Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.26, NO.2, Apr. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.2.523

개인정보보호 대책의 효과 및 인과관계: 기업 및 개인의 개인정보보호 행동에 대한 실증분석 및 그 시사점*

신 일 순†‡
인하대학교

Effects and Causality of Measures for Personal Information: Empirical Studies on Firm and Individual Behaviors and their Implications*

Ilsoon Shin†‡
Inha University

요 약

본 연구에서는 갈수록 심각해지는 개인정보 유·노출 문제에 대해 기업 및 개인의 개인정보와 관련한 행동 및 그 결과를 기존의 연구에 비해 보다 충실하고 풍부하며 이질적인 데이터를 이용하여 실증적으로 살펴봄으로써 개인정보 문제를 좀 더 미시적인 차원에서 이해하고 이를 기초로 해결책을 논의하는 시도를 하였다. 선택편이(selection bias)의 문제를 해결할 수 있는 성향점수 매칭(PSM) 방법을 통한 실증분석의 결과, 우리의 직관과는 달리 기업이 기술적 대책을 적극적으로 수립하고 정보보호를 위한 투자를 많이 할수록 오히려 개인정보 침해사고를 경험할 가능성이 높아지며, 개인들 역시 바이러스 검사를 더 자주할수록 침해사고를 경험할 가능성이 높아지는 것으로 나타났다. 또한 개인정보 침해를 경험한 기업이 개인정보보호 투자를 늘리는 경향이 있으며, 개인정보 침해를 경험한 사람들이 바이러스 검사를 더 적극적으로 할 가능성이 높아지는 역의 인과관계(reverse causality)의 결과가 도출되었다. 이러한 결과를 바탕으로 개인정보보호 대책에 대한 시사점을 논의하였다.

ABSTRACT

This paper studies the empirical relationship between various privacy protection measures and personal information invasion experience of firms and individuals using rich and heterogeneous survey data. By analyzing PSM models. we get the following results: first, the treatment group which have more technical measures and/or IS investment tends to experience more privacy invasion than the control group which have less of them. second, the reverse causality, that is firms and individuals with more experience of privacy invasion tends to take more measure for personal information protection, is found to exist. From these result, we discuss proper privacy policies implications in respects of attackers benefits and individual irrationality.

Keywords: Privacy, Personal Information, Invasion, Empirical Analysis, Propensity Score Matching, Reverse Causality

I. 서 론

최근 몇 년간 국내 여러 기업에서 크고 작은 개인정보 유출 사고가 발생함에 따라 이미 대부분 국민의 개인정보가 노출되었다고 할 정도로 상황이 심각하며, 명의도용, 스미싱(smishing), 피싱(phishing) 등의 2차적 피해를 포함하여 개인적 피해와 사회적 손실이 확대되고 있다. 개인정보 유출 및 이로 인한 피해의 심각성은 국내뿐 아니라 전세계적으로 나타나는 현상이며, 향후에 더욱 증가할 것으로 예측되고

Received(02. 29. 2016), Accepted(03. 12. 2016)
* 이 논문은 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2014S1A5A2A01015999)
† 주저자, ishin@inha.ac.kr
‡ 교신저자, ishin@inha.ac.kr(Corresponding author)

있다. 보안솔루션 및 서비스 업체인 Trustwave는 이미 2012년 전세계 보안 트렌드 보고서에서, 해커들의 주된 공격 목표가 고객의 개인정보(customer record)라고 언급하고 있으며[1], 최근 들어 인터넷 거버넌스 다자간 회의, ITU 전권회의 등의 국제회의에서도 개인정보 유출과 스팸 방지를 위한 사이버 보안 등이 주요 의제로 논의되고 있다.

그런데 기존의 개인정보보호에 대한 대책을 살펴보면 대부분 사후적·대증적이어서, 개인정보 유출 사고가 발생한 이후에 제도적으로 이를 보완하는 방식으로 문제를 해결하고 있는 것으로 판단된다. 개인정보 유출의 문제를 보다 근본적으로 해결하기 위해서는, 개인정보를 다루는 기업들과 이로 인해 피해를 입는 개인들이 개인정보와 관련해서 어떠한 유인 구조(incentive structure)를 가지고 어떻게 행동(behavior)하며, 만일 불합리한 행동을 하고 있다면 그 이유는 무엇이고, 이를 변화시키기 위해 어떠한 대책이 필요한지에 대한 논의가 요구된다. 이러한 논의를 위해서는 미시적인 차원에서 기업과 개인의 실제 데이터를 이용한 실증적(empirical) 연구와 이에 기초한 개인정보보호 대책의 수립이 필요하다.

