응용통계연구 (2010)
23(2), 263–274

기업경기실사지수의 통계적 성질 고찰

김규성[1]

[1]서울시립대학교 통계학과

(2010년 1월 접수, 2010년 2월 채택)

요 약

기업경기실사지수는 기업의 지난 실적과 기업가의 계획이나 판단 등을 기초로 하여 만들어지는 대표적인 경기 전망 지수이다. 이 지수는 경제 현장에서 널리 사용됨에도 불구하고 아직까지 규명된 통계적 성질은 많지 않다. 본 논문에서는 기업경기실사지수에 대한 통계적 성질을 고찰한다. 유한모집단에서 모집단 기업경기실사지수를 정의하고 단순확률표집을 가정한 후 모집단 기업경기실사지수 추정량을 구하며, 기업경기실사지수의 기댓값, 분산, 비편향 분산 추정량, 신뢰구간, 상대표준오차를 구한다. 그리고 기업경기실사지수의 평가지표로서 상대표준오차보다는 신뢰구간이 더 합리적임을 언급한다.

주요용어: 기업경기조사, 다변량 초기하분포, 모집단 기업경기실사지수, 자기 가중 표본.

1. 서론

기업경기동향지수(Business tendency index; BTI)라고도 부르는 기업경기실사지수(Business survey index; BSI)는 기업의 지난 시점의 실적과 기업가의 다음 시점에 대한 계획이나 경기 전망 등을 기초로 만들어진다 (OECD, 2003; 한국은행, 2006; 이상호, 2001a; 등). 기업경기실사지수는 간단한 설문조사에 의하여 만들어지기 때문에 조사가 간단하고 결과를 빠르게 발표할 수 있는 장점이 있다. 이러한 속보성 때문에 기업경기실사지수는 경기 전망 지수로서 널리 활용되고 있다(예를 들어 한국은행의 기업경기실사지수, 전국경제인연합회의 기업경기실사지수 등). 과거의 경제 데이터를 기초로 하여 수리적으로 경기를 전망하는 방법들과는 달리 기업경기실사지수는 기업의 지난 실적을 바탕으로 기업가의 계획이나 판단, 전망 등에 의존하여 만들어지므로 기업경기실사지수에는 객관적인 근거와 함께 주관적이고 심리적인 요인이 포함된다.

통상적인 기업경기실사지수는 다음과 같이 계산된다. 응답자는 기업의 현황과 전망을 묻는 조사 항목에 대하여 '긍정', '보통' 혹은 '부정' 중의 하나를 응답하게 된다. 전체 응답 중 긍정 응답 비율과 부정 응답 비율의 차이로 기업경기실사지수가 계산된다.

$$\text{기업경기실사지수} = \frac{\text{긍정 응답수} - \text{부정 응답수}}{\text{전체 응답수}} \times 100 + 100. \tag{1.1}$$

상수 100을 긍정 응답 비율과 부정 응답 비율의 차이에 곱하고 더함으로써 기업경기실사지수의 범위는 0에서 200까지가 된다. 지수가 100을 넘으면 지난달 혹은 지난 분기에 비하여 경기가 호전될 것이라고

이 논문은 2009년도 서울시립대학교 교내학술연구비에 의하여 연구되었음.
[1](130-743) 서울특별시 동대문구 시립대길 13, 서울시립대학교 통계학과, 교수. E-mail: kskim@uos.ac.kr

판단하고, 만일 지수가 100 미만이면 경기는 지난달에 비하여 악화될 것으로 판단한다 (예를 들어 한국은행, 2006; Curtin 등, 2000; 등).

기업경기실사지수는 기업경기조사(Business survey) 혹은 기업동향조사(Business tendency survey) 등을 통하여 작성된다(예를 들어 한국은행의 기업경기조사, 전국경제인연합회의 기업경기동향조사, 캐나다 통계청의 기업동향조사 등). 그리고 기업경기조사는 모두 표본조사이므로 식 (1.1)의 기업경기실사지수는 표본 추정량이다. 따라서 기업경기실사지수에 대하여 표본 추정량의 일반적인 성질에 대한 물음이 가능하다.

(i) 기업경기실사지수가 추정하고자 하는 것(모수, parameter)은 무엇인가?

(ii) 기업경기실사지수의 정확성에 대한 지표는 무엇인가?

첫 번째 질문에 대한 답은 모집단 기업경기실사지수일 것이다. 그렇다면 모집단 기업경기실사지수에 대한 정의가 필요하다. 두 번째 질문에 대한 답은 기업경기실사지수의 신뢰구간이나 표준오차 (혹은 상대표준오차)일 것이다. 여기에서는 신뢰구간이나 표준오차에 대한 추정이 필요하다.

