



국민계정리뷰

Quarterly National Accounts Review



- [논고] 표본의 대표성 평가와 무응답 대체
- [자료] 빅데이터의 경제통계 활용 현황 및 시사점
- [자료] 세계산업연관표(MOT)의 구조와 이해
- 한국은행 경제통계 안내
- 2017년 4/4분기 통계공표일정

2017년 제3호



한국은행 인터넷 홈페이지를 통해 신속하고 편리하게
한국은행 정보를 이용하실 수 있습니다.

▣ 인터넷 홈페이지 주소 : <http://www.bok.or.kr>

국 민 계 정 리 뷰

QUARTERLY NATIONAL ACCOUNTS REVIEW

◆ [논고] 표본의 대표성 평가와 무응답 대체	1
◆ [자료] 빅데이터의 경제통계 활용 현황 및 시사점	28
◆ [자료] 세계산업연관표(WIOI)의 구조와 이해	46
◆ 한국은행 경제통계 안내	71
◆ 2017년 4/4분기 통계공표일정	83



한국은행
THE BANK OF KOREA

표본의 대표성 평가와 무응답 대체*

모집단 전체를 조사하지 않는 표본조사는 일부 자료를 표본 추출하여 조사하기 때문에 필연적으로 표본오차(sampling error)가 발생한다. 표본오차는 표본 설계 당시 크기를 파악할 수 있으며 이를 근거로 표본 규모를 결정한다. 그러나 비표본오차(non-sampling error)는 크기 파악이 어려우며 매우 큰 부분을 차지하는 것이 무응답과 이상점에 의해 발생한 오차이다. 따라서 조사에서 발생하는 무응답과 이상점을 적절히 처리함으로써 추정의 정확성 및 정밀성을 향상시킬 수 있다.

본 연구에서는 무응답이 존재하는 CSI, BSI 그리고 외감기업 자료를 이용하여 R-지수를 구한 후, 월별 또는 분기별 변화를 확인하여 표본의 대표성을 진단하였다. R-지수는 편향(bias)과 관련된 정확성을 나타내는 지표로 R-지수가 높으면 대표성과 정확성이 높다고 평가된다. 반면 무응답이 일정 패턴을 갖는다면 편향이 발생하게 되므로 R-지수 값은 낮아진다. 따라서 R-지수를 이용해 대표성을 평가함으로써 무응답 발생 원인을 찾을 수 있다. 또한 외감기업 자료처럼 관심변수가 연속형 자료이고 관심변수와 상관관계가 높은 독립변수가 존재하는 경우 사용할 수 있는 무응답 대체법을 연구하였다. 특히 연속형 자료인 경우 이상점의 존재는 추정의 정확성에 매우 큰 영향을 주므로 이상점 탐지(outlier detection) 및 처리 방법도 연구하였다.

분석결과 모두 높은 수준의 대표성과 정확성을 유지하는 것으로 나타났다. 다만 외감기업 자료의 경우 이상점 처리 후 대체법을 사용한다면 더욱 우수한 결과를 얻을 수 있을 것으로 판단된다.

I. 서론

II. R-지수 및 대표성 기본 개념

1. R-지수
2. R-지수 분석에 사용된 통계 분석 방법
3. 분석에 사용된 비교통계량
4. 가중치 보정법과 대체법

III. R-지수를 이용한 정확성 진단

1. CSI 분석 결과

2. BSI 자료 분석 결과
3. 외감기업 자료 분석 결과

IV. 외감기업 자료의 이상점 처리와 무응답 대체법

1. 무응답 대체법에 사용된 통계 분석 방법
2. 이상점 처리 후 대체 결과
3. 외감기업 자료의 이상점 처리 및 대체법 결론

IV. 결론 및 시사점

* 본고는 한국외국어대학교 통계학과 신기일 교수, 한국은행 황희진 차장이 작성한 것으로, 본고의 내용은 집필자의 개인의견으로서 한국은행의 공식견해를 나타내는 것은 아님.

I. 서론

모집단 전체를 조사하지 않는 표본조사는 일부 자료를 표본 추출하여 조사하기 때문에 필연적으로 오차가 발생하며 이를 표본오차(표집오차)라고 한다. 대부분의 표본조사에서 표본오차는 표본 설계 당시에 어느 정도 그 크기를 파악할 수 있다. 반면 실제 조사에서 발생하는 비표본오차는 그 크기를 파악하기가 매우 어려우며 이 비표본오차의 매우 큰 부분을 차지하는 것이 무응답에 의해 발생한 오차이다.

무응답은 표본 수를 감소시켜 추정의 정밀도(precision)와 정확도(accuracy)를 떨어뜨린다. 따라서 무응답으로 인해 발생하는 비표본오차를 줄여 추정의 정확도와 정밀도를 향상시키기 위한 다양한 방법이 연구되었으며 이미 잘 알려진 것처럼 가중치 보정법과 무응답 대체법이 널리 사용되고 있다.

무응답이 작은 조사는 추정의 정도가 높기 때문에 조사의 품질을 무응답 비율 또는 응답률로 평가하고 있으며 이를 위해 많은 조사 결과 보고서에 무응답률 또는 응답률을 수록하여 이를 조사 품질의 지표로 사용하고 있다. 그러나 응답률을 조사 결과의 정밀도를 결정하는 지표로 사용하는 것은 타당하지만 정확도 지표로 사용할 경우 오류를 범할 수도 있는 것으로 알려져 있다. 이는 작은 수의 무응답이라도 특정 패턴에 따라 발생하게 되면 모수 추정 시에 편향이 발생하기 때문이다. 이와 관련된 내용은 Bethlehem(2008)을 참조하기 바란다.

모수 추정의 우수성은 평균제곱오차(MSE: mean square error)를 기준으로 판단하며 평균제곱오차는 분산(variance)과 편향(bias)으로 구성되어 있다. 분산이 작으면 정밀도(precision)가 높다고 하고, 편향이 작으면 정확도(accuracy)가 높다고 표현한다.

정밀도의 경우 자료 수가 늘어나면 일반적으로 높아지며 이는 모수 추정량의 분산 추정량에 의해 쉽게 이해될 수 있다. 따라서 무응답 비율이 작아지게 되면 분석에 사용된 자료의 수가 줄어들지 않기 때문에 표본 설계 당시 예상되었던 정밀도가 낮아지지 않는다. 따라서 무응답 비율 또는 응답률을 정밀도의 지표로 사용하는 것은 매우 타당하다.

그러나 정확도의 경우에는 자료의 수와 크게 관계가 없기 때문에 정확도를 평가할 수 있는 지표가 필요하게 된다. 결국 정확도는 편향과 관계가 있고, 조사된 표본이 모집단을 잘 대표하게 되면 편향은 없으므로 정확도는 표본의 대표성을 이용하여 측정하게 된다.

최근 표본의 대표성(representativeness)을 판단할 수 있는 지수가 개발되었으며 이를 R-지수(R-indicator: representativeness indicator)라 부른다. Kruskal and Mosteller(1979)는 표본의 대표

성 개념에 관하여 연구하였고 최근 Schouten et al.(2009), Schouten et al.(2011)이 응답 대표성(response representativeness)의 정의를 구체적으로 제시하였으며 이후 여러 연구에서 이 개념이 사용되고 있다. 이들 논문에서는 R-지수의 기본 개념과 R-지수의 추정 방법이 연구되었다. 이후 R-지수의 신뢰구간을 추정하기 위해 사용되는 붓스트랩 방법 대신 선형 근사 분산 추정량이 Schouten et al.(2012)에서 제안되었다. 또한 Ouwehand and Schouten(2014)은 R-지수를 이용한 자료 분석을 실시하였으며 Schouten et al.(2013)에서도 R-지수를 이용한 자료 분석을 실시하였다.

