

계량 경제학 모형의 부분 식별에 관한 연구⁽¹⁾

이 석 배

계량 경제학 모형의 부분 식별(partial identification of econometric models)은 계량 모형의 제약 하에서 모형의 모수가 부분적으로만 식별되는 경우를 의미한다. 이러한 상황에서는 통상적인 추론 및 검정 방법과는 다른 방법을 요구한다. 본 논문의 목적은 계량 경제학 모형의 부분 식별에 대한 최근의 계량 경제학 문헌을 선별적으로 소개하고 경제학에의 적용 사례를 통해 새로운 추론 및 검정 방법의 유용성을 살펴보는 데 있다.

주제어: 교집합 범위, 부분 식별, 식별 영역

1. 머리말

계량 경제학 모형의 부분 식별(partial identification of econometric models)은 계량 모형의 제약 하에서 모형의 모수(parameter)가 부분적으로만 식별되는 경우를 의미한다. 전통적인 모형 식별에 관한 입장은 모형의 모수가 모수 집합(parameter space)에서 한 점(single point)으로 정해지거나 그렇지 않은 경우로만 국한되어 왔었다. 그러나 Charles Manski가 선도하여 그의 저서들[Manski(2003, 2007)]에서 집대성된 부분 식별의 이론에서는 모수가 모수 집합의 순 부분집합(strict subset)이 되는 경우에도 유용한 정보를 얻을 수 있음을 보여 주었다. 다시 말해, 계량 모형에서 작위적인 가정이 약할수록 얻어진 결론이 제한적일 수는 있지만 좀 더 신뢰할 수 있다는 것이다. 이런 부분 식별 방법론의 입장을 Manski(2003, p. 1)는 다음과 같은 명제로 서술하였다.

“신뢰성 체감의 법칙(The Law of Decreasing Credibility): 추론의 신뢰성은 유지되는 가정의 강도에 따라 감소한다(The credibility of inference decreases with the strength of the assumptions maintained.)”

(1) 본 연구는 제원연구재단의 지원으로 이루어진 것임. 제원연구재단의 지원에 감사드린다.

일반적으로 말해, 자료 그 자체가 제공하는 정보만으로는 의미 있는 추론을 하는 것이 불가능한 경우가 많다. 따라서 계량 모형은 추론을 위해서 필수적인 경우가 대부분이라고 할 수 있다. 계량 모형들을 구성하는 가정들은 크게 두 가지 종류로 구분할 수 있다. 첫째로는 희소성과 합리성 등에 따른 경제 원리에 근거한 제약들이 있고 둘째로는 추정의 편의를 위해 추가된 가정들이 있다. 후자는 종종 함수의 형태(functional form)나 분포에 관한 가정(distributional assumption)으로 표현되는 경우가 많다.

부분 식별 방법론은 추론의 결과가 후자와 같은 작위적인 가정들에 얼마나 의존하는지를 분석하고자 한다. 다양한 가정들을 적용해 봄으로써 모형의 가정들이 어떻게 모수에 대한 추론을 바꾸는지를 살펴보는 데에 부분 식별 방법론이 유용하다고 할 수 있다. 예를 들면, 많은 실증 분석에서 특히 미시 계량 모형에서는, 개인들 간의 관측되지 않는 이질성(unobserved heterogeneity)이 중요하다고 인식되어 왔다. 경제 이론은 많은 경우에 있어서 개인들의 이질성에 관해서 어떻게 모형화해야 하는지에 대해서는 도움이 되지 않는다. 이런 경우에 이질성에 대한 확률분포 가정이 얼마나 추론에 영향을 미치는지는 중요한 실증 분석의 문제이다. 부분 식별 방법론은 이런 경우에 있어서 해답을 줄 수 있다.

부분 식별의 중요한 예로서, 많은 경우에 있어서 부분 식별된 계량 모형의 모수(parameter)는 교집합 범위(intersection bounds)의 형태로 표현될 수 있다. 예를 들어 모형의 모수가 θ^* 라고 하자. 그리고 모든 $v \in \mathcal{V} \subset \mathbf{R}^d$ 에 대하여, θ^* 가 범위 구간인 $[\theta^l(v), \theta^u(v)]$ 에 속한다고 가정하자. 그러면 θ^* 에 대한 식별 영역(identification region)은 다음과 같이 주어진다.

$$(1.1) \quad \Theta_I = \cap_{v \in \mathcal{V}} [\theta^l(v), \theta^u(v)] = [\sup_{v \in \mathcal{V}} \theta^l(v), \inf_{v \in \mathcal{V}} \theta^u(v)]$$

이러한 교집합 범위(intersection bounds)는 다양한 계량 경제학 모형들에서 도출될 수 있다. 그 예로는 Manski(2003, 2007)에서 많이 찾아 볼 수 있다. 이러한 상황에서는 통상적인 추론 및 검정 방법과는 다른 방법을 요구한다. 본 논문의 목적은 계량 경제학 모형의 부분 식별에 대한 최근의 계량 경제학 문헌을 소개하고 경제학에의 적용 사례를 통해 새로운 추론 및 검정 방법의 유용성을 살펴보는 데 있다. 현재까지 개발된 방법들을 개관하고 정리하여 비교·분석하고 새로운 연구 방향을 제시하고

자 한다. 부분 식별의 최근 동향을 잘 정리한 선행 연구로는 Tamer(2010)를 참조하기 바란다.

