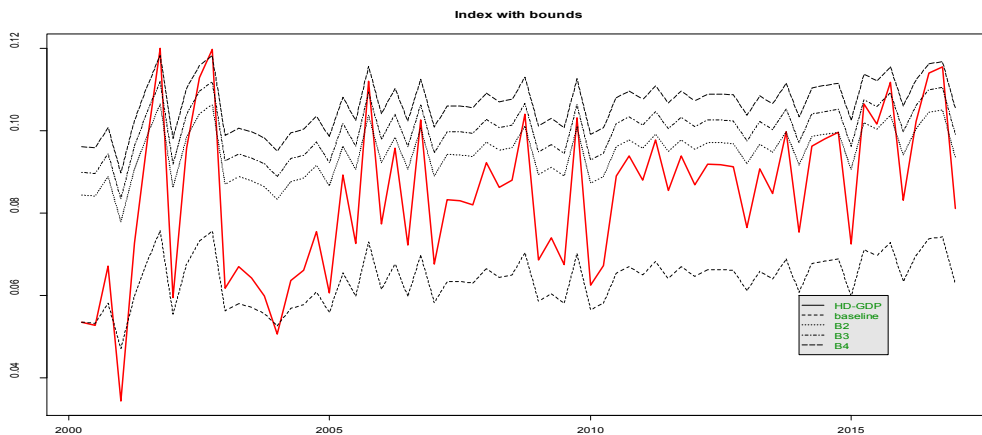
2) 구간변동(moving window)을 통해 추정된 시간에 따라 변화하는 표준오차를 사용하여도 결과에는 큰 변화가 없다.

3) 식 (3)에서 주어지는 위기의 기준선 대신 $Index_{2,t}$ 의 이동평균선도 고려가 가능하다.

<그림 6> $Index_{2,t}$ 와 위기의 기준선 (γ_t)

주 : 위에서 변동하는 실선이 $index_{2,t}$ 이고 아래에서 변동하는 선이 위기의 기준선 (γ_t)이다.

<그림 6>의 $Index_{2,t}$ 와 식 (3)에 표현된 위기의 기준선(γ_t)를 나타내고 있다.⁴⁾ γ_t 는 $Index_{2,t}$ 가 커질 때 커지며 작아질 때 작아지는 거의 비슷한 변동을 보이고 있으나 $Index_{2,t}$ 보다는 낮은 수준의 값을 가지고 있다. 이러한 비슷한 변동은 EWS지수의 필요치 않은 급격한 변동을 배제하는데 도움을 준다. <그림 7>은 위에서 고려된 각 위기의 경계선을 나타내고 있다.⁵⁾

<그림 7> $Index_{2,t}$ 와 위기의 경계선

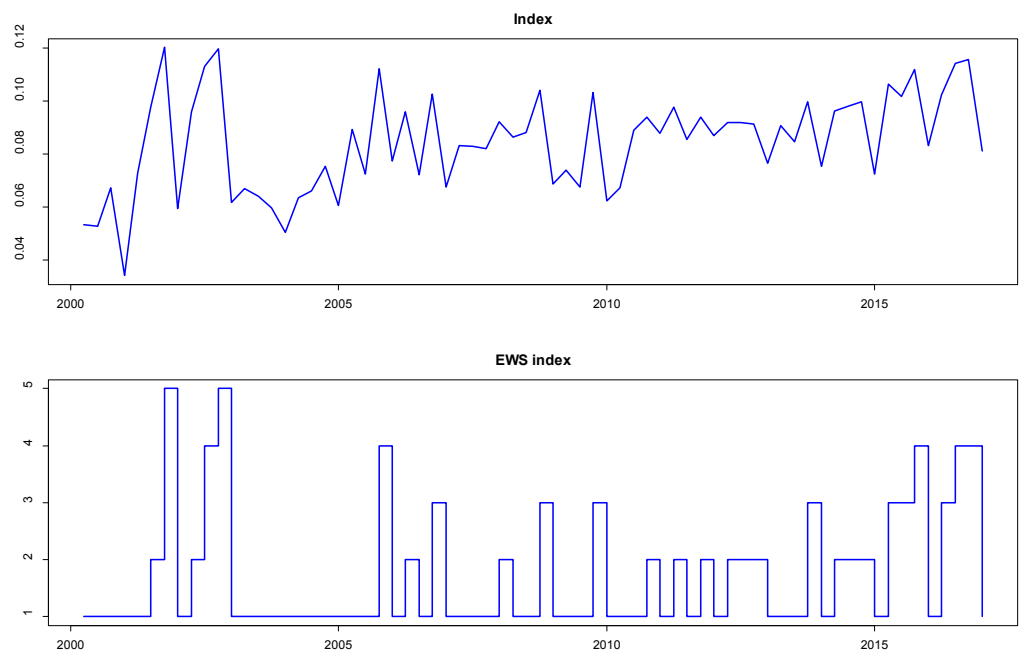
주 : 실선으로 표기된 HD-GDP가 $index_{2,t}$ 를 나타내며 baseline은 위기의 기준선 그리고 B2, B3, B4는 각각 단계 2-3, 3-4, 4-5를 구분하는 경계선을 나타낸다.

4) $Index_{1,t}$ 의 경우도 같은 분석을 하였으며 그에 대한 결과는 부록을 참고하라.

5) $Index_{1,t}$ 의 경우도 같은 분석을 하였으며 그에 대한 결과는 부록을 참고하라.

위에서 주어진 모든 모수값을 고려하여 나타낸 EWS지수가 $Index_{2,t}$ 와 함께 <그림 8>에 나타나 있다.⁶⁾ 위기가 발생한 단계 5는 궁극적으로 c_4 의 값에 의해 설정되며 이는 2002년의 신용위기를 기준으로 하였기에 신용위기 기간에만 단계 5의 위기가 나타나 있다. 2006년부터 단계 3과 단계 2를 반복하면서 최근 2015년과 2016년에 들어 단계 4로 EWS 지수가 상승하고 있음을 알 수 있다. 그리고 이러한 EWS 지수는 $Index_{2,t}$ 를 잘 반영하고 있음을 알 수 있다.

<그림 8> $Index_{2,t}$ 와 EWS지수



IV. 실증분석

본 장에서는 III장에서 고안된 가계부채 지수를 사용하여 가계부채 부실위험 조기경보 모형을 추정하고 이를 이용하여 2017년 2분기부터 2019년 2분기까지 조기경보지수를 예측 및 경보한다. 모형에 사용되는 변수들은 과거의 연구들에서 가계부채의 결정요인 분석 시 자주 사용되는 설명변수 뿐만이 아니라 인터넷 검색자료들이 같이 고려되어 모형

6) $Index_{1,t}$ 의 경우도 같은 분석을 하였으며 그에 대한 결과는 부록을 참고하라.

을 추정한다. 조기경보지수의 예측에 사용되는 모형은 종속변수와 설명변수의 복잡한 관계를 잘 근사시킨다고 알려져 있는 인공신경망(artificial neural network: ANN)모형을 사용한다.

1. 분석자료

실증분석에서 사용되는 자료는 2000년 1분기부터 2017년 1분기까지의 분기별 총량 시계열 자료를 사용하였다. 일반적으로 가계부채의 결정요인을 분석하는 많은 연구들에서는 한국노동패널자료나 가계금융복지조사 패널자료 또는 CB자료 등 가구별 미시자료를 이용해서 분석하고 있으나 이들 미시자료들은 연도별 자료로 주어지기 때문에 가계부채 부실위험을 추정하고 예측하는데 한계가 존재한다. 가계부채의 결정요인의 분석에 사용되는 설명변수 자료는 크게 나누어 보면 수요에 미치는 요인, 공급에 미치는 요인 그리고 수요와 공급 모두에 미치는 요인의 세 가지 유형이 존재한다 (Pouvelle(2012)와 김현정·손종철·이동렬·임현준·나승호(2013)). 수요에 미치는 요인변수 중에서는 실질 GDP 증가율, 대출금리 변동, 소비자물가상승률을 고려하였으며 공급에 미치는 요인변수는 금융기관수신 증가율, 고정이하여신비율, BIS기준 자기자본비율을 고려하였으며 수요와 공급에 모두 영향을 미치는 요인변수로는 주가지수 및 주택매매가격지수를 사용하였다. 모형의 추정에 사용된 가계부채 결정요인 변수는 <표 1>에 요약되어져 있으며 모든 자료는 한국은행의 경제통계시스템에서 획득하였다.