본 연구에서는 무응답이 존재하는 CSI, BSI 그리고 외감기업 자료의 R-지수를 구하여 현재의 자료가 대표성을 유지하고 있는지 진단하였다. 특히 각 자료에서 얻어진 R-지수는 통계간 비교에 사용하는 것이 아니라 각 자료의 시점 간에 R-지수의 변화를 파악하여 대표성의 하락 또는 상승을 진단하는데 사용되기 때문에 본 연구에서도 월별 또는 분기별 자료의 R-지수 변화를 파악하여 대표성을 진단하였다.

다음으로 무응답 대체를 위한 이상점 처리 방법을 연구하였다. 이상점은 보조변수와 관심변수의 관계를 왜곡시키고 다중대체 시 분산을 크게 하여 무응답 대체의 효율을 떨어뜨리는 요인이 된다. 따라서 무응답 대체 전에 이상점을 적절히 처리함으로써 대체의 성능을 향상시킬 수 있다. 본 연구에서는 무응답 대체 시 효과를 볼 수 있는 연속형 자료인 외감기업 자료의 V7, V8 자료를 관심변수로 하여 이상점 처리 방법을 연구하였다. 이때 보조변수로는 V6이 사용되었다. 먼저 외표준화잔차를 이용하여 이상점을 탐지한 후 탐지된 이상점을 적절히 처리하여 이상점의 영향력을 감소시켰다. 이상점의 영향력 축소에 사용된 방법에는 각 자료가 동일한 가중치를 사용할 경우를 가정하였으며 제안된 방법을 사용한 결과 무응답 대체 성능이 매우 향상된 것을 모의실험을 통하여 확인하였다.

본 연구의 구성은 다음과 같다. 먼저 2장에서 R-지수의 기본 개념을 설명하였다. 특히 R-지수를 구하기 위해 사용되는 통계 모형인 일반화선형모형(*generalized linear model*)을 설명하였고 R-지수를 추정하기 위해 사용되는 표본 R-지수도 설명하였다. 3장에서는 2장에서 설명한 R-지수를 CSI, BSI, 그리고 외감기업 자료에 적용하여 얻어진 R-지수 진단 결과를 설명하였다. 4장에서는 외감기업 자료를 이용하여 이상점의 처리 유무가 무응답 대체에 얼마나 효과적인지를 살펴보았다. 5장에 본 연구의 결론 및 시사점이 있다.

II . R-지수 및 대표성 기본 개념

1. R-지수(representativeness indicator)

가. R-지수(population R-indicator) 개념

대표성에 대한 정의는 두 가지로 나누어진다. 먼저 강한 대표성(strongly representative)의 정의는 다음과 같다.

$$\rho_k = \Pr(r_k = 1 | s_k = 1) = \rho, \quad k = 1, \dots, N$$

이는 모든 자료에서 특정 k 번째 자료가 표본으로 추출되었을 때 응답 확률이 모두 같은 경우를 의미한다. 이 경우는 MCAR(missing completely at random)에 해당되는 경우로 무응답으로 인해 발생한 결측이 모수 추정에 편향을 발생시키지 않아 대표성이 유지된다.

반면 약한 대표성(weakly representative)은 모집단이 L 개의 층으로 나누어졌다는 조건에서 다음과 같이 정의된다.

$$\bar{\rho}_h = \frac{1}{N_h} \sum_{k=1}^{N_h} \Pr(r_{hk} = 1 | s_{hk} = 1) = \rho, \quad k = 1, \dots, N_h, h = 1, \dots, L$$

즉 층별 응답 확률의 평균이 층과 무관하게 일정하다는 것이다. 이제 각각의 응답 확률 $\rho_k, k = 1, \dots, N$ 가 모든 원소에서 알려져 있다고 가정하면 응답 확률의 표준편차는 다음과 같이 구해진다.

$$S(\rho) = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (\rho_k - \bar{\rho})^2}$$

또한 이를 기반으로 작성된 R-지수는 다음과 같이 정의된다.

$$R(\rho) = 1 - 2S(\rho)$$

여기서 응답 확률 ρ_k 의 표준편차 $S(\rho)$ 는 $0 \leq S(\rho) \leq 0.5$ 를 만족하므로 R-지수의 범위는 항상 $0 \leq R(\rho) \leq 1$ 이 된다.

따라서 응답 확률 ρ_k 가 각 자료에 따라 많이 다르게 되면 표준편차 $S(\rho)$ 가 커지게 되므로 R-지수는 작아지게 되며 결국 ρ_k 의 크기에 상관없이 그 값이 일정하면 R-지수는 큰 값을 갖게 되고 대표성이 높은 것으로 평가된다.

나. 표본 R-지수(sample R-indicator)

실제 자료 분석에서 모든 응답 확률 $\rho_k, k=1, \dots, N$ 가 알려진 경우는 없으며 따라서 ρ_k 는 추정되어야 한다. 이를 위해 흔히 사용되는 통계 모형은 일반화선형모형(generalized linear model)인 로짓(logit) 모형과 프로빗(probit) 모형이다. 두 모형은 통계 분석에서 흔히 사용하는 모형으로 모형에 관한 간단한 설명을 2장에 수록하였으며 자세한 내용은 Agresti(2002)를 참조하면 된다. 여러 논문과 관련 서적에서 서로 다른 기호가 사용되기 때문에 본 연구에서는 Bethlehem 등(2008)과 Schouten 등(2009)에서 사용한 기호를 사용하여 표본 R-지수를 구하는 방법을 설명하였다.

먼저 추정된 응답 확률(response probability, response propensity)을 $\hat{\rho}_k$ 이라 하면 응답 확률의 평균 추정치는 다음 식 (II-1)과 같이 얻어진다.

$$\hat{\rho} = \frac{1}{N} \sum_{k=1}^N \rho_k \frac{s_k}{\pi_k} = \frac{1}{N} \sum_{k=1}^n \frac{\hat{\rho}_k}{\pi_k} = \frac{1}{N} \sum_{k=1}^n w_k \hat{\rho}_k \quad (\text{II-1})$$

여기서 s_k 는 응답인 경우 “1”이고 무응답이면 “0”인 지시변수이다. 또한 π_k 는 추출확률이고 따라서 $w_k = \frac{1}{\pi_k}$ 로 설계 가중치이다. 만약 모집단이 L 개의 층으로 이루어졌고 각 층의 층별 가중치가 동일하다면 $w_{k(k \in h)} = \frac{N_h}{n_h}$, 즉 층별 모집단 수를 층별 표본 수로 나눈 가중치가 사용된다. 결국 응답 평균 추정치는 가중치를 이용하여 계산된 가중 평균값으로 얻어진다.