기업경기실사지수는 경제 현장에서 널리 쓰이는 대표적인 지수임에도 불구하고 위의 두 가지 통계적인 질문에 대한 이론적인 연구결과는 거의 없다. 통상적인 지수나 추정치는 그 정확성이 궁금하기 때문에 표준오차 같은 평가 지표가 추정량과 함께 개발된다 그런데 기업경기실사지수는 전망 지수이기 때문에 다음 시점에 전 시점의 기업경기실사지수 전망치에 대한 검증이 어느 정도 가능하다. 따라서 경제 현장에서는 사후적으로 기업경기실사지수의 정확성을 이해하고 해석하는 것으로 보인다. 또한 기업경기실사지수는 경기 선행 지수로 간주되어 다른 주요 경제지표와의 상관관계분석 혹은 시차상관관계분석에 많이 활용된다 (이상호, 2001b; 유재준과 이상호, 2002; 심상달, 2002; 등). 그런데 사후적으로 검증하는 방법은 조사 오차(survey error)는 고려하지 않고 시계열 오차(time-series error)에 초점을 맞춘 것이다. 즉, 기업경기조사가 정확하게 수행되었다는 가정 아래 기업가의 이전 시점 전망을 검증하는 것이다. 그런데 기업경기실사지수의 예측력은 시계열 오차 뿐 아니라 조사오차로 부터도 영향을 받기 때문에 기업경기실사지수를 올바로 활용하기 위해서는 시계열 오차와 조사오차를 모두 평가할 수 있는 검증 방법이 필요하다.

본 논문은 기업경기실사지수의 조사오차에 관한 것이다. 조사오차 중에서 표집오차(sampling error)에 초점을 맞출 것이며 표집 방법 중 가장 기본이 되는 단순확률표집(simple random sampling)을 가정할 것이다. 기업경기조사의 조사방법은 OECD에서 출간된 보고서 (OECD, 2003) 등에 구체적으로 설명되어 있는데 표집오차에 대한 설명이 매우 부족하여 기업경기실사지수를 평가하기 어렵기 때문에 가장 기본적인 표집방법인 단순확률표집에서 출발하려는 것이다. 또한 기업경기실사지수의 형태도 가장 간단하면서도 널리 쓰이는 식 (1.1)의 형태를 가정할 것이다. 기업경기실사지수는 5점 척도로 하여 정의하기도 하고(예를 들어 한국은행의 소비자 동향조사, 중소기업협동조합중앙회의 월간중소기업경기전망조사 등) 계절조정을 하기도 하며(기업은행의 중소제조업경기전망조사 등) 업종별로 가중치를 주어 가중 기업경기실사지수를 정의하기도 한다 (OECD, 2003). 그러나 본 연구에서는 식 (1.1)의 기업경기실사지수로 한정하기로 한다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 모집단 기업경기실사지수를 정의하고 3절에서는 2절에서 정의한 모집단 기업경기실사지수를 단순확률표본을 기초로 추정한다. 또한 기업경기실사지수 추정량의 기댓값, 분산, 분산추정량을 유도하고 평가지표로서 신뢰구간과 상대표준오차를 제시하고 비교한다. 마지막으로 4절 결론에서는 본 논문을 간략하게 요약하고 향후 연구주제를 언급한다.

표 2.1. $N=6$인 경우 (A,B,C)가 취하는 값

B		C						
		0	1	2	3	4	5	6
A	0	6	5	4	3	2	1	0
	1	5	4	3	2	1	0	-*
	2	4	3	2	1	0	-	-
	3	3	2	1	0	-	-	-
	4	2	1	0	-	-	-	-
	5	1	0	-	-	-	-	-
	6	0	-	-	-	-	-	-

*: '-' 해당 없음.

2. 모집단 기업경기실사지수

2.1. 정의

크기가 N인 유한모집단 $U=\{1,2,3,\ldots,N\}$라고 하자. 어떤 조사항목에 대하여 각 구성원은 '긍정', '보통', '부정' 중의 하나를 응답한다고 하고 그 수는 각각 A,B,C라고 하자($N=A+B+C$). 그리고 각 비율은 $P_A(=A/N)$, $P_B(=B/N)$, $P_C(=C/N)$라고 하자. 이 모집단을 전수조사하면 기업경기실사지수는 다음과 같이 계산된다.

$$\mathrm{BSI}=\frac{A-C}{N}\times 100+100=(P_A-P_C)\times 100+100 \tag{2.1}$$

여기에서 $A=0,1,\ldots,N; C=0,1,\ldots,N-A; B=N-A-C$의 값을 갖는다. 식 (2.1)의 BSI는 유한모집단에서 계산되는 값이므로 '모집단 기업경기실사지수(BSI)'가 된다.

BSI가 정의되는 모수공간(parameter space)은 (A,B,C)가 취할 수 있는 값들의 모임이다. 즉,

$$\Omega=\{(A,B,C): A=0,1,\ldots,N;\ C=0,1,\ldots,N-A;\ B=N-A-C\}$$

그리고 모수공간의 크기는 (A,B,C)가 취할 수 있는 값들의 총수로서 그 값은 다음과 같다.

$$|\Omega|=1+2+\cdots+(N+1)=\frac{1}{2}(N+1)(N+2)\equiv T$$

예제 2.1: $N=6$인 유한모집단을 가정하자. 그러면 각 응답수는 다음의 값을 갖는다.

$$A=0,1,\ldots,6; \qquad C=0,1,\ldots,6-A; \qquad B=6-A-C$$

그리고 각 응답수 조합의 수는 $T=1+2+\cdots+7=28$이다. 표 2.1은 (A,B,C) 조합을 보여준다.

