

정책평가의 계량경제학적 접근



김 주 섭
인적자원정책연구실
책임연구원

경제적인 관점에서 보았을 때, 성공적인 정책평가는 평가를 통해 제공되는 정보의 기대가치가 평가에 소요되는 기대비용보다 같거나 클 때 달성된다.

평가의 대상이 되는 정책이 정부 정책일 경우 정책평가로부터 산출되는 정보의 사회적 기대가치와 사회적 기대비용이 비교되어야 하며, 기업정책의 경우 정보의 기대가치와 기대비용이 기업의 입장에서 비교되어야 할 것이다.

그런데 정책평가가 제공할 수 있는 정보의 기대가치는 '정보의 정확성'과 평가의 대상이 되는 '정책의 규모'에 의해 결정되며, 정책의 규모는 외생적으로 결정되는 것이므로 평가로부터 산출되는 정보의 정확성이야말로 정책평가의 성공을 결정하는 가장 중요한 요소라고 할 수 있다.

1980년대 이후 미국을 중심으로 각종 노동정책과 복지정책에 대한 정책평가에서 계량경제학적인 접근방법이 광범위하게 사용되고 있는 이유는 데

이터 생성에 막대한 비용이 소요됨에도 불구하고, 이러한 방법이 평가의 정확성을 제고하여 평가가 제공하는 정보의 가치를 현저히 높여줌으로써 정부투자의 효율성과 효과성을 신장시킬 수 있기 때문이다.

이 글에서는 정책평가에 적용될 수 있는 몇 가지 계량모형을 소개하고, 이러한 방법들의 장·단점과 정책평가에 적용될 수 있는 계량모형을 추정하기 위해서는 어떠한 종류의 데이터가 필요한가를 살펴보기로 한다.

1. 정책평가의 계량경제학적 문제

계량경제학적 관점에서 볼 때 모든 평가의 문제는 본질적으로 missing data 문제로 귀결된다. 정부의 훈련정책을 예로 들어보자. 편의상 훈련정책의 목표는 훈련참가자의 소득 상승이라고 하자.

Y_{it} 를 시간 k 에 훈련을 받은 개인 i 의 시간 t ($t > k$)에서의 소득이라 하

고, Y_{0t} 를 시간 k 에 훈련을 받지 않았다고 가정할 때의 개인 i 의 시간 t 에서의 소득이라고 하자. 여기서 주목해야 할 점은 Y_{1t} 는 관측될 수 있는(observable) 자료이지만, Y_{0t} 는 개인 i 의 가상상태(counterfactual state)에서의 소득이라는 점이다. 평가자는 정책이 실시되었을 경우의 개인의 소득과 훈련이 실시되지 않았다고 가정하였을 경우의 동일한 개인에 대한 소득의 차이를 추정해야만 할 것이다. 따라서 정책효과는 $\alpha_t = Y_{1t} - Y_{0t}$ 로 표현될 수 있다.

훈련이 실시되었을 경우 Y_{1t} 는 관측이 가능하지만 Y_{0t} 는 관측이 가능하지 않기 때문에 정책효과 α_t 를 추정하기 위해서는 missing된 데이터 Y_{0t} 에 대한 추정을 필요로 한다. 가상상태에서의 소득 Y_{0t} 는 개인의 특성벡터 X_t 에 의해서 결정된다면 다음의 계량경제학적 모형이 설정될 수 있다.

$$Y_{0t} = X_t \beta_t + U_t \quad (1)$$

정책효과 α_t 에 대한 정의를 이용하여 식(1)에서 가상상태의 소득 Y_{0t} 를 소거하면,

$$Y_{1t} = X_t \beta_t + d_t \alpha_t + U_t \quad (2)$$

를 얻는다(여기서 i 가 훈련참가자이면 $d_t=1$, 훈련비참가자이면 $d_t=0$ 으로 정의함).

Y_{1t} , X_t , d_t 는 모두 관측가능한 값들이므로 식(2)는 추정가능한 계량경제학적 모형이 되며, 적절한 데이터와 추정기법을 이용하여 정책효과 α_t 의 추정치 $\hat{\alpha}_t$ 를 얻을 수 있다.

