

석사학위논문

카드이용이 세수확보에 미치는 영향

2014년

한성대학교 대학원

경제학과

부동산경제학 전공

김시언

석사학위논문
지도교수 김상봉

카드이용이 세수확보에 미치는 영향

Effects of Using Amount of Cards on a Source of Taxation

2013년 12월 일

한성대학교 대학원

경제학과

부동산경제학 전공

김시언

석사학위논문
지도교수 김상봉

카드이용이 세수확보에 미치는 영향

Effects of Using Amount of Cards on a Source of Taxation

위 논문을 경제학 석사학위 논문으로 제출함

2013년 12월 일

한성대학교 대학원

경제학과

부동산경제학 전공

김시언

김시언의 경제학 석사학위논문을 인준함

2013년 12월 일

심사위원장 _____ 인

심 사 위 원 _____ 인

심 사 위 원 _____ 인

국 문 초 록

카드이용이 세수확보에 미치는 영향

한성대학교 대학원

경제학과

부동산경제학전공

김 시 언

본 연구는 2005년 1분기에서 2012년 4분기까지의 분기별 자료를 이용하여 카드이용이 세수확보에 어떠한 영향을 미치는 지에 대해 제도 변화에 따라 분석하였다. 더미변수를 포함한 다중회귀분석으로 어떤 변수가 부가가치세에 가장 큰 영향을 미치는지 분석하였고, LASSO(Least Absolute Shrinkage and Selection Operator)분석을 이용하여 다중회귀분석에서 발생할 수 있는 다중공선성을 제거하여 분석하였다. 또한, VECM을 이용하여 신용카드 관련 변수가 부가가치세에 미치는 영향을 분석하였다. 분석결과, 다중회귀분석에서 현금영수증이 한 단위 증가할 때 부가가치세가 가장 크게 증가하는 것으로 나타난다. LASSO모형 분석결과, 현금영수증이 부가가치세에 대해 가장 높은 값을 갖는 것으로 나타난다. VECM에서는 GDP, 체크카드 이용금액, 신용카드 이용금액, 현금영수증 금액의 순서는 부가가치세에 대해 설명력이 높은 것으로 나타난다.

【주요어】 신용카드, 현금영수증, 다중회귀분석, LASSO 분석, VECM

목 차

I. 서 론	1
1.1 연구의 목적	1
1.2 연구의 범위와 방법	3
II. 이론적 배경	5
2.1 선행연구의 검토	5
2.2 본 연구의 차별성	9
III. 연구설계	10
3.1 연구모형	10
IV. 연구결과	27
4.1 기술통계량	27
4.2 상관관계 분석	33
4.3 다중회귀 분석	34
4.4 LASSO모형 분석	36
4.5 VECM 분석	37
V. 결 론	44
참고문헌	46
ABSTRACT	49

표 목 차

〈표 1〉 신용카드 공제율 변화	3
〈표 2〉 체크카드 공제율 변화	4
〈표 3〉 현금영수증 금액 공제율 변화	4
〈표 4〉 표본자료의 변수별 기초통계량	27
〈표 5〉 연도별 변수들 금액	28
〈표 6〉 연도별 변수들의 증가율	29
〈표 7〉 소득공제율의 변화에 따른 증가율	30
〈표 8〉 급여비율의 변화에 따른 증가율	31
〈표 9〉 소득공제율 × 급여비율의 변화에 따른 증가율	32
〈표 10〉 부가가치세와의 상관관계	33
〈표 11〉 다중회귀분석 결과	34
〈표 11-1〉 다중회귀분석 결과	35
〈표 12〉 LASSO모형 분석 결과	36
〈표 13〉 단위근 검정	37
〈표 14〉 공적분에 대한 Johansen 검정	38
〈표 15〉 그랜저 인과관계	40
〈표 16〉 충격반응함수	41
〈표 17〉 동태적 영향에 대한 분산분해	42

그 립 목 차

〈그림 1〉 선형보간법의 의미	10
〈그림 2〉 LASSO의 회귀계수	19
〈그림 3〉 LASSO의 회귀계수	19
〈그림 4〉 신용카드, 체크카드, 현금영수증, 부가가치세 금액	28
〈그림 5〉 충격반응함수 결과	41
〈그림 6〉 동태적 영향에 대한 분산분해	43

I. 서 론

1.1 연구의 목적

신용카드 소득공제 제도는 상거래의 투명성 제고를 통한 세원 확보와 자영업자의 세 부담 완화를 위하여 1999년에 도입되어 3년 동안 한시적으로 시행되었다. 제도 도입 당시, 우리나라는 자영업자와 소비자 간 거래 시 현금 거래가 빈번하여 영수증 발행이 잘 이루어지지 않아 세원 파악이 어려웠다. 또한, 자영업자가 세금 신고 시 현금매출 부분을 누락시켜 세금 부담을 회피하는 경우가 많았다. 1990년대 말까지도 자영업자의 과표양성 비율은 약 30~50% 정도에 불과한 것으로 추정되어 세 부담 형평성 측면에서 문제가 많았다.¹⁾ 이러한 현실에서 정부는 자영업자와 봉급생활자 간의 세 부담의 형평성을 제고하고, 조세정의를 구현하여 상거래의 투명성을 높이기 위해 신용카드를 사용하는 소비자에게 세금 감면 혜택을 주는 신용카드 활성화정책을 추진하게 되었다. 동 정책은 시행 이후, 일몰기한이 4차례 연장되는 가운데 2011년 말 일몰폐지가 예정이었으나 정부는 신용카드 소득공제 제도의 적용시한을 3년 연장하였다. 동 제도의 일몰이 도래할 때마다 시민단체와 언론 등은 카드사용에 따라 세금감면 혜택을 받고 있는 중산층 서민의 세 부담이 증가하므로 신용카드 소득공제 제도를 존속시켜야 한다고 주장하였다. 학계에서 상기의 효과성 평가에 대한 논의가 제대로 이루어지지 못한 상황에서 행정부는 신용카드 소득공제 제도를 3년 연장하는 내용으로 2011년 세법개정안을 국회에 제출하였다. 반면, 일부 경제 전문가들은 조세부담 형평성 측면에서 서민보다 고소득층을 위한 것으로 동 제도가 당초 시행시의 취지를 달성하였다면 예정대로 폐지하는 것이 바람직하다는 견해를 보이고 있다.²⁾ 한편, 신용카드 소득공제 제도의 유지를 위해 직·간접적인 비용이 따르는데 직접적인 비용인 소득세 감면액(조세지출액)은 2010년에 약 7천억원에 달하는 것으로 추

1) 김재진(2013), 신용카드 활성화정책의 효과와 사회적 비용

2) 성명기(2011), 신용카드 소득공제의 소득계층별 귀착 및 세수효과와 시사점

계되고 있다. 카드 소득공제액 대비 조세지출액 비율은 공제대상자 층이 두터운 소득구간에서 하락폭이 상대적으로 크고 전체 하락세를 주도하는 경향을 보였다.³⁾

신용카드 산업은 정부의 신용카드의 활성화 정책에 힘입어 단기간 내에 양적인 성장을 이루었다. 총 민간소비지출 대비 신용카드 이용금액은 1999년 8.1%에서 2012년 64.4%로 크게 증가하였고, 신용카드 발급 수 또한 1999년 약 3,899만매 수준에서 2012년에는 약 1억 964만매로 증가하였다. 또한, 신용카드 거래를 위해 등록된 신용카드 가맹점도 매년 빠르게 늘어나 1999년 619만 점에서 2012년 2,016만 점으로 약 3.2배가 증가하였으며, 1만원 이하의 소액결제에 대한 결제도 증가하는 것 또한 신용카드 활성화 정책으로 인해 신용카드 거래가 소액결제까지 확산되고 있다. 이와 같이 신용카드 거래의 양적 증가는 종합소득세 신고인원의 증가, 종합소득세 장부 및 증빙에 의한 신고 비중의 증가, 부가가치세 신고인원 및 과세표준의 증가, 부가가치세 간이과세자의 비율 및 납부의무면제자 비율의 감소, 소비자상대 현금수입 업종의 부가가치세 신고인원 및 과세표준 증가는 정책목표인 과표양성화에 긍정적인 영향을 미치고 있다. 그러나 신용카드 거래를 활성화하기 위한 정책을 추진하는 과정에서 발생하는 조세지출 비용과 자영업자의 가맹점수수료 지출비용, 신용불량자 문제 등 국민이 부담해야 하는 사회적 비용도 적지 않다.

신용카드 사용율의 증가에도 불구하고 소득세 증가율은 상대적으로 낮아 카드사용에 따른 세수확보가 어려워지고 있다. 또한 동 제도로 인해 우리 지하경제 규모는 줄어들고 있으나 아직 높은 수준이며 자영업자 소득탈루율도 여전히 높은 상황임을 감안할 필요가 있다.

본 연구의 기여 및 목적은 다음과 같다. 첫째, 2005년 현금영수증 소득공제 제도가 새롭게 도입됨으로써 세원확보에도 영향을 주기 시작하였다. 본 논문에서는 신용카드, 체크카드 그리고 현금영수증의 소득공제율 변화에 초점을 맞추어 분석하였다.⁴⁾ 따라서 소득공제 제도변화가 세수확보에

3) 성명기(2011), 신용카드 소득공제의 소득계층별 귀착 및 세수효과와 시사점

4) 다중회귀분석에서 체크카드 이용금액은 신용카드 이용금액에 비해 비중이 크지 않기 때문에 공제율 변화에 대한 더미변수를 지정하지 않았다.

얼마나 영향을 줄 것인가를 분석하였다. 둘째, 기존 경제학에 도입되지 않은 LASSO 모형을 사용하여 OLS의 문제점으로 인식되는 다중공선성을 개선하여 모형을 설명하였다.

1.2 연구의 범위와 방법

자료의 시간적 범위는 현금영수증 소득공제 제도가 시작된 2005년 1분기부터 2012년 4분기까지 32개 분기로 하였다. <표 1>, <표 2>, <표 3>과 같이 연도별 공제율 변화를 고려하여서 세수의 흐름을 분석하였다.

부가가치세를 종속변수로 하고 독립변수인 신용카드 이용금액, 체크카드 이용금액, 현금영수증 발급금액, GDP의 상관관계를 분석하였다. 상관계수를 통하여 각각의 설명변수들이 부가가치세와 유의한 상관관계가 있는지를 분석한 후, 기간 더미 변수에 따른 다중회귀분석을 통하여 각각의 선행 지표들이 부가가치세에 미치는 영향을 분석하였다. 또한, LASSO 모형을 사용하여 계수 값의 영향력을 보다 정확하게 측정하였다. 마지막으로 VAR모형 또는 VECM을 이용하여 종합적으로 부가가치세에 미치는 영향을 분석하였다.