<표 1> 가계부채 결정요인변수

| 통계표 | 항목명 | 단위 |
|---------------------------------|------------|------------------|
| 가계부채 (십억원) | hdebt | 가계및 비영리단체 |
| 주가지수 (1980.1.4=100) | kospi | KOSPI |
| GDP(계절조정, 실질) | rgdp | 국내총생산(시장가격, GDP) |
| GDP(계절조정, 명목) | ngdp | 국내총생산(시장가격, GDP) |
| 소비자물가지수(2015=100, 가중치=1000)(전국) | cpi | 총지수 |
| 가계신용 (십억원) | hcredit | 가계신용 |
| 가계대출금리 (%) | hir_mean | 총대출평균 |
| | hir_house | 가계대출 |
| 예금은행 총수신 (말잔) | brecep | 수신합계 |
| 비은행기관 수신(말잔) | nbrecep | 비은행금융기관(말잔) |
| 시장금리 | call_unse | 무담보콜금리 전체 |
| | call_se | 담보콜금리전체 |
| BIS기준자기자본비율 (%) | bis_asset | BIS기준자기자본비율 |
| 고정이하 여신비율 (%) | loan_below | 고정이하 여신비율 |
| 주택매매가격지수(KB, 2015.12=100) | hidx | 총지수 |

자료 : 한국은행 경제통계시스템

위에서 고려한 변수들 이외에 인터넷 검색자료를 이용하였다. 최근의 연구들에서 인터넷 검색자료를 모형에 포함한 연구는 모형의 예측력이 대체로 통계적으로 유의하게 상승되었다는 공통된 특징을 보이고 있다. 인터넷 검색자료의 장점은 기존의 거시경제변수들이 측정하지 못하는 경제 주체의 활동을 변수로 나타낼 수 있으며, 기존의 변수들이 포함하지 못하는 정보를 포함하고 있다는 점이다. 인터넷 검색자료로는 한국언론진흥재단에서 제공하는 빅카인즈 (BIG KINDS)자료를 이용하였다. 이궁희(2017)는 빅카인즈 검색자료를 이용하여 거시경제 변수를 예측한 결과가 기존의 변수들을 사용한 경우보다 더 정확한 예측치를 가지는 것을 실증적으로 보인 바 있다.⁷⁾ 빅카인즈를 이용할 때 사용된 신문사로는 경향신문, 국민일보, 내일신문, 문화일보, 서울신문, 세계일보, 한겨레, 한국일보, 매일경제, 서울경제, 파이낸셜뉴스, 한국경제, 헤럴드경제 신문들을 이용하였으며 신문기사 검색어로는 가계대출, 은행대출, 주택담보대출, 신용대출, 신용대출금리, 대출한도, 신규대출, 2금융권대출, 담보대출의 9개 검색어를 사용하였다.

<표 2>는 빅카인즈에서 검색어별 취합된 자료와 가계부채, $Index_{2,t}$ 및 EWS지수와 같은 시점에서의 표본상관계수를 나타내고 있다. 전체적으로 볼 때 검색어 자료들은 가계부채와 아주 큰 상관관계를 지니고 있으며 특히 주택담보대출, 신용대출 그리고 2금융권대출은 가계부채와 상관관계가 각각 0.650, 0.763 그리고 0.849로 높은 상관관계를 지니고 있다. 하지만 III장에서 고려한 $Index_{2,t}$ 와 EWS지수와는 가계부채의 그것에 비해 상당히 낮은 상관관계를 보여주며 몇몇의 변수는 음의 상관관계를 보여주고 있다. 가계부채와의 높은 상관관계는 검색어 변수 역시 가계부채 변수와 같이 증가하는 추세를 가지는 것에 기인하고 있는 것으로 보인다.

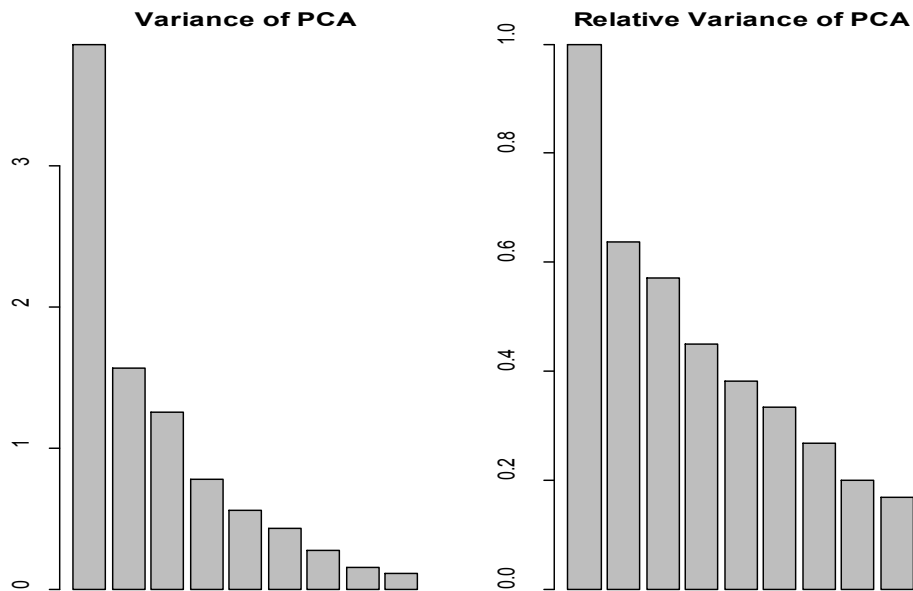
<표 2> 신문기사 검색어와 가계부채의 상관관계

| | 가계부채 | Index 2 | EWS 지수 |
|--------|-------|---------|--------|
| 가계대출 | 0.236 | 0.173 | 0.227 |
| 은행대출 | 0.647 | -0.104 | -0.128 |
| 주택담보대출 | 0.650 | -0.126 | -0.127 |
| 신용대출 | 0.763 | 0.078 | 0.001 |
| 신용대출금리 | 0.503 | 0.028 | -0.108 |
| 대출한도 | 0.550 | 0.082 | 0.056 |
| 신규대출 | 0.304 | -0.110 | 0.013 |
| 2금융권대출 | 0.849 | -0.103 | -0.153 |
| 담보대출 | 0.692 | -0.073 | -0.156 |

7) 인터넷 검색자료로 많이 사용되는 자료는 구글 트렌드 (Google Trend)와 네이버 검색자료 (Naver Lab)이 있다. 구글 트렌드는 월별 및 분기별 자료를 2004년부터 취합이 가능하고 또한 국내 상황을 다른 검색자료에 비해 명확히 반영하지 못하는 면이 있기에 제외하였다. 네이버 검색자료는 2016년부터 월별 및 일별 자료가 취합 가능하여 본 연구에서 사용하는데 한계가 있다.

9개의 신문기사 검색자료를 설명변수로 모두 사용할 수도 있지만 각 변수들의 시차변수들을 사용하게 되면 너무 많은 설명변수들을 모형에 사용하게 되어 본 연구에서는 주성분분석(principal component analysis: PCA)을 통한 9개 검색자료의 주성분을 모형에 사용하였다. <그림 9>은 신문기사 검색자료에 대한 주성분분석의 분산을 나타낸다. 첫 번째 성분의 분산이 다른 성분들보다 월등히 크게 나타나고 있음을 알 수 있다. 따라서 본 연구에서는 하나의 주성분을 신문기사 검색자료로부터 추출하여 사용하였다.⁸⁾

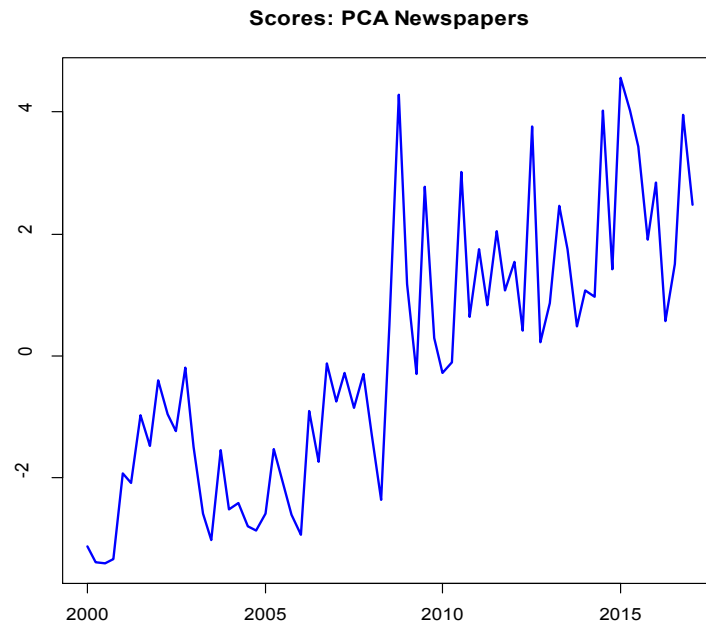
<그림 9> 신문기사 검색어 자료의 주성분 분석



<그림 10>는 PCA를 통해 추출된 신문기사 검색어 자료들의 주성분 시계열을 나타낸다. 주성분 시계열은 시간에 따라 증가하는 추세를 지니고 있으며 특히 2002년 신용위기 시점에서는 선형추세보다 월등히 높은 값을 지니고 있는 것을 볼 수 있다.

8) 카이저 기준 (Kaiser Criterion)에 따르면 특성근 (분산)이 1보다 큰 요인들을 선택하는 것으로 권고하였다. 본 연구에서는 주요인을 두 개로 선택하여 분석한 것이 하나를 선택하여 분석한 것과 결과에 차이가 없었다.