본 논문은 선행 연구의 결과를 정리하여 개괄적으로 소개하는 검토서이며, 다음과 같은 구조로 이루어져 있다. 2장에서는 경제학 실증 연구에 부분 식별을 이용한 사례들을 경제학 분야별로 분류해 소개할 것이다. 3장에서는 최근 계량 경제학에서 연구된, 부분 식별로 정해지는 식별 영역에 관한 추론 방법을 소개할 것이다. 마지막으로, 4장의 결론으로 본 논문을 마무리할 것이다.

2. 실증 분석 사례

이 장에서는 경제학 실증 연구에 부분 식별을 이용한 사례들을 경제학 분야별로 분류해 소개하고자 한다. 특히, Blundell, Gosling, Ichimura, and Meghir(2007)의 실증 분석 사례와 Ciliberto and Tamer(2009)의 실증 분석 사례에 관해서 자세히 소개할 것이다. 현재까지는 주로 응용 미시 경제학, 특히 노동경제학과 산업조직론에서 부분 식별이 가장 많이 응용되었다. 이 장의 마지막에서는 다른 선행 연구에 대해서도 간단하게 소개할 것이다.

2.1. Blundell, Gosling, Ichimura, and Meghir(2007)의 실증 분석 사례

임금 구조의 변화는 노동경제학에서 많이 연구되는 주제 중 하나이다. 특히 학력이나 성별에 따른 임금 격차의 변화 분석이나 같은 집단 내에서의 불평등(within-group inequality)의 변화를 분석하는 것이다. 이러한 연구에 있어서 어려운 문제 중 하나가 개인의 노동시장에서의 참여에 관한 선택(selection into work) 문제를 어떻게 고려하느냐는 것이다. 구체적으로 말해서 개인이 노동시장에 참여하느냐의 결정은 임의로(randomly) 결정되는 것이 아니라 개인의 선택에 인한 것이므로 노동시장의 참가율과 임금 구조가 동시에 변할 경우에는 임금 구조의 변화를 분석하기 쉽지 않다. 개인이 노동 시장에 참여하는 원인과 임금 수준을 결정하는 요인의 관계에 대해서 계량 모형을 수립하고자 할때 강한 가정 없이는 모형의 식별(identification)을 달성하기가 쉽지 않다. 이에 대한 대안으로서 Blundell, Gosling, Ichimura, and Meghir(2007)는 부분 식별 방법을 이용하여 영국에서의 임금 구조 변화에 대해서 연구하였다.

Blundell, Gosling, Ichimura, and Meghir(2007)의 모형을 설명하면 다음과 같다.

종속변수(W)는 로그 임금이고 설명변수들(X)은 성별, 연령, 교육 수준 등이다. 또한 더미변수(E)는 임금이 관측될 때에는 1의 값을 가지고 그렇지 않을 때에는 0의 값을 가진다고 하자. $X = x$ 로 주어져 있을 때 W 의 조건부 누적 분포 함수(conditional cumulative distribution function)를 $F(w|x)$ 라고 정의하자. 선택 편의(selection bias)의 문제 때문에

$$F(w|x) \neq F(w|x, E=1)$$

다시 말해서, 연구자가 알고자 하는 로그 임금의 조건부 누적 분포 함수는 관측된 임금의 조건부 누적 분포 함수로는 식별이 되지 않는다. Manski(2003, 2007)의 부분 식별은 다음과 같은 식에서 출발한다.

$$F(w|x) = F(w|x, E=1)P(x) + F(w|x, E=0)[1 - P(x)]$$

위 식에서 $P(x) = \Pr(W=1 | X=x)$, 즉 $X=x$ 로 주어져 있을 때 노동시장에서 고용될 조건부 확률을 뜻한다. $F(w|x)$ 가 식별이 어려운 이유는 $F(w|x, E=0)$ 가 관측되는 자료들을 통해서 직접 식별이 가능하지 않기 때문이다. 식별할 수 없는 $F(w|x, E=0)$ 가 확률이므로 언제나

$$0 \leq F(w|x, E=0) \leq 1$$

이라는 사실에 착안하여, Manski(2003, 2007)의 최악 경우 범위(worst-case bounds)는 다음과 같이 주어진다.

$$(2.1) \quad F(w|x, E=1)P(x) \leq F(w|x) \leq F(w|x, E=1)P(x) + [1 - P(x)]$$

이런 로그 임금의 조건부 누적 분포 함수에 대한 범위(bounds)를 조건부 분위수(conditional quantiles)로 바꾸어서 표현하면 다음 식과 같다.

$$(2.2) \quad Q(\tau|x) \leq Q_{w|x}(\tau|x) \leq Q_u(\tau|x)$$

여기서, $Q_{w|x}(\tau|x)$ 는 로그 임금의 τ -조건부 분위수를 뜻하고, $Q_l(\tau|x) \equiv w_l^*$ 와 $Q_u(\tau|x) \equiv w_u^*$ 는 각각 다음의 두 식을 만족하는 해들이다.

$$\tau = F(w_l^* | x, E = 1)P(x) \text{ and } \tau = F(w_u^* | x, E = 1)P(x) + [1 - P(x)]$$

위의 식 (2.1)에서 볼 수 있듯이, $P(x)$ 가 클수록 식별되는 구간이 좁아진다. 다시 말해, 많은 개인들이 고용되어 있을 때에만 Manski(2003, 2007)의 최악 경우 범위(worst-case bounds)는 유용한 정보를 줄 수 있다.

