

<연구노트>

Replication에 관한 소고

Replications in Economics and Business Research

황 민 (Hwang, Min)*

I. 서론

Research Misconduct의 문제는 과학자체만큼이나 오래된 것이지만¹⁾ 최근에 들어서 그 중요성이 다시 재조명되고 있다. Research Misconduct에는 여러 형태가 있지만 많은 경우 데이터나 결과의 조작이 행해지는 경우 중요한 방지 수단으로 꼽히는 것이 연구 결과에 대해 독립적인 replication을 요구하는 것이다. Replication이란 발표된 연구 결과가 제 3자에 의해 동일하게 재현될 수 있는가 하는 것인데 대상과 방법론에 따라 다르게 정의된다²⁾. 특히 근래의 많은 연구처럼 데이터가 커지고, 통계적 방법론이 복잡해지면서 컴퓨터에 대한 의존도가 커져서 Replication을 얼마나 엄밀하게 요구할 것인가에 대한 문제도 쉽지는 않다.

이 소고에서는 현재 부동산학을 포함한 경제-경영학부분에서 Replication의 중요성을 실례를 통해 살펴보고Replication을 위해 Open Data Policy 방침이 학술지들에게 얼마나 받아들여지고 있는지 간단히 정리해 보고자 한다.

Replication이 중요한 이유는 여러 가지가 있다. 첫째는 King(1995)에 의하면 실증연구를 이해하고 평가할 수 있는 유일한 방법은 데이터가 생성되고 분석이 이루어지는 과정을 정확하게 이해하는 것이며 이를 위해서는 논문에 데이터 생성과 분석과정에 대한 정확한 기술이 포함되어 있어야 한다는 것이다. 이 경우 Replication은 정확한 기술의 근거를 제공하는 기준이 된다. 이 기술이 정확해서Replication이 쉬워질수록 논

문에 대한 이해도도 더 높아지며 그 논문을 바탕으로 한 후속연구도 쉽고 효율적으로 이루어질 수 있다. 둘째는 연구하는 과정에서 부주의한 실수나 에러가 발생할 수 있고 그 실수가 연구 결과에 중대한 영향을 미칠 수 있기 때문이다. 더 나아가 불행하게도 연구 과정에 허위와 고의적인 조작이 있는 경우가 없지 않기 때문에 이러한 일들을 방지할 수 있는 중요한 수단 중의 하나가 Replication이다. 이 소고에서는 부주의한 실수에 의해 연구결과가 잘못 보고된 경우(Reinhart and Rogoff, 2010a)와 데이터 조작에 의한 경우(LaCour and Green, 2014)를 살펴보고 그 논문들의 오류를 시정하는 과정에서 Replication의 역할을 점검해 보고자 한다. 셋째의 이유는 저자들이 연구 과정에서 데이터의 취합과 어떤 통계분석 방법론을 어떻게 사용할 것인가에 대한 많은 자의적인 선택이 불가피한데 이 경우 논문 결과의 robustness가 문제되는 경우가 많다(Cochrane, 2015). 특히 저자들이 원하는 결론에 부합하지 않은 통계결과들은 보고하지 않으려고 하는 자연스러운 경향을 고려하면 replication에 의한 robustness의 확보는 중요한 문제이다. 이 부분은 첫째 이유와도 연관이 깊은데 결과의 robustness가 잘 확보된 논문일수록 후속 연구에 더 큰 도움을 줄 수 있기 때문이다.

Replication은 Working Paper와 학술지에 실린 논문 둘 다 중요하지만 학술지의 논문은 더 이상의 수정이 어려운 최종본인 점을 감안하면 학술지의 논문에 더 중요한 것이라고 할 수 있다. 그러면 학술지의 논문의 Replication이 원활하게 잘 이루어지기 위해서 필

* 조지워싱턴대학교 교수, min@gwu.edu

1) 역사적으로 잘 알려진 1900년대초의 과학증거위조 사례로는 1914년 Charles Dawson에 의한 Piltdown Man 사건 참조(Kaplan, 2016).
2) Fuess(1996)은 Replication을 reproduction과 reexamination으로 분류.

요한 조건은 무엇일까?

첫째는 학술지 개재가 결정된 논문들은 논문의 자료들, 특히 데이터와 컴퓨터 프로그램을 의무적으로 제출하게 하는 학술지의 방침이 필요하다. 논문저자가 제출해야 하는 자료는 데이터와 프로그램 이외에도 데이터의 여러 변수에 대한 설명이나 프로그램 실행 시 필요한 지시 사항들을 담은 readme 파일도 포함시키게 되어 있다. 제출해야 하는 데이터의 경우 cleanup된 최종 데이터도 중요하지만 cleanup 되기 전의 원 데이터가 제출되어야 하며 프로그램은 프로그램 버전 정보와 simulation이나 random number generation이 사용된 경우에는 seed 정보도 포함되어야 한다. 경제학의 경우 1980년대에는 Journal of Money, Banking and Credit만이 이 방침을 가진 학술지였지만, 이제는 상당수의 학술지들이 이 방침을 따르고 있다. 주목할 만한 것은 American Economic Review(AER)가 2005년부터 논문 자료 제출을 의무화 했는데 많은 학술지들이 AER의 방침을 모델로 삼아 유사한 방침을 확립했다. Journal of Political Economy(JPE)의 경우 학술지의 홈페이지에 “In an effort to promote consistent standards and requirements among general-interest journals in the field of economics, the Journal of Political Economy has adopted the AER data availability policy. We are grateful to the editors of the American Economic Review for their leadership in constructing this policy.”라고 명시하면서 JPE의 자료 제출 의무방침이 AER의 방침에 따라 만들어졌음을 밝히고 있다. 현재 AER과 JPE 이외에도 Review of Economic Studies, Econometrica, The Economic Journal, Quarterly Journal of Economics와 Review of Economics and Statistics 등의 대표적인 학술지들이 자료 제출 의무방침을 적용하고 있다. 경제학에 비해 경영학 특히 재무쪽이나 부동산 학술지들은 현재 자료 제출 의무방침을 세워놓은 학술지가 거의 없다³⁾. 이는 부분적으로는 재무 연구와 부동산 연구가 경제학에 비해 사적 데이터(proprietary data)에 대한 의존도가 큰 것도 이유지만 그것보다는 재무와 부동산 학계 전체적으로 경제학계에 비해 필