반면 개인정보 문제를 실증적으로 분석하기 위해서는 다른 분야와 달리 특징적인 어려움이 존재한다. 실증분석을 위해서는 개인정보보호를 위한 각종 대책 및 유·노출 사고에 대한 데이터의 존재가 필수적이지만, 개인정보 유출 및 침해 사실을 외부에 공개하기 꺼려하는 기업 특성 때문에 데이터의 공개가 근본적으로 제한된다는 점이다. 미국 기업을 대상으로 개인정보보호 투자의 효과를 추정하는 모델을 마련한 후 데이터 수집을 시도하였던 한 연구에서도 기업의 불충분한 답변으로 인해 분석을 완료하지 못하였으며, "개인정보보호와 관련된 정확한 정보를 기업의 도움 없이 수집하는 것은 거의 불가능하다"라는 결론을 내리고 있다[2].

그런데 이러한 개인정보 문제에 대한 실증 분석의 어려움과 특이성에도 불구하고, 국내의 경우 상당 기간에 걸쳐 한국인터넷진흥원에서 『개인/기업의 정보보호 실태조사』를 수행하여 왔다[3, 4]. 이 조사는 정보보호에 대한 여러 가지 내용을 설문 방법을 통해 다수의 기업(2010년의 경우 6,000개 이상) 및 개인(2010년의 경우 인터넷 이용자 5,422명)을 대상으로 체계적으로 수집한 데이터로 다른 나라에서는 유사 조사를 쉽게 발견할 수 없다. 따라서 실태조사의 원자료를 활용하면 기업 및 개인 수준에서 개인정보와 관련한 행동 및 그 효과에 대한 실증적인 연구가 가능하며, 특히 기존의 연구와 비교할 때 본 연구의 차별성은 분석을 위해 사용하는 데이터의 풍부함과 충실성에 있을 것으로 판단된다.

이에 본 연구에서는 기업과 개인을 구분하여 개인정보보호 대책이 정보보호 침해 사고에 미치는 효과를 추정하고자 하였다. 먼저 전통적인 실증분석 방법을 이용하여 효과를 추정하였으며, 다음으로 선택 편이(selection bias) 및 내생성(endogeneity) 문제를 해결하기 위해 성향점수 매칭(propensity score matching; PSM) 방법을 사용하여 분석하였다. 특히 정보보호 대책이 침해사고에 미치는 효과뿐 아니라 침해사고가 정보보호 대책 수립에 미치는 역의 인과관계(reverse causality)기 존재하는 지를 분석하고자 하였다. 최종적으로 실증분석의 결과를 활용하여 개인정보보호 대책에 대한 몇 가지 시사점을 논의하였다.

II. 실증연구 데이터 및 방법

2.1 데이터

본 연구는 기업 및 개인 분석으로 구성된다. 먼저 기업 분석에서 사용한 데이터는 [3]의 원자료이다. 한국인터넷진흥원은 민간부문의 정보보호 인식 제고를 위한 각종 정책 활동의 성과지표 산출, 국내 정보보호 수준 측정을 위한 각종 지수 산출, 민간부문의 정보보호 통계자료 제공 등을 목적으로 2001년부터 (2005년 이후에는 매년) 기업부문의 정보보호에 대한 조사를 실시하고 있다. 2010년 조사의 경우 종업원 5인 이상의 사업체 가운데 네트워크에 연결된 컴퓨터를 1대 이상 보유하고 있는 총 30만개의 사업체를 모집단으로 하고, 11개 업종과 4개 규모(종업원 기준)로 할당한 6,529개의 기업을 표본으로 하여 2010년 9월에서 10월까지 조사하였다.

조사 내용은 시점에 따라 약간의 차이는 있지만, 2010년의 경우 정보보호 정책 수립 및 정보보호 조직 구성 현황, 정보보안 환경 평가 및 위험 요소 평가, 임직원 대상 정보보호 교육 실시 및 정보보호에 대한 투자 현황, 정보보호 제품 사용 현황 및 정보보호 업무 수행 방식, IT관련 신규 서비스 도입 및 보안정책 수립, 보안패치 적용 방식 및 정보시스템 사용자 인증 기법, 웹사이트를 통하여 수집한 개인정보의 유/노출 방지를 위한 대응 현황, 개인정보처리시

스템 및 보안서버 구축 현황, 인터넷 침해사고 대응을 위한 활동 및 복구계획 수립/운영 현황, 인터넷 침해사고 및 개인정보 유/노출 피해 경험 및 현황 파악 등으로 이루어져 있다.

한편 개인 분석을 위해 사용한 데이터는 [4]의 원자료이다. 한국인터넷진흥원은 기업 데이터와 함께 인터넷 이용자의 정보보호 현황 및 관련 정책 수립의 기초 자료의 수집을 위해, 1998년부터(2006년 이후에는 매년) 조사를 실시하여 왔는데, 2010년 조사의 경우 지역별, 성별, 연령별로 층화변수화 하여 비례할당한 표본을 이용하여 전국 만 12~59세의 인터넷 이용자 5,422명을 대상으로 2010년 10월에 온라인 조사를 수행한 결과로 만들어진 자료이다.