2. 정책평가의 계량경제학적 방법

가. 회귀방법(Regression Approach)

1) covariate post-test 모형

식(2)를 covariate post-test 모형이라 한다. 이 모형을 추정하기 위해서는 모든 i, t 에 대하여 X_t 가 오차항 U_t 에 대해 독립적이라는 가정이 성립되어야만 한다. 즉 $E(U_t | X_t, d_t = 1) = 0$ 가 성립되어야 하며, 이 경우 식(2)는 횡단(cross-sectional) 데이터를 이용하여 추정할 수 있게 된다. 그러나 일반적으로 $E[U_t | X_t, d_t=1] = 0$ 이 성립하는 것은 극히 예외적인 경우인데, 이는 훈련참여가 완전히 무작위적으로 결정되어 훈련참여에 대한 '선택편의'(selection bias)가 존재하지 않을 경우에만 성립할 수 있는 것이다.

선택편의가 존재하는 일반적인 경우에 있어서는 식(2)의 추정방법으로는 α_t 에 대한 올바른 추정치를 산출할 수 없는데, 정책평가의 계량경제학적 방법에 관한 많은 연구가 바로 선택편의의 제거를 위한 방법론 개발에 집중되고 있다.

2) 일차차분 (first difference) 모형

선택편의를 제거하기 위한 방법 중의 하나는 식(2)의 차분방정식을 이용하는 방식이다. 즉, 식(2)에서 $t-1$ 년도의 i 의 소득은

$$Y_{i,t-1} = X_{i,t-1} \beta_{t-1} + U_{i,t-1} \quad (3)$$

으로 추정될 수 있으므로 식(2)에서 식(3)를 빼면 다음과 같은 방정식이 도출된다.

$$Y_{it} - Y_{it-1} = X_{it} \beta_t - X_{it-1} \beta_{t-1} + d_t \alpha_t + U_{it} - U_{it-1}$$

$$\begin{aligned} & (X_{t-1} \beta_t \text{를 빼고 더하면}) \\ & = X_t \beta_t - X_{t-1} \beta_t + X_{t-1} \beta_t - X_{t-1} \beta_{t-1} + d \alpha_t + U_t - U_{t-1} \\ & = (X_t - X_{t-1}) \beta_t + X_{t-1} (\beta_t - \beta_{t-1}) + d \alpha_t + U_t - U_{t-1} \end{aligned} \quad (4)$$

식 (4) 우변의 첫번째 항은 개인 i의 특성벡터 변화가 소득 변화에 미치는 영향을 나타내며, 두 번째 항은 계수의 변화로 인한 소득변화를 나타낸다. 훈련이 장기가 아닌 경우 개인의 특성변화는 영(0)에 가까우므로 $X_t - X_{t-1} \approx 0$ 이며(이 경우 F test를 이용해 검증), 따라서 식 (4)의 첫번째 항은 무시될 수 있다. 즉

$$Y_t - Y_{t-1} = X_{t-1} (\beta_t - \beta_{t-1}) + d \alpha_t + U_t - U_{t-1} \quad (5)$$

이 성립한다. 또한 계수변화 역시, 시간 불변적(time invariant)인 경우 $\beta_t - \beta_{t-1} = 0$ 가 성립되므로 식(5)는 다음과 같이 단순한 형태로 표현된다. (이 경우 χ^2 -test를 이용해 검증을 실시한다.)

$$Y_t - Y_{t-1} = d \alpha_t + U_t - U_{t-1} \quad (6)$$

식(4), (5), (6)이 추정가능하려면,

$E(U_t - U_{t-1} | X_t, d = 1) = 0$ 이 성립되어야만 하는데, 일반적으로 훈련참가에 대한 선택편의가 존재하며, $E(U_t | X_t, d = 1) \neq 0$ 이라 할지라도 참가자와 비참가자의 개인특성 분포의 변화가 동일한 경우 $E(U_t | X_t, d = 1) = 0$ 이 성립한다.