<표 1> 신용카드 공제율 변화

(단위: %, 만원)

	2005	2006	2007	2008	2009	2010	2011	2012
총 금액	15	15	15	20	20	25	25	25
공제율	20	15	15	20	20	20	20	20
공제한도	500	500	500	500	500	300	300	300

〈표 2〉 체크카드 공제율 변화

(단위: %, 만원)

	2005	2006	2007	2008	2009	2010	2011	2012
총 금액	15	15	15	20	20	25	25	25
공제율	20	15	15	20	20	25	25	30
공제한도	500	500	500	500	500	300	300	300

〈표 3〉 현금영수증 금액 공제율 변화

(단위: %, 만원)

	2005	2006	2007	2008	2009	2010	2011	2012
총 금액	15	15	15	20	20	25	25	25
공제율	20	15	15	20	20	20	20	20
공제한도	500	500	500	500	500	300	300	300

Ⅱ. 이론적 배경

2.1 선행연구의 검토

아직 우리나라에서 신용카드 사용과 세수에 대한 연구는 많지 않은 편이다. 먼저, 김재진(2013)은 정부가 자영업자의 과세표준을 양성화하여 근로소득자와의 세 부담 형평성을 제고하고자 추진한 신용카드 활성화 정책은 단기간에 목적을 달성했다고 보았다. 그러나 분석결과, 신용카드 활성화정책의 부정적 효과로 상당한 사회적 비용이 발생하는 것으로 보았다. 가맹점이 1999년~2011년까지 신용카드 결제로 지불한 가맹점수수료가 약 61조원, 신용카드 활성화를 위한 조세지출비용이 약 22조 2,300억원에 달하는 것으로 나타난다. 또한, 무분별한 카드발급으로 인한 신용불량자 문제로 인한 사회적 비용을 감안하지 않더라도, 신용카드 활성화정책으로 인하여 우리 사회가 부담한 사회적 비용은 총 83조 2,300억원에 이르는 것으로 나타난다. 앞으로는 정책 추진과정에서 나타난 여러 가지 부작용과 사회적 비용을 최소화할 수 있도록 정책방향을 전환할 필요가 있다고 보았다.

박춘래·이은호(2003)는 신용카드 공제제도의 방법적인 문제로 저소득근로자들이 받는 혜택이 적어 신용카드를 적극적으로 사용하지 않는다고 하였다. 또한, 납세자들을 대상으로 신용카드 소득공제 제도의 활용 실태를 분석하여 제도가 가야할 방향을 연구하였다. 결과적으로, 근로소득자의 신용카드사용에 대해서 혜택을 주는 방법 중에 하나인 세액공제 제도가 더 합리적인 방법이라고 보았다.

김상헌(2001)은 신용카드 영수증 복권제도와 신용카드 사용액 소득공제 제도의 경제적 효과를 분석하였다. 분석결과, 두 제도의 총편익은 부가가치세수와 사업자 소득세수의 증대로 인한 부분이 1조 6,290억 정도이고 내국세의 증대는 4조 7,900억으로 나타난다. 한편, 총 비용은 2,540억으로 추산되어, 총 편익 대비 비용의 비율은 6.08과 17.88로 나타난다. 따라

서 편익/비용 측면에서 볼 때 두 제도는 일단 성공적인 것으로 평가하였다. 두 제도의 효과를 비교하면 신용카드 영수증 복권제도가 신용카드 사용액 소득공제제도에 비해 효과적인 것으로 보았다.

송헌재·성명재(2012)는 신용카드 소득공제 제도가 가지고 있는 세수효과, 소득재분배 효과, 소득세 경감효과를 분석하였다. 모의실험결과, 신용카드 소득공제 제도는 소득공제 확대를 통해 1.4조원의 근로소득세 세수를 감소시켰으나, 과표양성화를 통해 사업소득세 세수가 2.3조원 증대되었다. 총체적으로는 9천억원 정도의 세수 순증효과를 나타낸 것으로 추정하였다. 소득재분배의 관점에서, 동 제도들은 사업소득세의 과표양성화를 통해 0.23%p의 소득재분배 확대 효과를 나타내지만, 근로소득세의 소득공제 확대를 통해서도 오히려 소득재분배 효과가 0.19%p 만큼 축소되어, 전체적으로는 0.04%p의 순증효과를 나타낸 것으로 분석되었다. 회귀분석결과, 총급여액이 약 1.5억원 수준에 근접하는 고소득층에 이르기까지, 평균적으로 근로소득세 경감 효과가 계속 확대됨으로써 세 부담의 수직적 형평성 제고라는 취지에 부합하지 않는 것으로 분석되었다. 즉, 회귀분석결과에서도 소득재분배 효과의 관점에서 동 제도가 바람직하지 않은 것으로 평가되었다. 신용카드 소득공제 제도의 주된 도입 목적이 근로소득자들의 세 부담 경감이 아니라 사업소득세 과표양성화에 있었던 만큼, 소득재분배효과 측면에서 부정적인 효과가 큰 신용카드 소득공제 제도는 소기의 목적이 달성될 때까지 한시적으로 운용하는 것이 바람직하다고 보았다.

이만우·고영일(2002)은 신용카드 지원제도의 세수증대 효과를 분석하였다. 부가가치세집계표를 이용하여 신용카드 지원제도로 인하여 부가가치세와 개인사업자 및 법인의 소득세가 얼마나 증가하는 지에 대해 분석하였다. 분석방법은 업종을 현금수입업종과 비현금수입업종으로 구분하고, 부가가치세, 개인사업소득세 그리고 법인세를 시간추세항과 신용카드 지원제도의 효과를 반영하는 2000~2001년 기간더미를 설명변수로 추정하여 기간더미가 유효한지에 대하여 검토하는 방식을 택하였다. 추정결과 비현금수입업종에 대한 기간더미의 추정치는 유효하지 않은 반면 현금수입업종의 추정치는 유효하여 신용카드 지원제도가 현금수입업종에 영향을

미친다고 보았다. 신용카드 지원제도에 따른 세수증가 효과는 각 세수에 대하여 시간추세항과 신용카드 지원제도의 효과를 반영하는 2000~2001년 기간더미가 존재함에 따라 기울기가 변화되므로 기간더미 존재 전후 추정된 세수의 차이로 파악하였다. 그 결과, 2000년 6,606억원, 2001년 1조 7,009억원 수준으로 추정되었다. 이 중, 부가가치세가 전체 세수변화의 56~59%, 사업소득세수가 25~27% 그리고 법인세수가 25~28%를 각각 차지하였다. 또한, 업종별로 2001년 18,345억원에 대하여 기타서비스업이 1조 481억원으로 57.1%를 차지하며, 소매업이 22.5%(4,129억원), 음식업이 19.8%(3,640억원)를 차지하고 있다. 한편, 신용카드 지원제도로 인한 비용은 2000년 5,603억원으로 추계되는데 근로세액공제액이 1,254억원, 사업자세액공제가 4,129억원, 그리고 영수증복권제 당첨금 및 운영비가 221억원으로 나타난다. 2001년에는 1조 136억원으로 늘어나는데 근로세액공제액이 3,600억원, 사업자세액공제가 6,316억원, 그리고 영수증복권제 당첨금 및 운영비가 220억원으로 나타난다. 따라서 신용카드 지원제도의 세수효과/비용 비율은 2000년 1.17, 2001년 1.67로 나타나 신용카드 지원제도가 효과는 있지만 그 효과가 비용대비 그리 높지 않으며, 향후 과표양성화가 진전됨에 따라 그 효과는 작아질 것으로 판단되므로 사업자에게 주는 세액공제제도 축소를 고려할 필요가 있다고 보았다.

박성배·홍성경(2004)은 2000~2002년을 자료와 서울시내 296개 음식사업자의 부가가치세 과세표준 신고내역서를 이용하여 선형회귀분석의 공분산분석을 시행하여 신용카드 활성화가 부가가치세 과세표준과 의제매입세액공제에 미치는 영향을 분석하였다. 분석결과, 신용카드 활성화로 인하여 신용카드 결제액이 증가할수록 매출과세표준도 증가그러나 업종별 및 과세유형별로는 차이가 나지 않으며, 매출과세표준 증가에 따라 매입과세표준과 의제매입액도 증가그러나 그 증가정도는 매출과세표준보다 낮게 나타난다. 음식사업자의 경우, 신용카드 매출증가로 매출액이 양성화되면서 매출과세표준은 증가하였지만 매입과세표준과 의제매입세액공제액은 그만큼 증가하지 못하여 부가가치세 부담이 현저히 확대되었다고 주장하였다. 음식사업자의 부가가치세 과세표준 신고서라는 미시자료를 활용하여 분석

하고 있으나 특정 2개 연도와 서울 시내 일부 음식사업자만을 대상으로 분석하였다는 한계가 존재하므로 이 역시 일반화된 결과로 해석하기는 곤란하다.

빈기범·박명호·우석진(2010)은 신용카드 사용이 지하경제 축소에 미치는 영향을 분석하였다. 1990~2009년 자료와 GDP 대비 지하경제 규모의 비율을 종속변수로 자영업자수, 국민부담률, 현금영수증터미, GDP 대비 신용카드사용액 비율을 설명변수로 하였다. 분석결과, GDP 대비 신용카드 사용액 비율만이 유의한 것으로 나타났으며, 이 비율이 1% 상승시 GDP 대비 지하경제가 0.11~0.13% 감소한다고 추정하였다. 따라서 신용카드거래의 일반화와 보편화로 인한 카드사용액의 증가는 지하경제 규모를 축소시키는 유의한 요인으로 작용하고 있다.

김상점·박범조·송재은(2011)은 신용카드 사용 활성화가 국가재정에 미치는 영향을 분석하였다. 1990~2009년 기간을 대상으로 상태공간모형(state space model)으로 분석한 결과, 신용카드 사용액 1% 증가는 부가가치세수를 0.75% 증가시키고 있었다. 한편, 소비변동을 억제하는 것으로 나타난다. 이를 근거로 신용카드 사용이 거래투명성을 제고하여 과세베이스(tax base)를 확대시켜 세수를 증대시키는 효과가 있는 것으로 나타난다. 카드사용의 여신제공기능에 따라 경기변동에 따른 소비변동성을 억제하여 소비평활화(consumption smoothing) 효과를 유발함으로써 세수확대 및 안정성을 높이게 된다고 보았다. 그러나 저자들은 같은 모형으로 부가가치세 이외 개별소비세, 교통세 그리고 소득세 신고분에 대하여 실증 분석한 결과 유의하지 않는다고 밝혔으나 추정결과는 제시하지 않았다.

송헌재(2011)는 조세부담의 수직적 형평성 관점에서 신용카드 소득공제 제도를 평가하였다. 국세청 국세통계 자료를 이용하여 현재의 카드소득공제제도에서 소득공제 한도가 설정되어 있음에도 불구하고 총급여가 증가하면 공제액도 증가하도록 설계되어 있으며, 고소득계층 일수록 카드사용액도 크고 한계세율도 높기 때문에 소득세 감면 혜택이 커짐을 보였다. 따라서 카드사용 소득공제제도가 서민을 위한 대표적인 제도로 사람들이 인식하고 있지만 실상은 고소득자에게 혜택이 돌아가는 제도이므로 조세부

담의 수직적 형평성 관점에서 폐지하는 것이 바람직할 수도 있다는 견해를 밝히고 당초 카드사용 공제제도의 도입 당시 목적이었던 거래투명성 확보를 위한 세원발굴과 근로소득자 세부담 증가 완화라는 문제가 달성되었는가에 대한 면밀한 평가를 해야 한다고 주장하였다.

송재은·박범조·김상겸(2012)은 신용카드사용의 활성화가 세수 증대에 어떠한 영향을 미치는지에 대해 이론적, 실증적으로 분석하였다. 1990년부터 2009년까지의 시계열 자료와 신용카드 사용액, 국내총생산액, 투자액, 재화 및 서비스수출액과 수입액, 민간소비액을 설명변수로 하였으며, 종속변수로는 부가가치세, 특별소비세, 교통세의 변화율로 하였다. 분석결과, 신용카드의 활성화는 간접세 중 부가가치세의 세수증대에 유의한 영향을 미친 것으로 나타난다. 신용카드의 활성화는 세수증대 및 세수안정성 측면에서 유의미한 효과를 발생시킨 것으로 평가되며, 이에 따라 우리나라 재정건전성 및 안정성 확보에도 긍정적 효과를 발생시킨 것으로 판단하고 있다.