조사 내용은 정보보호 및 개인정보보호의 중요성 및 역기능에 대한 인식, 정보보호 제품 이용 현황 및 정보보호 대책 수립 여부, 인터넷 침해사고 대응 현황, 데이터 보안 관련 대응 현황, 이메일 및 휴대전화 스팸 방지를 위한 대응 현황, 인터넷 역기능 피해 및 신고 현황, 신규서비스에 대한 정보보호 인식 및 활동 등으로 이루어져 있다.

2.2 실증분석 방법론

통제된 상태에서 무작위성(randomness)이 확보되는 실험실 데이터에서는 선택 편이(selection bias)의 문제를 통제할 수 있지만, 비실험 데이터인 사회과학 데이터에서는 이를 통제하기가 거의 불가능하다. A라는 특징이 B라는 성과에 어떠한 영향을 미쳤는지를 추정하기 위해서는, 다른 모든 조건이 동일하고 A라는 특징에서만 차이가 있는 두 개의 집단 - 흔히 A 특징의 여부에 따라 처방집단(또는 처리군, 분석집단)과 통제집단(또는 대조군)으로 부름 - 을 구분하고, 두 집단의 성과를 비교하여야 한다. 그러나 이렇게 두 개의 집단을 나누는 것은 비실험 데이터에서는 불가능하다.

만일 두 집단이 A 특징 이외에도 미관측된 C라는 특징에서 차이가 존재하는 경우에는 B의 차이가 A에 연유하는지 혹은 미관측된 C에 연유하는지를 알기가 어렵게 된다. 실증경제학에서는 이러한 문제를 '선택 편이(selection bias) 문제'라고 하는데, 선택 편이를 통제하지 않은 상태로 전통적인 실증분석을 하게 되면 내생성 때문에 추정의 편이성이 나타날 가능성이 높다. 최근의 여러 실증경제학 연구들은 이러한 문제를 비교적 간단하게 해결하는 방법들을 제시하고 있다. 본 연구에서는 이러한 방법 중에서 PSM 방법을 이용하여 선택편의 및 이에 따른 기존 방법의 내생성의 문제를 해결하고자 하였다. 즉 PSM 방법을 사용하면 선택 편이의 문제를 해소하여 서로 다른 두 집단의 효과를 분석할 수 있다는 점에서 집단을 구분하는 요인으로부터 효과가 발생하는 인과관계를 추정할 수 있게 된다. PSM 방법은 금융경제학에서 신규 상장기업의 효과를 분석하거나 노동경제학에서 임금효과를 분석하는 등 최근 들어 활발히 응용되고 있는 방법론이다.

PSM 방법에서는 정보보호 대책을 수행한 기업들을 처방집단으로 하여 이들의 성향점수(propensity score)를 추정하고, 특정한 매칭 방법을 사용해 정보보호 대책이 미비된 기업들 중 이와 유사한 성향점수를 가지는 기업(통제집단)들을 비교하여, 정보보호 대책에 따른 침해사고 감소 효과를 분석하게 된다.

PSM 방법을 통한 추정은 통상적으로 세 가지 단계를 통해 이루어진다. 첫 번째 단계에서 성향점수를 추정하고, 두 번째 단계에서는 성향점수를 이용하여 처방집단과 통제집단을 대응(매칭)시키며, 마지막으로 세 번째 단계에서 처방집단과 통제집단 간의 결과변수의 차이를 구하고 이를 통해 분석하고자 하는 효과를 추정한다.

먼저, 첫 번째 단계에서 성향점수는 주어진 결정특성(determining characteristics) 하에서 처방집단(D=1) - 본 연구에서는 정보보호대책을 수행한 기업 및 개인 - 에 속할 조건부 확률(conditional probability)로 정의된다. 즉 X를 관찰된 개별 특성 벡터(a vector of observed individual characteristics)라 하면 성향점수는 다음과 같이 정의된다.

$$\mathrm{P}(\mathrm{X}) \equiv \Pr(D=1 \mid X) = E(D \mid X)$$

성향점수는 처방집단과 통제집단을 구분하는 이항변수인 D를 종속변수로, 개별 특성들(X)을 독립변수로 하는 성향점수 회귀분석식을 구성하고, 이에 대해 프로빗(probit) 또는 로짓(logit) 회귀분석을 실시한다. 회귀분석의 결과 추정치의 예측 확률값(predicted probability value)을 성향점수로 부여한다.

PSM을 통해 추정된 값이 유의미하기 위해서는 다음의 두 가지의 조건이 충족되어야 한다. 첫째, 조건부 독립 가정(conditional independent assumption)과 둘째, 공통영역 가정(common

support assumption)이다. 전자는 관찰되지 않은 어떤 특성도 집단을 구분할 때 영향을 주지 않음을 의미하며, 후자는 처방집단과 통제집단의 확률분포가 동일한 범위 안에 속한다는 것을 의미한다. 이러한 가정을 식으로 표현하면 다음과 같다. 여기서 Y는 처방집단과 통제집단의 성과(결과)를 의미한다.