다음으로 R-지수의 추정값인 $\hat{R}(\rho)$ 는 식 (II-2)에 의해 구해진다. 여기서 N 은 모집단 총 수이고 n 은 표본 총 수이다.

$$\hat{R}(\rho) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{k=1}^N \frac{s_k}{\pi_k} (\hat{\rho}_k - \hat{\rho})^2} = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{k=1}^n w_k (\hat{\rho}_k - \hat{\rho})^2} \quad (\text{II-2})$$

2. R-지수 분석에 사용된 통계 분석 방법

가. 일반화선형모형(generalized linear model)

응답률은 일반적으로 알려져 있지 않다. 따라서 자료에서 응답률을 추정해야 하며 이때 사용되는 모형이 일반화선형모형이다. 일반화선형모형에서 흔히 사용하는 모형은 로짓모형(logit model)과 프로빗모형(probit model)이다. 이 두 모형 중 하나를 모형 선택 기준을 적용해 선택한 후 자료에 적합하여 응답률을 추정하게 된다. 잘 알려진 것처럼 두 모형은 자료가 성공과 실패 또는 응답과 무응답처럼 “0”과 “1” 자료로 얻어질 때 사용하는 방법이다. 모형을 적합하여 얻어진 예측확률을 경향점수(성향점수, propensity score)라고도 부른다. 본 연구에서 사용할 일반화선형모형은 다음과 같다.

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}, \quad i = 1, \dots, n$$

일반화선형모형에서 흔히 사용하는 변환 방법은 로짓 변환과 프로빗 변환이다. 로짓 변환은 로지스틱 모형을 가정할 경우에 사용되며 프로빗 모형은 정규분포의 누적확률함수(cdf, cumulative distribution function), Φ 를 이용한다. 따라서 로짓 모형인 경우는

$$\eta_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \text{ 변환을 사용하고 프로빗 모형인 경우는 } \eta_i = \Phi(\pi_i)^{-1} \text{ 변환을 사용한}$$

다. 여기서 π_i 는 응답 확률이다. 응답 확률 π_i 는 p 개의 독립변수 x_1, x_2, \dots, x_p 에 따라 달라진다. p 개의 독립변수는 층화 추출 시에 사용된 층화 변수와 응답 확률에 관련이 있다고 판단되는 공변량을 모두 포함한다.

동 분석은 SAS 등 기초적인 통계 패키지를 이용하면 쉽게 수행할 수 있다. Bethlehem 등(2008)과 Schouten 등(2009)에 의하면 R-지수의 변동 상황을 확인하기 위해서는 기준이 되는

같은 모형과 같은 독립 변수가 사용되어야 의미가 있다. 따라서 로짓 모형과 프로빗 모형 중에서 하나의 모형을 선택하고 자료를 잘 설명할 수 있는 독립 변수를 선택하여야 한다. 본 연구에서는 모든 월별 또는 분기별 자료에서 모형 적합이 이루어지도록 최적의 층화 변수와 공변량을 선택하였으며 모형 선택 기준 통계량으로 AIC(Akaike information criterion)와 BIC(Bayesian information criterion)를 사용하여 최적의 모형을 찾아낸 후 이를 기준으로 R-지수를 계산하였다.

나. 층별 최종 가중치

표본 설계는 모집단을 잘 대표할 수 있는 최적의 표본을 추출하기 위해 사용된다. 다양한 표본 설계 방법이 연구되었고 또한 사용되고 있다. 가장 강력하면서도 흔히 사용되는 방법이 층화추출법(stratified sampling)이다. 이 방법은 모집단을 여러 개의 층으로 나눈 후 각 층에서 표본을 추출하는 방법이다. 표본을 추출하기 위해 표본 틀(sampling frame)이 사용되는데 이 표본 틀에는 L 개의 층으로 층을 나눌 때 필요한 모든 정보가 포함되어 있다. 이제 주어진 표본 틀의 자료 수, 즉 h 번째 층의 조사 모집단 수(N_h)와 표본 수(n_h)를 이

용하면 층별 가중치 $w_h = \frac{N_h}{n_h}$ 가 구해진다.

다. 분석에 사용된 비교 통계량

본 연구에서는 R-지수가 변화하였을 때 편향이 어떻게 변하는지 또는 제곱근평균제곱오차(root mean squared error)가 어떻게 변화하는지 확인하였다. 편향과 RMSE는 각각 추정치의 정확성과 정밀성을 파악할 수 있는 대표적인 비교 통계량이다. 본 연구에서는 척도 변수인 CSI와 BSI를 위해 편향, 절대편향 그리고 RMSE를 사용하였으며 또한 관심변수가 연속형 자료인 외감기업 자료의 경우에는 상대편향을 추가하여 비교하였다. 사용된 비교 통계량의 정의는 다음과 같다.

$$\text{편향(Bias)} : Bias = \frac{1}{R} \sum_{r=1}^R (\hat{t}_r - t_r^{TRUE})$$

$$\text{절대 편향(Absolute bias)} : AB = \frac{1}{R} \sum_{r=1}^R |\hat{t}_r - t_r^{TRUE}|$$

제공근 MSE(Root mean squared error) : $RMSE = \left\{ \frac{1}{R} \sum_{r=1}^R (\hat{t}_r - t_r^{TRUE})^2 \right\}^{1/2}$

상대 편향(Relative bias) : $RB = \frac{1}{R} \sum_{r=1}^R (\hat{t}_r - t_r^{TRUE}) / t_r^{TRUE}$

여기서 \hat{t}_r 은 r 번째 반복에서 얻어진 추정값이고 t_r^{TRUE} 는 참값

라. 가중치 보정법과 대체법

(1) 가중치 보정법

모수 추정 및 R-지수에 사용되는 최종 가중치 w^f 는 표본 설계 당시에 주어진 표본 가중치(sample weight)에 여러 단계의 보정 과정을 거친 후 최종적으로 얻어진다. 여러 보정 과정 중에서 가장 중요하게 사용하는 것은 무응답 보정, 이상점 보정 그리고 사후 층화 보정(벤치마킹 보정)이다. 본 연구에서는 이상점 보정과 사후 층화 보정을 사용하지 않고 무응답 보정만을 고려하였다. 무응답 보정은 표본 설계 당시에 배정된 표본 수가 실제로 조사되지 않고 더 적은 수의 자료가 얻어진 경우에 사용된다.

이제 h 번째 층에 n_h 개의 표본 수가 배정되었으나 실질적으로 r_h 개의 조사 자료가 얻어졌다고 하면 무응답 보정인자는 $f_h^r = \frac{n_h}{r_h}$ 로 계산된다. 따라서 최종 가중치는 다음과 같이 얻어진다.

$$w_h^f = \frac{N_h}{n_h} \times f_h^r = \frac{N_h}{n_h} \times \frac{n_h}{r_h} = \frac{N_h}{r_h}$$

이는 층별 모집단 수를 실질적으로 얻어진 층별 자료수로 나누어준 결과이다. 이렇게 얻어진 최종 가중치 w_h^f 가 모수 추정 및 R-지수 계산에 사용된다.

(2) 무응답 대체법

무응답 대체법은 무응답으로 인해 발생한 결측값을 가장 타당하다고 생각되는 값으로 대체하는 방법이다. 매우 다양한 무응답 대체법이 개발되었으며 자료에 따라 최적의 대체

법을 사용하면 된다. 흔히 사용하는 방법 중의 하나가 회귀모형을 이용한 다중대체법이며 본 연구에서도 이 방법을 사용하였다. 이 방법은 SAS/ PROC MI와 같이 다양한 통계 패키지를 사용하여 수행할 수 있다.

III. R-지수를 이용한 정확성 진단

1. CSI 분석 결과

가. 자료 설명 및 일반화선형모형 분석

2015년 11월에서 2016년 3월까지 약 2,200개의 CSI 자료가 R-지수 분석에 사용되었다. 이 분석에서 독립변수로 사용된 자료는 가구주 연령, 성별, 직업, 수입, 가구원수, 지역이다. 특히 사용된 독립변수 중에서 가구주 연령이 연속형 변수인 공변량(covariate)으로 사용되었다. 개별 소비자동향지수를 종합하여 얻어지는 소비자심리지수(composite consumer sentiment index: CCSI)를 구하기 위해서는 모든 개별 자료가 사용되어야 하지만 본 분석에서는 일부 변수만 분석이 가능하여 CCSI의 R-지수 분석을 실시하지 않았다. 따라서 본 분석에서는 하나의 개별 소비자동향지수(consumer survey index)를 이용하여 R-지수를 분석하였다.