나. 실험적 방법(Experimentation Method)

회귀방법(Regression Approach)이 정책의 결과

에 대한 분포를 가정하여 가상상태(counterfactual state)를 추정하는 방식인데 반하여, 실험적 방법은 정책을 평가의 목적에 맞게 design하여 통계적인 기법을 이용하는 접근방식이다. 실험적 방법은 훈련참여자(실험집단)와 비참여자(비실험집단)의 선택방식과 수집한 자료의 성격에 따라 다음과 같이 재귀적 방법(Reflexive Method), 준 실험적(Quasi-Experimental) 방법 그리고 실험적 방법(Experimental Method)으로 나눌 수 있다.

1) 재귀적 방법(Reflexive Method)

회귀적 방법은 관측되지 않는 가상상태(Counterfactual state)는 정책이 실시되기 전의 상태(Baseline state)와 같은 것이라는 가정에 바탕을 둔다. 앞의 예를 들어 설명하면, 개인 i의 관측되지 않는 가상상태 Y_{it} 를 훈련이 시작되기 전의 개인 i의 소득관측치 $Y_{i,t-1}$ 과 같다고 가정한다. 이 경우 훈련정책의 효과는 식 (6)에 의해

$$E(\alpha_t) = E(Y_t) - E(Y_{t-1}) \quad (7)$$

이 되며 훈련의 기대효과는 훈련에 참가한 개인들의 훈련 후 소득과 훈련전 소득의 차이로 측정된다. 재귀적 방법(Reflexive Method)의 장점은 여타 실험적 방법에 비해 자료수집이 비교적 용이하기 때문에 평가비용이 저렴하다는 것이다. 때문에 정책 이외의 외생변수들의 변화가 크지 않는 안정적(stable)인 상태에서는 유용하게 적용될 수 있는 평가방법일 수 있다. 한편 재귀적 방법(Reflexive Method)의 가장 큰 단점은 평가시 비교집단을 설정하고 있지 않아 외부 환경변화로 인한 종속변수의 변화를 정책효과와 분리할 수 없다는 점이다.

2) 준 실험방법 (Quasi-experimental Method)

준 실험방법은 교정그룹(treatment group)과 대비되는 비교그룹(control group)을 설정하여 두 그룹간의 정책 전·후의 산출물의 변화를 비교함으로써 정책효과를 측정하는 방법이다. 이 경우 비교집단은 교정그룹과 가장 유사한 특성을 가진 집단을 인위적으로 설정해야 한다. 앞서 설명한 재귀적 방법(reflexive method)도 준 실험방법의 하나의 유형으로 볼 수 있는데, 재귀적 방법에서는 교정그룹(treatment group) 자신이 비교그룹(control group)의 역할을 한다. 준 실험방법에서 평가효과 α_t 의 추정치를 구하는 방법은 다음과 같다.

식 (2)에서 교정그룹과 비교그룹에 대해 $d = 1$ 과 $d = 0$ 을 각각 대입하여 기대값 $E(\cdot)$ 를 취하면,

$$\text{교정그룹: } E(Y_t^T) = E(X_t^T \beta_t^T) + E(\alpha_t) \quad (8)$$

$$\text{비교그룹: } E(Y_t^C) = E(X_t^C \beta_t^C) \quad (9)$$

식(8)에서 식(9)를 뺀 후 $E(\alpha_t)$ 에 대해 정리하면,

$$E(\alpha_t) = E(Y_t^T) - E(Y_t^C) - [E(X_t^T \beta_t^T) - E(X_t^C \beta_t^C)]$$

$$E(X_t \beta_t) = E(Y_{0t}) \text{이므로,}$$

$$= [E(Y_t^T) - E(Y_t^C)] - [E(Y_{0t}^T) - E(Y_{0t}^C)] \quad (10)$$

식(10) 우변 첫번째 항은 정책 후 두 그룹간의 평균 산출물(average outcome)의 차이를 나타내며, 두번째 항은 정책이 실시되지 않았다고 가정한 경우의 두 그룹간의 산출물(outcome)의 차이를 나타낸다.

그런데 앞서 설명한 바와 같이, 정책이 실시되지 않았다고 가정한 경우의 교정그룹의 산출물(즉, 가상상태에서의 산출물) Y_{0t} 는 관측될 수 없으므로 식(10)에 의한 $E(\alpha_t)$ 의 도출은 불가능하다. 따라서 일정한 가정 하에서만 $E(\alpha_t)$ 가 구해질 수 있는데, 준 실험방법에서는 정책이 실시되지 않았다고 가정할 경우의 시간 t에서의 두 그룹간의 평균 산출물의 차이와 같다고 가정한다.