<그림 10> 신문기사 검색어 자료의 주성분 시계열



주 : 9개의 신문기사 검색자료의 주성분분석을 통한 첫 번째 주성분 시계열자료임.

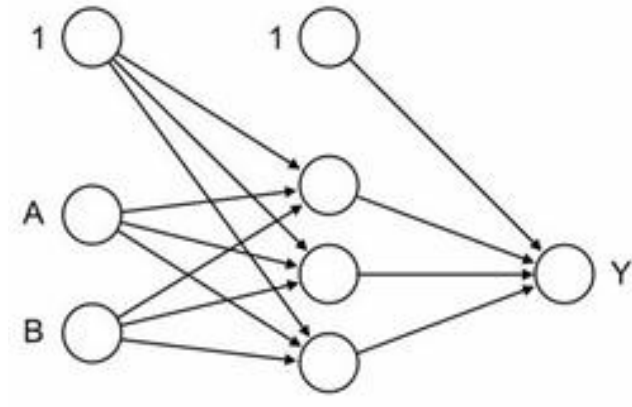
모형추정을 위한 모든 설명변수는 정상성을 만족하는 시계열을 사용하였으며 원시계열이 정상성을 만족하지 않는 경우 원 시계열의 증가율을 사용하여 정상시계열로 변환하였다.⁹⁾

2. 실증분석 모형: 인공신경망모형

많은 실증분석에서 종속변수와 설명변수의 함수적 관계에 대한 연구를 행하고 있다. 이러한 변수들 간 가장 기본적으로 그리고 많이 설정하는 관계는 선형관계이며 필요한 경우 이들 변수의 관계를 비선형으로 설정하여 분석하고 있다. 조기경보모형에서 많이 사용되는 모형은 신호접근법과 프로빗·로짓모형이 있다. (김상환(2001), 남준우(2003), 박원암(2001, 2015)) 본 연구에서는 인공신경망 모형을 사용하여 조기경보모형을 설정한다. 인공신경망모형은 선형적으로 종속변수와 설명변수의 관계를 설정할 필요가 없으며 많은 잘 알려진 비선형모형을 포함할 수 있으며 또한 예측성과도 우수하고 변수들 간 아주 복잡한 관계를 잘 근사시키는 것으로 알려져 있다 (Hornik, Stichcombe and White, 1989; Bishop, 1995).

9) ADF, PP와 DF-GLS 단위근 검정을 실시하였으며 모든 시계열에서 단위근이 존재한다는 귀무가설을 유의수준 1%에서 기각하였다. 단위근 검정결과는 지면 사정상 논문에 기재하지 않았다.

〈그림 11〉 간단한 인공신경망모형



〈그림 11〉은 두 개의 투입신경세포와 하나의 출력신경세포 그리고 세 개의 은닉신경세포로 이루어진 간단한 인공신경망모형을 나타내고 있다. 신경망에는 입력층, 은닉층 그리고 출력층이라는 층들이 있으며 〈그림 11〉의 가장 좌측에 위치하는 층이 A와 B를 신경세포로 가지는 입력층이며 중간의 층이 세 개의 신경세포로 이루어진 은닉층이며 가장 우측에 위치한 층이 출력층이다. 은닉층은 위의 그림과 같이 하나의 층으로만 이루어져 있지 않고 많은 은닉층을 가진 신경망을 고려할 수 있다. 각 층간 각 신경세포를 연결시켜주는 연결가중치라고 정의되는 선이 있으며 자료를 가지고 이를 추정하게 되는 된다. 각 층에는 활성화함수(activation function)가 있으며 이전 신경세포의 값과 연결가중치를 이용해 다음 층의 신경세포 값을 구하게 된다. 가장 간단한 인공신경망모형은 다음과 같이 나타낼 수 있다.

$$y = f(w_0 + \sum_{i=1}^n w_i x_i) = f(w_0 + w^T x).$$

이 모형은 은닉층이 존재하지 않는 간단한 모형이며 통계모형에서는 일반화된 선형모형(generalized linear model: GLM)과 동일한 모형이다. 따라서 GLM모형에서 본다면 위의 모형은 활성화함수의 역함수($f^{-1}(\cdot)$)를 링크함수로 가지는 모형이다. 이를 조금 더 일반화시키면 다음의 모형으로 나타낼 수 있다 (Hornik, Stichcombe and White, 1989).

$$\begin{aligned} y &= f(w_0 + \sum_{j=1}^J w_j \cdot g(w_{0j} + \sum_{i=1}^n w_{ij} x_i)), \\ &= f(w_0 + \sum_{j=1}^J w_j \cdot g(w_{0j} + w_j^T x)), \end{aligned} \quad (4)$$

여기서 $f(\cdot)$ 와 $g(\cdot)$ 은 활성화함수이고 w_0 는 출력신경세포의 상수항이며 w_{0j} 는 은닉신경세포의 상수항이다. 또한 w_j 는 j 의 은닉신경세포에서부터 출력신경세포로의 연결가중치이며 $w_j = (w_{1j}, w_{2j}, \dots, w_{nj})$ 는 j 번째 은닉신경세포로 연결되는 연결가중치이다. 위와 같은 모형은 GLM모형의 확장모형이라고 할 수 있으나 각 모수들은 GLM에서 해석되는 것과 같이 해석될 수는 없다.

일반적으로 위의 식 (4)와 같이 주어진 하나의 은닉층이 있는 모형이 경제학이나 재무학 연구에서 가장 많이 쓰이는 간단한 모형이다. 활성화함수로는 쌍곡탄젠트함수(hyperbolic tangent), 로지스틱함수(logistic) 혹은 ReLU(rectified linear unit)함수가 많이 사용되나 본 연구에서는 식 (4)와 같이 주어지는 모형을 사용하였으며 출력층은 선형 활성화함수 $f(x) = x$ 를 가정하였고 은닉층은 로지스틱함수 $g(x) = 1/(1 + \exp(-x))$ 를 사용하였다¹⁰⁾

가계부채 부실 조기경보모형을 추정하기 위해 본 연구에서는 식 (4)로 주어진 인공신경망모형에 세 개의 은닉신경세포를 $J=3$ 을 고려하여 추정하였다.¹¹⁾ 그리고 설명변수는 두 가지 경우를 고려하였다. 첫 번째 경우는 투입신경세포로는 앞서 고려한 총 9개의 설명변수들의 1분기전 시차변수를 사용하는 모형(L1)이고 두 번째 모형은 설명변수들의 1분기전과 2분기전의 시차변수를 사용하는 모형(L2)이다. 모형의 추정에 사용된 총 설명변수의 갯수는 첫 번째 모형(L1)은 9개로 주어지고 두 번째 모형의 총 설명변수의 갯수는 18개이다.

3. 실증분석 결과

인공신경망모형의 추정은 일반적인 오류역전파법(backpropagation)방법을 사용하여 추정된다. 여기서 역전파는 출력층으로부터 은닉층을 거쳐서 입력층까지 역으로 진행되는 과정이며 이 때 역전파되는 주체가 오차(오류)이어서 오류역전파법이라고 한다. 각 신경세포에서 그 세포에 상응되는 오차를 계산함으로써 손실함수의 경사(gradient)를 계산할 수 있으며 이 경사를 이용한 일반적인 경사하강법(gradient descent method)을 사용할 수 있게 된다. 본 연구에서는 오류역전파법을 좀 더 확장한 탄력적 오류역전파법(resilient backpropagation)을 사용하여 모형을 추정하였다. 오류역전파법에서는 모든 가중치에 대한 학습률(learning rate)이 동일하게 주어지지만 탄력적 오류역전파법의 학습률은 각 가중치에서 다 다르게 설정되어 신경망 전체에 적절한 학습률을 고려할 수 있다는 장점이 있다 (Riedmiller and Braun (1993)과 Riedmiller (1994)).

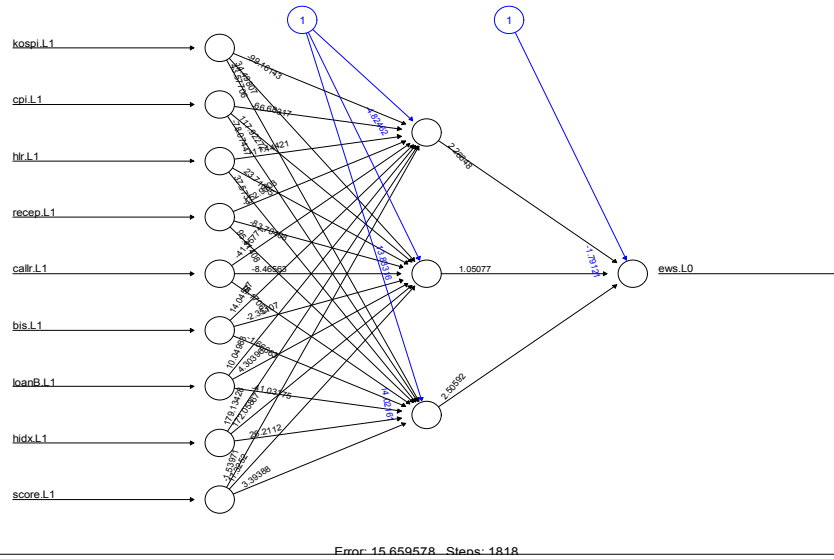
탄력적 오류역전파법을 사용하여 인공신경망모형 L1과 L2를 추정한 결과는 각각 <그림 12>와 <그림 13>에 나타나있다.¹²⁾ 두 모형의 평균제곱오차(MSE)는 L1이 15.66이고 L2

10) 인공신경망모형의 추정법과 이론적인 부분은 Bishop(1995), Anthony and Bartlett (1999)와 McNelis (2005)를 참고하라.