R-지수를 구하기 위해 고려된 모형은 프로빗 모형과 로짓 모형이며 최종 모형에 사용된 독립 변수 선택은 모형 및 변수 선택 기준인 AIC와 BIC를 사용하였다. 이때 직업, 지역, 가구원 수 등은 층이 너무 세분화되어 있어 일부 층을 합쳐 새로운 층을 만든 후 최종 독립 변수로 사용하였다. 즉 직업의 경우 근로자와 비 근로자 등 2개, 가구원 수의 경우는 1-2인 가구, 3인 이상 가구 등 2개, 지역은 특별시, 광역시, 기타시도의 3개 층으로 구분하여 사용하였다. 전체적으로 로짓 모형과 프로빗 모형의 AIC와 BIC 값이 유사하여 어떤 모형을 사용하여도 유사한 결과가 나올 것으로 판단되지만 프로빗 모형이 나쁘지 않고, 2016년 1월 자료에서 로짓 모형이 적합이 안 되었기 때문에 본 연구에서는 프로빗 모형을 사용하였으며 이를 기반으로 응답 확률인 경향점수, $\hat{\rho}_k$ 를 구하였다. <표 III-1>은 모형 선택을 위한 월별 AIC와 BIC 결과이다.

〈표 III-1〉 각 월별 모형별 AIC와 BIC

시점	모형	AIC	BIC
15년 11월	logit	-706,6727	-501,6091
	probit	-701,5379	-496,2943
15년 12월	logit	167,8263	372,8900
	probit	170,3568	375,4205
16년 01월	logit	모형 식별 안 됨.	
	probit	-432,0794	-232,7120
16년 02월	logit	202,8764	402,2438
	probit	202,3524	401,7199
16년 03월	logit	-267,3436	-62,2799
	probit	-265,3705	-60,3069

나. CSI 자료의 R-지수 분석 결과

다음의 <표 III-2>에 월별 R-지수 분석 결과를 수록하였다. 결과를 살펴보면 R-지수는 0.9에서 작아지다 0.85정도에서 멈추고 있는 것을 확인할 수 있다. 이는 동 자료가 매우 안정적으로 정확성이 유지되고 있음을 말해주고 있다. 그러나 무응답 비율이 증감을 보이는 반면 R-지수는 감소하는 경향을 보이고 있어 무응답이 특정 패턴을 갖고 발생하고 있는지와 같은 무응답 특성을 살펴볼 필요가 있다.

〈표 III-2〉 각 월별 CSI 무응답 비율과 R-지수

	조사 월				
	15_11	15_12	16_01	16_02	16_03
무응답 비율	0.04636	0.0700	0.05682	0.07273	0.06136
R-지수	0.90322	0.87601	0.87877	0.85346	0.85871

결론적으로 현재 CSI 자료는 일정 수준 이상의 정확성을 확보하고 있다고 판단된다. 다만 현재의 수준보다 높은 정확성을 유지하기 위해서는 무응답을 보이고 있는 자료의 응답률을 높이는 방안을 강구해야 하며 특히 특정 패턴을 갖고 무응답이 발생하는지 또는 특정 층에서 많은 무응답이 발생하는지 등을 확인할 필요가 있다.

다. CSI 자료에서 추가 응답이 있을 경우 정밀성과 정확성 비교 결과

다음의 <표 III-3>은 정확성과 정밀성 진단을 위해 2015년 11월 자료를 이용하여 얻어진 비교 통계량 분석 결과이다. 약 120개의 자료가 무응답으로 나타났으며 무응답 분석에서 얻어진 무응답의 예상 응답률 ρ_k 를 계산한 후 이중 ρ_k 가 작은 값 30개 중 각각 5, 15, 25개 자료, 그리고 50개 중 5, 15, 25, 35개 자료가 추가 응답일 경우의 R-지수를 계산하였다.

또한 R-지수가 향상되었을 때 추정의 정확성과 정밀성을 파악할 수 있는 비교 통계량인 편향(bias), 절대편향 그리고 제곱근 MSE도 계산되었다. 물론 비교 통계량을 구하기 위해서는 참값이 있어야 하지만 본 연구에서는 참값이 존재하지 않는다. 따라서 먼저 결측값을 이월대체법(carry-over imputation method)을 이용하여 대체한 후 완전자료(complete)를 구성하였다. 만들어진 완전자료를 이용하여 추정된 값이 모수의 참값이라는 가정 하에 편향과, 절대편향 그리고 제곱근 MSE를 구하였다. 이때 반복수는 $R=300$ 을 사용하였다. 따라서 대체값의 영향으로 본 모의실험 결과는 현실과 차이가 있을 수 있다.

추정 방법으로는 두 가지가 사용되었다. <표 III-3>은 층별로 무응답 가중치 보정을 통하여 얻은 결과이며 <표 III-4>는 가중치 보정법 대신 회귀모형을 이용한 무응답 대체법을 이용하여 얻은 결과이다. <표 III-3>의 결과를 살펴보면 R-지수는 0.9로 높은 수준이다. 또한 편향은 약 0.1, 절대편향은 약 0.15의 매우 작은 값을 보이고 있어 추정의 정확성 면에서 매우 우수하기는 하나 편향이 모두 양수값으로 현 상태에서는 약간 과소/과대 추정이 되고 있는 상태로 판단된다. 반면 대체법을 사용할 경우 편향은 거의 “0”에 가깝고 양수와 음수가 모두 나오기 때문에 편향은 없는 것으로 판단된다. 두 방법 모두 절대편향과 제곱근 MSE는 비슷한 결과를 주고 있다.

이제 추가 응답이 있을 경우 결과를 살펴보면 R-지수가 빠르게 상승하고 있으며 절대편향과 제곱근 MSE도 빠르게 향상되고 있다. 반면 편향의 경우 가중치 보정법을 사용한 경우에는 무응답이 응답으로 바뀌었을 때 결과가 향상되고 있으나 대체법을 사용한 경우에는 큰 향상을 보이고 있지 않다. 이와 같은 결과를 종합해 보면 동 자료는 무응답에 관한 대표성이 있으며 대체법을 사용한다면 편향은 감소할 것으로 판단된다. 그러나 CSI와 같은 척도자료의 경우 대체법 사용 효과가 크지 않고, 추정에서 편향의 크기가 작기 때문에 가중치 보정법을 사용하여도 문제가 되지 않을 것으로 판단된다.