Blundell, Gosling, Ichimura, and Meghir(2007)는 영국 자료를 이용해 분석했는데 여성과 같은 집단에 있어서는 고용률이 높지 않아 식 (2.1)이나 혹은 식 (2.2)로는 성별로 인한 임금 격차에 대해서 유용한 결과를 구할 수 없었다. 이런 문제를 극복하기 위해서, Blundell, Gosling, Ichimura, and Meghir(2007)는 Manski(2003, 2007)의 최악 경우 범위(worst-case bounds)에 경제학 이론을 기반으로 해서 추가적인 가정을 부가해 범위가 줄어들어지는지를 보았다.

가능한 추가적인 가정의 예로서, Blundell, Gosling, Ichimura, and Meghir(2007)는 확률적 지배(stochastic dominance) 가정을 고려하였다. 통상적인 경제 이론에 따르면, 개인의 노동시장에서의 참여에 관한 선택(selection into work)이 양(positive)의 관계를 가진다고 할 수 있다. 구체적으로 말하면 고용되어 있는 근로자들의 임금 분포가 고용되어 있지 않은 근로자들의 임금 분포를 1차 확률적으로 지배(first-order stochastic dominance)한다고 가정하는 것이다. 수식으로 표현하면, 모든 x 와 w 에 대해서, 다음이 성립한다.

$$(2.3) \quad F(w | x, E = 1) \leq F(w | x, E = 0)$$

식 (2.3)의 추가적인 가정 하에서는 범위가 다음 식으로 더 좁아진다.

$$(2.4) \quad F(w | x, E = 1) \leq F(w | x) \leq F(w | x, E = 1)P(x) + [1 - P(x)]$$

식 (2.1)과 식 (2.4)를 비교해 보면, 낮은 범위(lower bound)가 $F(w | x, E = 1)P(x)$ 에서 $F(w | x, E = 1)$ 로 커진 것을 확인할 수 있다.

양의 선택(positive selection)이 모든 근로자들에게 항상 유효한 가정이 아닐 수 있다. 다른 예로서 Manski(2003, 2007)를 바탕으로 하여, Blundell, Gosling, Ichimura, and Meghir(2007)는 제외 조건(exclusion restriction)이나 혹은 확률적 단조성(stochastic monotonicity)을 가정해서 범위를 줄이기도 한다. 먼저, 제외 조건을 이용한 방법을 소개하자. $X=x$ 로 주어져 있을 때 새로운 변수(Z)가 로그 임금(W)과 조건부 독립이라고 가정하자. 다시 말해, 모든 x, w 와 z 에 대해서 다음이 성립한다.

$$(2.5) \quad F(w|x, z) = F(w|x)$$

다음과 같은 두 함수를 정의하자.

$$F_1(w|x) := \max_z \{F(w|x, z, E=1) P(x, z)\}$$

그리고

$$F_2(w|x) := \min_z \{F(w|x, z, E=1) P(x, z) + [1 - P(x, z)]\}$$

여기서, $P(x, z) = \Pr(W=1 | X=x, Z=z)$ 이다. 그러면 다음과 같은 식별 범위가 주어진다.

$$(2.6) \quad F_1(w|x) \leq F(w|x) \leq F_2(w|x)$$

식 (2.6)으로 주어진 식별 범위의 약점은 제외 조건을 만족하는 변수를 실증 사례에서 찾기가 어렵다는 것이다. 따라서 좀 더 완화된 가정으로서, 임금의 조건부 분포가 z 에 따라 단조적으로 줄어든다(stochastic monotonicity)고 가정해 볼 수 있다. 수식으로 표현하면, 모든 x, w, z, z' ($z < z'$)에 대해서 다음이 성립한다.

$$(2.7) \quad F(w|x, z') \leq F(w|x, z)$$

이 가정 하에서는, $Z=z_1$ 에 대해서 다음과 같은 식별 범위가 주어진다.

$$(2.8) \quad F_3(w | x, z_1) \leq F(w | x, z_1) \leq F_4(w | x, z_1)$$

여기서,

$$F_3(w | x, z_1) := \max_{z \geq z_1} \{F(w | x, z, E = 1)P(x, z)\}$$

이고

$$F_4(w | x, z_1) := \min_{z \leq z_1} \{F(w | x, z, E = 1)P(x, z) + [1 - P(x, z)]\}$$

이다. 제외 조건이나 확률적 단조성 가정을 추가해서 식별 범위를 구하기 위해서는 Z 변수의 관측이 요구된다. Blundell, Gosling, Ichimura, and Meghir(2007)의 연구에서는 비근로 복지 수혜 임금(out-of-work welfare benefit income)을 Z 변수로 이용하였다. Blundell, Gosling, Ichimura, and Meghir(2007)의 연구 결과에 따르면 비근로 복지 수혜 임금(Z)은 제외 조건을 충족하지 않는다는 검정 결과를 얻은 반면, 완화된 가정인 확률적 단조성에 대해서는 그런 결과를 보이지 않았다.

Blundell, Gosling, Ichimura, and Meghir(2007)의 실증 분석 결과를 간략히 요약하면 다음과 같다. 영국 자료(구체적으로는 U.K. Family Expenditure Survey)를 이용하여 임금 불평등(wage inequality)의 변화에 관해 부분 식별 방법을 응용한 결과, 1970년대 후반에서 2000년 사이에 (1) 전체 임금 불평등과 교육 수준 내의 임금 불평등이 증가하였고, (2) 연령 간 차이도 증가하였으며, (3) 교육 수준에 따른 임금 격차도 증가하였고, (4) 25세 저숙련 여성 근로자들의 동일한 남성 근로자들과의 격차는 적어도 23% 이상 줄어 들었지만, 40세 저숙련 여성 근로자나 대졸 여성 근로자들에 대해서는 남성과의 임금 격차가 줄었다는 유의미한 근거를 찾지 못했다. Blundell, Gosling, Ichimura, and Meghir(2007)는 부분 식별 방법을 이용해서 임금 분포 변화에 있어서 유용한 실증 결과를 낼 수 있음을 보였다. 국내 자료를 이용해서 그들의 연구 방법을 응용 및 추가, 발전시키는 것이 흥미로운 미래의 연구 주제라고 할 수 있다.