요성에 공감하지 않은 것이라고 보인다. 많은 경제학 학술지들이 사적 데이터를 이용한 논문들은 데이터 제출을 면제해주고 컴퓨터 프로그램은 제출하도록 요구하는 방침을 세워 놓고 있지만 대부분의 재무와 부동산 학술지들은 그런 방침도 가지고 있지 않기 때문이다. 둘째는 replication 연구들이 정상적인 scholarly activity로 인정받아야 한다는 것이다. 이것은 현재 학술 풍토가 “난제해결”에 집중하고 있으며 기존의 연구를 재연하는 노력에는 오히려 “상상력의 부족”이라고 간주할 수 있기 때문이다(Dewald, Thursby and Anderson, 1986). “난제 해결”에 집중하는 학풍은 replication 연구들에 또 다른 장애를 가져오는데 다른 연구자들이 replication 연구들을 인용하는 경우가 많지 않을 것이며 이 경우 학술지의 편집장들은 페이지 당 인용 수에 근거한 순위를 무시할 수 없기 때문에 replication 연구들을 학술지에 실는 것을 꺼릴 수 있다는 것이다(Duvenack, Palmer-Jones and Reed, 2015). 이 문제는 replication 연구들이 어떤 논문을 대상으로 하는가에도 달려 있으며 LaCour and Green(2014) 나 Reinhart and Rogoff(2010a, 2010b)와 같이 영향력이 큰 논문들의 경우에는 replication 연구들이 상당한 파급력을 가질 수도 있다. 하지만 그러한 파급력에도 불구하고 현재의 학계 풍토에서는 학문적으로 높이 평가받고 있는 학술지들은 replication 연구들을 개재하는 유인이 많지 않을 것이다. replication 연구들에게 별로 우호적이지 않은 현재의 학풍은 연구원 개인적인 유인과도 연결되는데 많은 대학이나 정부 민간 연구기관들이 승진과 재임용 과정에 연구 성과를 강조하면서 학술지에 등재하기 위한 경쟁이 더 치열해졌기 때문이다. 결국 이 문제는 replication의 중요성을 학술계 전체적으로 어떻게 평가하는지에 달려 있는 것이며 한편 그 평가는 동시에 replication 연구들이 궁극적으로 어떤 학문적 업적을 남길 수 있는가에 달려 있다고 하겠다.

3) Ivo Welch가 편집장으로 있는 Critical Finance Review(CFR)는 재무 학술지 중 예외이다. Critical Finance Review는 replication에 근거한 critique 논문들을 집중적으로 개재한다. Welch는 학술지 홈페이지에서 CFR은 “views critiques as conversations about economic subjects and academic papers, the same way we view academic seminars as venues to understand ideas and evidence better.”라고 밝혔다.

II. Replication 의 중요성에 대한 실례: LaCour and Green(2014) 와 Reinhart and Rogoff(2010)

1. LaCour and Green(2014)

LaCour and Green(LG)의 2014년 Science 논문은 working paper때부터 많은 주목을 받았던 논문이다. 일반적으로 논란이 많은 주제에 대해서는 사람들이 자신의 믿음이나 편견을 거의 바꾸지 않는다는 것이 정설이었다. 따라서 편견을 감소시키는데 중요한 요인들에 대한 연구들이 있었는데 그중 직접적인 대면에 의한 의사소통이 편견을 줄이는데 훨씬 효과적이라는 주장은 오래 있어 왔다. 하지만 인과 관계를 명확하게 하기 위해 필요한 엄격한 실증연구가 쉽지 않아서 설득력 있는 연구가 많지 않았었는데 LG는 처음으로 그 실증적인 증거를 제시한 것으로 이미 출판 전부터 많이 알려져 왔던 논문이다. 논문의 실증연구의 핵심은 통제된 실험실에서의 실험과 유사하게 실행된 조사에서 동성애자인 선거운동원이 유권자와 20분 정도의 동성결혼의 당위성에 대해 나누는 대화가 유권자의 생각을 바꿀뿐 아니라 유권자의 가족의 생각까지도 바꾼다는 것이었다. 이 연구는 사람들의 믿음이나 편견을 바꿀 수 있는 중요한 요인에 대한 단서를 제공한 것으로 인정되었는데 즉 대화의 과정에서 전달하고자 하는 메시지도 중요하지만 특별히 누가 그 메시지를 전달하는가에도 크게 기여한다는 것이다.