2.2 본 연구의 차별성

선행연구의 연구모형은 단순회귀모형, 상관관계, VAR·VECM모형을 대표적으로 많이 사용하였다. 또한, 소득공제혜택 중에서 상대적으로 기간이 짧은 현금영수증 공제율은 포함시키지 않고 분석하였다. 그러나 본 연구에서는 다중공선성을 제거하여 계수를 추정하는 최소절대계수축소변수선택법 또는 라소방법(Least Absolute Shrinkage and Selection Operator, LASSO)라는 새로운 모형을 사용하였고 공제혜택으로 현금영수증 공제율을 추가하여 종속변수와 독립변수 간의 관계를 살펴본다.

Ⅲ. 연구 설계

3.1 연구모형

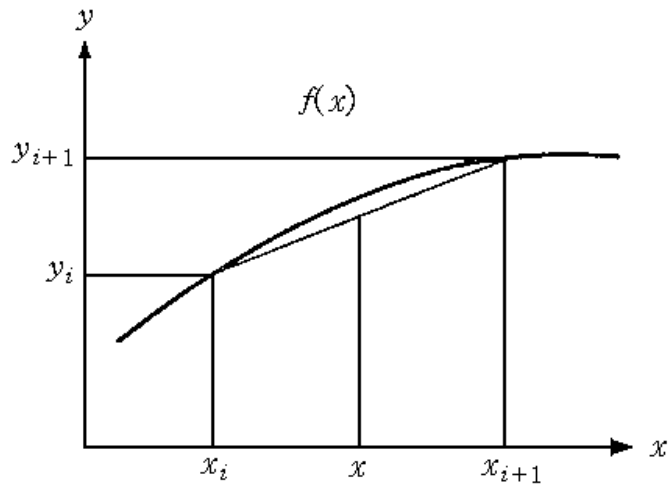
1. 선형보간법

선형보간(linear interpolation)법은 다음과 같다.

$$y = f(x) \quad (1)$$

식(1)의 곡선상의 두 점 (x_i, y_i) , (x_{i+1}, y_{i+1}) 이 주어진 경우, 임의의 점 x 에 대해서는 $f(x)$ 의 근사값 $p(x)$ 을, 이 두 점을 직선으로 연결된 x 의 1차 함수로서 구하는 방법이다.

〈그림 1〉 선형보간법의 의미



선형보간법 두 점을 맺는 직선은 다음과 같다.

$$\frac{y - y_i}{x - x_i} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \quad (2)$$

이것을 y 에 관하여 풀고, $y = p(x)$ 라고 놓으면 로 쓸 수 있다. 또한, 선형보간은 1차보간이라고도 한다.

$$p(x) = \frac{y_i(x - x_{i+1}) - y_{i+1}(x - x_i)}{x_i - x_{i+1}} \quad (3)$$

본 연구에서는 부가가치세와 현금영수증의 결측값을 재보정하기 위하여 선형보간법을 사용하였다.⁵⁾

5) 현금영수증은 2005년에 분기 데이터이고 2006년부터 2012년까지는 반기 데이터이다. 부가가치세는 2005년에서 2012년까지 연간 데이터이다.

2. 상관관계

상관계수란 두 변수가 관계되어 있는 정도를 나타내는 지수를 말한다. 즉, 상관관계를 수치적으로 나타낸 값을 말한다. 상관계수는 자료상의 두 변수의 가변성(variability)으로 간주되는 통계적 공변동의 표준화된 측정치로, 그 개념을 이해하기 위해서는 이변수(bivariate) 데이터와 두 임의 변수의 결합확률분포가 필요하다.

2.1 모상관계수 : ρ

측정단위나 대상에 관계없이 두 변수 사이의 선형관계의 강도를 나타내는 지표는 두 변수의 공분산을 표준화시키면 된다. 모집단의 상관계수를 모상관계수(population correlation coefficient) ρ 라고 한다. 일반적으로 이차원 유한모집단 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 인 경우에는 각 특성에 대한 값의 집합인 $\{x_1, x_2, \dots, x_n\}$ 과 $\{y_1, y_2, \dots, y_n\}$ 의 모평균 (μ_x, μ_y) 와 모표준편차 (σ_x, σ_y) 를 이용하여 모상관계수를 계산할 수 있다.

$$\rho = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right) \quad (4)$$

이는 공분산 $cov(X, Y)$ 를 각 변수들의 표준편차인 σ_x, σ_y 의 곱으로 나눈 값이기도 하다.

$$\rho = \frac{Cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\sigma_{xy}^2}{\sigma_x \cdot \sigma_y} \quad (5)$$

또한, 모상관계수 ρ 는 다음의 식으로부터 항상 $-1 \leq \rho \leq 1$ 임을 알 수 있다.

$$0 \leq \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} - \rho \frac{y_i - \mu_y}{\sigma_y} \right)^2 = 1 - \rho^2 \quad (6)$$

2) 표본상관계수 : r

모집단의 모수와 같이 모상관계수 ρ 도 확률표본을 사용하여 추정하는데, ρ 의 추정치로 가장 흔히 사용되는 것이 피어슨의 표본상관계수(sample correlation coefficient) r 이며 다음과 같이 정의된다.

$$\begin{aligned} r &= \frac{1}{(n-1)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \end{aligned} \quad (7)$$

여기서 $\bar{x}$, $\bar{y}$ 는 각각 모평균 μ_x , μ_y 의 추정치이고, s_x , s_y 는 각각 모표준편차 σ_x , σ_y 의 추정치이며 다음과 같이 추정한다.

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \\ s_x &= \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad s_y = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2} \end{aligned} \quad (8)$$

3. 다중회귀분석

종속변수 Y_i 가 $X_{2i}, X_{3i}, \dots, X_{ki}$ 와 상수항 1을 포함한 k 개의 독립변수에 의해 설명되는 다중회귀모형의 일반형은 다음과 같이 표현될 수 있다.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i \quad (9)$$

위 모형을 추정함은 곧 $\beta_1, \beta_2, \dots, \beta_k$ 의 k 개의 모수들을 추정함을 말한다. 이 때 표본에서의 추정식은 단순회귀분석에서와 마찬가지로 다음과 같다.

$$Y_i = b_1 + b_2 X_{2i} + \dots + b_k X_{ki} + e_i \quad (10)$$

여기서 $X_{ji} (j=1,2,\dots,k, i=1,2,\dots,n)$ 는 j 번째 변수에 대한 i 번째 관측치를 나타내며, 첫 번째 변수는 $X_{1i} = 1$ 로 상수항을 나타낸다.

다중회귀분석에서도 단순회귀분석과 마찬가지로 결정계수가 정의된다. 그런데 다중회귀분석에서는 독립변수의 수가 여럿인데, R^2 의 특성 가운데 하나는 독립변수의 수가 증가함에 따라 R^2 의 값도 증가한다는 점이다. 독립변수의 수가 증가함에 따라 만일 새로 추가된 독립변수가 종속변수를 설명하는데 중요한 변수라면 독립변수가 설명하지 못하는 잔차항 제곱합 $\left(\sum_{i=1}^n e_i^2\right)$ 이 크게 감소한다. 따라서 R^2 가 크게 증가하게 된다. 반면에 새로 추가된 독립변수가 종속변수를 설명하는데 중요하지 않은 변수라고 하더라도 잔차항 제곱합을 조금 감소시켜 R^2 가 조금 증가하게 된다. 따라서 단순히 결정계수 R^2 를 증가시키기 위해서 종속변수(Y)의 설명에 중요하든 안중요하든 상관없이 독립변수의 수를 증가시킬 수도 있다. 이러한 바람직하지 않은 경우를 방지하기 위해 R^2 를 산출할 때 포함된 독립변수의 수를 감안하여 그 수가 크면 그만큼 R^2 에 불이익을 주는 것을 고려할 수 있다. 이를 조정하기 위해 다음과 같은 조정된 결정계수(adjusted R^2 ; $\overline{R^2}$)를 사용한다.

$$\overline{R^2} = 1 - \frac{\sum_{i=1}^n e_i^2 / (n-k)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)} = 1 - (1 - R^2) \frac{n-1}{n-k} \quad (11)$$

$k > 1$ 이면 항상 $R^2 > \overline{R^2}$ 의 관계가 성립하며, 다른 조건이 일정할 때 독립 변수의 개수 k 가 증가함에 따라 $\overline{R^2}$ 의 값은 감소함을 알 수 있다. 따라서 종속변수 Y 의 설명에 유의적이지 않은 독립변수를 포함시켜 결정계수가 증가하는 것을 방지할 수 있다.

4. LASSO

4.1 이론적 배경

다음과 같은 회귀모형이 있다고 하자.

$$y_i = \alpha + x_i^T \beta + \epsilon, \quad i = 1, 2, \dots, p \quad (12)$$

일반적으로 이 문제를 해결하기 위해 잔차의 제곱합을 최소화하는 보통 최소제곱(Ordinary Least Squares, OLS)이 사용된다. 그러나 보통최소제곱은 독립변수의 개수가 증가하면, 독립변수들 사이의 강한 상관관계(correlation)로 인한 다중공선성(multicollinearity)이 존재할 수 있으며, 따라서 회귀계수의 분산이 커져서 추정회귀식의 예측력(prediction accuracy)이 떨어지는 문제점이 있다. 또한, 독립변수의 개수가 증가하면 변수에 대한 해석력(interpretation)이 떨어진다. 즉, 많은 독립 변수 중 어떤 변수가 중요한 역할을 하는지에 대한 판단이 어려워진다.

보통최소제곱의 문제점이 보완된 방법으로 부분집합선택방법(subset selection method)과 능형회귀(ridge regression)가 있다. 부분집합선택방법(subset selection method)은 해석 가능한 모형을 제공그러나 데이터가 조금만 바뀌어도 전혀 다른 모형이 나타날 수 있고 이것은 예측력(prediction accuracy)을 떨어뜨릴 수 있다. 능형회귀(ridge regression)는 회귀 계수들에 제약조건인 $\sum \beta_j^2 \leq t$ 로 수를 추정하는 방법으로 회귀계수를 축소(shrinkage)시켜 분산이 작아지지만 회귀계수를 완전히 0으로 추정하지는 못하므로 모형의 해석이 쉽지 않다.

LASSO(Least Absolute Shrinkage and Selection Operator)는 Tibshirani(1996)에 의해 개발된 방법으로, 회귀 계수들에 제약조건 $\sum |\beta_j| \leq t$ 을 주어 계수 추정치들의 크기를 축소(shrinkage)시키는 동시에 변수선택도 할 수 있는 추정방법이다.