2.2. Ciliberto and Tamer(2009)의 실증 분석 사례

Ciliberto and Tamer(2009)는 완전 정보(complete information) 하에서의 정적인

(static) 이산(discrete) 게임에서 플레이어들의 수익 함수(payoff function)를 추정하는 방법을 개발하였다. 특히 플레이어들의 이질성(heterogeneity)을 강조하여 시장에서 균형이 결정되는 과정에 대한 가정을 이전 연구에 비해 완화시킨 것이 특징이다. 그 결과 계량 모형이 부분 식별되는 것으로 표현되었고, Ciliberto and Tamer(2009)는 이런 방법론을 미국 항공운송업에 적용해 새로운 추론 방법의 유용함을 보였다.

Ciliberto and Tamer(2009)의 추론 방법을 구체적으로 소개하기 위해서 다음과 같은 기업의 이윤 함수를 정의하자.

$$(2.9) \quad \pi_{im} = S'_m \alpha_i + Z'_{im} \beta_i + W'_{im} \gamma_i + \sum_{j \neq i} \delta_j^i y_{jm} + \sum_{j \neq i} Z'_{jm} \phi_j^i y_{jm} + \varepsilon_{im}$$

여기서, 첨자 i 는 기업을 뜻하고 첨자 m 은 시장을 나타낸다. 또한, π_{im} 는 기업 i 의 시장 m 에서의 이윤을 뜻하고, y_{im} 은 기업 i 가 시장 m 에 진입했는지($y_{im} = 1$) 안 했는지($y_{im} = 0$)를 보여 주는 더미 변수이며, S_m 은 시장 m 에서 기업들에게 공통된 특성들을 나타내는 벡터이다. 두 변수 Z_{im} 과 W_{im} 은 시장 m 에 진입한 기업 i 의 고유한 특성들이다. 두 변수들 간의 차이는 Z_{im} 가 시장에 진입한 다른 기업들의 이윤에 영향을 미칠 수 있는 변수인 반면 W_{im} 은 시장에 진입한 다른 기업들에는 영향을 미치지 않는 변수라는 점이다. 예를 들면, Z_{im} 은 소비자의 선호에 영향을 미칠 수 있는 상품의 특성일 수 있는 반면 W_{im} 은 생산에 필요한 비용을 들 수 있다. 마지막으로, ε_{im} 은 실증 분석하는 연구자에게는 관측되지 않는 변수이다. Ciliberto and Tamer(2009)는 완전 정보를 가정함으로써 시장의 모든 플레이어가 ε_{im} 을 관측한다고 가정하고 있다. 시장 m 에 있어서 모든 관측 가능한 독립변수들을 모아서 $X_m = (S_m, \{Z_{im} : i = 1, \dots, K\}, \{W_{im} : i = 1, \dots, K\})$ 이라고 하자. 여기서, K 는 시장에 진출할 수 있는 모든 잠재적 기업들의 수이다.

이윤을 결정하는 위 식 (2.9)에서는 모든 계수(coefficient)들이 기업마다 고유한 값을 가지고 있다고 가정하고 있다. 즉, 모든 계수들이 첨자 i 에 의존한다. 특히 $\{\delta_j^i, \phi_j^i\}$ 는 다른 기업들의 시장 진입에 따라 기업 i 의 이윤이 어떻게 변하는지를 보여 주는 계수들이다.

Ciliberto and Tamer(2009)는 각 시장에서의 기업들의 수(K)는 작고 시장의 수(n)가 무한대로 커진다고 가정하며, 또한 ε_{im} 가 관측되는 독립변수들(S_m, Z_{im}, W_{im})과 독립이고 ε_{im} 의 확률분포가 유한차원의 모수 θ 에 의해 결정된다고 가정한다. 벡터 $\mathbf{y}_m = (y_{1m},$

\dots, y_{km})는 시장에 진입할 수 있는 잠재적 K 개의 기업들의 진입 여부를 나타내는 변수이고 $\mathbf{X}_m = (X_1, \dots, X_k)$ 라고 하자. \mathbf{y}_m 과 \mathbf{X}_m 이 무작위 포본으로부터 추출되었다고 가정하면, Ciliberto and Tamer(2009)는 이 경우에 기업들의 조건부 시장 진입 확률 ($P(\mathbf{y} | \mathbf{X})$)이 θ 에 의존하는 함수들에 의해서 부등식의 형태로 다음과 같이 나타나게 됨을 보였다.

$$\mathbf{H}_1(\theta, \mathbf{X}) \leq P(\mathbf{y} | \mathbf{X}) \leq \mathbf{H}_2(\theta, \mathbf{X})$$

다시 말해, 모형의 모수 θ 들이 조건부 부등식 제약 하에서 부분 식별되는 것이다. 실제 추정 시에는 \mathbf{H}_1 과 \mathbf{H}_2 를 직접 구하기는 쉽지 않다. 따라서, Ciliberto and Tamer (2009)는 몬테카를로 모의 시행(Monte Carlo simulations)을 통해 구하는 것을 제안하고 있다. 구체적인 내용은 Ciliberto and Tamer(2009)의 3.5절을 참조하기 바란다.