LG의 저자는 당시 UCLA 의 정치학과 박사과정 학생이었던 Michael LaCour와 Columbia 대학의 Donald Green 교수였는데, 특히 Donald Green은 미국 정치학계에서 학문적인 신망이 높은 사람이었다. 이 연구의 데이터 수집과 같은 실증적 분석 부분은 주로 Michael LaCour에 의해 수행되었다. 2008년 캘리포니아주에서 동성결혼을 금지하는 법(Proposition 8)이 통과되고 난후 Proposition 8에 찬성표를 많이 투표한 지역을 중심으로 LA LGBT Center에서 가하호호 방문을 통한 캠페인을 펼쳤다. 이 연구는 LA LGBT Center가 벌였던 캠페인 형식을 빌려 lab experiment와 유사한 환경에서 데이터를 수집하여 인과 관계를 증명할 수 있는 실증연구를 목표로 진행되었다. 먼저 캘리포니아의 선거인 명부에서 추출된

대상자들 중 설문조사에 동의한 4,000여명과 그들이 추천한 5,500여명을 합해 약 9500명의 설문조사 대상자를 확보한 후 일차 설문조사를 실시해서 동성결혼에 대한 태도를 조사했다. 여기서 사용된 설문조사는 Cooperative Campaign Analysis Project(Jackman et al., 2012)의 형식을 따라 수행되었다. 설문 조사 대상자들은 LG논문에 의하면 2008년의 동성결혼 금지법에 찬성했던 남캘리포니아의 주민 중 본인 외에 동거가족 중 한 명 이상의 유권자가 있는 사람들로 구성되었다. 설문조사 대상자들은 첫 설문조사 후 직접 자택을 방문한 선거운동원들과 대화를 나누게 되며 이후 그들의 동성결혼 합법화에 대한 입장의 변화는 후속 설문조사를 통해 기록되었다.

선거운동원들은 동성애자/이성애자 이외에는 여러 측면(나이, 인종, 성별, 선거운동 경험 등)에서 유사한 표본으로 구성되었으며 선거운동원들의 구성은 54%가 동성애자고 나머지는 이성애자로 구성되었다. 설문조사 대상자들은 선거운동원의 동성애자 여부와 선거운동원과 나누는 대화의 주제에 따라 다음의 5개의 그룹으로 무작위 할당되었다.

- (I) 동성애자 선거운동원과 동성결혼 합법화에 대해 대화하는 그룹
- (II) 동성애자 선거운동원과 폐기물 재활용에 대해 대화하는 그룹
- (III) 이성애자 선거운동원과 동성결혼 합법화에 대해 대화하는 그룹
- (IV) 이성애자 선거운동원과 폐기물 재활용에 대해 대화하는 그룹
- (V) 선거운동원과 대화를 하지 않은 그룹

이성애자 선거운동원과 동성애자 선거운동원은 동일한 대본을 사용하지만 인터뷰 중간에 동성애자 선거운동원은 자신이 동성애자임을 밝히고 법에 의해 결혼하고 있지 못하고 있다고 호소하도록 되어있다. 가가호호 방문조사에 의한 인터뷰가 끝난 뒤 6차례(3일후, 12일후, 23일후, 27일후, 45일후, 280일후)에 걸친 온라인 설문조사에 의해 동성결혼에 대한 생각이 변화했는지 1에서 5사이의 수치로 측정했다. 두 가지가 중점적으로 측정되었는데 하나는 선거운동원이 폐기물 재활용에 대해 이야기한 경우와 동성결혼에 대한 이야기를 나누는 경우의 비교 측정(I and III vs. II and IV)이었

고 또 하나는 동성결혼에 대한 이야기를 나누는 경우에 한해서 선거운동원이 동성애자인 경우와 이성애자인 경우의 비교 측정(I vs. III)이었다.

<표 1>의 주요 결과는 다음과 같다. (1) 방문조사 후 사흘이 지난 뒤에 이루어진 설문조사에 의하면 선거운동원이 폐기물 재활용에 대해 이야기했던 경우와 비교해서 동성결혼에 이야기했던 경우(I and III vs. II and IV), 설문조사 대상자들의 동성결혼에 대한 태도가 현격하게 변화했다. 이 결과는 선거운동원이 동성애자이든 이성애자이든 큰 차이가 없었다. 주목할 점은 차이의 크기가 보수적인 미국 중부 주민의 일반적 태도에서 진보적인 동부 주민의 태도의 차이만큼이나 큰 결로 나타났다. 따라서 약 20분의 선거운동원과의 대화는 보수적인 미국 중부주민을 진보적인 동부주민의 수준으로 바꾸어 놓을 수 있다는 것이다. (2) 선거운동원의 성적편향은 방문조사 사흘 뒤에 있던 온라인 설문조사에서는 거의 차이를 가져오지 않은 것으로 나타났다지만 장기적인 효과에 있어서는 큰 차이를 드러냈다. 이성애자인 운동원과 대화를 나누는 경우 90%의 초기 효과가 한 달 만에 사라지는 것으로 나타났지만, 동성애자 운동원과 대화를 나누는 경우에는 한 달 후에도 차이가 없었으며 동성애자 운동원과 폐기물 재활용에 대해 대화를 나누는 경우와 비교해서 9달 이후 오히려 더 강화되는 것으로 나타났다. (3) 본인의 직접적인 접촉이 없이 가족 중의 다른 사람이 동성애자 운동원과 접촉한 경우에는 이성애자 운동원과 접촉한 경우에 비해 동성결혼에 대한 태도의 변화가 단기(사흘)와 장기(아홉달) 두 가

지 경우 다 현격한 차이를 가져오는 것으로 나타났다.