4.2 모형의 설정

p개의 독립변수 $x_i = (x_{i1}, \dots, x_{ip})^T$ 와 종속변수 y_i 를 가지는 다음과 같은 모형이 있을 때,

$$y_i = \alpha + x_i^T \beta + \epsilon, \quad i = 1, 2, \dots, p \quad (13)$$

회귀계수 추정치를 $\hat{B} = (\hat{B}_1, \hat{B}_2, \dots, \hat{B}_p)^T$ 이라고 하면, LASSO에 의한 회귀계수 추정 $(\hat{\alpha}, \hat{\beta})$ 은 다음을 만족한다.

$$\begin{aligned} \operatorname{argmin} \{ & \sum_{i=1}^N (y_i - \alpha - \sum_j \beta_j x_{ij})^2 \} \\ \text{subject to } & \sum_j |\beta_j| \leq t \end{aligned} \quad (14)$$

여기서 $t (\geq 0)$ 는 회귀계수의 축소의 정도를 조절하는 조절모수(tuning parameter)이다. 독립변수 x_{ij} 는 평균이 0, 분산이 1로 표준화(standardization)되었다고 가정하면, 모든 t 에 대해서 다음을 만족한다.

$$\begin{aligned} \frac{\partial \left[\sum_{i=1}^N (y_i - \alpha - \sum_{j=1}^P \beta_j x_{ij})^2 \right]}{\partial \alpha} \Big|_{\alpha = \hat{\alpha}} &= -2 \sum_{i=1}^N (y_i - \alpha - \sum_{j=1}^P \beta_j x_{ij}) \Big|_{\alpha = \hat{\alpha}} \\ &= -2 \sum_{i=1}^N (y_i - \alpha - \sum_{j=1}^P \beta_j x_{ij}) \\ &= 0 \end{aligned} \quad (15)$$

모든 j 에 대해서 $\sum_{i=1}^N x_{ij} = 0$ 이므로, $\hat{\alpha}$ 은 다음과

$$\begin{aligned}\hat{\alpha} &= \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^P \hat{\beta}_j \left(\sum_{i=1}^N x_{ij} \right) \\ &= \frac{1}{N} \sum_{i=1}^N y_i\end{aligned}\tag{16}$$

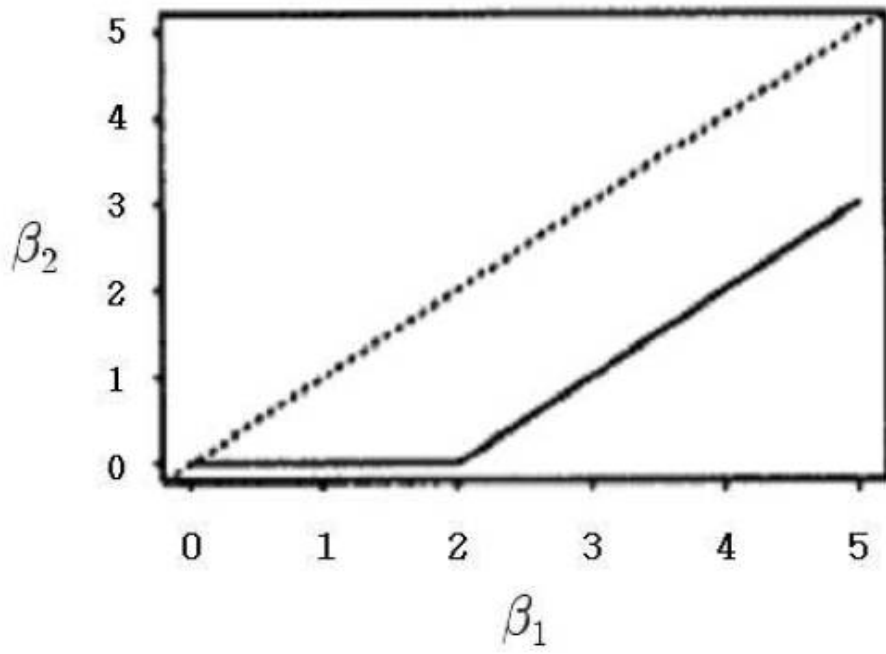
일반화손실(loss of generality)없이 $\bar{y}=0$ 을 가정할 수 있으므로 α 는 생략할 수 있다. 따라서 LASSO에 의한 회귀계수 추정치는 다음을 만족하는 $\hat{\beta}$ 을 찾는 것이다

$$\begin{aligned}\operatorname{argmin} \{ & \sum_{i=1}^N (y_i - \sum_j \beta_j x_{ij})^2 \} \\ \text{subject to } & \sum_j |\beta_j| \leq t\end{aligned}\tag{17}$$

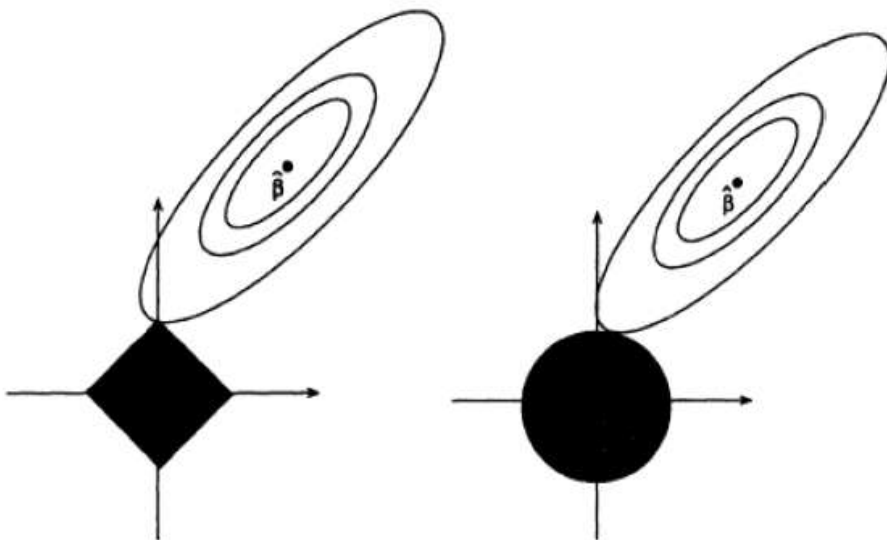
LASSO 추정 방법의 제약조건을 L1-제약조건(L1-constraint)이라고 부르며, 제약조건의 특성상 t 가 줄어들에 따라 중요하지 않은 변수들의 회귀계수들부터 차례로 0을 만들게 된다. 조절 상수 t 는 그 값이 충분히 커질 때 계수에 대한 아무런 제약을 주지 않게 되어 일반적인 회귀모형(regression model)과 같은 결과를 주게 된다.

〈그림 2〉과 〈그림 3〉은 LASSO가 회귀계수를 0으로 축소시키는 것을 보여준다. 그림1-1의 점선은 보통최소제곱에 의한 추정치를, 실선은 LASSO에 의한 추정치를 나타낸다. 그림1-2는 능형회귀의 제약조건과 비교한 것으로 왼쪽의 사각형은 LASSO의 제약 조건 $|\beta_1| + |\beta_2| \leq t$ 을, 오른쪽의 원모양은 능형회귀의 제약조건 $\beta_1^2 + \beta_2^2 \leq t$ 을 의미한다. 등고선 중심에 있는 $\hat{\beta}$ 은 보통최소제곱추정치이고, 타원형등고선(elliptical contours)은 잔차제곱합(residual sum of squares)을 나타낸다. LASSO의 경우 사각형과 닿는 지점에서 추정된 회귀계수가 0이 될 수 있다.그러나 능형회귀의 경우에는 추정된 회귀계수가 0이 될 수 없다는 것을 보여준다.

〈그림 2〉 LASSO의 회귀계수



〈그림 3〉 LASSO의 회귀계수



4.3 조절모수의 추정

LASSO에서 조절모수 t 를 추정하기 위해 교차타당성(Cross-Validation, CV)이 사용된다. 교차타당성은 예측오차(prediction error)를 추정하는데 널리 사용되는 방법이다. 자료를 임의의(random) k 개로 나누어 $k-1$ 개의 자료는 훈련용 자료(training set)로 사용되고, k 개의 자료 중 하나는 검증용 자료(test set)로 사용된다. 즉, $k-1$ 개의 훈련용 자료로 모형을 만들고 검증용 자료로 모형의 타당성을 검정하며, 이것을 k 번 반복하여 평균제곱오차(Mean Squared Error, MSE)를 계산한다. 조절모수 t 를 변화하면서 교차타당성을 실시한 후 가장 작은 평균제곱오차를 가지는 t 를 선택하거나, 1-표준편차 원칙을 적용하여 가장 작은 평균제곱오차의 표준오차만큼의 범위에 포함되는 t 중에서 가장 단순한 모형을 갖는 t 를 선택한다.

4.4 회귀계수의 추정

LASSO는 회귀계수들의 절대값의 합이 조절모수 t 보다 작거나 같도록 제약조건을 주어 회귀계수를 추정하게 된다. 그러나 이것은 미분 가능하지 않기 때문에 비선형 알고리즘(nonlinear algorithm)등을 이용하여 해를 구해야 한다. 본 연구에서는 LASSO의 회귀계수를 추정하기 위한 두 가지 알고리즘을 소개하기로 한다.

LASSO의 회귀추정은 β 에 대해서 가능한 부호(sign)의 조합에 해당되는 2^p 개의 부등식 제약조건(inequality constraints)을 가진 최소제곱에 관한 문제로 표현될 수 있다. 즉, 제약조건을 행렬(matrix)로 나타내면 $G\beta \leq t$ 와 같이 나타낼 수 있는데, $G_{2^p \times p}$ 는 $(\pm 1, \pm 1, \dots, \pm 1)$ 로 구성되어 있다. 이 부등식 제약조건을 다음과 같이 순차적으로(sequentially)해결함으로써 LASSO의 해를 구할 수 있다(Lawson and Hansen, 1974).

$$g(\beta) = \sum_{i=1}^N (y_i - \sum_j \beta_j x_{ij})^2, \quad i = 1, 2, \dots, 2^p \text{에 대해서 } \delta_i \text{는 } (\pm 1, \pm 1, \dots, \pm 1) \text{로 구성}$$

되어 있다고 할 때, 제약조건 $\sum |\beta_j| \leq t$ 은 $\delta_i^T \beta \leq t$ 와 같게 된다. 이 제약조

건을 다음과 같이 E 와 S 로 나타낸다.

$$E = \{i : \delta_i^T \beta = t\}, \quad S = \{i : \delta_i^T \beta < t\} \quad (18)$$

E 는 회귀계수의 절대값의 합이 조절모수 값과 같은 등식제약식집합(equality set)이고, S 는 회귀계수의 절대값의 합이 조절모수 값보다 작은 집합(slackset)을 의미한다.

$i \in E$ 에 대해서 G_E 는 행(row)이 δ_i 인 행렬을 나타내고, 1은 행렬 G_E 의 행의 수와 같은, 1로 이루어진 벡터(vector)를 나타낼 때 LASSO의 회귀계수 추정을 위한 알고리즘을 다음과 같이 정리할 수 있다.

(1) $\delta_{i_0} = \text{sign}(\hat{\beta}^0)$ 인 $E = \{i_0\}$ 로 시작한다. $\hat{\beta}^0$ 은 최소제곱추정치(overall least squares estimate)이다.

(2) $G_E \beta \leq t \mathbf{1}$ 을 제약조건으로 하는 $g(\beta)$ 를 최소화하는 $\hat{\beta}$ 을 찾는다.

(3) $\sum |\hat{\beta}_j| \leq t$ 을 만족하는 경우, 회귀계수 추정을 끝낸다.

$\sum |\hat{\beta}_j| \leq t$ 을 만족하지 않는 경우, 위배된 제약조건 i 를 E 에 추가한다.

(4) $G_E \beta \leq t \mathbf{1}$ 을 제약조건으로 하는 $g(\beta)$ 를 최소화하는 $\hat{\beta}$ 을 찾는다.

위의 과정은 2^p 개의 제약조건을 가지므로 2^p 번의 반복(iteration)이 일어난 후 수렴하게 된다.

두 번째 알고리즘은 David Gay에 의해 제안되었다. β_j 를 $\beta_j^+ - \beta_j^-$ 로 표현하는 방법이다. 즉, 제약조건 $\beta_j^+ \geq 0$, $\beta_j^- \geq 0$ 와 $\sum_{j=1}^p (\beta_j^+ + \beta_j^-) \leq t$ 를 가진 최소

제공이 되어 첫 번째 알고리즘에서의 변수 p 개, 제약조건 2^p 개를 가진 추정이 변수 $2p$ 개, 제약조건 $2p+1$ 개를 가진 추정으로 바뀌게 된다. 따라서 $2p+1$ 번의 반복이 일어난 후 수렴하게 되며, 회귀계수를 추정하기 위한 과정은 첫 번째 알고리즘과 같다.