Ciliberto and Tamer(2009)의 실증 분석 예를 간략히 정리하면 다음과 같다. 미국 항공운송업에서 시장 진입과 관련하여 부분 식별 방법을 적용한 결과, 경쟁의 효과가 대형 항공사들(American, Delta, United)인지 혹은 Southwest와 같은 저가 항공사 인지에 따라서 달라졌음을 보였다. 또한 정책 실행에 관한 모의 실험을 계량 모형 결과를 바탕으로 수행할 수 있음을 보였다. 특히, 모형의 식별이 제한된 구조적 모형(structural model)에서도 정책 실험(policy experiment)이 가능함을 보였다는 점에서 큰 시사점이 있다고 할 수 있다.

2.3. 기타 경제학 분야에서의 실증 분석 사례

Blundell, Gosling, Ichimura, and Meghir(2007)의 연구뿐만 아니라 그 외의 다양한 노동경제학의 주제들에 대해서 부분 식별 방법이 적용되었다. 몇 가지 예를 들면, Manski and Pepper(2000)는 단조적 도구 변수(monotone instrumental variables)를 이용해 교육 수준으로 인한 임금 격차를 부분 식별 방법으로 연구하였으며, Gonzalez (2005)는 언어 습득 기술(language skills)에 대한 노동시장에서의 보수(returns)를 비모수적 부분 식별 방법으로 분석하였다. 또한 Kreider and Pepper(2007)는 장애와 고용 간의 관계에 관하여, Lee and Wilke(2009)는 독일에서의 실업 보상에 대한 정책 변화의 효과에 대해서, Nicoletti, Foliano, and Peracchi(2011)는 소득빈곤도 추정에 있어서 각각 부분 식별 방법을 이용하여 실증 분석하였다. 또한 Chetty(2011)는 부분

식별의 아이디어에 착안하여 노동 공급의 탄력성에 대한 기존 연구들을 종합하였다.

산업조직론에 있어서도 Ciliberto and Tamer(2009)의 연구 결과 이외에도 여러 선행 연구가 있다. 예를 들면 Haile and Tamer(2003)는 영국식 경매(English auction)에 부분 식별을 적용해 분석하는 방법을 제시했고, Pakes(2010)는 산업조직론에서 경제행위의 선택 모형들이 어떻게 부등식을 유도해 내는지를 보여 주고 있다.

최근에는 노동경제학이나 산업조직론 이외에도 부분 식별의 방법론이 전파되고 있다. 거시경제학 실증 분석에서의 예를 들면, Moon, Schorfheide, Granziera, and Lee(2011)는 구조적 벡터 자기 회귀 모형(structural vector autoregressions, 줄여서 SVARs)에 있어서 부호 제약(sign restriction)에 따른 부분 식별에 관해서 연구하였다.

3. 부분 식별에 있어서의 추론 방법

이 장의 목적은 부분 식별로 정해지는 식별 영역에 관한 추론 방법을 소개하는 것이다. 최근 계량 경제학에서 가장 많이 연구되고 있는 분야 중의 하나가 식별 영역에 관한 추론 방법들이다. 그 중 가장 활발히 연구가 진행된 분야는 부분 식별된 변수가 유한한 수의 무조건부 부등식 제약(unconditional inequality restrictions)으로 정의되는 경우이다. 예를 들면, 최근의 참고 문헌으로는 Andrews and Guggenberger(2009), Andrews and Jia(2008), Andrews and Soares(2010), Beresteanu and Molinari(2008), Bugni(2010), Canay(2010), Chernozhukov, Hong, and Tamer(2007), Galichon and Henry(2009), Romano and Shaikh(2008, 2010), Rosen(2008) 등이 있다.

유한한 수의 부등식 제약을 다루는 문헌에 추가하여 가장 최근에는 연속체 부등식(a continuum of inequalities)이 주어져 있을 때의 추론 방법들이 개발되고 있다. 예를 들면 Andrews and Shi(2009), Chernozhukov, Lee, and Rosen(2009), Kim(2009), Menzel(2009) 등이 있다.

본 논문에서는 선행 연구 중에서 특히 Chernozhukov, Hong, and Tamer(2007)와 Chernozhukov, Lee, and Rosen(2009)에서 개발된 추론 방법들을 자세히 소개한다.

3.1. Chernozhukov, Hong, and Tamer(2007)의 추론 방법

먼저 Chernozhukov, Hong, and Tamer(2007)의 추론 방법을 소개한다. Chernozhukov, Hong, and Tamer(2007)의 분석은 계량 모형의 제약들을 담은 모집단의 판정 기준 함

수(criterion function)로부터 시작한다. $\theta \in \Theta \in \mathbf{R}^d \mapsto Q(\theta)$ 를 판정 기준 함수라고 표현해 보자. 여기서 편의상 판정 기준 함수가 항상 음수가 아니라고 가정하자. 그러면 식별 영역은

$$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$$

으로 주어진다. Chernozhukov, Hong, and Tamer(2007)는 식별 영역을 추정하기 위해서 표본 판정 기준 함수 $Q_n(\theta)$ 의 경로 집합(contour set)을 이용한다.