2.2. BSI의 유한모집단 분포

이제 모수공간에서 BSI의 분포를 구해본다. BSI는 $(A-C)$의 값이 같으면 동일한 BSI가 계산되므로 BSI의 분포를 구하기 위해서는 $(A-C)$의 값과 빈도를 구하여야 한다. 편의상 $D=(A-C)$라고 하고 $N(D)$는 모수공간에서 D값을 갖는 (A,B,C)조합의 수라고 하자. 그러면 D의 값의 범위는 $\{-N,-N+1,\ldots,N-1,N\}$이고 $N(D)$값의 합은 모수공간 크기인 T와 같다. 즉,

$$\sum_{D=-N}^{N} N(D)=\frac{1}{2}(N+1)(N+2)=T.$$

표 2.2. $N=6$인 경우 D의 범위와 BSI, $f(D)$의 값

D	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
BSI	0	16.7	33.3	50	66.7	83.3	100	116.7	133.3	150	166.7	183.3	200
$f(D)$	1/28	1/28	2/28	2/28	3/28	3/28	4/28	3/28	3/28	2/28	2/28	1/28	1/28

D의 상대도수(relative frequency)를 $f(D)$로 표시하면 $f(D)=N(D)/T$가 되고 $f(D)$를 구체적으로 구하면 다음과 같이 된다.

$$f(D)=\begin{cases}\dfrac{N-|D|+2}{(N+1)(N+2)}, & (N\text{: 짝수, } D\text{: 짝수) 또는 } (N\text{: 홀수, } D\text{: 홀수}),\\ \dfrac{N-|D|+1}{(N+1)(N+2)}, & (N\text{: 홀수, } D\text{: 짝수) 또는 } (N\text{: 짝수, } D\text{: 홀수}),\\ \dfrac{1}{N+1}, & N\text{: 짝수, } D=0,\\ \dfrac{1}{N+2}, & N\text{: 홀수, } D=0.\end{cases}$$

$f(D)$는 N과 D가 홀수인 경우와 짝수인 경우에 따라 다르게 표현된다. BSI는 D와 일대일 대응관계가 있으므로, $\mathrm{BSI}=D/N\times 100+100$, BSI의 상대도수는 대응되는 D의 상대도수와 같다.

예제 2.2: $N=6$인 유한모집단에서 D의 범위는 $\{-6,-5,\ldots,5,6\}$이고, D의 각 값에 대한 상대도수는 아래와 같다.

$$f(D)=\begin{cases}\dfrac{6-|D|+2}{(6+1)(6+2)}=\dfrac{8-|D|}{56}, & D=-6,-4,-2,0,2,4,6,\\ \dfrac{6-|D|+1}{(6+1)(6+2)}=\dfrac{7-|D|}{56}, & D=-5,-3,-1,1,3,5,\\ \dfrac{1}{6+1}=\dfrac{1}{7}, & D=0\end{cases}$$

그리고

$$\mathrm{BSI}=\frac{D}{6}\times 100+100, \qquad D=-6,-5,\ldots,5,6$$

이며 BSI의 상대도수는 대응되는 D의 상대도수와 같다. 아래 표 2.2에 D, BSI, $f(D)$의 구체적인 값이 계산되어 있다.

위의 예제 2.2에서도 나타나듯이 동일한 BSI에는 여러 가지 경우의 수가 생긴다. 예를 들어, 표 2.2에서 BSI가 100이 되는 경우는 $(A=0,B=6,C=0)$, $(A=1,B=4,C=1)$, $(A=2,B=2,C=2)$, $(A=3,B=0,C=3)$의 4가지 경우이다. 즉, 긍정과 부정이 없이$(A=C=0)$ 전체가 보통이라고 전망하는 경우$(B=6)$도 BSI는 100이며, 반대로 긍정이 3명, 부정이 3명, 보통이 0명이 경우도 BSI는 100이 된다. BSI는 이러한 세부적인 상황은 표현하지 않는다.

3. BSI 추정

3.1. BSI 추정량

2절에서 가정한 크기 N인 유한모집단에서 n명을 단순확률추출(simple random sampling)한다고 하자. 그리고 무응답은 없다고 가정하자. 그러면 긍정 응답수 a, 보통 응답수 b, 부정 응답수 c가 취할 값은 다

변량 초기하 분포(Multivariate hypergeometric distribution)를 따른다 (Johnson과 Kotz, 1969).

$$P(a,b,c) = \frac{\binom{A}{a}\binom{B}{b}\binom{C}{c}}{\binom{N}{n}},$$

여기에서 (a, b, c)가 취하는 값의 범위는 다음과 같다:

$$a + b + c = n, \qquad 0 \leqq a \leqq A, \qquad 0 \leqq b \leqq B, \qquad 0 \leqq c \leqq C$$

다변량 초기하 분포에서 응답수 a, b, c의 분포는 초기하 분포를 따르므로 응답수에 대한 기댓값과 분산은 쉽게 얻을 수 있다. 예를 들어 긍정응답수 a의 분포는 다음과 같으므로,

$$P(a) = \frac{\binom{A}{a}\binom{N-A}{n-a}}{\binom{N}{n}}, \quad \max(0, n-N+A) \leqq a \leqq \min(n, A).$$

긍정응답수 a의 기댓값과 분산은 아래 식 (3.1)과 같다.

$$E(a) = nP_A; \qquad \mathrm{Var}(a) = \left(\frac{N-n}{N-1}\right) nP_A(1-P_A). \tag{3.1}$$

또한 세 응답수 a, b, c에 대한 쌍별 공분산도 구할 수 있다. 예를 들어 (a, b)의 공분산은 다음과 같다.

$$\mathrm{Cov}(a,b) = -n\left(\frac{N-n}{N-1}\right) P_A P_B. \tag{3.2}$$

다른 조합인 (a, c)와 (b, c)의 공분산도 유사한 형태로 얻을 수 있다.