〈표 III-3〉 CSI의 비교 통계량 결과(가중치 보정법 사용)

n_sam	R-지수 (2015년 11월)		편향		절대편향		제공근 MSE	
	30	50	30	50	30	50	30	50
0	0.904	0.904	0.118	0.088	0.155	0.143	0.192	0.175
5	0.915	0.913	0.110	0.080	0.149	0.135	0.185	0.166
15	0.937	0.931	0.086	0.066	0.131	0.124	0.163	0.152
25	0.959	0.949	0.063	0.058	0.114	0.121	0.142	0.148
35		0.965		0.040		0.103		0.128

〈표 III-4〉 CSI의 비교 통계량 결과(대체법 사용)

n_sam	R-지수 (2015년 11월)		편향		절대편향		제공근 MSE	
	30	50	30	50	30	50	30	50
0	0.904	0.904	0.000	-0.016	0.175	0.170	0.213	0.227
5	0.915	0.913	0.016	-0.043	0.173	0.161	0.213	0.234
15	0.937	0.931	0.002	-0.020	0.156	0.154	0.197	0.214
25	0.959	0.949	0.023	0.002	0.147	0.158	0.195	0.195
35		0.965		-0.018		0.140		0.188

2. BSI 자료 분석 결과

BSI 분석은 2015년 11월에서 2016년 3월까지의 약 3,300개 자료 중에서 하나의 관심 변수를 선택하여 수행되었다. 기업경기실사지수(business survey index: BSI)는 업종별 BSI와 이들의 가중평균인 산업별(제조업, 비제조업) BSI로 나누어진다. 본 연구에서는 제조업과 비제조업으로 구분된 산업별 BSI의 R-지수를 구하였다

R-지수를 구하는 방법은 CSI와 동일하다. 다만 CSI 자료에 비해 보조변수로 사용되는 변수의 수가 작아 R-지수 계산에 필요한 일반화선형모형의 변수 선택과정이 생략되었다. 표본 R-지수를 구할 때에는 가중치를 사용해야 한다. BSI는 층별 설계 가중치와 사업체 규모 정보를 인자화한 사업체 규모인자를 보정인자로 사용하여 얻어진 최종 가중치를 사용하며 이를 기반으로 최종 가중치가 구해진다. 이렇게 얻어진 최종 보정가중치가 “1”이다. 본 연구에서는 가중치를 “1”로 사용하였다. 다음의 <표 III-5>는 산업별 BSI 결과이다.

<표 III-5>

월별 산업별 BSI R-지수

	15_11	15_12	16_01	16_02	16_03
전산업	0.944	0.950	0.932	0.928	0.915
제조업	0.932	0.931	0.914	0.925	0.909
비제조업	0.905	0.923	0.901	0.907	0.886

<표 III-5>를 살펴보면 모두 0.9 이상으로 높은 R-지수를 보이고 있으나 지수가 점차 감소하고 있음을 파악할 수 있다. 제조업의 경우 산업별 R-지수보다 낮은 결과를 주고 있으나 역시 0.9 이상이며 비제조업의 경우도 우수한 결과를 주고 있다. 다만 제조업과 비제조업 모두 R-지수가 점차 감소하는 경향을 보이므로 주의해서 무응답 특성을 살펴볼 필요가 있다. BSI 분석에서는 R-지수를 구하기 위해 사용되는 보조변수의 수가 작아 R-지수를 관리하는데 충분한 정보를 제공하지 못하였다.

다음의 <표 III-6>과 <표 III-7>은 2016년 2월의 R-지수와 그에 따른 편향, 절대 편향, 그리고 제공근 MSE 결과이다. CSI 분석과 같이 비교 통계량을 구하기 위해서는 참값이 있어야 하지만 BSI 분석에서도 참값이 존재하지 않는다. 따라서 CSI에서 사용한 이월 대체법으로 결측값을 대체하여 완전자료(complete)를 만들었다. 이후 완전자료를 이용하여 추정된 참값이라는 가정 하에 편향과, 절대편향 그리고 제공근 MSE를 구하였다. 이때 반복수는 $R=300$ 을 사용하였다. 따라서 모의실험 결과는 현실과 차이가 있을 수 있다.

<표 III-6>

BSI 비교 통계량 분석(가중치 보정 방법 사용)

n_sam	R-지수 (16_02)		편향(bias)		절대 편향		제공근 MSE	
	제조업	비제조업	제조업	비제조업	제조업	비제조업	제조업	비제조업
0	0.925	0.907	0.332	0.632	0.538	0.688	0.664	0.817
20	0.965	0.929	0.142	0.593	0.435	0.645	0.553	0.765
40	0.980	0.952	0.093	0.495	0.354	0.544	0.453	0.648

<표 III-7>

BSI 비교 통계량 분석(무응답 대체법 사용)

n_sam	R-지수 (16_02)		편향(bias)		절대 편향		제공근 MSE	
	제조업	비제조업	제조업	비제조업	제조업	비제조업	제조업	비제조업
0	0.925	0.907	0.061	0.109	0.699	0.544	0.879	0.671
20	0.965	0.929	0.060	0.085	0.590	0.534	0.747	0.673
40	0.980	0.952	0.038	0.107	0.479	0.478	0.608	0.601

<표 III-6>과 <표 III-7>의 결과는 다음과 같다. 먼저 편향과 절대편향을 살펴보면 모두 0.7보다 작다. 따라서 정확성 면에서 매우 안정적인 결과를 주고 있다고 판단된다. 또한 편향이 모두 양수로 나와 현재 자료는 무응답으로 인해 약간 과소/과대 추정되고 있다고 판단된다. 물론 정확한 참값을 알 수 없기 때문에 해석에 유의할 필요가 있다. 가중치 보정법을 사용한 경우와 무응답 대체법을 사용한 결과를 비교해 보면 편향을 기준으로 무응답 대체법이 보다 우수한 것을 확인할 수 있다. 그러나 절대편향과 제곱근 MSE를 비교해보면 두 방법에는 큰 차이가 없다.

CSI와 같이 BSI도 척도자료이기 때문에 무응답 대체효과가 크지 않고 분석결과 편향도 크지 않다. 따라서 현재 사용하고 있는 방법인 가중치 보정법은 추정에 있어 큰 무리가 없다고 판단된다. 물론 모의실험의 완전자료 생성을 위해 사용된 무응답 대체에 충분한 정보를 사용하지 않았기 때문에 실질적으로 두 방법은 큰 차이를 보이지 않는 것이 타당하다.

3. 외감기업 자료 분석 결과

외감기업 자료에 대한 R-지수 분석은 2015년 1/4분기부터 3/4분기까지 약 3,300개의 자료 중 V1, V2, V3와 V6, V7, V8 변수가 사용되었다. 이중 V1, V2, V3는 동일한 가중치를 사용하여 분석하였으며 V6, V7, V8은 가중치 설계 당시의 층별 가중치를 사용하였다. 이때 각 변수는 연속형 자료이다.

가. 동일한 가중치를 사용한 경우

외감기업 자료의 R-지수 분석은 약 3,300개의 표본 수가 있는 2015년 1분기부터 3 분기까지의 자료 중에서 V1, V2, V3 변수가 사용되었으며 비교 통계량 계산을 위해 각 변수의 총합이 이용되었다. 층은 산업분류와 사업체 규모로 분류되었기 때문에 이 자료는 보조 변수로 사용되었다. 이 자료에서는 층화변수(산업분류, 사업체 규모) 외에 연속형 자료로 사용할 수 있는 보조변수가 없는 상태이다. 따라서 외감기업 분석 결과도 BSI 경우와 같이 R-지수를 이용한 정확성 진단에 충분한 보조정보를 제공하지 못하고 있다.

<표 III-8>에는 R-지수 분석 결과를 수록하였다. 결과를 살펴보면 0.726~ 0.767 사이에서 얻어졌다. 각 변수 V1, V2, V3의 결과를 보면 세 자료 모두에서 결측이 동시에 발생하는 자료가 많아 R-지수는 비슷하거나 같은 값을 주고 있다. 또한 각 분기별로도 유사한 값을 보여주고 있다. 한편 각 분기별로 R-지수를 비교하였을 때 R-지수가 감소하지 않고 있어

대표성은 유지되고 있음을 확인할 수 있다.