이 논문은 당시 미국사회에서 큰 이슈였던 동성결혼 합법화 논쟁과 결부되어 큰 관심을 끌었는데 학계의 반응은 동성애 조사원과의 약 20분간의 대화가 가져오는 효과가 믿기 어려울 만큼 크다는 것이었다. 컬럼비아 대학의 앤드류 겔만(Andrew Gelman)은 워싱턴 포스트지에 실린 기사에서 LaCour와 Green 의 연구에서 발견한 태도의 변화는 다른 연구에서는 찾아보기 힘든 규모라고 지적하면서 동성결혼에 대한 사회의 관점이 과거 15년간 변하고 있었으면 특히 캘리포니아 주에서는 이미 많은 사람들이 관점을 바꾸고 있는 중이었으며 실험 당시의 사회적인 조건이 그 추세를 빠르게 만들었던 것이라고 추론하였다(Gelman, 2014). 공저자인 Green은 자신도 이런 결과를 기대하지 못했다고 하면서 이 연구를 통해서 자신의 생각이 바뀌었다고 언급했다(Carey, 2014). 스탠포드 대학의 Krosnick 교수는 연구 결과가 기존의 연구와 상당히 다르지만 Green 교수가 공저자이기 때문에 결과를 신뢰한다고 하였다(Singal, 2015). LG는 논문의 결과를 재확인하는 차원에서 1년후 Southern California 의 다른 지역에서 동일한 방식으로 데이터를 수집하여 분석하여 동일한 결과가 얻었다고 보고하였다.

LG 논문은 2014년 12월 Science지에 실렸는데 약 5개월후 David Broockman, Joshua Kalla 와 Peter Aronow는 통계분석(Broockman, Kalla and Aronow, 2015)을 통해 LG에 사용된 데이터는 위조된 것이라고 주장하였다. 문제제기의 계기는 아이러니컬 하게도

<표 1> LaCour and Green 결과

Treatment assigned		Direct contact		Secondhand contact	
		Time since treatment			
Message	Messenger	5 days	9 months	5 days	9 months
Same-sex marriage	Gay	0.46 (0.05)	0.91(0.14)	0.21(0.04)	0.52(0.09)
Recycling	Gay	0.06 (0.05)	0.14(0.11)	0.00(0.02)	0.10(0.07)
Treatment effect for gay messengers		0.41 (0.07)	0.77(0.18)	0.21(0.04)	0.41(0.11)
n		242	184	437	324
Same-sex marriage	Straight	0.50(0.05)	0.23(0.10)	0.03(0.02)	0.17(0.08)
Recycling	Straight	-0.07(0.07)	0.15(0.11)	-0.02(0.02)	0.15(0.09)
Treatment effect for gay messengers		0.56(0.08)	0.07(0.15)	0.04(0.03)	0.02(0.12)
n		225	182	371	282

출처: LaCour and Green (2014)

LG 논문을 확대하는 후속연구에서 시작되었다. 당시 UC Berkeley 박사과정에 있던 David Broockman은 대학원 동료였던 Joshua Kalla와 같이 LG의 후속 연구에 착수하면서 두 가지 난관에 부딪치게 되었다. 첫 번째 난관은 연구 예산의 문제였다. Broockman이 LG에서 사용한 방법에 따라 데이터수집하는 과정에서 그 비용이 생각보다 훨씬 큰 백만불에 가까운 액수라는 것에 의심을 품기 시작했다. 정치학과 대학원생으로서 그 정도의 연구 자금을 지원받는 것은 사실상 거의 불가능한 일이었기 때문이다. 두 번째 난관은 Broockman과 Kalla가 LG에서 사용한 방법에 따라 설문조사를 한 결과는 유사한 경제적 유인에도 불구하고 설문조사 응답률이 1%도 확보하기 어려웠는데 이에 반해 LG의 응답률은 약 12% 였다는 것이다(Singal, 2015). 결국 이 문제를 해결하기 위해 LG의 연구에서 이용했다고 했던 회사(uSamp)에 문의한 결과 그 회사에서는 LG의 연구 참가한 적이 없다는 걸 발견했다. Broockman과 Kalla는 며칠 뒤 다른 경로를 통해서 LaCour 를 위해 uSamp 회사에서 설문조사를 실시했다고 하는 직원은 실제로 존재하지 않는 가공의 인물임을 발견했다. 이에 따라 Broockman과 Kalla는 Yale 대학의 Peter Aronow 교수와 함께 Science 지의 Open Data Policy에 의해 공개되어있던 LG의 데이터를 분석하기 시작했다. 분석 결과, 그들은 LG에 사용된 설문조사에 근거한 데이터는 2012년의 Cooperative Campaign Analysis Project(CCAP) 데이터를 사용해서 만들어낸 가공의 데이터라는 결론을 내리고 Irregularities in LaCour(2014)라는 보고서를 작성해서 분석에 사용된 데이터와 통계 프로그램의 코드와 함께 LG의 공저자인 Green에게 보내게 된다. Green은 이것을 받고 LaCour에게 설명을 요구하고⁴⁾ 만족할만한 대답을 얻지 못하자 보고서를 받은지 이틀 후에 Science 지에 논문취소요청서를 보내어 논문을 공식적으로 취소하였다.