11) 은닉신경세포(J)가 3개, 4개 그리고 5개의 경우를 모두 고려하여 모형을 추정하였으나 3개를 고려했을 때와 평균제곱오차(MSE)값에는 통계적으로 유의한 차이가 존재하지 않았다.

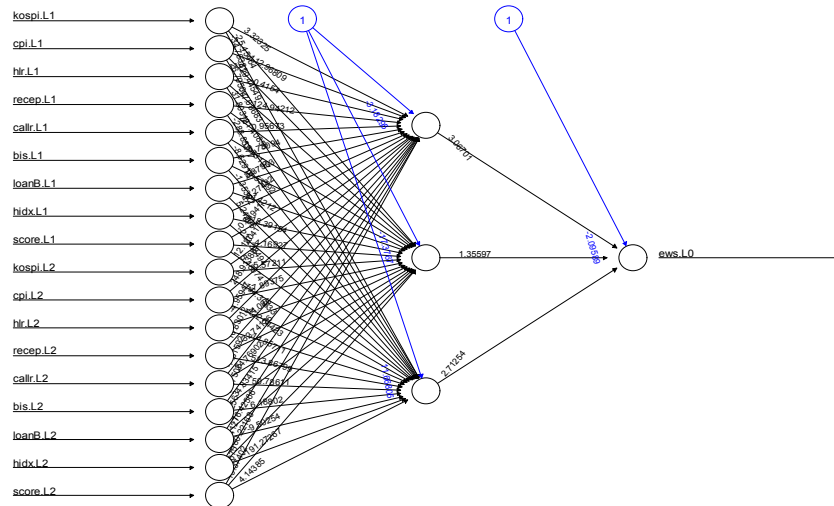
가 10.89로 L2의 모형에 대한 설명력이 높게 나타났다.

<그림 12> 인공신경망모형 추정결과 (L1)



주 : kosp.L1, cpi.L1, hir.L1, recep.L1, callr.L1, bis.L1, loanB.L1, hidx.L1, 그리고 score.L1 각각 주가지수, 소비자물가지수, 가계대출금리, 총수신, 시장금리, BIS기준 자기자본비율, 고정이자 여신비율, 주택매매가격지수 그리고 인터넷 검색지수의 1차 시차변수이다.

<그림 13> 인공신경망모형 추정결과 (L2)

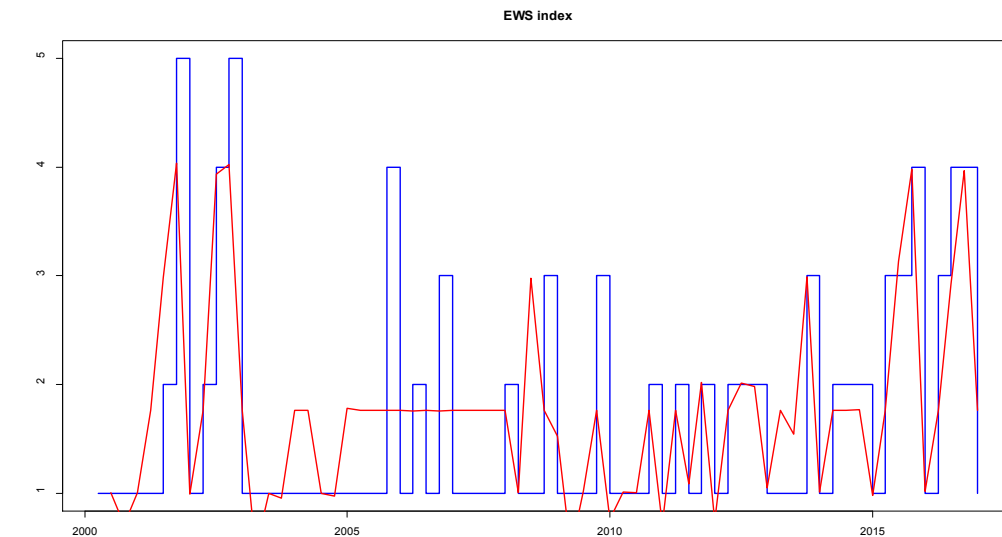


주 : kosp.L1, cpi.L1, hir.L1, recep.L1, callr.L1, bis.L1, loanB.L1, hidx.L1, 그리고 score.L1 각각 주가지수, 소비자물가지수, 가계대출금리, 총수신, 시장금리, BIS기준 자기자본비율, 고정이자 여신비율, 주택매매가격지수 그리고 인터넷 검색지수의 1차 시차변수이다.

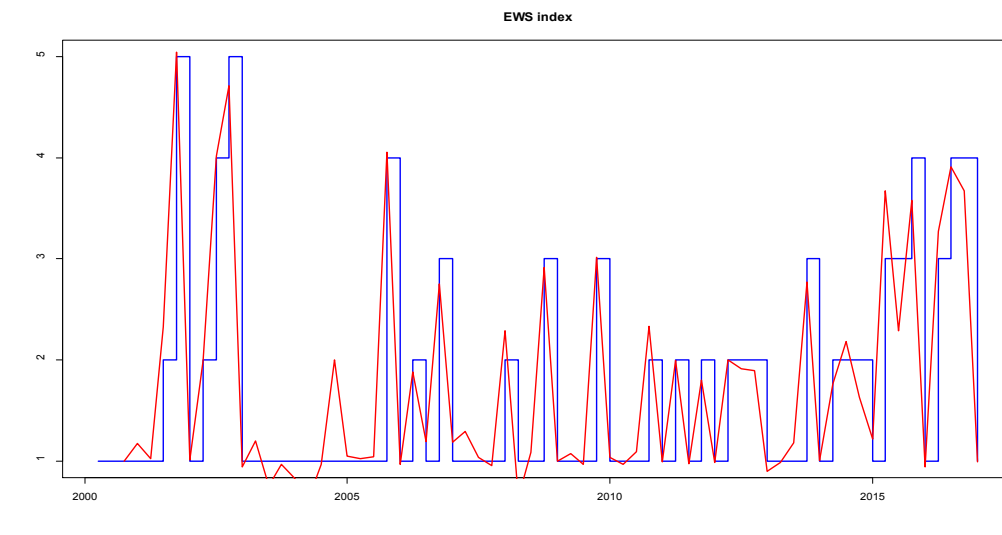
- 12) 인공신경망모형의 추정에 사용된 프로그램은 R이고 Fritsch, Guenther, Suling and Mueller (2016)에 의해 제공된 R의 “neuralnet” 라이브러리를 사용하였다. 이 라이브러리의 간략한 설명과 이용방법은 Gunther and Fritsch (2010)에 잘 나타나있다. 그리고 $Index_{1,t}$ 을 사용하여 모형을 추정하였으며 그에 대한 결과는 부록을 참고하라.

<그림 14>와 <그림 15>는 L1과 L2 모형추정결과를 사용된 표본 내 적합도를 EWS지수와 비교하여 나타내고 있으며 그림에서 계단식으로 나타난 선은 III장에서 구성된 EWS지수를 나타내고 나머지 한 실선은 인공신경망모형을 사용하여 추정된 값을 나타낸다.¹³⁾

<그림 14> 인공신경망모형 표본 내 결과 (L1)



<그림 15> 인공신경망모형 표본 내 결과 (L2)



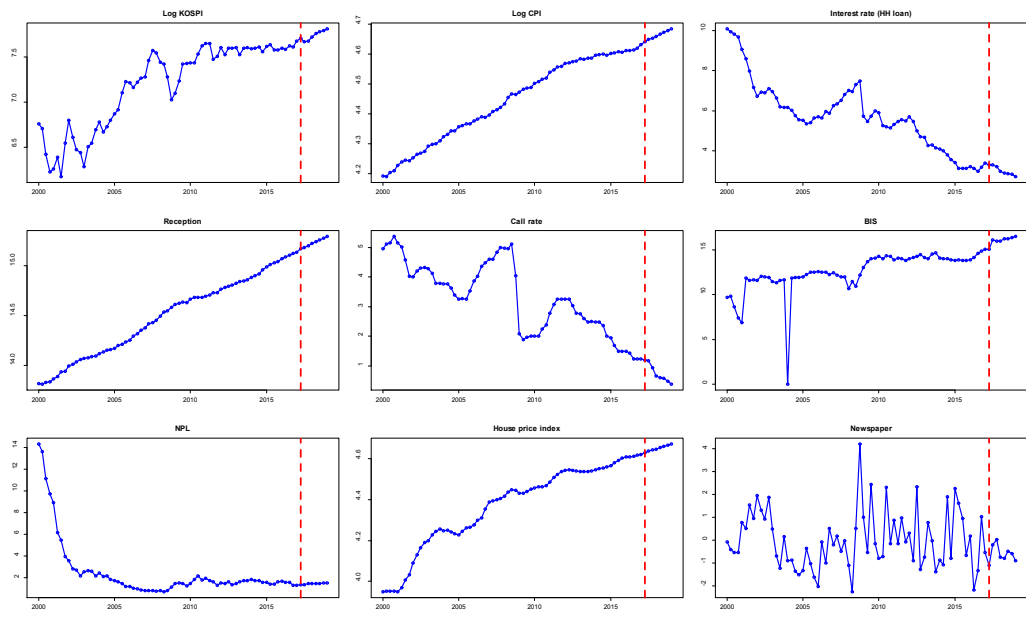
13) $Index_{1,t}$ 을 사용하여 모형을 추정하였으며 그에 대한 결과는 부록을 참고하라.