이제 표본 응답수 (a, b, c)를 활용하여 모집단 경기실사지수 BSI를 추정하자. 식 (3.1)로부터 a/n는 P_A의 비편향 추정량임을 알 수 있으므로 이 사실을 BSI에 대입하면 다음과 같은 기업경기실사지수 추정량을 얻을 수 있다. 즉,

$$\widehat{\mathrm{BSI}} = \left(\frac{a-c}{n}\right) \times 100 + 100 = (p_a - p_c) \times 100 + 100, \tag{3.3}$$

여기에서 $p_a = a/n$, $p_c = c/n$이다. 위의 식 (3.3)에서 구한 값은 기업경기실사지수 추정치인데, 앞의 식 (1.1)에서 나타난 통상적으로 말하는 기업경기실사지수이다. 본 논문에서는 편의상 BSI는 모집단 기업경기실사지수(식 (2.1)) 그리고 $\widehat{\mathrm{BSI}}$는 기업경기실사지수 추정량(식 (3.3))으로 표현하기로 한다.

3.2. $\widehat{\mathrm{BSI}}$ 분산 및 분산추정

이제 $\widehat{\mathrm{BSI}}$의 통계적 성질을 고찰한다. 먼저, 식 (3.1)의 기댓값의 성질을 이용하면 $\widehat{\mathrm{BSI}}$는 BSI의 비편향 추정량임을 쉽게 보일 수 있다. 즉,

$$\begin{aligned} E\left(\widehat{\mathrm{BSI}}\right) &= \left(\frac{E(a)-E(c)}{n}\right) \times 100 + 100 \\ &= (P_A - P_C) \times 100 + 100 = \mathrm{BSI}. \end{aligned}$$

또한 식 (3.1)과 (3.2)의 결과를 이용하면 $\widehat{\text{BSI}}$의 분산도 구할 수 있다. 즉,

$$\begin{aligned}\text{Var}\left(\widehat{\text{BSI}}\right) &= \frac{100^2}{n^2}\left[V(a)+V(c)-2\text{Cov}(a,c)\right] \\ &= \frac{100^2}{n^2}\left(\frac{N-n}{N-1}\right)\left[P_A+P_C-(P_A-P_C)^2\right] \qquad (3.4)\end{aligned}$$

그리고 위의 식 (3.4)을 비편향 추정하는 분산 추정량은 다음과 같이 얻을 수 있다.

$$v\left(\widehat{\text{BSI}}\right) = \frac{100^2}{n-1}\left(1-\frac{n}{N}\right)\left[p_a+p_c-(p_a-p_c)^2\right]. \qquad (3.5)$$

다음의 정리 3.1은 이제까지의 내용을 요약한 것이다.

정리 3.1 모집단은 N명으로 구성되어 있고 각 구성원은 '긍정', '보통', '부정' 중 하나의 특성값을 갖는다고 하자. 그리고 N중 n명을 단순확률추출한다고 하자. 그러면

(I) 식 (3.3)의 추정량 $\widehat{\text{BSI}}$는 식 (2.1)의 기업경기실사지수 BSI의 비편향 추정량이다.

(II) 추정량 $\widehat{\text{BSI}}$의 분산은 식 (3.4)과 같다.

(III) 식 (3.5)의 추정량은 $\widehat{\text{BSI}}$의 비편향 분산추정량이다.

증명: (i)과 (ii)는 쉽게 계산되므로 증명은 (iii)만 하기로 한다. 식 (3.4)에서 모수 P_A, P_B, P_C를 표본비율로 대체한 분산추정량 v_0를 고려하자.

$$v_0 = \frac{100^2}{n}\left(\frac{N-n}{N-1}\right)\left[p_a+p_c-(p_a-p_c)^2\right]$$

그리고 v_0에 기댓값을 취하면 다음을 얻을 수 있다.

$$E(v_0) = \text{Var}\left(\widehat{\text{BSI}}\right) - \frac{100^2}{n}\frac{N-n}{N-1}\text{Var}(p_a-p_c)$$

그런데 $\text{Var}(p_a-p_c) = \text{Var}(\widehat{\text{BSI}})/100^2$이므로, 계산 후 정리하면 다음을 얻는다.

$$E(v_0) = \text{Var}\left(\widehat{\text{BSI}}\right) \times \frac{(n-1)N}{n(N-1)}$$

따라서 $v_0 \times n(N-1)/(n-1)N = v(\widehat{\text{BSI}})$는 $\text{Var}(\widehat{\text{BSI}})$의 비편향추정량이 된다. □

예제 3.1: 기업경기조사에서 $n = 1{,}000$을 조사하여 긍정응답 $a = 450$, 보통응답 $b = 150$, 부정응답 $c = 400$을 얻었다고 하자. 그러면 기업경기실사지수 추정치는

$$\left(\widehat{\text{BSI}}\right) = \left(\frac{450}{1{,}000} - \frac{400}{1{,}000}\right) \times 100 + 100 = 105$$

가 된다. 그리고 $\widehat{\text{BSI}}$의 분산추정치는

$$v\left(\widehat{\text{BSI}}\right) = \frac{100^2}{999}\left[0.45+0.40-(0.45-0.40)^2\right] = 8.48$$

이다. 여기에서 모집단 크기 N은 큰 것으로 간주하여 $n/N \approx 0$으로 계산하였다.