$Y_0, Y_1 \perp D|X$: 조건부 독립 가정(conditional independent assumption)

$0 < \Pr(D=1|X) < 1$: 공통영역 가정(common support assumption)

이 두 가정들이 충족될 경우, 성향점수를 통해 표본을 처방집단과 통제집단으로 나누는 것은 무작위 실험을 통해 집단을 나누는 것과 같기 때문에[5] 선택적 편의 문제로부터 자유로울 수 있다. 즉, 두 가지 가정이 충족되면 성향점수를 통한 두 집단 간의 분류는 임의성을 띠게 되어 두 집단 간에는 정보보호 대책을 제외한 다른 특성들의 분포는 같게 된다.[1)]

다음으로 두 번째 단계에서는 부여된 성향점수를 이용하여 처방집단과 통제집단간의 매칭을 실시한다. 매칭은 다양한 방식으로 가능한데, 본 연구에서는 처방집단과 통제집단간의 성향점수를 비교하여 가장 가까운 성향점수군끼리 매칭을 하는 Nearest Neighbor Matching을 사용하였다.[2)]

마지막으로 세 번째 단계에서는 처방의 순효과인 ATT(Average treatment on the treated)를 추정한다. 본 연구의 주목적인 정보보호 대책에 따른 침해사건 감소 효과는 PSM을 통해서 정보보호 대책 기업의 실제 침해사건과 만일 이 기업이 정보보호 대책을 수립하지 않았다면 나타났을 침해사건과의 차이를 통해 분석된다. ATT는 다음과 같이 정의된다.

$$\begin{aligned} ATT &= E(Y_1 - Y_0|D=1) \\ &= E(Y_1|D=1) - E(Y_0|D=1) \\ &= E(Y_1|D=1) - E_X[E(Y_0|D=1,X)|D=1] \end{aligned}$$

ATT의 정의 중에서 우변의 두 번째 항은 실제로 관측될 수 없다는 문제가 존재한다. 여기서 조건부 독립성 가정이 성립하면 관측할 수 있는 값(개별 기업의 각종 특성들)으로 X을 통해 처리 여부(D)와 상관없이 Y_0의 결과를 구할 수 있고 이를 가상사실(counter-factual)로 사용할 수 있다. 즉, 조건부 독립성 가정에 의하여 다음의 식이 성립한다.

$$\begin{aligned} &E_X[E(Y_0|D=1,X)|D=1] \\ &= E_X[E(Y_0|D=0,X)|D=1] \end{aligned}$$

결과적으로 ATT는 다음과 같이 표현할 수 있다.

$$\begin{aligned} &ATT \\ &= E(Y_1|D=1) - E_X[E(Y_0|D=0,X)|D=1] \end{aligned}$$

위의 식을 추정하기 위해서 PSM 방법은 다차원의 매칭문제를 1차원 매칭 문제로 축소시키는 장점을 갖고 있다. 즉, 모든 특성 변수들에 대해 기업들을 매칭하는 것이 아니라 성향점수, $P(X)$ 만을 이용하여 매칭한다.[3)]

$$\begin{aligned} ATT = &E(Y_1|D=1) \\ &- E_{P(X)}[E(Y_0|D=1,P(X))|D=1] \end{aligned}$$

PSM 반법은 Monte Carlo 시뮬레이션을 통하여 잘 활용할 경우 상당히 신뢰할만한 처방효과를 추정할 수 있는 것으로 나타났으며[7], 편향을 완화시키는 것으로 나타났다[8]. 그러나 성향점수 매칭법에 대한 논쟁이 아직까지 존재하는 만큼 성향점수 매칭법만을 단독으로 사용하여 추정한 결과를 신뢰하기 보다는 전통적인 회귀분석 방식에 추가로 성향점수 매칭법을 이용하여 활용하는 것이 보다 보수적으로 결과를 추론할 수 있다[9].

1) 본 연구에서는 첫 번째 가정을 충족시키기 위해 성향점수에 대한 균형검증(balancing test)을 하고, 두 번째 가정을 충족시키기 위해 분석집단 관측치들의 성향점수 최댓값보다 크거나 최솟값보다 작은 성향점수를 가지는 통제집단의 관측치들은 표본에서 제외시켰다.

2) 다양한 매칭방법의 장단점에 대해서는 [6]을 참조할 것.