판정 기준 함수가 계량 모형으로부터 어떻게 도출되는지를 보여 주기 위해서, 다음과 같은 적률 부등식 제약을 고려하자.

$$(3.1) \quad E[m(\theta, W)] \leq 0$$

여기서, $m(\theta, W)$ 는 모수(θ)와 확률 벡터(W)에 의존하는 적률 함수들의 벡터이다. 따라서 식 (3.1)로부터 정의된 식별 영역은

$$\Theta_I = \{\theta \in \Theta : E[m(\theta, W)] \leq 0\}$$

이다. 이제 다음과 같은 판정 기준 함수를 정의하자.

$$(3.2) \quad Q(\theta) := \| E[m(\theta, W)]' W^{1/2}(\theta) \|_+^2$$

여기서, $\|x\|_+ = \|\max(x, 0)\|_+$ 이고 $W(\theta)$ 는 θ 에 대해서 연속이며 대각선 원소들이 모든 θ 에 대해서 양인 대각선 행렬이다. 그러면 Θ_I 는 식 (3.2)에서 정의된 $Q(\theta)$ 를 최소화하는 θ 값들의 집합으로 정의할 수 있다. 따라서 Θ_I 에 대한 추론은 다음과 같은 표본 판정 기준 함수 $Q_n(\theta)$ 를 바탕으로 할 수 있다.

$$(3.3) \quad Q_n(\theta) := \left\| \left[n^{-1} \sum_{i=1}^n m(\theta, W_i) \right]' W_n^{1/2}(\theta) \right\|_+^2$$

여기서, $W_n(\theta)$ 는 $W(\theta)$ 를 균일하게 일치하는(uniformly consistent) 추정량이다. 실제 응용에 있어서는 $W_n(\theta)$ 를 단위 행렬(identity matrix)로 두거나 각각의 표본 적률의 표준 오차로 둘 수 있다.

Chernozhukov, Hong, and Tamer(2007)는 식별 영역을 추정하기 위해서 다음과 같은 경로 집합(contour set)을 제시한다.

$$C_n(c) := \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}$$

여기서, 상수 c 는 기각값(critical value)이며, a_n 는 표준화하는 수열이다. 후자에 대해서 구체적으로 언급하면, a_n 는

$$C_n := \sup_{\theta \in \Theta} a_n Q_n(\theta)$$

가 확률적으로 유계(stochastically bounded)이고 정상적인 대표본 분포를 가지도록 보장할 수 있도록 선택되어야 한다. 가령, 많은 경우에 있어서 $a_n = n$ 이다.

Chernozhukov, Hong, and Tamer(2007)는 상수 c 의 결정에 관해서 여러 방법을 언급하고 있는데 한 방법은 c 가 C_n 의 α -분위수(quantile)의 추정량이 되도록 설정하는 것이다. 이럴 경우, $C_n(c)$ 는 Θ 의 α 수준의 신뢰 영역이 된다. Chernozhukov, Hong, and Tamer(2007)는 C_n 의 α -분위수(quantile)의 일반적인 추정 방법으로 부표집(subsampling)을 추천하고 있다. Chernozhukov, Hong, and Tamer(2007)는 부분 식별 추정 방법의 일반론을 광범위하게 발전시킨 논문으로서, 추가적인 논문들에 강한 영향을 주었다. 자세한 이론적 결과들과 구체적인 예들은 Chernozhukov, Hong, and Tamer(2007)를 참조하기 바란다.

3.2. Chernozhukov, Lee, and Rosen(2009)의 추론 방법

이제 Chernozhukov, Lee, and Rosen(2009)의 추론 방법을 소개하도록 하자. 머리말의 식 (1.1)에서 정의된 식별 영역에 관한 추론을 위해서는 양 끝점(end-point)에 대한 추론이 요구된다. 구체적으로 언급하면, $\sup_{v \in \mathcal{V}} \theta'(v)$ 와 $\inf_{v \in \mathcal{V}} \theta''(v)$ 에 대한 추론이 필요하다. 양 끝점(end-point)에 대한 논의가 대칭적이므로, 이 장에서는 상단 범위(upper bound)에 관해서만 논의하도록 하자. 즉, 다음과 같은 상단 범위값(θ_0)을 고

려한다.

$$(3.4) \quad \theta^* \leq \theta_0 := \inf_{v \in \mathcal{V}} \theta(v)$$

여기서, $v \mapsto \theta(v)$ 는 범위를 결정하는 함수이고 \mathcal{V} 는 하한(infimum)을 정하는 집합이다.

상단 범위값(θ_0)에 대한 한 방법은 $\theta(v)$ 에 대한 추정량($\hat{\theta}(v)$)을 이용하여 표본 최소값($\inf_{v \in \mathcal{V}} \hat{\theta}(v)$)을 고려하는 것이다. 그러나 표본 최소값은 실제 응용에 있어서 문제가 있을 수 있다. 첫째, 실제 표본에서는 표본 최소값은 하방 편향(downward biased)되기 쉽다. 이럴 경우, 표본에서 정해지는 식별 영역이 모집단(population)에서 정의되는 식별 영역보다 훨씬 더 좁게 정해질 수 있다. 둘째, 추정량($\hat{\theta}(v)$)은 v 값이 변함에 따라 다른 분산값을 가질 수 있다. 다시 말해, 추정량($\hat{\theta}(v)$)의 정밀도가 v 값에 따라 크게 변한다면 표본 최소값($\inf_{v \in \mathcal{V}} \hat{\theta}(v)$)이 모집단에서 정의되는 식별 영역에 대한 왜곡된 정보를 주기 쉽다.