3.3. $\widehat{\mathrm{BSI}}$의 신뢰구간

통상적인 표본조사에서는 통계결과를 발표할 때 추정치뿐만 아니라 추정치를 평가할 수 있는 측도들, 예를 들어 신뢰구간이나 상대표준오차를 병기하여 발표하기를 권장하고 있다. 본 소절에서는 $\widehat{\mathrm{BSI}}$의 신뢰구간을 구하는 방법을 알아보기로 한다.

2.1절에서 가정한 모집단에서 n명을 단순확률추출한다고 하자. 구성원은 '긍정', '보통', '부정' 중 하나를 대답한다. 새로운 변수 x_j를 도입하자.

$$x_j = \begin{cases} 1, & \text{긍정응답} \\ 0, & \text{보통응답}, \quad j = 1, \ldots, n. \\ -1, & \text{부정응답} \end{cases}$$

그러면 단순확률표본 s에서 조사된 표본평균은

$$\bar{x}_s = \frac{1}{n} \sum_{j \in s} x_j = p_a - p_c$$

가 되어 $\widehat{\mathrm{BSI}}$는 다음과 같이 표현된다.

$$\widehat{\mathrm{BSI}} = \bar{x}_s \times 100 + 100.$$

그런데 유한모집단의 단순확률표집에서 표본평균은 일정한 조건에서 표본의 크기가 증가함에 따라 정규분포로 수렴하는 성질이 있으므로 (예를 들면 Hajek, 1960; Lehmann, 1999, 116쪽) 다음의 결과를 얻을 수 있다.

$$\frac{\widehat{\mathrm{BSI}} - \mathrm{BSI}}{\sqrt{v\left(\widehat{\mathrm{BSI}}\right)}} \quad \rightarrow \quad N(0,1). \tag{3.6}$$

식 (3.6)를 이용하면 $\widehat{\mathrm{BSI}}$의 근사신뢰구간은 다음과 같이 얻는다.

$$\left[\widehat{\mathrm{BSI}} - z_{\frac{\alpha}{2}} \sqrt{v\left(\widehat{\mathrm{BSI}}\right)}, \quad \widehat{\mathrm{BSI}} + z_{\frac{\alpha}{2}} \sqrt{v\left(\widehat{\mathrm{BSI}}\right)}\right],$$

여기에서 $z_{\alpha/2}$는 표준정규분포에서 상위 $(1-\alpha/2)$에 해당하는 값이다.

이제 식 (3.6)의 $\widehat{\mathrm{BSI}}$의 점근 정규성(asymptotic normality)를 만족하기 조건을 살펴본다. 점근 정규성을 언급하기 위해서는 유한모집단 수열과 표본 수열에 대한 가정이 필요하다. ν번째 모집단을 $U_\nu = \{x_1, x_2, \ldots, x_{N_\nu}\}$라고 하고 대응되는 단순확률표본은 $s_\nu = \{x_{i_1}, \ldots, x_{i_{n_\nu}}\}$라고 하자. 모집단은 점차적으로 증가한다고 하고, 즉, 만일 $\nu < \nu'$이면 , $U_\nu \subset U_{\nu'}$모집단 크기와 표본크기는 점점 커진다고 하자. 즉, 만일 $\nu \to \infty$이면, $n_\nu \to \infty$, $N_\nu - n_\nu \to \infty$이다. 그러면 표준화된 표본평균이 표준정규분포로 수렴할 충분조건은 다음과 같다 (Lehmann, 1999, 116쪽, 정리 2.8.2):

$$\frac{\bar{x}_\nu - E(\bar{x}_\nu)}{\sqrt{\mathrm{Var}(\bar{x}_\nu)}} \to N(0,1), \qquad \nu \to \infty, \tag{3.7}$$

여기에서 $\bar{x}_\nu = \sum_{j \in s_\nu} x_{\nu j}/n_\nu$는 표본평균이다.

(a) ν가 커짐에 따라 표본추출비율 $f_\nu = n_\nu/N_\nu$는 0이나 1로 가까이 가지 않는다. 즉, 임의의 실수 δ_1, δ_2에 대하여 $0 < \delta_1 < f_\nu < \delta_2 < 1$.

(b) ν가 커짐에 따라

$$\frac{\max_{j=1,\ldots,N_\nu}\left(x_{\nu j}-\bar{X}_\nu\right)^2}{\sum_{j=1}^{N_\nu}\left(x_{\nu j}-\bar{X}_\nu\right)^2}\to 0, \qquad \nu\to\infty, \tag{3.8}$$

여기에서 $\bar{X}_\nu=\sum_{j=1}^{N_\nu}x_{\nu j}/N_\nu$는 ν번째 모집단의 모평균이다.