추정된 결과에서 보면 L1으로 추정된 결과는 L2의 결과보다 모형의 적합도가 떨어지는 것을 볼 수 있다. 예를 들어, L1모형은 2002년에 발생한 신용위기의 위기정도를 과소 추정하고 있으며 2006년에서 2007년의 높은 EWS지수를 적절하게 추정하고 있지 못하다. 그와 비교해서 L2의 경우는 2002년의 신용위기 뿐만 아니라 고려된 구간에서 EWS 지수를 적절하게 추정하고 있는 것으로 나타났다. 이는 한 분기 전 시차변수로는 가계 부채의 위험정도를 적절하게 예측하고 있지 못함을 의미한다.

4. 조기경보지수 예측 (2017Q2-2019Q2)

앞서 설정된 EWS지수와 인공신경망 모형의 추정치를 사용하여 2017년 2분기부터 2019년 2분기까지 조기경보를 실시하였다. 모형에 포함되는 설명변수는 총 9개로 주어져 있으며 L1의 경우는 1차시차변수 그리고 L2의 경우는 1차와 2차시차변수가 포함되어 있어 EWS지수를 예측하기 위해서는 이들 설명변수의 예측치가 필요하다. 따라서 가계 부채 부실에 대한 조기경보를 위해서는 9개의 설명변수의 미래값을 먼저 모형을 통하여 예측한 다음 그 예측치를 조건으로 하여 조건부 예측을 실시하였다.

〈그림 16〉 VAR(2) 모형을 사용한 설명변수의 예측값



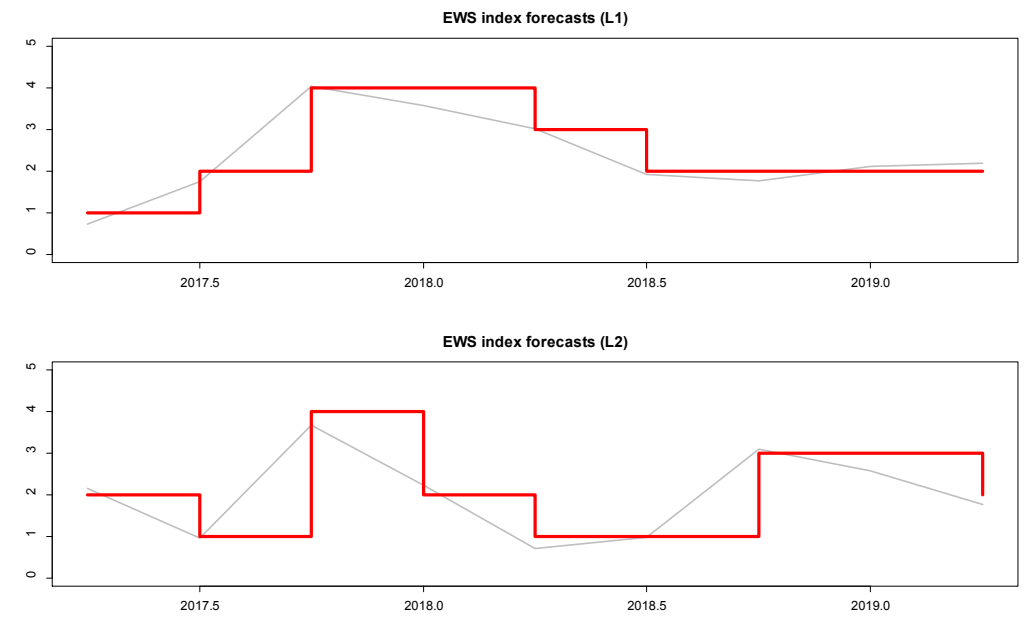
주 : 실선으로 주어진 수직선은 2017년 1분기 시점을 나타냄.

모형에서 사용한 9개의 설명변수는 모두 정상성을 만족하는 변수이고 서로 상호연관이 되어 있어 어떤 변수가 다른 변수에 대해 외생변수인지 혹은 내생변수인지 판가름하기에 한계가 있다. 따라서 시계열 변수의 예측에서 자주 쓰이는 p 차 벡터자기회귀

(VAR(p)) 모형을 사용하였다.¹⁴⁾ VAR모형의 차수 p 는 아카이케 정보기준(AIC)로 선택하였으며 이 때 선택된 차수는 2이다. <그림 16>는 설명변수 9개의 원 시계열의 예측값을 보여주고 있다. 수직인 실선은 2017년 1분기를 나타내며 실선 이후의 값들이 설명변수의 예측치를 나타낸다. VAR모형을 통한 예측치는 고려된 시계열 변수들의 증감율을 나타내지만 <그림 16>에서는 이 예측된 증감율을 이용하여 원시계열 예측치로 표현하였다.

<그림 17>은 조건부 예측방법을 이용하여 2017년 2분기부터 2019년 2분기까지 가계부채 위험 조기경보지수를 예측한 값이 나타나 있다.¹⁵⁾ L1모형과 L2모형의 예측치가 다르게 나타나고 있지만 공통적으로는 단계 5에 이르는 가계부채 부실위험은 없는 것으로 나타났다. 또한 L1과 L2모형에 따르면 2017년 3분기까지는 보통단계인 단계 1-2를 유지하다가 2017년 4분기에 들어서 위기 단계 4까지 상승하는 것을 볼 수 있으며 위기 단계 4가 L1모형에서는 2분기동안 지속되며 L2모형에서는 2018년 1분기에 다시 보통단계로 줄어드는 것을 볼 수 있다. 따라서 2017년 4분기에 적극적으로 가계부채의 추이와 관련 변수들을 주의 깊게 살펴 볼 필요가 있다고 하겠다.

<그림 17> 조기경보지수 예측치 (2017Q2-2019Q2)



14) 조기경보모형을 사용하기 위해 설명변수의 예측을 좀 더 정교한 여러 가지 다른 예측모형을 사용할 수 있으나 본 연구에서는 간단하면서 표본 외 예측성도가 좋은 VAR모형을 사용하였다.

15) $Index_{1,t}$ 을 사용하여 모형을 추정하였으며 그에 대한 결과는 부록을 참고하라.

V. 결론

본 연구는 최근 중요한 이슈로 대두되고 있는 가계부채 부실에 대한 조기경보지표를 개발하고 여러 거시금융자료와 인터넷 검색자료를 이용하여 가계부채 위험을 예측하였다. 최근에 들어 가계부채의 국내총생산과 총처분가능소득 대비 증가율이 꾸준히 상승하여 그 위험수위에 다다르고 있는 바, 가계부채 부실위험에 대한 정확한 조기경보는 위험에 대한 대비와 향후 경제정책 설정에 아주 중요한 도움을 주는 것임에는 틀림없다.

가계부채의 결정요인을 분석하는 연구는 국내적으로 꾸준히 진행되어져 왔지만 가계부채 부실에 대한 조기경보모형은 아직 없는 실정이다. 이는 시간에 따라 꾸준히 증가하는 가계부채의 위험수준에 대한 임계값을 객관적으로 설정하는 것에 한계가 있기 때문이다. 몇몇 국외 연구에서는 국내총생산 대비 가계부채의 비율이 50%나 85%가 넘어서면 위험수준이라고 제안하였지만 국내 가계부채 비율은 이미 그 수준을 넘어 있어 그 임계값을 현 상황에 적용하는 것은 현실적이지 않다고 볼 수 있겠다.