표본 최소값의 이런 문제를 극복하기 위해서, Chernozhukov, Lee, and Rosen(2009)은 다음과 같은 편향과 정밀도의 문제가 고려된 추정량을 제시하고 있다.

$$(3.5) \quad \hat{\theta}(p) := \min_{v \in \mathcal{V}} [\hat{\theta}(v) + k(p) \cdot s(v)]$$

여기서 $s(v)$ 는 $\hat{\theta}(v)$ 의 표준오차이고 $k(p)$ 는 기각값(critical value)이다. 기각값의 선택에 관해서는 아래에서 언급할 것이다. 다시 말해, Chernozhukov, Lee, and Rosen(2009)의 추정량($\hat{\theta}(p)$)은 편향과 정밀도를 수정한 후 최소값을 구한다.

위 식 (3.5)에서 정의된 추정량을 이용하기 위해서는 기각값($k(p)$)을 정해야 한다. Chernozhukov, Lee, and Rosen(2009)은 기각값을 정하기 위해 다음과 같은 정규화된(standardized) 통계량을 고려한다.

$$Z_n(v) = \frac{\theta(v) - \hat{\theta}(v)}{\sigma(v)}$$

여기서 $\sigma(v)$ 는 $s(v)$ 에 대해 균일하게 일치하는(uniformly consistent) 표본 추정량이다. 일반적으로, 유한 표본(finite sample)에서는 통계량 Z_n 의 확률분포를 알 수 없다. 그

러나 표본의 크기(n)가 증가함에 따라, Z_n 는 다음과 같이 가우시안 확률 과정들(Z_n^*)의 수열로 균일하게 근사할 수 있다.

$$(3.6) \quad \bar{a}_n \sup_{v \in \mathcal{V}} |Z_n(v) - Z_n^*(v)| = o_p(1)$$

여기서 \bar{a}_n 은 표본의 크기가 증가함에 따라 무한대로 가는 수열이다. 그러면 어떠한 콤팩트 집합($V \in \mathcal{V}$)에 대해서도, $\sup_{v \in V} Z_n^*(v)$ 의 확률분포를 점근 근사(asymptotic approximations)에 따른 분석 방법이나 좀 더 일반적으로는 모의 실험을 통해 근사할 수 있다. 따라서 Chernozhukov, Lee, and Rosen(2009)은 기각값을 $\sup_{v \in V} Z_n^*(v)$ 의 p -분위수로 정할 것을 제안하고 있다.

Chernozhukov, Lee, and Rosen(2009)에서는 이론적 근거 및 컴퓨터 모의 실험(simulation) 결과를 통해 새로운 추정량의 유용성을 제시하고 있다. 구체적인 통계량의 이행 방법(implementation method) 및 다양한 예들을 Chernozhukov, Lee, and Rosen(2009)에서 참조하기 바란다.

이 장에서는 마지막으로 2.1.장에서 소개된 사례들을 Chernozhukov, Lee, and Rosen(2009)에 근거한 추론 방법을 이용해 어떻게 분석할 수 있는지 논의하도록 하겠다. 예를 들어, 식 (2.8)에서 정의된 식별 범위에서 상한 범위값을 고려하자. 식 (2.8)에서 w, x, z_1 의 값들이 주어져 있다고 하면, 다음과 같이 상한 범위값(θ_0)이 결정된다.

$$\theta_0 = \min_{v \in \mathcal{V}} \theta(v)$$

여기서, $\mathcal{V} = \{v \in Z: v \leq z_1\}$, Z 는 비근로 복지 수혜 임금(Z)의 받침(support)이고 $\theta(v) = F(w | x, v, E = 1)P(x, v) + [1 - P(x, v)]$ 이다.

Chernozhukov, Lee, and Rosen(2009)의 추정량을 적용하기 위해서는 $\theta(v)$ 의 추정량이 필요하다. 여기서 $\theta(v)$ 는 다음과 같은 조건부 기대값(conditional expectation)으로 표현할 수 있다.

$$(3.7) \quad \theta(v) = E[1(W \leq w) 1(E = 1) + 1(E = 0) | X = x, V = v]$$

여기서, w, x 의 값들은 이미 주어져 있기 때문에 $\theta(v)$ 는 단지 v 의 함수로만 표시되어 있다. 종속변수에 대해서 $Y := 1(W \leq w) 1(E = 1) + 1(E = 0)$ 로 정의하면 $\theta(v)$ 는 $X = x, V = v$ 로 주어져 있을 때의 Y 의 조건부 기대값이 된다. 따라서, $\theta(v)$ 는 통상적인 조건부 기대값을 추정하는 방법으로 추정하면 된다. 예를 들면, 비모수적 회귀 모형(nonparametric regression model)을 이용해서 구할 수 있다. 구체적인 방법들은 Chernozhukov, Lee, and Rosen(2009)과 이 논문에서 인용하고 있는 참고 문헌을 참조하기 바란다.