3) 성향점수 매칭법에는 두 가지 단점이 존재한다. 첫째, 강한 가정인 조건부 독립 가정의 무결점 조건의 충족이다. 조건부 독립 가정을 잘못 세울 경우에는 정확한 추정이 어려울 뿐만 아니라 오히려 결과에 편향(bias)이 증가하는 경우가 발생할 수 있다. 둘째, 매칭방법에 대한 논란이다. 앞서 설명과 같이 성향점수 매칭법은 다양한 매칭방법이 존재하지만 이러한 매칭방법 중에서 어떠한 매칭방법이 가장 효율적인지 이론적 배경이 존재하지 않는다.

III. 실증모형

3.1 기업/개인의 실증분석 모형

본 연구는 성향점수 매칭법(PSM)을 이용하여 기업/개인의 개인정보보호 대책이 정보보호 침해 사고에 미치는 순효과, 즉 ATT(Average treatment effect on The treated)를 추정하는 것을 목적으로 하고 있다. 이를 위한 실증 모형은 다음과 같다.

$$Y_i = \alpha + \beta X_i + \gamma PM_i + \epsilon_i$$

좌측항의 종속변수 Y_i는 2010년 기준 개인정보 침해사고를 경험한 기업/개인에 대한 더미변수이며, 우측항의 X_i는 통제변수 벡터, PM_i는 기업/개인의 개인정보보호 대책에 대한 더미변수이며, ϵ_i는 오차항이다. 위 식에서 우리가 관심 있는 추정계수는 γ의 값이다.

3.2 실증분석 결과: 기업 분석

먼저 성향점수 매칭법에 앞서 로짓(logit) 모형을 이용한 실증분석을 실시하였다. 여기서 로짓 분석을 수행한 이유는 종속변수가 침해경험을 나타내는 것으로 이항분포를 가지기 때문이다.

〔표 1〕의 결과를 유의미한 추정계수을 기준으로 살펴보면, 우리의 직관에 부합하는 특징과 그렇지 않은 특징으로 구분할 수 있다. 먼저 개인정보 업무를 직접처리하지 않고 아웃소싱 할수록, 또한 수집한 개인정보를 제3자에게 제공하는 기업일수록 침해 경험이 높은 것은 직관에 부합하는 결과로 볼 수 있다. 그러나 정보보호 투자의 비중이 높을수록, 개인정보를 전담하는 책임자가 있을수록 침해 경험이 오히려 높은 것은 기존의 직관과는 괴리가 있는 결과이다.

이러한 결과가 위에서 설명한 것처럼 다소간의 편향성을 가지도 있을 가능성을 고려하여 다음으로는 PSM 방법을 이용하여 기업의 개인정보보호 대책의 효과를 추정하였다.[4)]

Table 1. Logit Estimation (Firm-level)

variable	explanation	coefficient (standard error)
scale_1	employee 1000+	0.2466 (0.3636)
scale_2	employee 50~1000	-0.5955 (0.4922)
ind_m	manufacture	0.0076 (0.4093)
ind_t	telecommunications	0.4925 (0.3712)
ind_f	Financial	-0.2967 (0.4292)
area_s	Seoul located	0.1457 (0.2887)
itsinv2009	Proportion of IS investment to Total IT investment	0.1026** (0.0457)
itsos	IS outsourcing	0.8199* (0.2940)
pto1	IS chief manager	0.7526** (0.3154)
pte	IS education	0.0166 (0.3260)
pto2	IS organization	0.0514 (0.3133)
ptt1	IS technical measure	-0.0146 (0.0553)
ptt2	Encryption Data Storage	0.8423* (0.4350)
pt3p	third-party data provision	1.9221*** (0.3005)
ptt3	secure server	0.24153 (0.3235)
cons	constant	-5.1181*** (0.4810)
# of obs	Number of obs	1380
Pseudo R2	Pseudo R2	0.2817

(note) *, **, *** denotes P-value is less than 1*. 5%. 10%. same in below.

〔표 2〕에 의하면 정보보호 기술 및 투자만이 t-value로 판단하였을 때 각각 5% 및 1%의 유의수준에서 통계적으로 유의한 결과가 초래되었다.[5)] 그런데 흥미로운 점은 여전히 우리의 직관과는 반대로 기술적 대책을 더 수행하고 정보보호 투자를 증가시킬수록 유의미하게 개인정보 침해의 경험이 늘어난다는 결과가 나타났다는 점이다.

이에 따라 역의 인과관계, 즉 기업의 개인정보보호 대책이 침해를 줄이는 방향이 아니라 반대로 침해를 경험한 기업이 개인정보보호 대책을 마련하는 방향의 효과를 살펴보기 위하여 역시 PSM 방법을 이용하여 역방향의 실증분석을 하였다. 그 결과는 〔표

4) 일반적으로 매칭이 잘 이루어졌는지에 대한 평가는 첫째, 표준화된 편차의 퍼센트 감소, 둘째, 처방집단과 통제집단 간의 매칭 전후의 각 공변량들의 평균값의 차이가 0인 귀무가설에 대한 t값 검정, 셋째, 매칭 전후의 Pseudo-R2값 등을 이용하여 판단한다〔10〕. 구체적으로는 STATA의 pstest 명령어를 이용하게 된다. 본 연구에서는 지면 관계상 이에 대한 추가 보고를 하지 않았으나 모든 PSM 분석에서 해당 조건이 만족하였다.