4.5 LASSO의 로지스틱 회귀모형

종속변수가 이분형 변수일 때에는 LASSO를 로지스틱 회귀모형(Logistic regression model)에 적용할 수 있다. 로지스틱 회귀모형은 종속변수가 두 가지 값만 취하는 질적인 이분형 변수일 때, 확률에 대해 로짓변환(logit transformation)을 한 것이다.

다시 살펴보면, p 개의 독립변수에 대해서 종속변수가 1을 가질 확률을 $P(Y=1|x_1, x_2, \dots, x_p)$ 라고 할 때, 로지스틱 반응함수는 다음과 같이 정리할 수 있다.

$$P(Y=1|x_1, x_2, \dots, x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_p x_p)} \quad (19)$$

로지스틱 반응함수의 계수들을 추정하기 위해 최대우도추정(Maximum Likelihood Estimation, MLE)을 사용하는데, LASSO는 β 에 대해 제약 조건이 있으므로 회귀계수 추정은 다음과 같이 나타낼 수 있다(Duffy and Santner, 1989).

$$\begin{aligned} & \operatorname{argmin} l(\beta) \\ & \text{subject to } \sum_j |\beta_j| \leq t \end{aligned} \quad (20)$$

$l(\beta)$ 는 로그우도함수(log-likelihood function)이고, t 는 조절모수이다. 이 식은 다시 다음과 같이 나타낼 수 있다.

$$l^\lambda(\beta) = l(\beta) - \lambda |\beta| \quad (21)$$

$|\beta| = \sum_j |\beta_j|$ 이며, λ 는 라그랑주 승수(lagrange multiplier)로써 t 와 일대일 대응되는 조절모수이다. LASSO에서 회귀계수의 최대우도 추정은 뉴턴-랩슨

방법(Newton-Raphson method)이나 반복재가중최소제곱(Iteratively Reweighted Least Squares, IRLS)과 같은 반복적인 방법(iterative method)에 의해서 해의 근사값을 구하며, 5.4의 알고리즘을 이용하여 제약조건의 수만큼 반복하는 과정을 거쳐 LASSO의 로지스틱 회귀모형의 해를 추정할 수 있다.

5. VAR모형과 VECM⁶⁾

본 연구는 동태적인 설명을 위하여 시계열 분석 모형인 VAR모형(Vector Auto Regressive Model) 또는 VECM(Vector Error Correction Model)을 이용하여 분석하기로 한다. 먼저, 변수의 불안정성 분석을 위하여 단위근 검정(unit root test)을 시행한다. 단위근을 가지는 불안정한 시계열이라면 평균이나 분산이 시간 가변적이므로 변수들의 과거 자료는 예측에 아무런 도움이 되지 못한다.

만일, 시계열이 불안정함에도 불구하고 일반적인 계량분석을 사용하면 변수 사이에 아무런 상관관계가 없는데도 불구하고 유의성이 높은 것처럼 나타나는 가성적 회귀 현상(spurious regression)이 발생하게 된다. 공적분 관계(cointegration relationship)는 불안정 시계열의 경우에도 이들 사이에 안정적인 시계열을 생성하는 선형결합이 존재하면 회귀분석을 비롯한 전통적 분석이론을 적용할 수 있는 특별한 경우 나타나게 된다. 즉, 적분계열의 특징이 공통적인 추세를 공유하는 경우, 일련의 경제변수들이 단기적으로는 상호괴리를 보이지만 장기적으로는 일정한 관계 유지하게 된다. 본 연구에서는 Johansen(1995)은 최우도 검정을 사용한 공적분 관계를 살펴보도록 한다. 만일, 불안정한 시계열을 결합했을 경우, 안정적인 선형결합이 존재한다면 하나의 변수로 다른 하나의 변수를 예측할 수 있다. 또한, 변수들 사이에 공적분 관계가 존재한다면 이후의 VAR모형이 아니라 VECM으로 추정해야 한다.

VAR이나 VECM을 이용하기 전에 변수의 순서를 정해야 한다. 계수에 대한 간단한 가설 검정을 통해 인과관계를 통하여 변수의 순서를 정하기로 한다. 즉, VAR모형의 계수에 대한 제약을 고려하여, VAR모형에 포함된 변수들 사이의 인과관계 분석이 가능하다. $n \times 1$ 인 y_t 를 $n_1 \times 1$ 인 y_{1t} 와 $n_2 \times 1$ 인 y_{2t} 로 나누면 다음이 성립한다.

6) 김상봉(2012), 대손충당금이 신용카드산업에 미치는 동태적 영향 분석

$$A_p(L)y_t = \epsilon_t \text{에서 } \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \quad (22)$$

여기서 L 은 시차연산자(lag operator)를 나타낸다. $A_{12}(L)=0$ 이면 y_{2t} 의 과거값이 y_{1t} 값을 설명할 수 없어 예측에 도움이 되지 않는다. 따라서 변수 y_{1t} 는 y_{2t} 에 대해 외생적(block-exogenous)이다. $A_{21}(L) \neq 0$ 이면 y_{2t} 의 값이 y_{1t} 의 과거값에 영향을 받게 되어 내생적(endogenous) 변수가 된다.

n 차원 벡터 $y_t = (y_{1t}, \dots, y_{nt})$ 에 대한 VAR모형은 일종의 연립방정식 모형으로 다음과 같이 표시될 수 있다.

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \epsilon_t \quad (23)$$

여기서 A 는 계수행렬이고 이후의 분석에서 y_t 는 변수이다. VAR모형에 포함된 변수가 불안정한 시계열인 경우, 차분한 자료를 대상으로 VAR모형을 추정하여 불안정성을 해결할 수 있다. 그러나 VAR모형에 포함된 변수들 사이에 공적분 관계가 있는 경우, 차분된 변수를 이용하면 두 변수 사이의 장기적인 관계에 대한 정보를 이용하기 위해 다음과 같은 VECM을 추정한다.

$$\Delta y_t = \Pi y_{t-1} + A_1^* \Delta y_{t-1} + A_2^* \Delta y_{t-2} + \dots + A_{p-1}^* \Delta y_{t-p+1} + \epsilon_t \quad (24)$$

여기서 Π 는 VAR모형에 포함된 수준(level) 변수들 사이의 관계에 관한 정보를 반영하고 $\Pi=0$ 이 아닌 경우에 단순 차분한 VAR(p-1)이 된다. 만일, Πy_{t-1} 를 누락하면 모형설정의 오류를 범하게 된다.

충격반응분석(impulse response analysis)을 이용하면 VAR모형이나 VECM의 계수에 대한 정보를 이용하여 모형에 포함된 여러 변수 중, 하나의 변수에 충격이 나타났을 때 시간이 경과함에 따라 다른 구성 변수들에 어떤 영향을 미치는 지 파악할 수 있다. 즉, 변수 y_j 에 대한 충격(impulse) ϵ_j 가 s 기 후의 변수 y_i 에 미치는 영향은 변수 y_i 가 변수 y_j 의 변화에 반응하는 정도로서 (i,j) 번째 원소 $\partial y_{i,t+s} / \partial \epsilon_{jt}$ 로 나타낼 수 있으며 출레스키 분해를 이

용하여 직교 충격반응함수를 구한다.

한 변수의 움직임에 대한 예측오차를 각 변수에 의해서 발생하는 비율로 분할하여 VECM내의 각 변수들의 상대적 중요성의 정도를 파악 가능한 분산분해분석(variance decomposition analysis) 방법은 다음과 같다. 예측오차 $e_{t+s} = y_{t+s} - \hat{y}_{t+s} = \epsilon_{t+s} + C_1 \epsilon_{t+s} + \dots + C_s \epsilon_{t+1}$ 라 하면 예측치 $\hat{y}_{t+s}$ 의 분산은 다음과 같이 표현된다.

$$\begin{aligned} MSE(\hat{y}_{t+s}) &= MSE(e_{t,s}) = E[(y_t - \hat{y}_{t+s})(y_t - \hat{y}_{t+s})'] \\ &= \Omega + C_1 \Omega C_1' + \dots + C_{s-1} \Omega C_{s-1}' \end{aligned} \quad (25)$$

여기서 변수 y_j 에 대한 s 기 후의 예측치 $\hat{y}_{i,t+s}$ 의 예측오차에서 j 번째 변수에 나타나는 충격이 차지하는 비중을 $\omega_{ij}(i, j = 1, \dots, n)$ 이라 하면 다음과 같이 계산된다.

$$\omega_{ij} = \frac{\{Var(u_{jt})[b_j b_j' + C_1 b_j b_j' C_1' + \dots + C_{s-1} b_j b_j' C_{s-1}']\}_{jj}}{\{MSE(\hat{y}_{t+s})\}_{ii}} \quad (26)$$

여기서 $\sum_{j=1}^n \omega_{ij} = 1$, $\{\cdot\}_{ii}$ 는 행렬 $\{\cdot\}$ 의 i 번째 대각원소이다.

IV. 연구결과

4.1 기술통계량

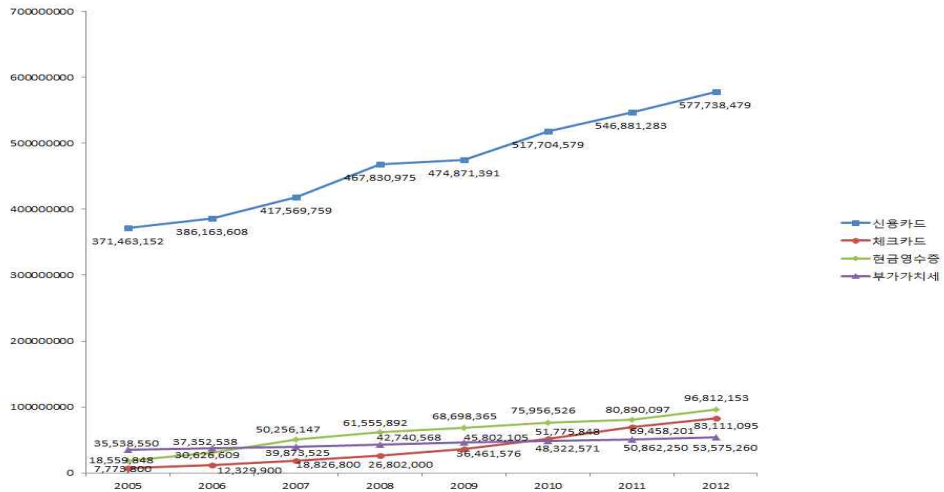
신용카드 이용금액은 평균 117조 5천억 정도이고, 체크카드 이용금액은 평균 9조 5천억 정도로 나타난다. 현금영수증 발급금액은 평균 15조 1천억 정도로 나타났으며 부가가치세는 평균 11조원, 카드이용합계 평균 127조원 정도로 나타난다.

〈표 4〉 표본자료의 변수별 기초통계량

(단위: 백만원)

	N	최소값	최대값	평균	표준편차
GDP	32	211,629,000	276,906,700	247,559,525	19,949,820
신용카드실적	32	88,805,487	148,972,336	117,506,976	18,409,910
체크카드실적	32	1,441,900	21,691,711	9,579,351	6,589,259
현금영수증	32	2,553,521	24,412,766	15,104,864	6,345,763
부가가치세	32	8,739,625	13,510,591	11,064,605	1,550,779
카드이용합계	32	90,247,387	170,664,047	127,086,326	24,821,804

〈그림 4〉 신용카드, 체크카드, 현금영수증, 부가가치세 금액



〈그림 4〉과 〈표 5〉은 신용카드 이용금액, 체크카드 이용금액, 현금영수증 금액, 부가가치세 금액을 나타낸다. 신용카드 이용금액은 2005년도에 371조원에서 2012년 577조원으로 증가하였다. 체크카드 이용금액은 2005년도에 7조원에서 2012년 83조원으로 증가하였다. 현금영수증 금액은 2005년도에 18조원에서 2012년 96조원으로 증가하였다. 부가가치세 금액은 2005년도에 35조원에서 2012년 53조원으로 증가하였다. 〈그림 4〉에서 나타나는 모든 변수들이 증가하는 추세를 보이고 있다.