4. 맺음말

지금까지 이 논문에서는 계량 경제학 모형의 부분 식별에 관해서 구체적으로 살펴 보았다. 부분 식별 방법론은 최근 10년간에 계량 경제학 내에서 가장 활발히 연구가 진행되고 있는 분야이며, 이론적 발전의 속도가 매우 빠른 분야이기도 하다. 그뿐만 아니라, 노동경제학 및 산업조직론 등에서 응용되기 시작하였고 현재 추세가 계속 된다면 가까운 미래에는 중요한 실증 분석들 중의 하나로서 자리매김할 확률도 높다. 부분 식별 방법은 자료의 미비함이나 계량 모형의 불완전성을 편의적 혹은 작위적 가정으로 대체하지 말고 모형의 모수가 부분적으로만 식별되는 상황 하에서도 유용한 결론을 얻을 수 있는나에 초점을 맞추고 있다. 따라서, 부분 식별 계량 모형에서 얻어진 결론은 상대적으로 좀 더 신뢰성이 높다고 할 수 있다. 이러한 부분 식별 계량 모형을 분석하기 위해서는 기존의 계량 이론과는 다른 새로운 이론이 요구되는데 최근 계량 경제학 이론에서 다양한 새로운 방법론을 제시해 왔다. 이 논문에서는 그런 연구 문헌을 소개하고 그 중에서 일부를 자세히 소개하였다. 이런 새로운 계량 기법을 이용하여 국내 자료를 분석하여 국내 실증 연구에 기여하는 것이 중요한 미래의 연구 주제 중의 하나이다. 또 다른 연구 주제는 새로 제시되고 있는 기법을 비교·분석하여 어느 방법이 우월한지 규명하는 것이다. 마지막으로, 연구자들이 편리하게 사용할 수 있도록 컴퓨터 프로그램을 개발하는 것도 중요 과제라고 할 수 있겠다.

서울대학교 경제학부 부교수

151-746 서울특별시 관악구 관악로 1

전화: (02) 880-6494
팩스: (02) 886-4231
E-mail: sokbae@snu.ac.kr

참고 문헌

- Andrews, D.W.K., and P. Guggenberger(2009): “Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities,” *Econometric Theory*, **25**, 669–709.
- Andrews, D.W.K., and P. Jia(2008): “Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,” Working Paper, Cowles Foundation.
- Andrews, D.W.K., and X. Shi(2009): “Inference for Parameters Defined by Conditional Moment Inequalities,” Working Paper, Cowles Foundation.
- Andrews, D.W.K., and G. Soares(2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, **78**, **1**, 119–157.
- Beresteanu, A., and F. Molinari(2008): “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, **76**, **4**, 763–814.
- Blundell, R., A. Gosling, H. Ichimura, and C. Meghir(2007): “Changes in the Distribution of Male and Female Wages Accounting for Employment Composition,” *Econometrica*, **75**, 323–363.
- Bugni, F.(2010): “Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified set,” *Econometrica*, **78**, **2**, 735–753.
- Canay, I.(2010): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” *Journal of Econometrics*, **156**, **2**, 408–425.
- Chernozhukov, V., H. Hong, and E. Tamer(2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, **75**, **5**, 1243–1284.
- Chernozhukov, V., S. Lee, and A. M. Rosen(2009): “Intersection Bounds: Estimation and Inference,” Working Paper, arXiv:0907.3503, Available at <<http://arxiv.org/abs/0907.3503v3>>.

- Chetty, R.(2011): “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, Forthcoming.
- Ciliberto, F., and E. Tamer(2009): “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, **77**, **6**, 1791–1828.
- Galichon, A., and M. Henry(2009): “A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters,” *Journal of Econometrics*, **159**, **2**, 186–196.
- Gonzalez, L.(2005): “Nonparametric Bounds on the Returns to Language Skills,” *Journal of Applied Econometrics*, **20**, 771–795.
- Haile, P. A., and E. Tamer(2003): “Inference with an Incomplete Model of English Auctions,” *Journal of Political Economy*, **111**, **1**, 1–51.
- Kim, K. I.(2009): “Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities,” Working Paper, University of Minnesota.
- Kreider, B., and J. Pepper(2007): “Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors,” *Journal of the American Statistical Association*, **102**, **478**, 432–441.
- Lee, S., and R. Wilke(2009): “Reform of Unemployment Compensation in Germany: A Nonparametric Bounds Analysis Using Register Data” *Journal of Business and Economic Statistics*, **27**, **2**, 193–205.
- Manski C. F.(2003): *Partial Identification of Probability Distributions*, New York, Springer.
- _____ (2007): *Identification for Prediction and Decision*, Cambridge, MA, Harvard University Press.
- Manski, C. F., and J. V. Pepper(2000): “Monotone Instrumental Variables: With an Application to the Returns to Schooling,” *Econometrica*, **68**, **4**, 997–1010.
- Menzel, K.(2009): “Estimation and Inference with Many Weak Moment Inequalities,” Working Paper, MIT.
- Moon, H. R., F. Schorfheide, E. Granziera, and M. Lee(2011): “Inference for VARs Identified with Sign Restrictions,” NBER Working Paper **17140**.
- Nicoletti, C., F. Foliano, and F. Peracchi(2011): “Estimating Income Poverty in the

- Presence of Missing Data and Measurement Error Problems,” *Journal of Business and Economic Statistics*, **29**, **1**, 61–72.
- Pakes, A.(2010): “Alternative Models for Moment Inequalities,” *Econometrica*, **78**, 1783–1822.
- Romano, J. P., and A. M. Shaikh(2008): “Inference for Identifiable Parameters in Partially Identified Econometric Models,” *Journal of Statistical Planning and Inference*, **138**, **9**, 2786–2807.
- _____(2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *Econometrica*, **78**, **1**, 169–211.
- Rosen, A. M.(2008): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” *Journal of Econometrics*, **146**, 107–117.
- Tamer, E.(2010): “Partial Identification in Econometrics,” *Annual Review of Economics*, **2**, 167–195.

Abstract

Partial Identification of Econometric Models: A Selective Survey

Sokbae Lee

Partial identification of econometric models refers to settings where parameters of interest are not point-identified but only known to lie within a subset of the parameter space. This paper surveys the latest developments in the literature on partial identification in econometrics and illustrates the usefulness of this new approach by reviewing a few examples in economics.

Keywords: Intersection bounds, Partial identification, Identification region