2. Reinhart and Rogoff(2010b)

Reinhart and Rogoff(2010b)는 어떤 의미에서는 LG의 논문보다 더 유명세를 탔던 논문이다. 당시 금융위기와 유로 위기 사태 가운데 정부의 재정정책 확대

의 장기적인 효과와 함의에 대해 미국과 유럽에서 많은 논쟁이 벌어졌는데 당시 하원 예산위원회 위원장이었던 미국 공화당의 폴 라이언이나 재무 장관이었던 영국 보수당의 조지 오스본과 같은 보수적인 정치인들의 긴축 재정정책의 논거로 자주 사용되면서 언론을 통해 일반인에게 널리 알려져 있었던 논문이다.

논문의 핵심은 공적부채와 국내 총생산의 비율과 국내 총생산의 성장률은 부의 관계를 갖고 있다는 것이다. 즉 공적부채와 국내 총생산의 비율을 근거로 해서 4개의 그룹(30% 이하, 30%에서 60%, 60%에서 90% 그리고 90% 이상)으로 나눠서 국내총생산의 성장률을 비교해 보면 첫째로 부채 비율과 성장률은 부의 관계가 있으며 둘째로는 30% 이하, 30%에서 60%, 60%에서 90%의 세 그룹 간에 어느 정도 비례적인 관계가 존재하지만 90% 이상의 경우의 국가들은 다른 세 그룹 간의 선형적인 관계에서보다 훨씬 낮은 성장률을 보인다는 것이다.

<표 2>는 Reinhart와 Rogoff의 NBER Working Paper의 Appendix Table 1을 재인용한 것인데 선진국의 경우 1946년에서 2009년까지 성장률 평균치를 보면 30% 이하, 30%에서 60%, 60%에서 90%의 세 그룹은 4.1%, 2.8%, 2.8% 이지만, 90% 이상의 국가들의 경우 -0.1%를 기록한 것을 볼 수 있다. 중앙값에 근거한 성장률이나 개발도상국의 경우는 하락폭이 조금 작지만 역시 동일한 모습을 보이고 있다.

<표 2> Reinhart and Rogoff(2010a) 결과

Public debt/ GDP ratios	GDP growth rates
≤30%	4.1%
30-60%	2.8%
60-90%	2.8%
>90%	-0.1%

출처: Reinhart and Rogoff(2010a).

이 연구는 당시 경제위기에 대응하는 바람직한 경제정책의 방향에 대한 많은 논쟁에서 확대 재정정책을 반대하고 긴축재정정책을 옹호했던 보수적인 정치가들에게 중요한 근거를 제시해 주었다. 하지만 위의 표는 부채와 성장률의 관계에 대한 인과관계가 아닌 상

4) 여러 신문 보도에 의하면 Green은 원자료(source data)에 대한 access는 없었으며 다만 무기명화(anonymized) 된 데이터만 볼 수 있었던 것으로 보인다.

관관계를 제시해 줄 뿐이며 저자들 자신도 논문에서 두 변수와의 관계를 associated(연관된 것으로만 표현했으며 인과관계로 주장하지는 않았다. 하지만 저자들이 정치인들이 그들의 논문을 인과관계로 해석하는 주장을 했을 때 반박한 적이 없다는 것과 관련해 보면 저자들이 오해의 여지를 남겨두었던 것에는 반론하기 힘들 것 같다.

학문적으로 이 논문이 기여한 것은 시계열적으로나 횡단적으로 기존에 사용했던 데이터보다 훨씬 광범위한 데이터가 이용되었다는 것이다. 기존의 연구가 선진국을 중심으로 20년에서 30년 정도의 데이터에 국한되었다면 이 논문은 선진국과 개발도상국을 포함한 44개국의 약 200년에 걸친 데이터를 사용하여 여러 정치 사회 제도, 환율체제, 화폐 정책 등의 다양한 역사적인 경험을 포함할 수 있었다(Reinhart and Rogoff, 2010). 하지만 오랜기간에 걸친 패널 데이터를 구축한다는 것이 어렵고 불가피하게 데이터 누락이 일어나면서 비대칭적 패널 데이터(unbalanced panel data)가 되기 쉽고, 연구자 재량으로 가중치 계산이 달라지면서 다른 결과를 가져올 수 있는 논란의 여지가 있다

Henderson, Ash and Pollin(2014)은 Reinhart and Rogoff의 데이터를 분석하여 2013년 4월에 “Does High Public Debt Consistently Stifle Economic Growth? A critique of Reinhart and Rogoff”라는 working paper를 발표하여 전 세계적으로 큰 논란을 불러왔다. Henderson, Ash and Pollin(2014)의 결론은 Reinhart and Rogoff의 분석이 상당부분 문제의 소지가 있는 가정과 데이터 편집 그리고 특히 엑셀 계산상의 오류에 근거한 것이며 이 부분들을 수정하고 나면 Reinhart and Rogoff이 내렸던 결론보다 훨씬 약해진다는 것이다.

Henderson, Ash and Pollin(2014)이 제기한 세가지 문제는 다음과 같다.