<표 III-8> 각 분기별 외감기업 자료의 R-지수

변수	15_01	15_02	15_03
V1	0.726	0.767	0.757
V2	0.726	0.767	0.756
V3	0.730	0.767	0.760

다음의 <표 III-9>와 <표 III-10>은 비교 통계량을 이용한 V1 자료의 분석 결과이다. 여기서는 간단한 방법으로 완전자료를 만들기 위해 SAS Proc MI의 EM 방법을 사용하였다. <표 III-9>의 결과인 보정 가중치를 이용한 상대편향을 살펴보면 약 6% 내외의 결과를 보여주고 있다. 또한 모든 편향과 상대편향은 양수를 보이고 있어 추정값은 과소 추정되고 있음을 알 수 있다. 반면 대체를 이용하여 구한 상대편향 결과를 살펴보면 약 0.1%에서 0.5%로 나타났으며 음수와 양수가 모두 나타나고 있다. 따라서 대체법을 사용하면 과소 추정되는 것을 방지할 수 있게 된다. 물론 상대편향과 제공근 MSE는 두 방법에서 유사한 결과를 주고 있다.

본 연구에서도 충분한 보조변수와 공변량 자료가 없어 정확한 R-지수 분석이 어렵다. 또한 모의실험을 위한 완전자료 생성과정도 정보부족으로 인해 현실과 차이가 있을 수 있다. 그러나 외감기업 자료와 같은 연속형 자료인 경우에는 충분한 보조정보가 있다면 척도자료에 비해 대체효과가 매우 크게 나타난다. 따라서 대체법을 사용하는 것이 타당하다.

<표 III-9> 외감자료 V1의 비교 통계량 분석(가중치 보정 방법 사용, 15_01)

n_sam	R-지수		bias		상대 편향		절대편향		제공근 MSE	
	300	600	300	600	300	600	300	600	300	600
0	0.726	0.726	-134E6	-127E6	-0.061	-0.059	138E6	129E6	156E6	144E6
100	0.815	0.780	-150E6	-101E6	-0.069	-0.047	151E6	105E6	165E6	119E6
200	0.884		-167E6		-0.077		168E6		180E6	
300		0.884		-77E6		-0.036		80E6		93E6
500		0.948		-59E6		-0.028		61E6		68E6

〈표 III-10〉 외감자료 V1의 비교 통계량 분석(대체법 사용, 15_01)

n_sam	R-지수		bias		상대 편향		절대편향		제곱근 MSE	
	300	600	300	600	300	600	300	600	300	600
0	0.726	0.726	-130E5	494E3	-0.0048	0.0047	140E6	131E6	181E6	165E6
100	0.815	0.780	-41E5	-207E3	-0.0009	-0.0005	119E6	115E6	149E6	146E6
200	0.884		-77E5		-0.0028		104E6		130E6	
300		0.884		-531E3		0.0019		91E6		115E6
500		0.948		-402E3		0.0011		58E6		73E6

나. 층별 가중치를 이용한 경우

설계 가중치를 이용한 경우의 외감기업 자료 R-지수 분석은 3개 분기의 V6, V7, V8 변수가 사용되었다. 층화변수인 산업분류와 사업체 규모가 보조변수로 사용되었으며 동일한 가중치를 사용한 경우와 큰 차이를 보이고 있지 않다. 다음의 <표 III-11>은 R-지수 분석 결과이다.

〈표 III-11〉 각 분기별 외감기업 자료의 R-지수

변수	15_01	15_02	15_03
V7	0.724	0.777	0.775
V8	0.724	0.763	0.761
V9	0.725	0.754	0.761

IV. 외감기업 자료의 이상점 처리와 무응답 대체법

외감기업 자료의 이상점 처리와 무응답 대체를 위해 2015년 1분기에서 3분기 자료 중 V6, V7, V8 자료가 사용되었다. 분석은 변수 변환, 이상점 탐지 및 처리 그리고 무응답 대체 순으로 이루어졌다. 이때 각 자료의 가중치는 동일하다고 가정하였다.

1. 무응답 대체법에 사용된 통계 분석 방법

가. 변환

분석에 사용된 이상점 처리 및 무응답 대체 방법은 자료가 정규분포를 따르거나 대칭인 분포를 따를 경우에 매우 효율적이다. 따라서 자료가 정규분포 또는 대칭이 아닌 경우 이를 정규분포 또는 대칭성을 유지하도록 변환할 필요가 있다. 외감기업 자료의 경우 오른쪽으로 꼬리가 긴 분포이기 때문에 변환을 고려할 수 있다.

변환을 위해서는 Box-Cox 변환과 Yeo-Johnson 변환을 사용할 수 있으며 가장 흔히 사용하는 방법이 Box-Cox 변환이다. 이때 자료에 음수가 있다면 충분히 큰 양수 c 를 더하여 모든 자료를 양수로 만든 후 사용한다. Box-Cox 변환 공식은 다음과 같다.

$$Z = (X + c)^\lambda, \lambda > 0$$

$$Z = \log(X + c), \lambda = 0$$

다음으로 자료에 음수가 있는 경우 적당한 양수 c 값을 사용하지 않고 자료에 직접 사용할 수 있는 변환 방법이 Yeo-Johnson 변환이다. Yeo-Johnson 변환 공식은 다음과 같다.

$$Y = \begin{cases} \frac{(X+1)^\lambda - 1}{\lambda}, & 0 < X, \lambda \neq 0 \\ \log(X+1), & 0 \leq X, \lambda = 0 \\ -\log(-X+1), & X=0, \lambda = 2 \\ \frac{(-X+1)^{2-\lambda} - 1}{2-\lambda}, & 0 < X, \lambda \neq 2 \end{cases}$$

이상의 변환 공식을 이용하여 본 자료에 변환을 적용하는 것이 타당한지를 확인하였다.

나. 이상점 탐지법

CSI 및 BSI와 같은 3점 또는 5점 척도 자료에서는 이상점이 존재하지 않기 때문에 이상점 처리는 필요하지 않다. 그러나 외감기업 자료와 같이 연속형 자료인 경우에는 이상점이 존재할 수 있으므로 이 경우에는 적절한 이상점 처리 방법이 필요하다.

본 연구에서는 이상점을 탐지하는 방법으로 흔히 사용되고 있는 외표준화잔차를 사용하였다. 최근 개발된 θ -IPOD 방법은 아직 SAS를 이용하여 얻을 수 없기 때문에 본 연구에서는 제외하였다. θ -IPOD에 관한 내용은 She and Owen (2011)을 참조하기 바란다.

외표준화잔차(externally studentized residual, studentized deleted residual) 정의는 다음과 같다.

$$t_i = \frac{r_i}{s(d_i)} = \frac{r_i}{s(i) \sqrt{(1-h_{ii})}}$$

여기서 $r_i = y_i - \hat{y}_{(i)}$ 이고 y_i 는 관측값, 그리고 $\hat{y}_{(i)}$ 는 i 번째 관측값을 제거한 후 얻어진 예측값을 의미한다. 또한 h_{ii} 는 지렛값 또는 레버리지이다. t_i 는 자유도 $(n-p-1)$ 인 t 분포를 따르는 것으로 알려져 있다. 외표준화잔차는 SAS/Proc REG의 출력결과(Rstudent)에서 쉽게 얻을 수 있는 결과이다.