〈표 5〉 연도별 변수 금액

(단위: 백만)

	신용카드	체크카드	현금영수증	부가가치세
2005년	371,463,152	7,773,800	18,559,848	35,538,550
2006년	386,163,608	12,329,900	30,626,609	37,352,538
2007년	417,569,759	18,826,800	50,256,147	39,873,525
2008년	467,830,975	26,802,000	61,555,892	42,740,568
2009년	474,871,391	36,461,576	68,698,365	45,802,105
2010년	517,704,579	51,775,848	75,956,526	48,322,571
2011년	546,881,283	69,458,201	80,890,097	50,862,250
2012년	577,738,479	83,111,095	96,812,153	53,575,260

〈표 6〉은 연도별 변수들의 증가율을 나타낸다. GDP는 05~06년 5.2%의 증가율을 보이고, 서브프라임 시기인 08~09년도에 0.3%로 낮은 증가율을 보이며 11~12년도에는 2.0%의 증가율을 나타낸다. 신용카드 이용금액은 05~06년 4.0%의 증가율을 보이면서 계속 증가하다가 서브프라임 시기인 08~09년도에 1.5%로 낮은 증가율을 보이며 11~12년도에는 5.6%의 증가율을 나타낸다. 체크카드 이용금액은 05~06년 58.6%의 증가율을 보이며 11~12년도에는 19.7%의 증가율을 나타낸다. 현금영수증 금액은 05~06년 65.0%의 증가율을 보이며 11~12년도에는 19.7%의 증가율을 나타낸다. 부가가치세 금액은 05~06년 5.1%의 증가율을 보이며 11~12년도에는 5.3%의 증가율을 나타낸다. 카드이용 합은 05~06년 5.1%의 증가율을 보이고, 서브프라임 시기인 08~09년도에 3.4%의 낮은 증가율을 보이며 11~12년도에는 7.2%의 증가율을 나타낸다.

〈표 6〉 연도별 변수들의 증가율

(단위: %)

	GDP 증가율	신용카드 증가율	체크카드 증가율	현금영수증 증가율	부가가치세 증가율	카드이용 합증가율
05~06년	5.2	4.0	58.6	65.0	5.1	5.1
06~07년	5.1	8.1	52.7	64.1	6.7	9.5
07~08년	2.3	12.0	42.4	22.5	7.2	13.3
08~09년	0.3	1.5	36.0	11.6	7.2	3.4
09~10년	6.3	9.0	42.0	10.6	5.5	11.4
10~11년	3.7	5.6	34.2	6.5	5.3	8.2
11~12년	2.0	5.6	19.7	19.7	5.3	7.2

〈표 7〉은 소득공제율의 변화에 따른 증가율을 보여주고 있다. 두 표에서 신용카드 이용금액과 현금영수증 이용금액은 소득공제율의 변화 비율이 같다. 신용카드 이용금액은 05~06년에 20%의 소득공제율을 보이고 증가율은 4.0%의 증가율을 나타내고, 06~07년과 07~08년에는 15%로의 소득공제율을 보이고 각각 8.1%, 12.0%을 증가율을 나타낸다. 08~12년에는 20%의 소득공제율을 보이고 11~12년에는 5.6%의 증가율을 보인다. 현금영수증 금액은 05~06년에 65.0%의 증가율을 나타내고, 06~07년과 07~08년에는 각각 64.1%, 22.5%의 증가율을 보인다. 11~12년에는 19.7%의 증가율을 보인다. 체크카드 이용금액은 05~06년에 20%의 소득공제율을 보이고 증가율은 58.6%의 증가율을 나타내고, 06~07년과 07~08년에는 15%로의 소득공제율을 보이고 각각 52.7%, 42.4%을 증가율을 나타낸다. 08~10년에는 20%의 소득공제율을 보이며 11~12년에는 25%의 소득공제율을 보이고 증가율은 19.7%이다. 전체기간에 신용카드 이용금액은 55.5%증가율을 보이고, 현금영수증 이용금액은 421.6%의 증가율을 보인다. 또한 체크카드 이용금액은 969.1%의 증가율을 보인다.

〈표 7〉 소득공제율의 변화에 따른 증가율

(단위: %)

	전체 기간	소득공제율						
		20%	15%		20%			
		05~06년	06~07년	07~08년	08~09년	09~10년	10~11년	11~12년
신용카드	55.5	4.0	8.1	12.0	1.5	9.0	5.6	5.6
현금영수증	421.6	65.0	64.1	22.5	11.6	10.6	6.5	19.7
	전체 기간	20%	15%		20%		25%	
		05~06년	06~07년	07~08년	08~09년	09~10년	10~11년	11~12년
체크카드	969.1	58.6	52.7	42.4	36.0	42.0	34.2	19.7
부가가치세	50.8	5.1	6.7	7.2	7.2	5.5	5.3	5.3

〈표 8〉은 총 급여대비 비율을 변화에 따른 증가율로 보여주고 있다. 신용카드 이용금액, 현금영수증 이용금액, 체크카드 이용금액은 총 급여대비 비율이 같다. 신용카드 이용금액, 현금영수증 이용금액, 체크카드 이용금액은 05~08년에 총 급여대비 비율이 15%로 나타나고 증가율은 신용카드 이용금액의 증가율은 각각 4.0%, 81.%, 12.0%로 나타난다. 현금영수증 금액의 증가율은 각각 65.0%, 64.1%, 22.5%로 나타난다. 또한, 체크카드 이용금액의 증가율은 58.6%, 52.7%, 42.4%로 나타난다. 08~09년도에 총 급여대비 비율은 20%이고 10~12년도의 총 급여대비 비율은 25%이다. 신용카드 이용금액의 증가율은 10~12년 사이에 각각 5.6%, 5.6%였고 현금영수증 금액은 6.5%, 19.7%로 나타난다. 또한 체크카드 이용금액은 각각 34.2%, 19.7%의 증가율을 보였다.

〈표 8〉 총 급여대비 비율의 변화에 따른 증가율

(단위: %)

	전체 기간	총 급여대비 비율						
		15%			20%		25%	
		05~06년	06~07년	07~08년	08~09년	09~10년	10~11년	11~12년
신용카드	55.5	4.0	8.1	12.0	1.5	9.0	5.6	5.6
체크카드	969.1	58.6	52.7	42.4	36.0	42.0	34.2	19.7
현금영수증	421.6	65.0	64.1	22.5	11.6	10.6	6.5	19.7
부가가치세	50.8	50.8	5.1	6.7	7.2	7.2	5.5	5.3

〈표 9〉는 소득공제율 × 총 급여대비 비율의 변화에 따른 증가율을 보여주고 있다. 신용카드 이용금액과 현금영수증 이용금액은 소득공제율 × 총 급여대비 비율의 변화 비율이 같다. 신용카드 이용금액은 05~06년에 300%의 소득공제율 × 총 급여대비 비율을 보이고 증가율은 4.0%의 증가율을 나타내고, 06~08년에는 225%로의 소득공제율 × 총 급여대비 비율을 보이고 각각 8.1%, 12.0%를 증가율을 나타낸다. 08~10년에는 400%의 소득공제율 × 총 급여대비 비율을 보이고 10~12년에는 500%의

소득공제율 × 총 급여대비 비율이 나타나고 각각 5.6%, 5.6%의 증가율을 보인다. 현금영수증 금액은 05~06년에 65.0%의 증가율을 나타내고, 06~07년과 07~08년에는 각각 64.1%, 22.5%의 증가율을 보인다. 11~12년에는 19.7%의 증가율을 보인다. 체크카드 이용금액은 09년까지는 신용카드 이용금액과 현금영수증 금액과 같은 소득공제율 × 총 급여대비 비율을 보인다. 05~06년의 증가율은 58.6%이고 06~08년에는 각각 52.7%, 42.4%을 증가율을 나타낸다. 또한 10~12년에는 625%의 소득공제율 × 총 급여대비 비율을 보이고 증가율은 각각 34.2%, 19.7%이다.

〈표 9〉 소득공제율 × 총 급여대비 비율의 변화에 따른 증가율

(단위: %)

	전체 기간	소득공제율 × 총 급여대비 비율						
		300%	225%		400%		500%	
		05~06년	06~07년	07~08년	08~09년	09~10년	10~11년	11~12년
신용카드	55.5	4.0	8.1	12.0	1.5	9.0	5.6	5.6
현금영수증	421.6	65.0	64.1	22.5	11.6	10.6	6.5	19.7
	전체 기간	300%	225%		400%		625%	
		05~06년	06~07년	07~08년	08~09년	09~10년	10~11년	11~12년
체크카드	969.1	58.6	52.7	42.4	36.0	42.0	34.2	19.7
부가가치세	50.8	50.8	5.1	6.7	7.2	7.2	5.5	5.3

4.2 상관관계 분석

1. 상관관계 분석

<표 10> 부가가치세와의 상관관계

	GDP	신용카드 이용금액	체크카드 이용금액	현금영수증 금액	부가가치세	카드이용 합계
GDP	1	.967**	.960**	.956**	.972**	.972**
신용카드 이용금액	.967**	1	.971**	.963**	.985**	.985**
체크카드 이용금액	.960**	.971**	1	.921**	.979**	.979**
현금영수증 금액	.956**	.963**	.921**	1	.969**	.969**
부가가치세	.972**	.985**	.979**	.969**	1	1.000**
카드이용 합계	.972**	.985**	.979**	.969**	1.000**	1

주: **는 유의수준 1%에서 유의함을 나타낸다.

<표 10>은 부가가치세와의 상관관계가 나타나 있다. 신용카드 이용금액($r=0.967$), 체크카드 이용금액($r=0.960$), 현금영수증 금액($r=0.956$), GDP($r=0.972$), 카드이용합계($r=0.972$)는 모두 부가가치세와 강한 상관관계를 나타낸다.

4.3 다중회귀분석

2005년 1분기부터 2012년 4분기까지의 제도 변화에 따라 더미변수를 포함하여 회귀분석을 실시한다. 2005년 1분기부터 2007년 4분기까지의 신용카드 공제율, 2008년 1분기부터 2011년 4분기까지의 신용카드 공제율과 2012년 1분기에서 2012년 4분기까지의 신용카드 공제율이 각각 다르기 때문에 더미변수를 추가하였다.⁷⁾

분석결과, GDP가 100만원 증가하면 부가가치세 금액은 2만 5천원 증가하고, 신용카드 이용금액이 100만원 증가하면 부가가치세의 금액은 3만 4천원 증가하고, 현금영수증 금액이 100만원 증가하면 부가가치세 금액은 6만원 증가한다. 공제율에 대한 비율은 한 단계 늘 때마다 21만원 부가가치세가 증가한다.

〈표 11〉 다중회귀분석 결과

변수	추정계수(B)	표준오차	t 통계량	확률(p)
(상수)	-180130.064	1545801.413	-.117	.908
GDP	.025	.010	2.446	.021
신용카드	.034	.014	2.405	.023
현금영수증	.060	.027	2.229	.034
공제율 더미	210132.999	145088.874	1.448	.159
$R^2=0.993$ $F=735.829$ $p=0.000$				

7) 더미변수의 기준은 2005년을 0, 2006~2007년을 1, 2008~2012년을 2로 하였다. 그리고 체크카드 이용금액의 비중은 크지 않기 때문에 제외하였다.