첫째, Reinhart and Rogoff(2010a, 2010b)의 분석에 있어서 주요한 데이터는 1946-2009년까지 사이의 데이터인데 국가들마다 데이터의 시작이 다르고 어떤 국가들은 중간에 데이터가 없거나 데이터의 오류가 의심되는 경우가 있었다는 점이다. 더군다나 일부국가(오스트레일리아와 뉴질랜드, 캐나다)의 경우2차 대전 직후의 데이터는 분석에 제외되었는데 다른 나라와 비

<표 3> Reinhart and Rogoff (2010a) 오류 수정

	Public debt/ GDP category			
	≤ 30%	30-60%	60-90%	>90%
Recalculated results				
All data with country-year weighting	4.2	3.1	3.2	2.2
Replication elements				
Separate effects of RR calculations				
Spreadsheets error only	4.2	3.0	3.2	1.9
Selective years exclusion only	4.2	3.1	3.2	1.9
Country weights only	4.0	3.0	3.0	1.9
Interactive effects of RR calculations				
Spreadsheets error + selective years exclusion	4.2	3.0	3.2	1.7
Spreadsheets error + country weights	4.1	2.9	3.4	1.4
Selective years exclusion + country weights	4.0	3.0	3.0	0.3
Spreadsheets error + selective years exclusion + country weights	4.1	2.9	3.4	0.0
Spreadsheets error + selective years exclusion + country weights + transcription error	4.1	2.9	3.4	-0.1
RR published results				
RR (2010A, 2010B, Figure 2) (approximated)	3.8	2.9	3.4	-0.1
RR (2010B, Appendix Table 1)	4.1	2.8	2.8	-0.1

출처: Henderson, Ash and Pollin (2014)

교해서 일관성이 없다고 지적하였다.

둘째. Reinhart and Rogoff (2010a, 2010b)가 사용한 엑셀 프로그램에서 사용한 수식의 오류가 발견되었다는 점이다. 이 오류에 의해 오스트레일리아, 오스트리아, 벨기에, 캐나다와 덴마크의 데이터가 통계치 계산에 전부 제외되었다. 제외된 5개 국가가 알파벳상으 가장 앞에 있는 다섯 국가라는 점에서 고의적인 제외가 아니라 우연한 실수 (random error)로 보인다.

셋째. Reinhart and Rogoff(2010a, 2010b)가 국내 총생산 증가율의 평균을 공적부채/국내 총생산 비율에 근거한 네 그룹으로 나눠서 계산할 때 먼저 각국의 평균을 구한 뒤 국가별 평균을 다시 산술평균으로 계산했다는 점이다. 따라서 데이터가 많은 국가와 그렇지 않은 국가들이 같은 가중치를 가지고 평균이 계산되었다. 이 경우 데이터가 작은 국가의 관찰치가 상대적으로 많은 가중치를 차지하게 되는데 Henderson, Ash and Pollin(2014)는 관찰치의 수에 따라 가중치가 다르게 적용되는 것이 일반적인 방법이라고 주장하였다.

Henderson, Ash and Pollin(2014)는 이 세 가지의 오류를 수정한 방법으로 다시 국내총생산 증가율을 공적부채/국내 총생산 비율에 근거한 네 그룹으로 나눠서 계산했는데 그 결과는 <표 3>과 같다. 이 표에 의하면 세 가지 오류가 수정되었을 경우 공적부채/국내 총생산 비율에 근거한 네 그룹의 경제성장률이 조금씩 다르게 계산되지만, 특히 90%이상의 그룹에 차이가 크게 나타난다. Reinhart and Rogoff(2010a, 2010b)의 원 계산에 의하면 -0.1%에 불과하던 경제성장률이 2.2%로 올라가게 되는데, 이 수치는 Reinhart and Rogoff(2010a, 2010b)가 주장했던 “non-linearity”를 옹호하기 어려운 수치이다. 다시 말하면 세 가지 오류가 수정되고 나서 계산된 수치는 국내총생산 증가율과 공적부채/국내 총생산 비율간의 관계가 부의 관계이지만, 가장 공적부채가 높은 국가들이 특별히 더 낮은 경제성장률을 갖고 있다고 주장하기 어렵다는 것이다.

III. 학술지의 논문 자료 제출 방침과 연구

위에서 언급된 두 경우는 처음 논문이 발표되었을 때부터 사회 정치적인 이유로 주목을 받았었기 때문에

Replication에 의한 반박과 논문 취소도 언론으로부터 원 논문들 못지않은 주목과 관심을 끌었다. 하지만 그렇지 않은 많은 경제학이나 경영학 계통의 논문들은 어떨까? 경제학과 경영학에서는 아직 다른 과학 분야에서 만큼은 Replication이 중요한 문제로는 주목받고 있지는 못한 것 같다. 하지만 그에 대한 조직적인 연구는 소수이긴 하지만 이어져 왔는데 그 논문들의 주요 결과를 간단히 소개하기로 한다.

Replication에 대한 비교적 먼저 발표된 연구는 1986년 American Economic Review에 발표된 Dewald, Thursby and Anderson의 “Replication in Empirical Economics: The Journal of Money, Credit and Banking Project(JMCB Project)”이다. The Journal of Money, Credit and Banking(JMCB)은 국립과학재단의 연구기금으로 JMCB Data Storage and Evaluation Project를 시작했는데 프로젝트의 핵심은 논문 저자들에게 논문에 사용된 데이터와 컴퓨터 프로그램을 요구해서 다른 연구자들이 사용할 수 있게 하는 것이었다. 이 프로젝트에서 제출된 데이터와 프로그램으로 Dewald, Thursby and Anderson은 9개의 논문을 대상으로 통계분석 결과에 대한 Replication을 시도했다. 이 중에서 4개의 논문은 거의 동일하거나 유사한 결과를 얻었지만 나머지 논문들에서는 최소한 부분적으로 상당히 다른 결과를 얻었는데, 이 중에는 원저자가 적극적으로 도와줬음에도 불구하고 Replication이 불가능한 경우도 있었다. 또 논문에서 사용된 데이터가 공적 데이터이지만 데이터가 수정되면서 원래의 데이터를 복구할 수 없는 경우도 있었으며 이 경우 수정된 데이터를 사용했을 때 원래 논문의 결과를 도출할 수 없었다.