그런데 BSI 계산에서는 x_j가 -1, 0, 1 중 하나의 값을 취하고 $\bar{X}_\nu=P_{A_\nu}-P_{C_\nu}$이므로 식 (3.8)을 BSI 문제에 맞추어 계산하면 다음과 같다.

(a) 분자: $\max\left\{[1-(P_{A_\nu}-P_{C_\nu})]^2, (P_{A_\nu}-P_{C_\nu})^2, (1+P_{A_\nu}-P_{C_\nu})^2\right\}\leqq 4$.

(b) 분모: $\sum_{j=1}^{N_\nu}(x_{\nu j}-\bar{x}_{\nu\cdot})^2=N_\nu\left[P_{A_\nu}+P_{C_\nu}-(P_{A_\nu}-P_{C_\nu})^2\right]$.

따라서 식 (3.8)의 좌변이 0으로 수렴하기 위해서는 분자가 4보다 크지 않으므로 분모가 무한대로 커지면 되는데, 그러기 위해서는 $P_{A_\nu}+P_{C_\nu}-(P_{A_\nu}-P_{C_\nu})^2$의 값이 0으로 수렴하지 않아야 한다. 이를 만족시키는 충분조건은 $P_{A_\nu}\neq 1$, $P_{C_\nu}\neq 1$, $P_{E_\nu}=1-P_{A_\nu}-P_{C_\nu}\neq 1$이다. 즉, (i) 표본추출비율 f_ν가 일정한 상수로 수렴하고 (ii) 세 응답 중 어느 하나도 100%가 아니면 $\widehat{\text{BSI}}$는 정규분포로 수렴한다.

예제 3.2: 예를 들어, 세 응답 중 한 응답이 전체 응답이 되면 식 (3.8)의 분모는 0이 된다. 즉,

$$(P_A=1, P_B=P_C=0),\ \text{혹은}\ (P_B=1, P_A=P_C=0),\ \text{혹은}\ (P_C=1, P_A=P_B=0)$$

인 경우를 말한다. 이 경우 식 (3.8)은 0으로 수렴하지 않는다. 따라서 식 (3.8)이 수렴하기 위해서는 응답이 세 응답 중 한 응답으로 집중되어서는 안 된다. 그러나 그 경우는 응답자 모두가 긍정하거나, 부정하거나 혹은 보통이라고 응답하는 경우이므로 현실성은 높지 않다.

마지막으로 식 (3.6)가 성립하기 위해서는 분산추정량 $v(\widehat{\text{BSI}})$이 분산 $\text{Var}(\widehat{\text{BSI}})$의 일치추정량임을 보여야 한다. 채비쉐프(Chebychev) 부등식을 이용하면

$$P\left[\left|v\left(\widehat{\text{BSI}}\right)-\text{Var}\left(\widehat{\text{BSI}}\right)\right|\geqq\epsilon\right]\ \leqq\ \frac{\text{Var}\left[v\left(\widehat{\text{BSI}}\right)\right]}{\epsilon^2}$$

이 되므로 $v(\widehat{\text{BSI}})$의 일치성을 보이기 위해서는 $\text{Var}[v(\widehat{\text{BSI}})]\to 0$임을 보이면 충분하다 (예를 들어 Casella와 Berger, 2002). 이를 보이기 위하여 $v(\widehat{\text{BSI}})$의 형태를 일부 변형하자.

$$v\left(\widehat{\text{BSI}}\right)=\frac{100^4}{n_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)s_\nu^2,\quad s_\nu^2=\frac{1}{n_\nu-1}\sum_{j\in s_\nu}\left(x_{\nu j}-\bar{x}_\nu\right)^2$$

그리고

$$\begin{aligned}\text{Var}\left[v\left(\widehat{\text{BSI}}\right)\right]&=\frac{100^4}{n_\nu^2}\left(1-\frac{n_\nu}{N_\nu}\right)^2\text{Var}\left(s_\nu^2\right)\\&=\frac{100^4}{n_\nu^2}\left(1-\frac{n_\nu}{N_\nu}\right)^2\left[E\left(s_\nu^4\right)-\left(E\left(s_\nu^2\right)\right)^2\right]\\&\leqq\frac{100^4}{n_\nu^2}\left(1-\frac{n_\nu}{N_\nu}\right)^2\left[\frac{\mu_{\nu 4}}{n_\nu}-\frac{\sigma_\nu^4}{n_{\nu-1}}\left(\frac{n_\nu-3}{n_\nu}\right)\right],\quad \mu_{\nu 4}=\frac{1}{N_\nu}\sum_{j=1}^{N_\nu}\left(x_{\nu j}-\bar{X}_\nu\right)^4,\end{aligned} \tag{3.9}$$

여기에서 $\sigma_\nu^4 = [\sum_{j=1}^{N_\nu}(x_{\nu j} - \bar{X}_\nu)^2/(N_\nu - 1)]^2$, $\mu_{\nu 4} = \sum_{j=1}^{N_\nu}(x_{\nu j} - \bar{X}_\nu)^4/N_\nu$이다. 식 (3.9)의 부등호는 비복원 단순임의표본의 분산보다 복원 단순임의표본의 분산이 크기 때문에 발생한다 (Hansen 등, 1953, vol II, p.99). 이제 $x_{\nu j} = -1, 0, 1$인 모집단에 적용하면,

$$\mu_{\nu 4} = (P_{A_\nu} + P_{C_\nu}) - 4(P_{A_\nu} - P_{C_\nu})^2 + 6(P_{A_\nu} - P_{C_\nu})^2(P_{A_\nu} + P_{C_\nu}) - 3(P_{A_\nu} + P_{C_\nu})^4 \leqq 8$$
$$\sigma_\nu^4 = \left(P_{A_\nu} + P_{C_\nu} - (P_{A_\nu} + P_{C_\nu})^2\right)^2 \leqq 4$$