〈표 11-1〉은 총 급여액에 대한 비율을 더미로 하여 분석한 결과이다⁸⁾. 분석결과는 GDP가 100만원 증가하면 부가가치세 금액은 1만 3천원 증가하고, 신용카드 이용금액이 100만원 증가하면 부가가치세의 금액은 3만 3천원 증가하고, 현금영수증 금액이 100만원 증가하면 부가가치세 금액은 6만 2천원 증가한다. 총 급여액에 대한 비율은 한 단계 늘 때마다 32만원 부가가치세가 증가한다.

〈표 11-1〉 다중회귀분석 결과

변수	추정계수(B)	표준오차	t 통계량	확률(p)
(상수)	2834409.257	1530738.313	1.852	.075
GDP	.013	.008	1.559	.131
신용카드	.033	.011	3.163	.004
현금영수증	.062	.024	2.598	.015
총급여액 더미	366191.158	119731.102	3.058	.005
$R^2=0.993$ $F=735.829$ $p=0.000$				

8) 총 급여액에 대한 비율은 신용카드 이용금액, 체크카드 이용금액, 현금영수증 금액 모두 같다. 더미변수의 기준은 2005~2007년은 0, 2008~2009년은 1, 2010~2012년은 2로 하였다.

4.4 LASSO모형 분석

〈표 12〉 LASSO모형 분석 결과

Parameter Estimates		
Parameter	DF	Estimate
Intercept	1	475728
현금영수증	1	0.05245
신용카드	1	0.04791
GDP	1	0.01747

LASSO모형을 이용하고 Schwarz Bayesian Criteria(SBC)⁹⁾ 정보량을 기준으로 선택되어진 변수는 신용카드, 체크카드, 현금영수증, GDP등 4개의 변수이다. OLS에서 유의하지 않았던 변수가 유의해짐을 알 수 있다. 회귀 계수를 보면 현금영수증이 100만원 증가할 때 부가가치세는 5만 2천원 증가하고, 신용카드 이용금액이 100만원 증가할 때 부가가치세는 4만 7천 원 증가하고, GDP가 100만원 증가할 때 부가가치세는 1만 7천원 증가한다.

9) $SBC = n \ln \left(\frac{SSE}{n} \right) + p \ln(n)$

n=관찰치 수, p=추정 모수 개수, SSE=오차제곱합

4.5 VECM 분석

1. 단위근 검정

단위근 검정을 위하여 본 연구는 Augment Dickey-Fuller 검정을 사용하였으며 결과를 <표 13>에 표시하였다. 적정 시차(lag)를 찾기 위해 Ng and Perron(1995)은 연속적 시차 적용을 이용하여 적정 시차를 찾는 방법을 시행하였다. 즉, 시차는 최대 2차부터 시작하여 t-통계량이 1.645보다 작으면 차수를 하나씩 감소시켰다. 본 연구에서 사용되는 변수 중, 모든 시계열은 불안정한 시계열이다.

<표 13> 단위근 검정

변수	ADF	
	수준 변수	일차 변수
GDP	-1.110	-4.127*
신용카드 이용금액	2.809	-3.858*
체크카드 이용금액	0.055	-7.028*
현금영수증	-1.162	-5.692*
VAT	0.650	-3.408*
카드이용합계	0.656	-6.919*

주1: ADF의 귀무가설은 ‘단위근을 가진다’임

주2: *는 5%유의수준에서 유의함을 나타내며 ADF의 -2.87임

2. 공적분 검정

〈표 14〉은 공적분에 대한 Johansen 검정 결과를 나타낸다. 〈표 13〉에서 단위근 검정에서 사용된 카드이용합계 변수는 공선성의 문제로 제외시켰다. 공적분 관계 유무에 대한 가설검정은 트레이스(trace) 통계량 및 최대고유치(maximum eigenvalue) 통계량이 우도비검정(likelihood ratio test)를 사용하였다. 트레이스 통계량은 많아야 r 개의 공적분 벡터가 존재한다는 귀무가설과 공적분 벡터가 귀무가설보다 1개 더 존재한다는 대립가설을 검정한다. 최대고유치 통계량 r 개의 공적분 벡터가 존재한다는 귀무가설과 $r+1$ 개의 공적분 벡터가 존재한다는 대립가설이다. 트레이스 검정에서 공적분 개수가 2개보다 작거나 같다는 귀무가설을 기각하지 못하였고 최대고유치 검정에서는 공적분이 1개 존재한다는 귀무가설을 기각하지 못하였다. 따라서 변수 간 공적분이 2개 존재하는 것으로 알 수 있다.

〈표 14〉 공적분에 대한 Johansen 검정

Trace	$r \leq 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r \leq 4$
	83.462*	50.380*	20.466	8.302	2.682
Maximum Eigenvalue	$r=0$	$r=1$	$r=2$	$r=3$	$r=4$
	33.081	29.914*	12.164	5.620	2.682

주1: r 은 공적분 벡터의 수를 나타냄

주2: *는 5% 유의수준에서 유의함을 나타냄

3. 인과관계 검정

VAR모형 또는 VECM을 추정하기 위하여 그랜저 인과관계(Granger causality)에 의해 외생적인 변수부터 순서를 정해주어야 한다. 그랜저 인과관계에 있어 “Y가 X를 Granger Cause 하지 않는다”는 것은 “X가 Y에 대해 외생적이다.”라는 표현과 동일하다. <표 15>은 인과관계에 대한 F 검정결과를 보여주고 있다.

체크카드 이용금액은 신용카드 이용금액을 인과그러나 신용카드 이용금액은 체크카드 이용금액을 인과하지 못하므로 체크카드 이용금액이 신용카드 이용금액보다 외생적이다. 체크카드 이용금액은 부가가치세를 인과하고 부가가치세도 체크카드 이용금액을 인과하므로 체크카드 이용금액과 부가가치세는 상호 외생적이다. GDP는 신용카드 이용금액을 인과그러나 신용카드 이용금액은 GDP를 인과하지 못하므로 GDP가 신용카드 이용금액보다 외생적이다. 또한, GDP는 현금영수증 금액을 인과그러나 현금영수증 금액은 GDP를 인과하지 못하므로 GDP가 현금영수증 금액보다 외생적이다. VAT는 신용카드 이용금액을 인과그러나 신용카드 이용금액은 VAT를 인과하지 못하므로 VAT가 신용카드 이용금액보다 외생적이다. 현금영수증 금액은 신용카드 이용금액을 인과그러나 신용카드 이용금액은 현금영수증 금액을 인과하지 못하므로 현금영수증 금액이 신용카드 이용금액보다 외생적이다. 따라서 최종 목적변수는 부가가치세이기 때문에 GDP, 체크카드 이용금액, 신용카드 이용금액, 현금영수증의 순서대로 외생적이게 된다.

〈표 15〉 그랜저 인과관계

결과변수	원인변수				
	체크카드 이용금액	신용카드 이용금액	GDP	VAT	현금영수증
체크카드 이용금액		0.114	2.228	7.451***	1.504
신용카드 이용금액	8.866***		17.694***	26.421***	6.273**
GDP	1.709	0.002		3.318*	0.837
VAT	5.876**	0.664	0.484		2.115
현금영수증	1.979	1.558	5.440**	3.733*	

주1: *는 10% 유의수준에서 유의함을 나타냄

**는 5% 유의수준에서 유의함을 나타냄

***는 1% 유의수준에서 유의함을 나타냄

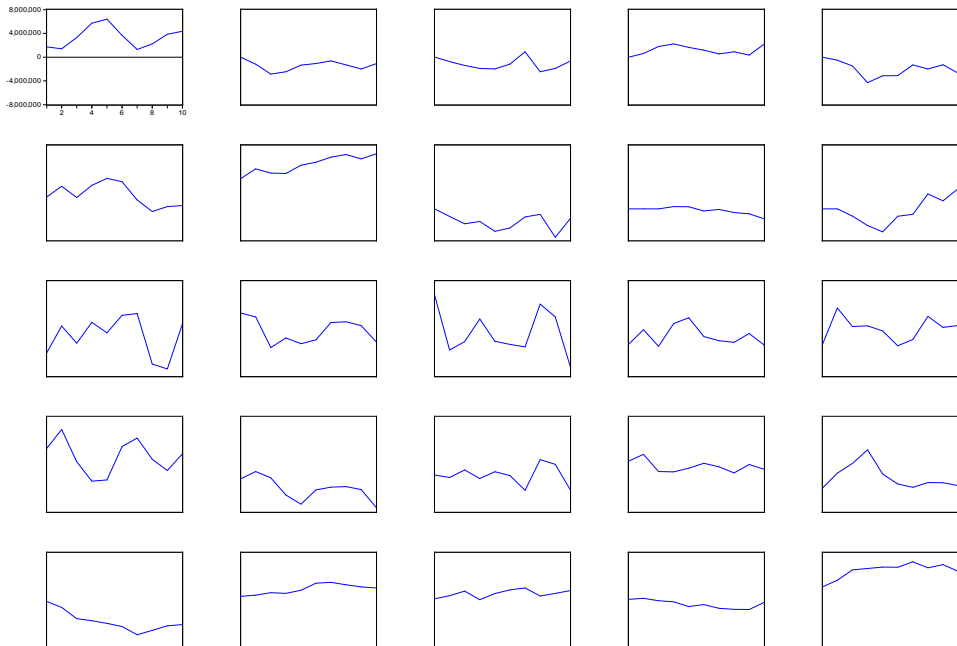
4. VECM

공적분관계가 존재하므로 VECM으로 분석하기로 한다. 충격반응함수의 결과는 〈그림 5〉와 같다. GDP가 1단위 표준편차만큼 상승하면 부가가치세는 음(-)으로 반응한다. 체크카드 이용금액이 1단위 표준편차만큼 상승하면 부가가치세는 양(+)으로 반응한다. 신용카드 이용금액이 1단위 표준편차만큼 상승하면 부가가치세는 양(+)으로 반응한다. 현금영수증 금액이 1단위 표준편차만큼 상승하면 부가가치세는 처음에 양(+)으로 반응했다가 2분기 후에 음(-)으로 전환한다.