Dewald, Thursby and Anderson의 연구의 중요한 동인이었던 JMCB Project는 1984년에 끝나고 JMCB는 논문 자료 제출을 더 이상 논문저자들에게 요구하지 않다가 1996년 다시 학술지의 방침으로 채택하게 되었다. McCullough, McGeary and Harrison(2006)은 JMCB에 1996년과 2002년 사이에 실린 논문들 중 실제로 데이터와 프로그램이 제출된 경우, 그리고 데이터와 프로그램이 제출된 경우 논문의 결과와 일치하는 Replication이 가능한 경우를 조사했다. 1996년과 2002년 사이에 실렸던 논문은 전부 266개였는데 그 중 193개의 논문이 JMCB의 방침에 의해 데이터와 프로그램을 제출하게 되어 있었지만 69개의 논문만 데이터

나 혹은 프로그램을 제출했으며, 그중 11개는 데이터만 제출한 것으로 드러났다. 그리고 프로그램이 제출된 경우에도 프로그램 전체가 제출된 것이 아니라 일부만 제출된 경우도 있었고 다수와 프로그램과 다수의 데이터가 동시에 제출되었지만 어떤 프로그램이 어떤 데이터와 결합되어야 하는지에 대한 기술이 없는 경우, 그리고 프로그램의 결과가 너무 길어서 실지로 어떤 부분이 논문에서 사용되었는지 알기 힘든 경우도 있었음을 보고했다. Replication을 할 수 없는 제약이 있었던 7개의 논문을 제외하고 McCullough, McGeary and Harrison(2006)은 62개의 논문 중 14개만 Replication 할 수 있었다.

Glandon의 2011년 American Economic Review (AER) 논문은 2005년부터 AER에 실리는 모든 논문에 의무사항이 된 데이터와 프로그램 제출에 대한 중간평가 형식으로 이루어진 프로젝트에서 나왔다. 프로젝트의 목적은 (1) 논문저자들이 의무사항을 얼마나 준수하는지 (2) 논문 결과들이 얼마나 손쉽게 replication될 수 있는지 (3) 제출된 자료들이 논문 결과를 replication 할 때 논문저자들의 도움을 받지 않고도 가능한지에 대한 조사였다. 이를 위해 6명의 박사학위 과정학생들이 선발되어 각자의 관심사에 따라 2006년 3월부터 2008년 3월까지 실린 논문 중 replication 할 논문을 골라내었는데 포함 36개의 논문(선택 가능한 논문의 29%)이 선택되었다. 39개 중 36개의 논문들이 replication에 도움이 되는 readme 파일이 있었으며 31개의 논문들이 저자의 도움 없이 replication이 가능할 것이라고 생각되었다. 이 중 사적 데이터(proprietary data)를 사용한 11개 논문을 제외한 20개 중 9개의 논문에 대해 실제적인 replication이 시도되었다. 통계분석 결과 수치의 정확성을 강조한 기존의 연구와는 달리 이 논문에서는 replication의 결과가 원 논문에 가까운지를 1에서 5사이의 숫자로 표현했다. 5개의 논문이 4를 받았고 나머지 4개의 논문이 3을 받았다(4는 거의 완벽하게 replication이 가능했던 경우에 주어진 점수였다).

가장 최근에 발표된 Replication에 관한 논문은 Chang and Li(2015)인데 이 논문은 2008년 7월부터 2013년 10월 사이에 출간된 잘 알려진 13개의 경제학 학술지를 대상으로 Replication을 조사했다. 이 논문은 기존의 논문과 방법론상 많은 차이가 있는데 첫째 이전의 모든 연구들이 한 학술지(Journal of Money,

Credit and Banking 이나 American Economic Review)에 국한되었던 데 반해 Chang and Li(2015)는 학술지의 수를 13개로 늘림으로 경제학 전반에 걸쳐 Replication 추세를 고려할 수 있었다. 둘째, 학술지를 13개로 대폭 확대시켰지만 조사 대상 논문은 미국 데이터를 사용하는 논문에 국한시켰고, 특히 사용하는 데이터 중에 국내 총생산(GDP) 자료가 포함되어 있는 경우에 한정시켰다. 셋째, Replication 여부를 가리는데 있어서, 논문에서는 훨씬 완화된 기준을 적용시켰다. 즉 기존의 논문들이 엄격하게 모든 통계치들이 다 부합되어야 하는 걸 요구했지만 이 논문에서는 질적으로(qualitatively) 논문의 주요 결과들을 재현할 수 있으면 replication이 되는 걸로 간주했으며 주요 결과들도 국내 총생산(GDP)에 관련된 결과들로 국한시켰다. 이 기준을 바탕으로 67개의 논문 중 29개의 논문의 Replication을 확인하였다. 학술지 중 데이터와 프로그램 제출이 의무화 되어있는 학술지들은 39개 중 23개의 논문(59%)이 Replication을 확인하였지만 의무화되어 있지 않은 학술지들은 28개 중 6개(21%)만이 Replication을 확인되었다. 저자들은 Replication이 안 되는 이유를 크게 넷으로 나눠 제시했는데 가장 큰 이유는 데이터나 프로그램이 없는 경우(55%), 데이터와 프로그램에 오류가 있는 경우(24%), 사적인 데이터(proprietary data)가 사용된 경우(16%), 그리고 마지막으로 저자(Chang and Li)들에게 제공된 프로그램을 실행할 수 있는 소프트웨어가 없는 경우(5%)였다.