이므로 이를 대입하면 증가하는 모집단 수열에서 ν가 증가함에 따라 $v(\widehat{\mathrm{BSI}})$의 분산은 0으로 수렴함을 보일 수 있다. 즉, $\nu \to \infty$이면,

$$\mathrm{Var}\left[v\left(\widehat{\mathrm{BSI}}\right)\right] \leqq \frac{100^4}{n_\nu^2}\left(1 - \frac{n_\nu}{N_\nu}\right)^2\left[\frac{8}{n_\nu} - \frac{4}{n_\nu - 1}\left(\frac{n_\nu - 3}{n_\nu}\right)\right] \to 0$$

따라서 $v(\widehat{\mathrm{BSI}})$는 $\mathrm{Var}(\widehat{\mathrm{BSI}})$의 일치추정량이다. 결과적으로 식 (3.6)의 점근 정규성은 성립한다.

예제 3.3: 앞의 예제 3.1에서 $\widehat{\mathrm{BSI}}$의 95% 신뢰구간을 구하면 다음과 같다.

$$\left[105 - 1.96\sqrt{8.48},\ 105 + 1.96\sqrt{8.48}\right] = [99.3, 110.7].$$

3.4. $\widehat{\mathrm{BSI}}$의 상대표준오차

$\widehat{\mathrm{BSI}}$의 상대표준오차 추정량은 다음과 같다.

$$\mathrm{cv}\left(\widehat{\mathrm{BSI}}\right) = \frac{\sqrt{v(\widehat{\mathrm{BSI}})}}{\widehat{\mathrm{BSI}}} \times 100$$

예제 3.1에서 $\widehat{\mathrm{BSI}}$의 상대표준오차 추정치를 계산하면 $\mathrm{cv}(\widehat{\mathrm{BSI}}) = \sqrt{8.48}/105 \times 100 = 2.77\%$이 된다.

앞에서 설명한 $\widehat{\mathrm{BSI}}$의 신뢰구간과 상대표준오차는 모두 $\widehat{\mathrm{BSI}}$의 분산의 함수이다. 따라서 두 측도의 성질을 알아보기 위해서는 $\widehat{\mathrm{BSI}}$의 분산의 성질을 알아볼 필요가 있다. 식 (3.8)에 주어진 $\widehat{\mathrm{BSI}}$의 분산식을 $\widehat{\mathrm{BSI}}$와 대응하여 그린 산점도가 그림 3.1에 있다. 편의상 그림은 $(n \times \mathrm{Var}(\widehat{\mathrm{BSI}})/100^2, \widehat{\mathrm{BSI}})$를 그렸다. 마찬가지로 $\widehat{\mathrm{BSI}}$와 상대표준오차와의 관계를 보기 위하여 그린 산점도가 그림 3.2에 있다.

그림 3.1에서 알 수 있는 것은 BSI = 100 부근에서 분산의 폭이 가장 크게 나타나고 양 끝으로 갈수록 분산의 폭은 줄어든다는 사실이다. 이는 BSI 분포 $f(D)$에서 나타나듯이 BSI = 100이 되는 상황이 가장 많음을 말해준다. 그리고 BSI = 100을 중심으로 분산은 대칭을 이룬다.

그림 3.2에서 보면 BSI와 상대표준오차는 매우 밀접한 연관이 있다는 사실을 알 수 있다. BSI가 클수록 상대표준오차는 낮게 나타나고, BSI가 작을수록 상대표준오차는 크게 나타난다. 이는 그림 3.1에서도 나타났듯이 분자의 표준오차는 BSI = 100을 중심으로 대칭인데, 분모의 상대표준오차는 대칭이 아니기 때문에 생기는 현상이다. 이는 BSI에서는 상대표준오차가 적절한 평가 지표가 아님을 시사한다.

기업경기실사지수 이용자의 입장에서는 다음 시점의 경기가 좋아질 것인지 아닌지가 가장 궁금한 사항이다. 예를 들어 예제 3.1에서와 같이 $\widehat{\mathrm{BSI}} = 105$가 나왔다고 하자. $\widehat{\mathrm{BSI}}$가 100을 넘었으므로 다음 시점에 경기가 좋아진다고 볼 수 있는가? 통계 이론이 제시하는 답은 '그렇지 않다'이다. 왜냐하면 95% 신뢰수준에서 105는 신뢰구간 안에 포함되기 때문이다.

$$105 \in [99.3,\ 110.7]$$

따라서 비록 $\widehat{\mathrm{BSI}} = 105$로서 100은 넘지만 확률적으로 다음 시점에 경기가 나아질 것이라고 결론짓기는 어렵다.