〈표 16〉 충격반응함수

	GDP	체크카드 이용금액	신용카드 이용금액	현금영수증 금액	VAT
1	-2466.03	7960.23	3002.06	1869.96	27819.06
2	-15267.40	10814.84	9735.37	4019.11	42124.61
3	-38630.93	15672.57	19048.24	-1288.41	63713.87
4	-42943.41	14320.13	1165.60	-3275.99	66721.22
5	-48802.63	21147.87	13859.30	-13252.18	69678.30
6	-55404.05	35712.73	21717.71	-9296.56	69137.49
7	-72703.04	37540.95	25809.80	-17099.21	80797.55
8	-63478.98	32559.91	8847.87	-19048.35	68192.75
9	-53952.66	28141.19	14352.05	-19620.66	74715.78
10	-51015.33	25730.13	20226.97	-4321.41	60691.69

〈그림 5〉 충격반응함수 결과

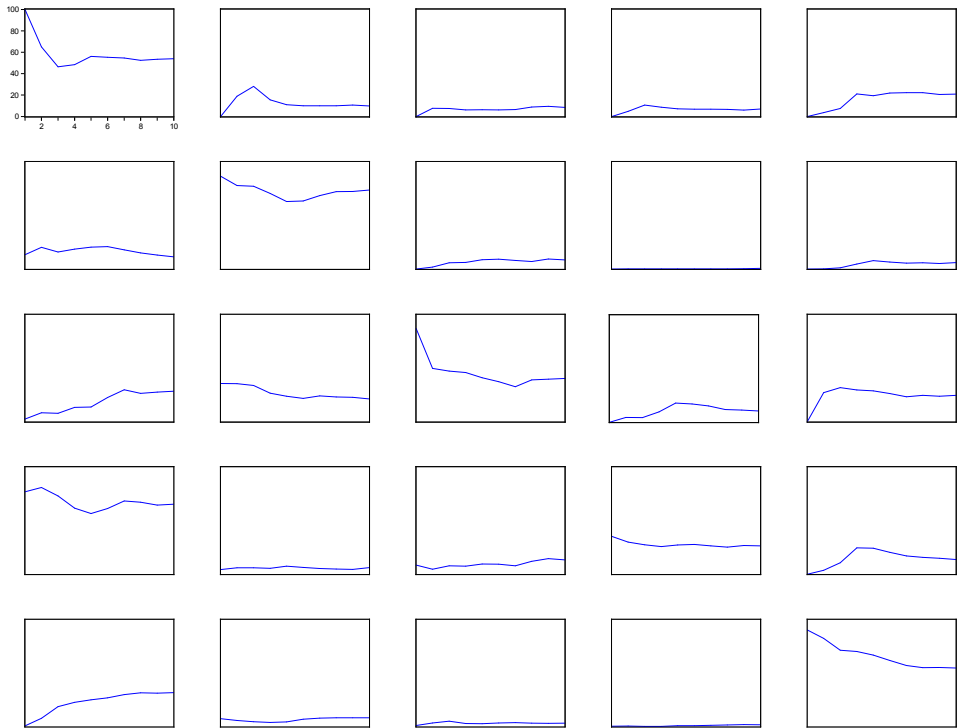


부가가치세의 예측오차는 자기 자신인 부가가치세에 의해 1분기 후 90.4%에서 이후 감소를 하여 10분기 후 54.7%에 의해 설명된다. GDP은 부가가치세의 예측오차에 대해 1분기 후 0.7%에서 시작하여 점차 증가하다가 10분기에는 31.7%로 설명 된다. 체크카드 이용금액은 부가가치세의 예측오차에 대해 1분기 후 7.4%에서 시작하여 증가하다가 10분기에는 8.3%로 설명된다. 신용카드 이용금액은 부가가치세의 예측오차에 대해 1분기 후 1.1%에서 시작하여 점차 증가하여 10분기에는 3.3%가 된다. 현금영수증 금액은 부가가치세의 예측오차에 대해 1분기 후 0.4%에서 시작하여 증가하다가 10분기에는 1.8%가 된다. 따라서 부가가치세의 예측오차의 설명력은 부가가치세 자신, GDP, 체크카드 이용금액, 신용카드 이용금액, 현금영수증 금액의 순서로 높게 나타난다.

〈표 17〉 동태적 영향에 대한 분산분해

	S.E.	GDP	체크카드 이용금액	신용카드 이용금액	현금영수증 금액	VAT
1	29255.01	0.710552	7.403734	1.053022	0.408569	90.42412
2	55599.67	7.736969	5.833296	3.357454	0.63565	82.43663
3	96193.84	18.71259	4.603306	5.042821	0.230297	71.41099
4	125563.8	22.67918	4.002353	2.968259	0.203232	70.14698
5	154330.7	25.01203	4.527063	2.771285	0.871873	66.81775
6	183032.9	26.94532	7.025611	3.378171	0.877848	61.77305
7	218364.1	30.01643	7.891672	3.77047	1.229941	57.09149
8	240549.3	31.69896	8.335278	3.242351	1.640591	55.08282
9	260269.5	31.37453	8.289097	3.073707	1.969704	55.29296
10	274073.1	31.75849	8.356519	3.316553	1.801153	54.76729

〈그림 6〉 동태적 영향에 대한 분산분해



V. 결론

본 연구는 카드이용과 현금영수증의 공제율 변화가 세수확보에 미치는 영향을 분석하였다. 현금영수증 제도가 2005년도부터 시행되어 2005년 1분기부터 2012년 4분기까지의 데이터를 사용하였으며, 각 연도별 공제율이 변화하였기 때문에 더미 변수를 사용하여 제도 변화를 고려하여 분석하였다. 다중회귀분석, VECM모형을 통해서 선행여부를 살펴보았다. 또 LASSO모형으로 다중회귀분석에서 생길 수 있는 다중공선성의 문제를 해결하여 계수를 추정하였다.

변수들 간의 상관관계를 살펴본 결과 신용카드 이용금액, 체크카드 이용금액, 현금영수증 금액, GDP 변수 모두가 종속변수인 부가가치세와 강한 상관성이 있는 것으로 나타난다.

다중회귀분석으로 분석한 결과 GDP, 신용카드 이용금액, 현금영수증 금액이 부가가치세에 유의한 것으로 분석되었다. 공제율 더미와 총 급여액 더미를 포함한 다중회귀 결과, 현금영수증이 제도 변화에 따른 세원확보에 가장 큰 영향을 미치는 것으로 나타난다.

LASSO를 사용하여 모형을 추정한 결과 SBC 정보량을 기준으로 선택되었던 변수는 신용카드, 현금영수증, GDP등 3개의 변수이다. 이는 OLS에서 유의하지 않았던 변수가 유의해짐을 알 수 있다. 분석결과 현금영수증 금액이 100만원 증가할 때 부가가치세가 5만 2천원으로 가장 높은 값을 보였다. VECM 분석결과 10분기 후 체크카드 이용금액이 가장 높은 설명력을 가진다.

세수를 확보하는 방법으로는 여러 가지가 있지만, 그 중에서 서민들에게 체크카드 사용과 현금영수증 발행을 장려하는 것이 세수확보에 효과적임을 본 연구에서는 보여주었다. 체크카드를 사용과 현금영수증을 장려하는 것이 개인이 현금 사용 시 자영업자들이 매출 신고를 누락시키는 결과를 줄일 수 있고 거래의 투명성도 제고할 수 있다.

그러나 본 연구에는 한계점이 존재한다.

데이터를 구하는데 있어서 현금영수증 금액과 부가가치세 금액이 연간

데이터와 반기 데이터로 나와 있었다. 연간 및 반기 데이터로 논문을 설명하기에는 설명력이 부족하다. 그래서 모형의 설명력은 다소 떨어지지만 보간법으로 데이터를 보정하여 사용하였다. 그러나 기존 논문에서 보지 못했던 공제율의 변화를 더미변수로 하여 분석함으로써 서민과 정부가 세수확보를 두고 나아가야 할 방법을 제시한다는 것에 의의를 두고 있다. 위에서 언급한 한계는 관련 데이터 및 연구 자료가 충분히 축적되는 시점에서 다시 검증개선훈 수 있을 것이라 판단된다.

참 고 문 헌

1. 국내문헌

- 김상겸, 박범조, 송재은. (2011). 『신용카드 활성화가 국가재정에 미치는 영향에 관한 연구』. 서울: 여신금융협회.
- 김상봉. (2012). 신용카드 가맹점 수수료 체계 개편에 관한 연구. 『신용카드리뷰』. 6(1), 107-142.
- _____. (2012). 대손충당금이 신용카드산업에 미치는 동태적 영향 분석. 『신용카드리뷰』. 6(2), 93-109.
- 김상헌. (2001). 신용카드 복권제도와 소득공제제도의 경제적 효과. 『공공경제』. 6(2), 111-135.
- 김성은. (2006). 『LASSO를 이용한 강경변 발생 예측 모형 연구』. 연세대학교 대학원 석사학위논문.
- 김재진. (2013). 신용카드 활성화정책의 효과와 사회적 비용. 『신용카드리뷰』. 7(1), 51-79.
- 노아라. (2012). 『고등학생들의 상관관계 개념의 이해도 조사』. 서울대학교 대학원 석사학위논문.
- 문권순. (1997). 『벡터자기회귀(VAR)모형의 이해』. 서울: 통계청.
- 박성배, 홍성경. (2004). 신용카드 활성화가 부가가치세 과세표준과 의제 매입세액공제에 미치는 영향. 『세무학연구』. 21(1), 133-171.
- 박춘래, 이은호. (2004). 신용카드 소득공제 제도의 활용실태와 개선방안. 『전산회계연구』. 2(2), 87-107.
- 반기범, 박명호, 우석진. (2010). 『신용카드가 지하경제 축소에 미치는 영향에 관한 연구』. 서울: 여신금융협회.

- 성명기. (2011). 『신용카드 소득공제의 소득계층별 귀착 및 세수효과와 시사점』. 서울: 대한민국의회.
- 송재은, 박범조, 김상겸. (2012). 신용카드사용이 세수에 미치는 영향에 대한 연구: 소비세를 중심으로. 『유라시아연구』. 9(2), 159-182.
- 송헌재. (2011). 조세부담의 수직적 형성성 관점에서 본 신용카드 소득공제제도. 『재정포럼』. 16(184), 6-18.
- _____, 성명재. (2012). 신용카드 사용액 소득공제 제도의 효과 분석. 『재정학연구』. 5(2), 157-194.
- 시승도, 강기훈. (2010). 일반화가법모형에서 축소방법의 적용연구. 『응용통계연구』. 23(1), 207-218.
- 이만우, 고영일. (2002). 신용카드 지원제도의 세수증대효과 분석. 『공공경제』. 7(2), 165-194.
- 전태영. (2008). 한국의 지하경제추이와 통제방안. 『국제회계연구』. 22, 35-56.
- 정찬우. (2008). 『신용카드업 제도 개선방안』. 서울: 한국금융연구원.
- 황상인. (2006). 벡터오차수정모형(VECM)을 이용한 한국의 외환보유고 분석. 『무역통상학회지』. 5(2), 109-126.
- 국세청, 『국세통계연보』, 각 연도.
- 기획재정부, 『조세지출보고서』, 각 연도.
- 통계청, 『국가통계포털』, 각 연도.
- 한국은행, 『경제통계시스템』, 각 연도.
- 금융감독원, 『금융통계정보시스템』, 각 연도.

2. 국외문헌

Hosmer, D.W. & Lemeshow, S. (2004). 『Applied Logistic Regression』 . Canada: John Wiley & Sons, Inc.

Duffy, D.E. & Santner, T.J. (1989). 『On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models』 . Ithaca, N.Y : Mathematical Sciences Institute, Cornell University.

Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. 『Journal of Statistical Software』 . 33(1), 1-22.

Johansen, S. (1995). 『Likelihood-based Inference in Cointegrated Vector Auto Regressive Models』 . Oxford: Oxford University Press.

Lawson, C. & Hansen, R. (1974). 『Solving Least Squares Problems』 . Englewood Cliffs: Prentice Hall.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. 『Journal of the Royal Statistical Society』 . 58(1), 267-288.

Park, T. & Casella, G. (2008). The Bayesian Lasso. 『Journal of the American Statistical Association』 . 103 (482). 681-686.

ABSTRACT

Effects of Using Amount of Cards on a Source of Taxation

Kim, Si-Eon

Major in Real Estate Economics

Dept. of Economics

The Graduate School, Hansung University

In this paper, we investigate how using amount of credit card in system changes affects a source of taxation with the first quarter of 2005 to the fourth quarter of 2012. We employ Ordinary Least Squares(OLS) with dummy variables. In addition, we implement the LASSO model, which eliminates the multicollinearity in OLS. Lastly, we analyze the effect of related credit card variables on Value Added Tax(VAT). As results, we find that VAT increases highest when cash receipt increases in OLS. In the LASSO model, the cash receipt has the most influence on VAT. In addition, GDP, using amount of debit card, using amount of credit card, using amount of the cash receipt show the order with the high explanatory power.

Keywords : Credit Card, Cash Receipts, Multiple Regression Analysis,
Lasso Analysis, VECM