IV. 결론

이 연구노트에서 경제 경영학에서 replication 논의에 대해 살펴보았다. Replication이 중요한 이유는 첫째 Replication은 논문을 이해하고 후속연구를 용이하게 할 수 있는 중요한 기준을 제공한다는 것이다. 둘째는 연구하는 과정에서 발생하기 쉬운 실수나 고의적인 조작에 의해 연구 결과가 잘못된 경우를 교정할 수 있는 필수적인 장치라는 것이다. 세 번째 이유는 학술지에 실린 논문 결과의 robustness를 확인할 수 있는 수단이라는 것이다. 또한 이 노트에서는 언론에서 많이 주목을 받았던 두 논문이 replication에 의해 오류와 조작이 밝혀진 실례를 보여주어 replication이

실제로 얼마나 중요한 역할을 하는지 강조하였다.

Replication의 실제적으로 학술지의 공식 방침으로 인정되어 오는 과정은 점진적이었다. 1980년대 Journal of Money, Credit and Banking이 미국 국립과학 재단의 기금을 지원받아 Replication 이 얼마나 잘 준수되고 있는지와 논문 저자들이 제출한 자료(데이터와 프로그램)을 가지고 얼마나 쉽게 replication을 할 수 있는지 연구를 시작한 이래, 적은 수이지만 꾸준히 연구들이 진행되어 왔다. 최근 American Economic Review의 조사 결과에 의하면 학술지의 논문 자료 제출방침은 비교적 잘 지켜지고 있으며 많은 경우 논문 결과를 replication 하는데 있어서 저자의 도움이 필요하지 않았다는 보고를 종합해 보면 방침의 양과 질에 있어서 많은 발전이 있었던 걸로 보인다. 하지만 replication의 학문적인 필요에 대한 학계 전반에 공감의 형성되어 있지 않으며 이를 위해서는 현재의 학술지 평가 시스템이나 재임용 그리고 승진에 관한 논의가 필요하다.

참고문헌

- Broockman, David, Joshua Kalla and Peter Aronow, 2015. Irregularities in LaCour (2014). Working Paper
- Carey, Benedict, Gay Advocates Can Shift Same-Sex Marriage Views. New York Times. 11 December, 2014
- Chang, Andrew and Phillip Li, Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not." Federal Reserve Board. Working Paper 2015-083, 2015
- Cochrane, John, Secret Data. The Grumpy Economist. <http://johnhcochrane.blogspot.com/2015/12/secret-data.html> . 29 December, 2015
- Dewald, William, Jerry Thursby and Richard Anderson, "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project," *American Economic Review* , Vol. 86, 1986 pp. 587-603
- Duvenack, Maren, Richard Palmer-Jones and W. Reed, "Replication in Economics: A Progress Report," *Econ Journal Watch*, Vol. 12, 2015, pp. 164-191
- Fuess, Scott, "On Replication in Business and Economics Research: The QJBE Case," *Quarterly Journal of Business and Economics*, Vol. 35, 1996, pp. 3-13
- Gelman, Andrew, Pushing at an open door: When can personal stories change minds on gay rights? Washington Post. 19 December, 2014
- Glandon Philip, "Appendix to the Report of the Editor: Report on the American Economic Review Data Availability Compliance Project," *American Economic Review*, Vol. 101, 2011, pp. 695-699
- Henderson, Thomas, Michael Ash and Robert Pollin, "Does High Public Debt Consistently Stifle Economic Growth? A critique of Reinhart and Rogoff," *Cambridge Journal of Economics*, Vol. 38, 2014, pp. 257-279
- Jackman, J. Sides, M. Tesler, L. Vavreck, Common Content Codebook, Advance Release 1, Cooperative Campaign Analysis Project (YouGov, Palo Alto, CA, 2012)
- Kaplan, Sarah, We finally know who forged Piltdown Man, one of science's most notorious hoaxes. Washington Post. 11 August, 2016
- King, Gary, "Replication, Replication," *Political Science & Politics*, Vol. 28, 1995, pp. 444-452
- LaCour, Michael and Donald Green, "When Contact Changes Minds: an Experiment on Transmission of Support for Gay Equality," *Science*, Vol. 346, 2014, pp. 1366-1369
- McCullough, B.D., Kerry McGeary and Teresa Harrison, "Lessons from the JMCB Archive," *Journal of Money, Credit and Banking*, Vol. 38, 2006, pp. 1093-1107
- Reinhart, Carmen and Kenneth Rogoff, "Growth in a Time of Debt," *American Economic Review*, Vol. 100, 2010b, pp. 573-578
- Reinhart, Carmen and Kenneth Rogoff, Growth in a time of Debt. NBER Working Paper w15639, 2010a
- Singal, Jesse, The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud. New York Magazine. May, 2015