따라서 식(10)은 다음과 같이 변형될 수 있다.

$$E(\alpha_t) = \underbrace{[E(Y_t^T) - E(Y_{0t}^T)]}_{\text{(정책실시 후 두 그룹간의 outcome 차이)}} - \underbrace{[E(Y_{0t}^T) - E(Y_{0t}^C)]}_{\text{(정책실시 전 두 그룹간 outcome 차이)}} \quad (10')$$

(정책실시 후 두 그룹간의 outcome 차이) (정책실시 전 두 그룹간 outcome 차이)

준 실험적 방법이 현실적인 유용성을 갖기 위해서는 $E(\alpha_t)$ 를 계산하기 위해 '설정된 가정이 과연 현실성이 있는가'가 입증되어야 한다. 그런데 이러한 가정의 현실성은 설정된 비교그룹이 얼마만큼 교정그룹을 모사(mimic)하는가에 좌우된다. 가장 완전한 비교그룹은 교정그룹을 완전모사(perfect mimic)하는 그룹이 될 것이다. 이 경우는 교정그룹의 가상상태가 비교그룹의 관찰을 통해 파악될 수 있으므로, 정책효과에 대한 완전한 정보를 획득할 수 있을 것이다.

준 실험방법의 장점은 실험방법(experimental method)보다는 비교적 저렴한 비용으로 정확한 평가결과를 측정할 수 있다는 점이다. 특히, 목표 모집단(target population)이 적은 경우에는 교정그룹과 유사한 특성을 지닌 비교집단의 설정이 용이하므로 평가결과의 정확성을 한층 높일 수 있다. 반면에, 이 방법은 목표 모집단

(target population)이 큰 정책의 경우, 비교집단을 구축하기가 힘들 뿐 더러, 선택편의(selection bias)를 적절히 제거할 수 없다는 단점을 지닌다.

3) 실험 방법 (experimental method)

계량경제학의 관점에서 볼 때, 이 실험방법(experimental method)은 가장 정확한 평가결과를 제공하는 정책평가 방법이다. 실험방법은 여타 평가방법과는 달리 교정그룹의 가상상태에 대한 가정을 두지 않고, 교정그룹과 비교그룹의 설정 과정에서 선택편의를 원천적으로 배제하는 방식을 취한다.

이 방법에서는 사회적 실험(social Experiment)을 통해 교정그룹과 비교그룹이 무작위적으로 배정되므로 정책의 목표 모집단(target population)에 속한 모든 개인이 교정그룹과 비교그룹에 속할 확률은 동일하다 - 이러한 과정을 무작위 배정(randomized assign)이라 한다.

이를 수식으로 살펴보면, 식(10)에서 두 그룹은 무작위적으로 배정되었으므로 $E(Y_{it}^T) = E(Y_{it}^C)$ 가 성립되어 우변의 두번째 항이 소거된다. 따라서 실험방법에서의 정책효과의 기대값은 정책 종료 후 두 그룹간 산출물(outcome)의 차이로 측정될 수 있다. 실험방법의 장점은 지금까지 고안된 여타 평가기법에 비해 가장 정확한 평가결과를 제공할 수 있다는 점이다.

이러한 이유 때문에 미국을 중심으로 한 선진국에서는 사회적 실험을 이용한 정책평가의 실행빈도가 점점증하고 있는 추세이다. 특히, 통계기법 및 통계처리에 대한 발달로 사회적 실험에 소요되는 비용이 낮아지고 있어 향후 정책평가의 중요한 기법으로 광범위하게 활용될

전망이다. 그러나 실험방법이 비록 이론적으로는 가장 정확한 평가결과를 산출한다고 할지라도 현실적인 적용에 있어서는 많은 어려움이 따르고 있다.

예를 들어, 실험대상이 되는 교정그룹에 속한 개인들의 '실험'에 대한 심리적 거부감과, 이로 인한 정책효과의 왜곡 등이 문제점으로 지적될 수 있다.