다. 기초 분석 결과

(1) Box-Cox 변환 결과

자료가 “0”이상이고 “0”이 포함되어 있는 V6의 경우와 같은 자료에서는 흔히 $c=1$ 을 사용한다. 또한 V7와 V8처럼 자료에 음수가 있으면 자료의 최소값에 해당되는 값의 절대값에 “1”을 더한 값을 c 로 사용하여 변환한다. 이때 양수 c 가 매우 크게 되면 변환의 효과가 없다는 것은 이미 잘 알려져 있다. 본 자료의 경우 자료에 매우 큰 음수 값이 있어 이를 양수로 바꾸기 위해서는 매우 큰 c 값을 사용해야하므로 Box-Cox 변환을 본 자료에 적용하는 것은 의미가 없다고 판단된다.

(2) Yeo-Johnson 변환 결과

Yeo-Johnson 변환에서 원 자료의 분포 형태를 기준으로 살펴보면 log 변환 등 $\lambda < 0.5$ 인 값을 사용하는 것이 타당하다고 판단된다. 그러나 $\lambda < 0.5$ 인 값을 본 자료에 적용한 결과 변환 후 분포가 봉우리가 두 개인 양봉 형태의 분포로 나타났다. 따라서 $\lambda < 0.5$ 인 값을 사용하는 것은 의미가 없다. 또한 $\lambda \geq 0.5$ 인 값을 적용할 경우 변환된 분포의 왜도가 매우 커서 대칭성을 만족하지 못한다. 이러한 결과를 종합해 볼 때 Yeo-Johnson 변환을 본 자료에 적용하는 것은 타당하지 않다.

이상의 결과를 종합할 때 본 자료 분석에는 변환을 실시하지 않은 원자료를 사용하는 것이 가장 타당하다고 판단된다.

(3) 원자료를 이용한 회귀분석 결과 및 이상점 탐지

결측값을 대체하는 대체 방법에서 가장 중요한 것은 사용하는 보조변수 또는 독립변수 정보량의 크기이다. 현재 V7을 위해 사용하는 보조 변수는 V6이며 상관분석을 통해 두 변수 간에 높은 상관관계가 있는 것을 확인하였다. 그러나 자료에 이상점이 있는 경우에는 무응답 대체 효과가 매우 떨어질 수 있다. 이를 위해 본 연구에서는 이상점을 처리하지 않은 경우와 처리한 경우의 결과를 살펴보았다.

먼저 <표 IV-1>는 V6와 V7을 이용한 분기별 회귀분석 결과이다. 결과를 살펴보면 기울기 추정값은 0.067에서 0.090정도이며 p-값은 모두 0.0001보다 작아 유의한 것을 확인할 수 있다. 또한 R^2 도 0.5에서 0.63 정도로 높은 것을 확인할 수 있다. 따라서 V7의 대체를 위해 보조변수로 V6를 사용하는 것은 매우 타당하다. 유사한 결과가 <표 IV-2>인 V8에서도 나타난다. 다만 3분기 V8의 R^2 이 상대적으로 낮게 나온 것은 주목할 필요가 있다.

〈표 IV-1〉 V6와 V7을 이용한 분기별 회귀분석 결과

분기	추정값	p- 값	R^2
2015Q1	0.06772	<0.0001	0.6373
2015Q2	0.09910	<0.0001	0.5193
2015Q3	0.09096	<0.0001	0.6278

〈표 IV-2〉 V6와 V8을 이용한 분기별 회귀분석 결과

분기	추정값	p- 값	R^2
2015Q1	0.06772	<0.0001	0.6373
2015Q2	0.09910	<0.0001	0.5193
2015Q3	0.09096	<0.0001	0.6278

다음으로 V6와 V7 자료의 산점도를 확인한 결과 각 분기별로 많은 수의 이상점이 탐지되었으며 따라서 이상점을 적절히 처리하여 이상점의 영향을 줄일 필요가 있다. 이는 이상점이 존재하면 자료의 분산이 커지게 되고 또한 다중대체법에서 대체된 값의 분산을 매우 크게 만드는 요인이 되기 때문이다.

물론 무응답 대체에서는 층별로 이상점이 탐지되고 처리되기 때문에 본 결과를 그대로 적용할 수는 없지만 특별한 경우가 아니면 먼저 이상점을 처리한 후 무응답 대체하여야 한다.

흔히 일반 회귀분석에서는 이상점 처리 방법으로 이상점을 제거하는 경우가 있지만 표본조사에서는 이상점을 제거하지 않고 이상점을 타당한 값으로 대체하여 사용하는 것이 일반적인 방법이다.

2. 이상점 처리 후 대체 결과

가. 모의실험 자료 생성 방법

이 절에서는 이상점을 처리한 후 결측값을 대체하였을 때 이상점 처리 효과를 살펴보았다. 이를 위해 먼저 V6, V7, V8 자료에서 결측값을 제거한 완전자료를 만든 후 완전자료를 정상자료와 이상점 자료로 나누었다. 여기서 이상점 탐지 방법으로 1.가.절의 이상점 탐지

법을 사용하고 이때 이상점 판단기준은 외표준화잔차 2.54 이상을 사용하였다.

정상자료를 22회 복원추출하여 유사 모집단(pseudo population)을 만들고 여기에 이상점 자료 세트를 추가하여 이상점이 포함된 최종 유사모집단 자료를 만들었다. 만들어진 유사 모집단에서 2,000개의 표본을 랜덤으로 추출하였다. 추출한 표본에 이상점을 포함시키기 위해 이상점 자료에서 100개를 랜덤으로 추출한 후 이 자료를 표본에 추가하였다.

따라서 표본은 2,100개의 자료로 구성된다. 이때 자료의 중복을 피하기 위해 강제로 추가한 이상점 자료와 2,000개 자료에 포함된 이상점 자료가 동시에 표본에 있는 경우에는 이상점 자료를 제거하였다. 따라서 정확한 표본 수는 2,100개보다 약간 작아지게 된다. 최종적으로 각 자료의 평균이 추정되어 유사모집단의 평균과 비교되었다.

나. 이상점 처리 후 결과 비교

(1) 이상점을 신뢰구간의 상, 하한으로 대체한 경우의 무응답 대체 결과

다음의 <표 IV-3>은 V7 자료에 결측이 있을 때 먼저 이상점을 탐지한 후 이상점을 신뢰 상한 또는 신뢰하한 값으로 대체하여 평균을 추정한 결과이다. 즉 예측값보다 큰 이상점이면 상한값으로 대체하고 이상점이 예측값보다 작으면 하한값으로 대체한다.

결과를 살펴보면 평균 추정값은 참값에 비해 과대 추정되며 따라서 편향은 항상 양수가 된다. 무응답 대체법을 사용하였기 때문에 결측 개수에 거의 무관한 결과가 나왔으나 결측 개수가 많게 되면 일반적으로 편향, 절대편향, 제곱근 MSE가 커지는 경향이 나타난다. V8 자료의 결과는 <표 IV-4>에 수록되었으며 이러한 경향은 <표 IV-4>에서도 확인된다.

결론적으로 이상점을 처리하지 않은 것보다 이상점을 상한 또는 하한 값으로 대체하는 것이 모든 비교 통계량을 기준으로 하였을 때 우수한 결과를 주는 것을 확인할 수 있다.