5) t-value는 그 값이 일반적으로 2.58, 1.96, 1.64 이상인 경우 1%, 5%, 10%의 유의수준을 갖는다.

Table 2. PSM estimation and ATT (Firm-level)

dep.var: invasion experience	treat-ment group (A)	control group (B)	A-B	stand-ard error	T-stat
IS organization	0.0710	0.0964	-0.0254	0.0162	-1.57
IS education	0.0743	0.0642	0.0101	0.0179	0.57
IS technical measure	0.0812	0.0484	0.0328	0.0150	2.19
IS investment	0.0477	0	0.0477	0.0070	6.72

3]과 같다. 유의미한 추정계수만으로 표현하면 침해 경험이 높은 기업이 개인정보보호에 대한 투자를 5% 유의수준에서 증가시키는 것으로 나타난다. 이 결과는 앞서 살펴본 결과와 함께 매우 흥미로운 해석을 가능하게 한다. 즉 기업의 개인정보보호에 대한 투자가 선행적으로 이루어져 침해 방지의 효과를 가지는 것이 아니라, 오히려 개인정보의 침해사고를 경험한 기업이 후행적으로 관련 투자를 늘리는 방식으로 나타나는 것으로 해석된다.

Table 3. PSM estimation and ATT (Firm-level, Reverse Causality)

indep.var: invasion experience	treat-ment group (A)	control group (B)	A-B	stand-ard error	T-stat
IS organization	0.5522	0.5011	0.0511	0.0741	0.69
IS education	0.3134	0.3258	-0.0123	0.0672	-0.18
IS technical measure	0.6912	0.6077	0.0835	0.0699	1.19
IS investment	5.8060	4.7774	1.0285	0.4812	2.14

3.3 실증분석 결과: 개인 분석

개인의 경우도 기업의 경우와 유사한 방식과 순서로 분석하였다. 먼저 이항변수인 개인정보 침해 경험을 종속변수로 하여 로짓분석을 수행한 결과는 다음과 같다.

통제변수를 제외하고, 유의미한 추정계수를 보이는 결과를 해석하면 다음과 같다. 기업의 경우와 유사하게, 개인의 경우에도 주 1회 이상 바이러스 검사하는 사람 및 파일 다운로드 시 바이러스 검사를 실시하는 사람이 더 높은 확률로 개인정보 침해를 경험하는 것으로 나타났다. 이 역시 편향성을 내포하고 있을 가능성이 있기 때문에 PSM 방법을 이용하여 분석하면 다음과 같다.

Table 4. Logit Estimation (Individual-level)

variable	explanation	coefficient (standard error)
age	age	-0.0215*** (0.0046)
sex	sex	-0.0100 (0.0819)
intfreq	frequent use of the Internet	0.5272*** (0.0970)
studum	student	-0.7175*** (0.1220)
hwdum	housewife	-0.2904** (0.1364)
unedum	unemployed	-0.4036** (0.1639)
rec02	recognition of privacy importance	0.3716*** (0.0813)
act01	use of antivirus SW	-0.3535 (0.3346)
act02	realtime antivirus inspection	-0.0281 (0.1434)
act03	more than once a week antivirus inspection	0.2672*** (0.0786)
act04	antivirus inspection when file download	0.3288** (0.1337)
act05	scheduled antivirus inspection	0.1989** (0.0824)
act06	automaic antivirus update	0.2749 (0.2546)
act07	IS protection measure	-0.0568 (0.0784)
cons	constant	-1.6415*** (0.3009)
# of obs	Number of obs	5422
Pseudo R2	Pseudo R2	0.0365

위에서 특이한 점은 바이러스 검사를 수행하는 개인이 그렇지 않은 개인에 비해 다른 조건을 동일하게 만든 이후에도 개인정보 침해의 경험을 할 확률이 높다는 점이다.

앞서와 같이 역의 인과관계, 즉 개인의 개인정보보호 노력이 침해를 줄이는 방향이 아니라 반대로 침해를 경험한 사람들이 노력을 강화하는 방향의 영향을 살펴보기 위하여 역시 PSM 방법을 이용하여 실증분석을 하였다. 그 결과는 [표 6]과 같다. 유의미한 추정계수만으로 표현하면 침해 경험이 높은 개인

Table 5. PSM estimation and ATT (Individual-level)

dep.var: invasion experience	treat-ment group (A)	control group (B)	A-B	stand-ard error	T-stat
use of antivirus SW	0.1574	0.0034	0.1540	0.1142	1.35
more than once a week antivirus inspection	0.2141	0.1727	0.0414	0.0121	3.43