일반적으로 많은 연구에서는 가계부채의 결정요인에 따라 가계부채의 위험도가 달라진다고 보고 있으므로 미시적 패널자료를 가지고 가계부채의 위험정도를 분석하는 것이 적절하다고 볼 수 있으나 그러한 미시적 가계패널자료는 거의 연도별 자료로 주어져 있기 때문에 이러한 자료를 가지고 조기경보모형을 설정하는 것은 큰 의미를 가지지 못한다. 따라서 본 연구에서는 가계부채와 그 결정요인들의 총량 시계열 자료를 사용하였다. 계속 꾸준히 증가하는 가계부채 원 시계열 자료를 이용하여 위험수준을 예측한다는 것은 한계가 있으므로 본 연구에서는 가계부채 시계열을 적절히 변환한 지수를 사용하였다. 강종구 (2017)의 연구는 가계부채의 저량적 그리고 유량적인 성격에 따라 국내총생산 및 소비에 미치는 영향이 상이하다는 것을 보였다. 따라서 가계부채의 유량적인 성격과 저량적인 성격을 함께 고려한 지수를 작성한다면 가계부채 부실위험을 좀 더 효율적으로 측정할 수 있기에 본 연구에서는 적절한 가중치를 설정한 유량변수와 저량변수의 블록결합을 통하여 새로운 가계부채 부실지수를 작성하였으며 이렇게 작성된 지수로 가계부채 위험 조기경보모형을 설정하였다.

금융위기에 대한 조기경보모형은 지금까지 여러 연구들에서 모형들이 제시되어져 왔다. 많이 쓰이는 모형은 신호접근법과 프로빗·로짓 모형 등이 있다. 이 모형들은 위기에 대한 변수로 이항확률변수를 사용하여 분석한다. 이는 위기상황이면 1의 값을 안정적 상황이면 0의 값을 가지는 변수를 설정하여 모형을 분석함을 의미한다. 하지만 현실에서 위기상황은 안정적인 상황으로부터 갑자기 발생하지 않는다. 따라서 본 연구에서는 가계부채의 위기를 5단계로 나누어 단계별 위기상황을 설정하였다. 또한 앞서 사용된 모형들은 종속변수와 설명변수의 복잡한 관계를 지나치게 단순화 시킨 선형모형을 설정하여 분석하고 있다. 물론 가계부채 부실지수를 잘 설명할 수 있는 명확한 그리고

많은 수의 설명변수가 존재하고 취득이 가능하다면 부실 위험에 대한 선형모형의 예측력은 다른 복잡한 비선형모형의 그것에 비해 높을 수 있다. 하지만 현실에서는 그러한 많은 수의 명확한 설명변수들을 선정하는데 한계가 있다. 따라서 본 연구에서는 사용되는 모형의 함수를 선형적으로 설정할 필요가 없는 인공신경망 모형을 사용하였다. 많은 비선형 모형은 인공신경망 모형의 특별한 경우이다. 인공신경망 모형은 종속변수와 설명변수의 복잡한 관계를 잘 근사시킨다고 알려져 있다. 뿐만 아니라 본 연구에서는 많은 수의 명확한 설명변수 획득이 어려우므로 최근에 많이 쓰이는 인터넷 검색자료를 이용하여 모형을 설정하였다. 최근의 많은 연구에서 인터넷 검색자료를 설명변수로 포함시킴으로써 표본 외 예측력의 향상이 통계적으로 유의하게 상승하였다고 보고하고 있다. 따라서 인터넷 검색자료를 설명변수로 포함시킴으로서 가계부채 부실 위험정도에 대한 예측력을 향상시키고자 하였다.

인공신경망 모형을 이용하여 추정된 결과를 가지고 본 연구에서는 2017년 2분기부터 2019년 2분기까지 가계부채 위험 단계를 예측을 하였다. 조건부 예측방법을 이용하여 분석한 결과 고려된 두 모형에서 공통적으로 예측기간에 단계 5에 이르는 가계부채 부실위험은 없는 것으로 나타났다. 또한 모형에 따르면 2017년 3분기까지는 보통단계인 단계 1-2를 유지하다가 2017년 4분기에 들어서 위기 단계 4까지 상승하는 것을 볼 수 있으며 따라서 2017년 4분기에 적극적으로 가계부채의 추이와 관련 변수들을 주의 깊게 살펴 볼 필요가 있다고 하겠다.

본 연구를 바탕으로 몇 가지 흥미로운 연구방향을 제시하면 다음과 같다. 첫 번째, 본 연구에서 사용하고 있는 빅데이터는 아주 협소한 의미의 빅데이터이다. 모형의 설명변수로서 일반적으로 사용되는 인터넷 검색자료만을 사용하고 있다. 가계단위의 미시자료와 가계부채(대출)에 일별자료 그리고 네이버 랩 등에서 구현될 수 있는 각 개인의 검색자료와 개인들의 특성 및 성격들이 주어진다면 가계부채 위험에 대한 조기경보지수를 더 잘 예측할 수 있을 것이라 사료된다. 둘째, 최근에 크게 대두되고 있는 딥러닝(deep learning)의 주요한 모형도 인공신경망 모형이다. 이미 몇몇의 경제 및 재무변수들에 대한 예측을 딥러닝 알고리즘을 사용하여 행하고 있으며 이때 많이 사용되는 모형들은 잡음제거 오토인코더 (stacked denoising autoencoder (Vincent et al, 2008, 2010)) 신경망 모형, 순환 신경망 (recurrent neural network) 모형이나 배깅 (bootstrap aggregation: bagging (Breiman, 1996)) 등이 사용되고 있는 실정이다. 따라서 본 연구에서 사용하고 있는 모형을 딥러닝 방법을 사용하여 가계부채 부실 조기경보모형의 예측력을 높일 수 있을 것이라 생각된다. 셋째, 정책적인 입장에서는 가계부채 부실에 대한 위험도 예측이 큰 위험을 대처하는데 유용할지라도 그 예측결과만으로는 큰 의미를 지니기 힘들다. 가용할 수 있는 설명변수들이 정해져있다는 가정 하에 일반적으로 모형의 예측력을 높이는데 여러 눈에 보이지 않는 변수들과 예측모형들의 복잡한 모형설정이 동반되기 마련이다. 예를 들면, 인공신경망 모형에서 은닉층의 갯수와 신경세포들의 숫자들이 그러하다. 이는 가끔 정책입안자들이 중요하게 고려하는 변수들 간의 인과성의 인지 및 식

별능력을 약화시킨다. 따라서 예측력이 뛰어난 모형을 사용하여 고려되는 변수들을 사용한 충분한 시뮬레이션 연구를 실시하여 변수들 간 정량적인 인과관계를 정립하는 것이 향후 연구방향이 되어야 할 것이다.

참고문헌

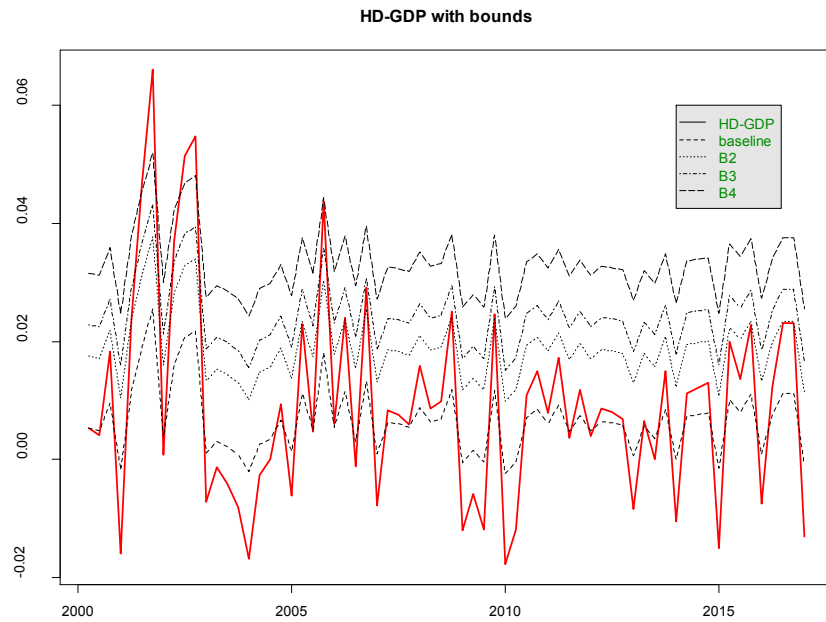
- 강종구(2017), “가계부채가 소비와 경제성장에 미치는 영향 - 유량효과와 저장효과 분석”, 「BOK 경제연구」, 한국은행, 2017-1.
- 김상환(2001) “은행위기 조기경보시스템 개발”, 한국금융연구원.
- 김영일·유주희(2013), “가계부채 부실위험에 대한 스트레스 테스트 : 가구자료를 중심으로”, 「경제분석」, 제19권, 제2호, pp. 59-95.
- 김우영·김현정(2009), “가계부채의 결정요인 분석”, 「금융경제연구」, 한국은행, 제380호.
- 김현정·김우영(2009), “가계부채가 소비에 미치는 영향 : 미시자료를 중심으로”, 「경제분석」, 제15권, 제3호, pp. 1-36.
- 김현정·김우영·김기호(2009), “한국노동패널자료를 이용한 가계부채 분석”, 「금융경제연구」 제366호, 한국은행 금융경제연구원.
- 김현정·손종철·이동렬·임현준·나승호(2013), “우리나라 가계부채의 증가 원인 및 지속가능성 분석”, 「BOK 경제리뷰」, 한국은행, No. 2013-4.
- 남준우(2003), “준모수 추정방법에 의한 기업의 부도 예측모형 추정”, 「KDIC 금융연구」, 제4권, 제3호, pp.31-60.
- 박대근·최우주(2015), “가계부채의 결정요인에 대한 패널자료 분석 : 주택가격과 대출심사기준을 중심으로”, 「경제연구」, 제33권 제1호, pp. 75-98.
- 박원암(2015), “우리나라 금융위기와 비모수적 및 모수적 조기경보모형의 성과 비교”, 「한국경제의 분석」, 제21권, 제3호, pp. 1-42.
- 박원암(2001), “한국 외환위기의 조기경보모형”, 「국제경제연구」, 제7권, 제1호, pp. 55-79.
- 서상원·오권영(2016), “가계부채의 부실위험성 예측 및 평가 : 가구자료를 활용한 지역별 분석”, 「대구경북 금융경제 조사연구 자료집」, 한국은행 대구경북본부, 제2장, pp. 29-71.
- 성태윤·박기영·안현미(2017), “가계부채 관리 제도의 방향 설정 : 국제 사례를 통한 정책적 시사점을 중심으로”, 「한국경제의 분석」, 제23권, 제1호, pp. 111-183.
- 이궁희(2017), “뉴스 빅데이터를 이용한 경기 판단 : 빅카인즈 뉴스 경기지수의 개발”, 「통계연구」, 제22권, 제2호, 67-94.
- 이동걸·전성인·정재욱·변동준(2014), “우리나라 가계부채의 연체결정요인 및 취약성 연구”, 「금융연구」, 제28권, 제2호, pp. 137-178.
- 이완수·박양수(2016), “경제 정보에 대한 비대칭적 반응 : 경제 뉴스에 대한 경제 주체의 심리와 행위”, 「한국언론학보」, 제60권, 제1호, 165-201.
- 정운영·정세은(2013), “소득계층별 부채가계의 특성과 결정요인의 비교: 부동산 요인을 중심으로”, 「사회보장연구」, 제29권, 제3호, pp.415-439.
- 함준호·조현철·권영철(2015), “가계부채의 경기순응성과 거시정보결합 신용평가에 대한 연구”, 「한국경제의 분석」, 제21권, 제1호, pp. 1-63.
- Abbas, O.M.(2017), “A Comparative Study on Business Forecasting Accuracy among Neural Networks and Time Series ”, International Journal of Computer, 26(1), pp. 175-183.
- Abbas, O.M.(2015), “Neural Networks in Business Forecasting”, International Journal of Computer, 19(1), pp. 114-128.
- Anthony, M., and P.L. Bartlett(1999), “Neural Network Learning: Theoretical