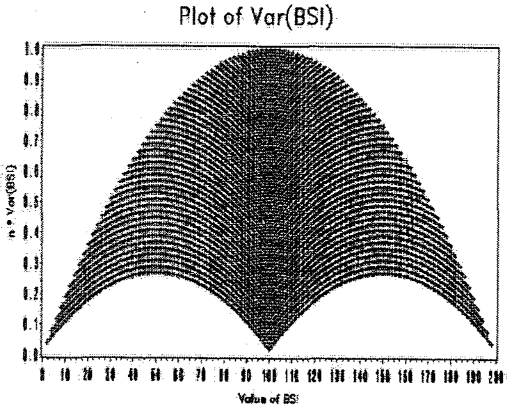

그림 3.1. BSI 분산의 산점도

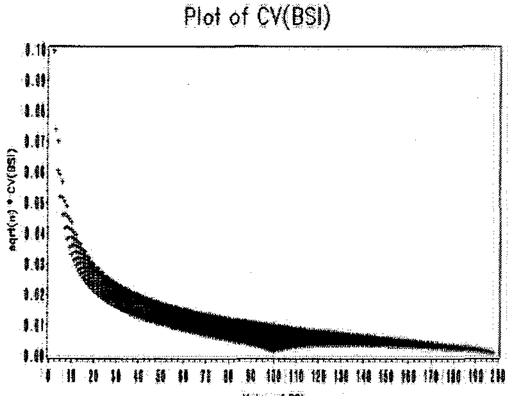

그림 3.2. BSI의 상대표준오차의 산점도

4. 결론

본 논문에서는 경기전망조사의 대표적인 지수인 기업경기실사지수에 대한 통계적 성질을 고찰하였다. 유한모집단에서 모집단 기업경기실사지수를 정의하였고, 단순확률표집을 가정한 후, 모집단 기업경기실사지수 추정량을 구하였다. 통상적으로 언급되는 기업경기실사지수는 사실상 기업경기실사지수 추정량이다. 그리고 기업경기실사지수의 기댓값, 분산, 비편향 분산추정량, 신뢰구간, 상대표준오차를 구하였다. 그리고 기업경기실사지수의 평가지표로서 상대표준오차보다는 신뢰구간이 더 합리적임을 언급하였다.

기업경기실사지수는 경제 현장에서는 매우 빈번하게 활용되는 지수임에도 불구하고 이에 대한 통계적 성질은 이제까지 규명된 바가 거의 없었다. 앞에서도 언급하였듯이 경제 현장에서는 다음달, 혹은 다음 분기에 전달 혹은 전 분기의 전망 지수를 확인할 수 있기 때문에 기업경기실사지수에 대한 통계적 성질을 규명할 필요성을 느끼지 못하였을 수도 있다. 그러나 엄밀히 말하면 단순확률표본과 같은 자기가중(self-weighting) 표본이 아니면 식 (3.3)의 기업경기실사지수 추정량은 비편향 추정량이 아니다. 따라서 원천적으로 편향된 결과를 보고 경기를 전망해야 하는 상황도 종종 발생한다.

본 연구에서는 가장 기본적인 단계로서 단순확률표집을 전제로 기업경기실사지수에 대한 통계적인 성질을 고찰하였다. 그런데 실제로는 복합설계를 기반으로 기업경기실사조사가 이루어지므로 일반적인 표집방법에 대한 기업경기실사지수에 대한 통계적 성질 고찰이 필요하다. 이는 향후 연구로 남긴다.

참고문헌

심상달 (2002). 기업경기실사지수의 경기 예측력 평가 분석, <한국개발연구원 보고서>.

유재준, 이상호 (2002). 합성 BSI 도출방안 연구, <전국경제인연합회 보고서>, CER-2002-28.

이상호 (2001a). 경기실사지수의 유용성 및 제고방안에 관한 연구, <전국경제인연합회 보고서>, CER-2001-22.

이상호 (2001b). 경기실사지수의 확률적 접근, <전국경제인연합회 보고서>, CER-2001-30.

한국은행 (2006). <경제지표해설>, 한국은행.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, Duxbury.

Curtin, R., Presser, S. and Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment, *Public Opinion Quarterly*, **64**, 413–428.

Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*, Houghton Mifflin.

Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population, *Publication of the Mathematical Institute of the Hungarian Academy of Science*, **5**, 361–374.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory*, Vol II. Theory, Wiley.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*, Springer.

OECD (2003). *Business Tendency Surveys*, A handbook.

Statistical Properties of Business Survey Index

Kyu-Seong Kim[1]

[1]Department of Statistics, University of Seoul

(Received January 2010; accepted February 2010)

Abstract

Business survey index(BSI) is an economic forecasting index made on the basis of the past achievement of the company and enterpriser's plan and decision for the future. Even the index is very popular in economic situations, only a little research result is known to the public. In the paper we investigate statistical properties of BSI. We define population BSI in the finite population and estimate it unbiasedly. Also we derive the variance of the estimated BSI and its unbiased estimator. In addition, confidence interval of the estimated BSI is proposed. We asserte that confidence interval of the estimated BSI is more reasonable than the relative standard error.

Keywords: Business survey index, multivariate hypergeometric distribution, self-weighting sample.

This work was supported by the University of Seoul 2009 Research Fund.

[1]Professor, Department of Statistics, University of Seoul, Seoul 130-743, Korea. E-mail: kskim@uos.ac.kr