더욱이 John Newman(1994)이 지적한 바와 같이, 후진국일수록 정책평가에 대한 인식의 부족, 전문가의 부족, 정치상황의 불안정으로 인한 정책 담당자의 잦은 교체 등은 장기간에 걸친 정책평가를 더욱 어렵게 만드는 경향이 있다.

3. 결론

Paul Ryan(1999)은 한국에서 실시한 실업자 재취업 교육에 참가한 사람들의 취업률에 대한 자료의 비평을 통해 "이러한 증거들로는 가상상태에 대한 아무런 정보도 얻을 수 없기 때문에 본질적으로 무의미하다"라고 결론 지은 바 있다. 앞에서 언급한 바와 같이, 정책평가의 성패여부는 평가가 제공하는 '정보의 가치'에 의해 결정된다는 것을 상기할 때, 그의 비평은 매우 신랄한 의미를 담고 있다고 보아야 할 것이다.

John Newman 외(1994)에 의하면, 미국 등 선진국을 제외한 개발도상국들에서 정책평가에 대한 계량적 접근(econometric approach)은 매우 드물게 이루어지고 있다.

더욱이 계량적 접근이 수행되고 있는 경우에 있어서도 많은 경우 횡단자료(cross-sectional


data)에 의존하고 있으며, 정책평가의 계량분석에 있어 필수적이라고 할 수 있는 비교그룹(control group 또는 reference group)에 대한 정보가 누락된 횡단자료가 사용되는 경우가 많다. 계량경제학적 관점에서 볼 때, 정책 평가에서 횡단자료의 사용으로는 선택편의(selection bias)를 제거할 수 없기 때문에 왜곡된 평가결과를 가져올 가능성이 많다. Rossi 외(1980)의 연구에서 추적조사(longitudinal survey)를 이용한 고용정책에 대한 횡단자료를 분석한 결과, 추적조사에서는 성공적이라고 평가된 정책이 횡단조사에서는 상반된 결과를 보이고 있음을 밝혔다. 이러한 상반된 정책평가 결과가 도출되는 이유는 횡단자료에 의한 분석의 경우 평가자가 관심을 가지는 산출물에 영향을 주는 중요한 설명변수들이 누락될 수 있기 때문이다.

비록 비교집단이 포함된 횡단자료의 경우라 할지라도 두 집단간의 선택편의를 제거할 방법이 없으므로 평가결과의 부정확성을 초래할 가능성이 매우 높다.

정책평가의 계량경제학적 분석을 위해 구축될 데이터가 가져야 하는 필수적인 요소들은 다음과 같다.

첫째, 정책 전 데이터(Baseline data)가 구축되어야만 한다. 정책이 시행 중이거나 종료된 후에 구축된 자료만으로는 정책으로 인한 종속변수의 변화를 측정할 수 없다. 데이터 구축의 시점이 정책 시작 후라 할지라도 회상자료(retrospect data)를 이용한 데이터 구축이 가능하지만, 이 경우에도 측정오류(measurement error)를 최소화하는 방법이 강구되어야 할 것이다.

둘째, 교정그룹(treatment group)과 유사한 특성을 가지는 비교그룹(control group)이 구축되어야 한다. 정책의 영향을 받는 하나의 그룹에 대한 자료로는 그 그룹의 산출물의 변화가 정책에 의한 것인지 또는 여타 외부환경에 의한 것인지를 식별할 수 없게 된다.

따라서 이러한 자료는 정책효과의 상향·하향 편(upward or downward bias)을 초래하기 쉽다. 

참고문헌

- Peter H. Rossi & Howard E. Freeman(1992). Evaluation : A Systematic Approach, SAGE Publication.
- Jooseop Kim et.al(1999). Can Private School Subsidies Increase Enrollment for the Poor ?, The World Bank Economic Review
- Jean Baldwin Grossman(1994). Evaluating Social Program: Principles and U.S. Experiences, World Bank Research Observer.
- John Newman, Laura Rawlings & Paul Gertler(1994). Using Randomized Control Designs in Evaluating Social sector Programs in Developing Countries, The World Bank Research Observer.
- 한국직업능력개발원(1999). 실직자 재취업 훈련평가에 대한 국제적 경험 및 한국적 시사점