Table 6. PSM estimation and ATT (Individual-level, Reverse Causality)

indep.var: invasion experience	treat-ment group (A)	control group (B)	A-B	stand-ard error	T-stat
use of antivirus SW	0.9778	0.9820	-0.0042	0.0057	-0.73
more than once a week antivirus inspection	0.5048	0.4445	0.0603	0.0181	3.33

이 주 1회 이상 바이러스 검사를 할 가능성이 1% 유의수준에서 높아진다는 것으로 나타나다. 이 결과는 앞서서 살펴본 결과와 함께 매우 흥미로운 해석을 가능하게 한다. 즉 개인의 개인정보보호에 대한 노력이 선행적으로 이루어져 효과를 가지는 것으로 나타나지 않고, 오히려 개인정보의 침해사고를 경험한 사람들이 후행적으로 바이러스 검사를 더 자주 하는 방식으로 관련 노력을 늘리는 방식으로 실증분석이 나타나는 것이다.

IV. 결론: 요약 및 시사점

기업 및 개인의 개인정보보호 대책의 침해사고 감소 효과에 대한 실증분석 결과는 다음과 같이 요약할 수 있다.

첫째, 종속변수를 개인정보 침해 경험 여부를 나타내는 이항변수(binary variable)로 하고, 설명변수를 다양한 기업/개인의 개인정보보호 대책으로, 기타 통제변수를 포함하여 로짓(logit) 분석을 하면 대부분의 개인정보보호 대책이 개인정보 침해 사고를 낮추기 보다는 오히려 이를 높이는 것으로 나타난다.

둘째, 기업/개인들에 미관측되는 고유한 속성이 존재하고 이 고유한 속성에 의하여 개인정보 침해가 영향을 받는다면, 이 결과가 개인정보 노력 및 대책에 것인지 혹은 기업/개인에 고유한 속성에 의한 것인지 구분할 필요가 있다. 이를 위해서는 기업/개인들을 그냥 비교하는 것이 아니라 이 고유한 속성이 유사한 기업들을 매칭하여 비교함으로써 고유한 속성에 의한 영향을 차단한 채 개인정보보호 대책의 영향만을 추출해 내는 것이 필요한 바, 성향점수매칭(PSM) 방법은 가능한 한 유사한 기업들을 매칭하여 비교함으로써 표본선택편의로 인해 발생하는 내생성 문제를 최소화하는 방법이다. PSM 방법을 이용하여 기업/개인의 개인정보보호 노력의 순효과를 추정하면, 우리의 직관과는 달리 기업이 기술적 대책을 적극적으로 수립하고 정보보호를 위한 투자를 더 많이 할수록 오히려 침해사고를 경험할 가능성이 높아지며, 개인들이 바이러스 검사를 더 자주할수록 오히려 침해사고를 경험할 가능성이 높아지는 결과가 도출되었다.

셋째, 역의 인과관계(reverse causality)의 가능성을 추정하기 위해 개인정보 침해 사고 경험을 기준으로 개인 및 기업을 매칭하여 분석하면, 개인정보 침해를 경험한 기업이 개인정보보호 투자를 늘릴 가능성이 높으며, 개인정보 침해를 경험한 사람들이 바이러스 검사를 더 적극적으로 할 가능성이 높아지는 결과가 도출되었다.

우리는 이러한 다소 비직관적인 실증분석의 결과로부터 개인정보보호 이슈 및 정책에 대해 다음과 같은 시사점을 얻을 수 있다.

첫째, 개인정보보호 대책의 강도가 높아짐에도 불구하고 침해가 줄어드는 것이 아니라 상식과는 다르게 더 늘어난다는 결과로부터 우리는 잠재변수(latent variable)의 가능성을 추론해 볼 수 있다. 개인정보보호와 관련하여 생각할 수 있는 잠재변수로는 개인정보를 획득을 목적으로 하는 소위 공격자(attacker)의 상황이다. 본 연구에서 자세히 분석되지는 않았지만, 일반적으로 공격자는 공격의 용이성이나 공격 성공의 확률과 함께 공격 성공에 따른 금전적/비금전적 이득을 기준으로 개인정보에 대한 공격을 결정할 것이다. 만일 어떠한 개인 또는 기업이 개인정보보호 대책을 충분히 수행하여 공격의 성공이 낮아지더라도, 공격이 성공할 경우 이득이 크다면 실제로 빈번한 공격에 따른 침해 경험이 초래될 수 있다. 이 경우 본 연구의 실증분석과 유사한 결과가 초래될 수 있다.