- Foundations” , Cambridge University Press.
- Arcand, J.L., E. Berkes and U. Panizza(2015), “Too much Finance?” , *Journal of Economic Growth*, 20(2), pp. 105-148.
- Artola C., and E.M. Galán(2012), “Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches” , Banco de Espana Occasional Paper No. 1203.
- Bean, C. (2016), “Independent Review of UK Economic Statistics” , Independent Report for HM Treasury and Cabinet Office, UK, March 2016.
- Berg, A., and C. Pattillo(1999), “Predicting Currency Crises : The Indicators Approach and an Alternative” , *Journal of International Money and Finance*, 18(4) pp. 561-586.
- Bishop C.M.(1995), “Neural Networks for Pattern Recognition” , Clarendon Press.
- Breiman, L.(1996), “Bagging Predictors” , *Machine Learning*, 24, 123-140.
- Chadwick, M., and G. Sengul(2012), “Nowcasting Unemployment Rate in Turkey: Let’s Ask Google” , Working Papers, Research and Monetary Policy Department, Central Bank of the Republic of Turkey.
- Carrière-Swallow, Y., and F. Labbe(2013), “Nowcasting with Google Trends in an Emerging Market” , *Journal of Forecasting*, 32(4), 289-298.
- Cecchetti, S., M. Mohanty and F. Zampolli(2011), “The Real Effect of Debt” , BIS Working Papers, no 352.
- Choi H., and H. Varian(2012), “Predicting the Present with Google Trends” , *Economic Record*, 88, 2-9.
- D’ Amuri, F., and J. Marcucci(2017), “The Predictive Power of Google Searches in Forecasting US Unemployment” , *International Journal of Forecasting*, 33(4), pp. 801-816.
- Davis, E.P., and D. Karim(2008), “Comparing Early Warning Systems for Banking Crises” , *Journal of Financial Stability*, 4(2), 89-120.
- Ghosh A., J. Ostry and N. Tamirisa(2009), “Anticipating the Next Crisis,” Finance and Development, 46(3), International Monetary Fund.
- Glostein, M., G. Kaminsky, and C. Reinhart(2000), “Assessing Financial Vulnerability: An Early Warning System for Emerging Markets” , Institute for International Economics.
- Hornik, K., M. Stinchcombe, and H. White(1989), “Multilayer Feedforward Networks are Universal Approximators” , *Neural Networks*, 2(5), pp. 359-366,
- Kaminsky, G., S. Lizondo, and C.M. Reinhart(1998), “Leading Indicators of Currency Crises” , *IMF Staff Papers*, 45(1), 1-48.
- Kim, H.H., G. Koop, and S. Chung(2016), “Macroeconomic Nowcasting using Naver Probabilities” , Mimemo.
- Koop G., and L. Onorante. “Macroeconomic Nowcasting using Google Probabilities” , Mimemo.
- Lawrence, M.J., R.H. Edmundson, and M.J. O’Connor(1985), “An Examination of the Accuracy of Judgemental Extrapolation of Time Series” , *International Journal of Forecasting*, 1, 25-35.
- Li, X.(2016), “Nowcasting with Big Data : Is Google useful in Presence of other Information?” , London Business School, Mimeo.
- McNeils, P.D.(2005), “Neural Networks in Finance: Gaining Predictive Edge in the

- Market” , Academic Press.
- Onorante, L., and K. Gary(2016), “Macroeconomic Nowcasting Using Google Probabilities” , Mimeo.
- Pouvelle, C.(2012), “Bank Credit, Asset Prices and Financial Stability: Evidence from French Banks” , IMF Working Paper No.12/103.
- Riedmiller M.(1994), “Advanced Supervised Learning in Multi-layer Perceptrons From Backpropagation to Adaptive Learning Algorithms” , *Computer Standards and Interfaces*, 16(3), pp. 265-278.
- Riedmiller, M. and H. Braun(1993), “A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm” , IEEE International Conference on Neural Networks, 1, pp. 586-591.
- Suhoy, T.(2009), “Query Indices and a 2008 Downturn” , Bank of Israel Discussion Paper 06.
- Tobback, E., and H. Naudts, W. Daelemans, E. Fortuny, and D. Martens(2016), “Belgian Economic Policy Uncertainty Index : Improvement through Text Mining” , *International Journal of Forecasting*, In Press.
- Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A.(2008), “Extracting and Composing Robust Features with Denoising Autoencoders” , *Proceedings of the 25th International Conference on Machine Learning*, ACM, Helsinki, Finland, 1096-1103.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A.(2010), “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion” , *Journal of Machine Learning Research*, 11, 3371-3408.

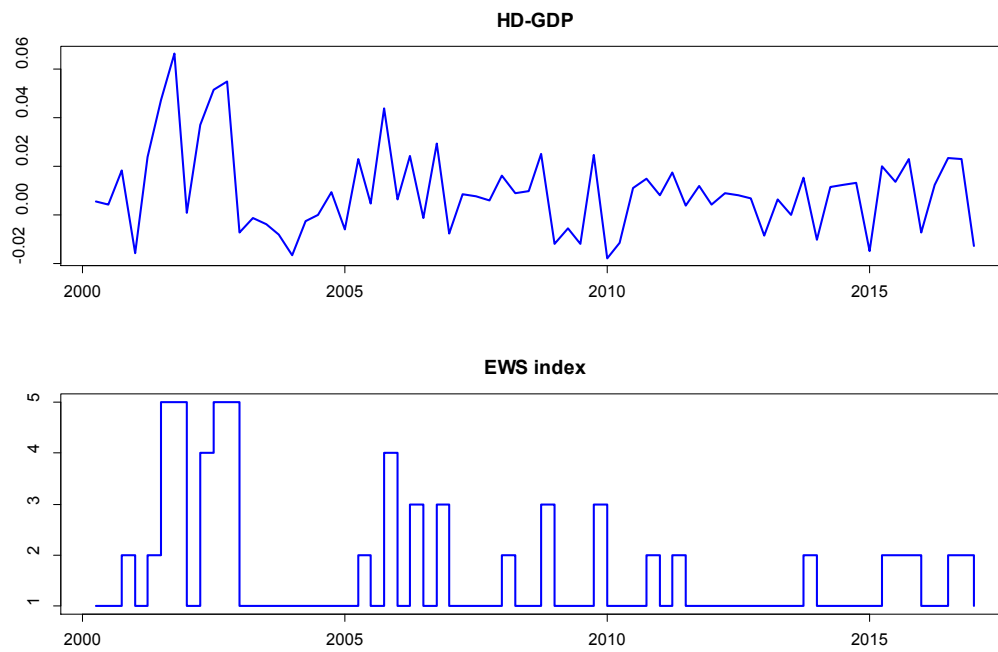
부 록

<그림 A1> $Index_{1,t}$ 와 위기의 경계선 [$Index_{1,t}$ 경우]

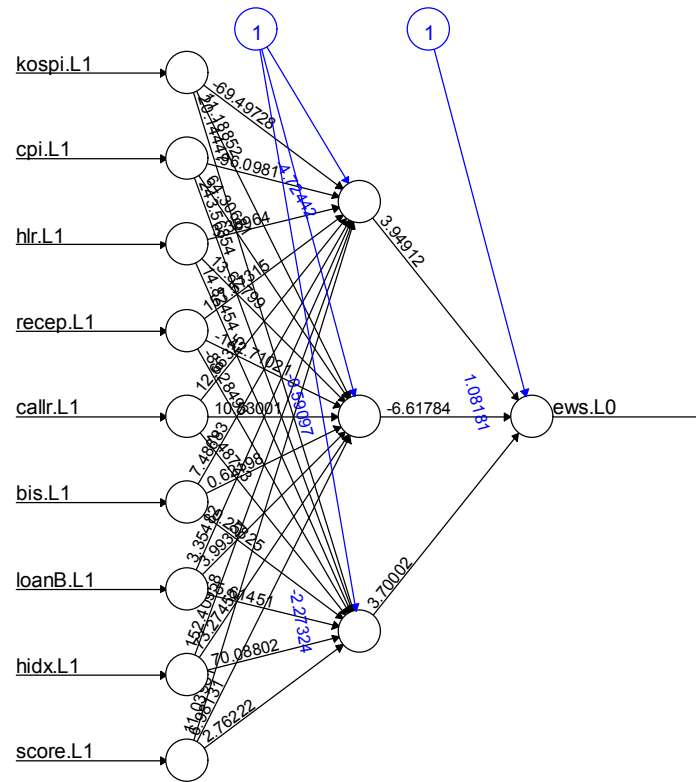


주 : 실선으로 표기된 HD-GDP가 $index_{1,t}$ 를 나타내며 baseline은 위기의 기준선 그리고 B2, B3, B4는 각각 단계 2-3, 3-4, 4-5를 구분하는 경계선을 나타낸다.

<그림 A2> $Index_{1,t}$ 와 EWS지수 [$Index_{1,t}$ 경우]



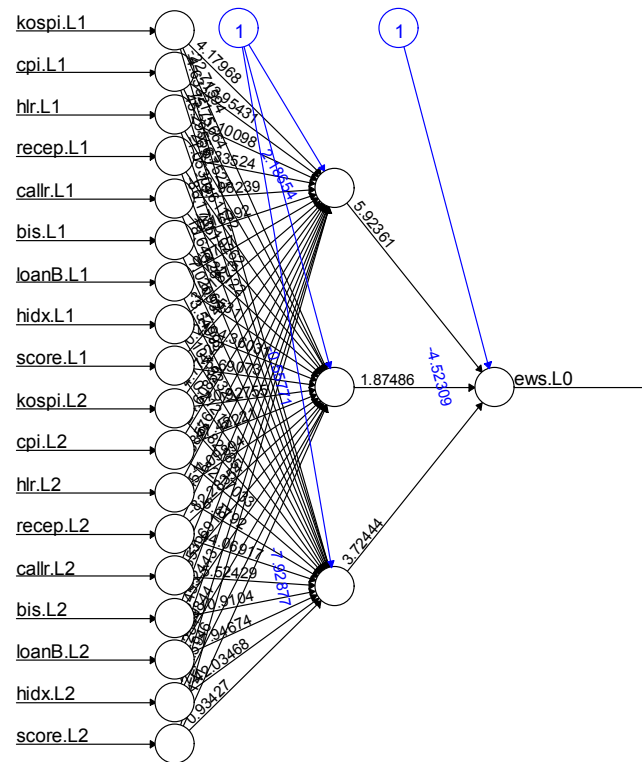
<그림 A3> 인공신경망모형 추정결과 (L1) [$Index_{1,t}$ 경우]



Error: 3.718909 Steps: 5585

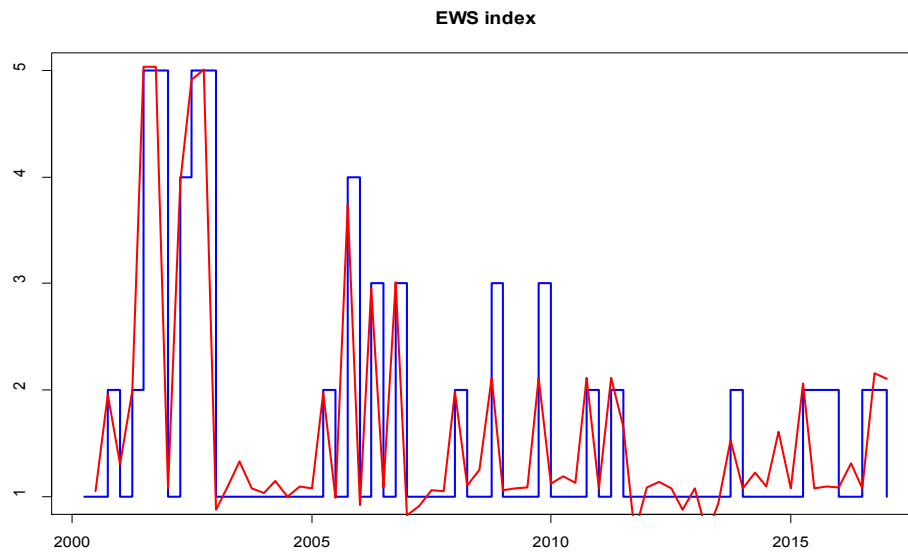
주 : kospi.L1, cpi.L1, hir.L1, recep.L1, callr.L1, bis.L1, loanB.L1, hidx.L1, 그리고 score.L1 각각 주가지수, 소비자물가지수, 가계대출금리, 총수신, 시장금리, BIS기준 자기자본비율, 고정이하 여신비율, 주택매매가격지수 그리고 인터넷 검색지수의 1차 시차변수이다.

<그림 A4> 인공신경망모형 추정결과 (L2) [$Index_{1,t}$ 경우]

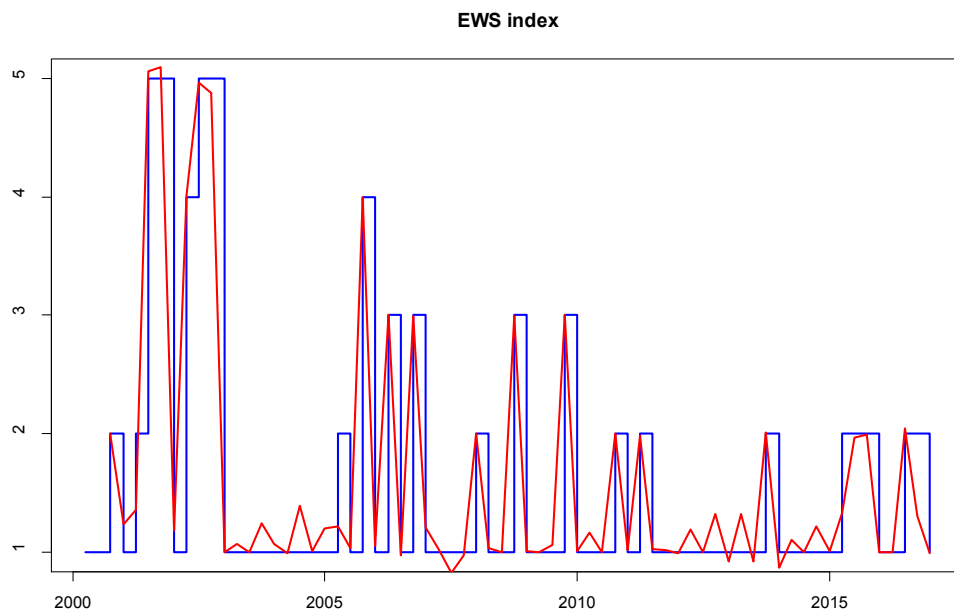


주 : kospi.Li, cpi.Li, hlr.Li, recep.Li, callr.Li, bis.Li, loanB.Li, hidx.Li, 그리고 score.Li 각각 주가지수, 소비자물가지수, 가계대출금리, 총수신, 시장금리, BIS기준 자기자본비율, 고정이하 여신비율, 주택매매가격지수 그리고 인터넷 검색지수의 i 차 시차변수이다.

<그림 A5> 인공신경망모형 표본 내 결과 (L1) [$Index_{1,t}$ 경우]



<그림 A6> 인공신경망모형 표본 내 결과 (L2) [$Index_{1,t}$ 경우]



<그림 A7> 조기경보지수 예측치 (2017Q2-2019Q2) [$Index_{1,t}$ 경우]