그런데 현재까지 개인정보보호에 대한 대책은 주로 공격자의 입장에서 공격 성공의 확률을 낮추는 방법에만 집중하였고, 공격 성공에 따른 이득을 줄이는 방법에 대해서는 그다지 신중한 고려를 하지 않은 것으로 판단된다. 따라서 본 연구의 실증분석으로부터 얻을 수 있는 하나의 시사점은 개인정보보호의 문제를 해결하기 위해서는 개인 및 기업의 시스템을 보호하는 여러 가지 대책이 필요하지만, 이와 더불어 비록 침해사고가 발생하더라도 침해의 공격자가 그다지 많은 이득을 취하지 못하게 하는 방안을 고려해야 한다는 것이다. 하나의 예를 들면, 개인의 주민번호를 공격자가 가지더라도 이 번호가 영속적이지 않다면 이를 통해 취할 수 있는 이득은 현저히 줄어들 것으

로 예상할 수 있다.

둘째, 침해 경험을 한 그룹의 개인정보보호 대책이 증가한다는 결과로부터 우리는 개인정보보호에 대하 노력이 사전적·예비적 행동이라기보다 사후적 보완책으로 이루어지는 것으로 이해할 수 있다. 이는 마치 홍수 피해 내지 대규모 부실 공사 등을 대비하여 사전적으로 이를 대비하는 노력을 행하는 것이 아니라, 사고가 터진 후에야 이에 대한 대책을 마련하는 것과 상당히 유사한 상황으로 볼 수 있다. 이러한 결과는 또한 기업이나 개인들이 개인정보 침해 사고가 발생하기 이전에는 침해의 부정적인 효과에 대한 중요성을 충분히 고려하지 않고 있는 것으로 해석할 수 있다. 다른 말로 표현하면, 개인정보보호 노력 및 침해의 결과에 대해 경제주체들이 완벽하게 합리적인 판단을 한다기보다는 다소 비합리적으로 행동하고 있는 것으로 볼 수 있다. 여기서 얻을 수 있는 하나의 시사점은 개인정보보호의 문제를 해결하기 위해서는 애초에 개인 및 기업의 합리성을 가정하지 말고, 넛지(nudge) 방식으로 합리성을 유도하는 방법의 모색이 필요하다는 점이다.[6] 예를 들면, 개인의 경우 다양한 정보기기에 개인정보보호를 위한 일종의 디폴트 셋팅(default setting)을 다소 강한 수준으로 강제하는 방법, 개인 스스로의 노력에 기대지 말고 기관, ISP 등이 조직구성원이나 고객의 개인정보 및 패스워드 관리 등에 개입하는 방법 등이 필요할 것으로 판단된다.

References

〔1〕 Trustwave, 2012 Global Security Report, Trustwave, 2012.

〔2〕 A. G. Kotulic and J. G. Clark, "Why there aren't more information security research studies," *Information & Management*, Vol. 41, Issue 5, pp. 597-607, May, 2004.

〔3〕 KISA, 2010 Information Security Survey: Firm-level, KISA, Seoul, 2011.

〔4〕 KISA, 2010 Information Security Survey: Individual-level, KISA, Seoul, 2011.

〔5〕 P. Rosenbaum and D. Rubin, "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, Vol. 79, No. 387, pp.516-524, Sep, 1984.

〔6〕 H. Park and I. Shin, "Estimating the Effect on firm's productivity of Enterprise Software using PSM," *Journal of Economics and Commerce*, Vol. 28, No. 1~2. pp.1-28, Dec, 2014.

〔7〕 P. C. Austin, "Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations," *Biometrical Journal*, Vol. 51, No. 1. pp.171-184, Feb, 2008.

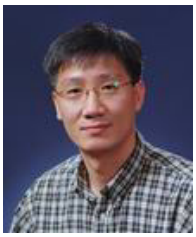
〔8〕 R. H. Dehejia and S. Wahba, "Propensity score matching methods for non-experimental causal studies," *Review of Economics and Statistics*, Vol. 84, No. 1. pp.151-161, Feb, 2002.

〔9〕 R. B. d'Agostino, Jr, "Tutorial in Biostatistics Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group," *Stat Med*, Vol. 17, No. 19, pp.2265-2281, Oct, 1998.

〔10〕 W. Lee, "Propensity score matching and variations on the balancing test," *Empirical Economics*, Vol. 44, No. 1, pp.47-80, Feb, 2013.

6) 넛지(nudge)'는 원래 '팔꿈치로 슬쩍 찌르다' 또는 '주의를 환기시키다'라는 뜻의 단어로, 최근 행동경제학 분야에서 '타인의 선택을 유도하는 부드러운 개입' 또는 '금지와 명령이 아닌 팔꿈치로 옆구리를 툭 치는 것처럼 부드러운 권유로 타인의 바른 선택을 돕는 것'이라는 의미로 사용되고 있다.

〈저자소개〉



신 일 순 (Ilsoon Shin) 정회원
1983년 2월: 서울대학교 경제학과 졸업
1995년 5월: University of Rochester 경제학 박사
2003년 3월~현재: 인하대학교 경제학과 교수
〈관심분야〉 인터넷경제, 기술경제, 정보보호