<표 IV-3> V7 자료에서 이상점을 상하한으로 대체 한 경우의 무응답 대체 결과

결측 개수	편향	편향_처리	절대편향	절대편향_처리	제곱근 MSE	제곱근 MSE_처리
100	4592.16	4185.33	4592.16	4185.33	24257090	20357528
200	4647.11	4210.56	4647.11	4210.56	24903524	20608314
300	4653.31	4204.71	4653.31	4204.71	24864230	20542742

〈표 IV-4〉 V8 자료에서 이상점을 상하한으로 대체 한 경우의 무응답 대체 결과

결측 개수	편향	편향_처리	절대편향	절대편향_처리	제공근 MSE	제공근 MSE_처리
100	5241.89	5148.95	5241.89	5148.95	31848084	30866933
200	5271.47	5172.84	5271.47	5172.84	32092166	31259361
300	5238.98	5165.63	5238.98	5165.63	31910823	31085169

(2) 이상점을 예측값으로 대체한 경우의 무응답 대체 결과

다음의 <표 IV-5>는 V7 자료에 결측이 있을 때 이상점을 탐지한 후, 이상점을 예측값으로 대체하여 평균을 추정한 결과이다. 결과를 살펴보면 앞의 분석결과와 유사하다. 평균 추정값은 참값에 비해 과대 추정되며 따라서 편향은 항상 양수가 된다. <표 IV-3>과 <표 IV-5>를 비교하면 이상점을 예측값으로 대체하였을 때 우수한 결과를 주고 있다.

이와 같은 경향은 V8자료의 결과인 <표 IV-6>에도 나타난다. 다만 본 모의실험의 결론은 이상점을 예측값으로 대체하는 것이 우수한 결과를 준다는 것이지만 본 자료가 실제 모집단 자료를 이용한 것이 아니며 설계 가중치를 사용한 것이 아니기 때문에 본 모의실험 결과를 외감기업 자료에 바로 적용하는 것은 삼가야 한다. 당연한 결론이지만 자료에 따라 상, 하한을 대체하는 것이 우수한 결과를 줄 수도 있기 때문에 어떤 값으로 이상점을 처리하느냐 하는 것은 자료에 따라 선택하여야 한다.

〈표 IV-5〉 V7 자료에서 이상점을 예측값으로 대체 한 경우의 무응답 대체 결과

결측 개수	편향	편향_처리	절대편향	절대편향_처리	제공근 MSE	제공근 MSE_처리
100	4692.15	4123.48	4692.15	4123.48	25468857	20136124
200	4733.85	4143.90	4733.85	4143.90	25873339	20217158
300	4772.39	4143.57	4772.39	4143.57	26376217	20340325

〈표 IV-6〉 V8 자료에서 이상점을 예측값으로 대체 한 경우의 무응답 대체 결과

결측 개수	편향	편향_처리	절대편향	절대편향_처리	제공근 MSE	제공근 MSE_처리
100	5543.60	5345.33	5543.60	5345.33	35395398	33373850
200	5567.77	5373.58	5567.77	5373.58	35860296	33748950
300	5584.73	5383.29	5584.73	5383.29	36175450	34050078

3. 외감기업 자료의 이상점 처리 및 대체법 결론

외감기업 자료에는 다수의 이상점이 존재하는 것을 확인하였다. 따라서 정확한 무응답 대체를 통한 평균 추정을 위해서는 적절한 이상점 처리가 필요하다. 본 연구에서는 이상점 탐지법으로 간단한 방법인 외표준화잔차법을 사용하였으나 이미 다수의 이상점 탐지법이 개발되어 있다. 따라서 다양한 연구 및 분석을 통해 최적의 이상점 탐지법을 찾아낼 필요가 있다.

또한 모의실험 결과 이상점을 처리 한 후 무응답 대체를 실시하는 것이 이상점을 처리 하지 않은 것에 비해 우수한 결과를 준다는 것도 확인하였다. 물론 탐지된 이상점을 대체 하는 다양한 방법도 이미 개발되어 있으므로 타당한 방법을 선택하여 사용해야 한다.

V. 결론 및 시사점

본 연구에서는 CSI, BSI 그리고 외감기업 자료의 표본 대표성을 살펴보았다. CSI 자료의 경우 다양한 보조변수를 독립변수로 활용하여 R-지수 분석을 실시하였으며 정확성 면에서 매우 우수한 결과를 얻었다. 즉 무응답으로 인해 발생한 결측이 있음에도 대표성에는 큰 문제가 없다. 다만 월별로 R-지수가 감소하고 있으므로 무응답 특성을 분석하여 R-지수가 감소하는 이유를 파악할 필요가 있다.

추정을 위해 무응답 대체법을 사용하게 되면 편향을 줄일 수 있으나 절대편향과 제곱근 MSE는 가중치 보정법을 사용한 결과와 유사한 값을 주고 있고 척도 자료의 경우 무응답 대체 효과가 크지 않기 때문에 가중치 보정법을 사용하여도 문제가 없을 것으로 판단된다. 물론 비교 통계량을 구하기 위해 사용된 참값은 단순한 이월 대체 방법이 사용되었으므로 이 결과를 실제 자료 분석 결과로 확대하는 것은 주의해야 한다.

BSI 자료의 경우 상대적으로 높은 R-지수를 보이고 있다. 또한 편향과 절대편향을 고려하였을 때 본 자료는 충분히 대표성이 있다고 판단된다. 다만 제조업과 비제조업 모두 R-지수가 감소하는 경향을 보이고 있으므로 무응답 특성을 지속적으로 살펴볼 필요가 있다. 또한 구해진 R-지수에 의미를 부여하기 위해서는 충분한 보조자료, 즉 독립변수가 필요하나 본 연구에서는 충분한 수의 독립변수가 존재하지 않았다.

비교 통계량을 구하기 위해서 BSI 분석에서도 이월 대체법을 사용하여 완전자료(무응답이 없는 자료)를 생성하였다. 따라서 본 연구에서 얻어진 결과를 그대로 적용하기에는 한계가 있다. 그러나 가중치 보정법과 대체법을 비교한 결과 큰 차이를 보이지 않고 있으므로 가중치 보정 방법을 사용하는 것은 타당하다.

외감기업 자료의 경우 상대적으로 낮은 R-지수를 보이고 있지만 다른 두 통계와 R-지수를 직접 비교하여 대표성을 평가하는 것은 의미가 없다. R-지수를 구하기 위해 사용된 보조변수의 정보가 충분하지 않고 모의실험을 위한 완전자료 생성 시 간단한 무응답 대체법을 사용하여 결측값을 대체하였기 때문에 분석 결과에 큰 의미를 부여하기는 어렵다. 그러나 분기별 R-지수가 감소하지 않고 오히려 증가하는 경향을 보이고 있어 대표성에는 큰 문제가 없다고 판단된다.

또한 무응답 대체법의 성능이 가중치 보정 방법에 비해 우수한 결과를 주고 있으므로 무응답 처리를 위해서는 무응답 대체법을 사용할 필요가 있으며 특히 무응답 대체법 중에서 연속형 자료에 흔히 사용되는 보조변수를 이용한 대체법을 사용한다면 추정의 정확성

및 정밀성을 충분히 확보할 수 있을 것으로 판단된다.

외감기업 자료의 무응답 대체 분석에서는 이상점 탐지와 이상점 처리, 특히 이상점 대체 방법이 연구되었다. 이미 다양한 이상점 탐지법과 이상점 대체법이 개발되었기 때문에 많은 연구와 분석을 통하여 본 자료에 맞는 최적의 방법을 찾아낼 필요가 있다. 찾아낸 방법으로 이상점을 처리한 후 무응답 대체법을 사용한다면 더욱 우수한 결과를 얻을 수 있을 것으로 판단된다.

참고문헌

- Agresti, A. (2002). Categorical data analysis, Wiley Series in probability and Statistics, New York: John Wiley & Sons. INC.
- Bethlehem, J., Cobben, F., and Schouten, B. (2008), Indicator for the representativeness of survey response, Proceedings of Statistics Canada Symposium, Data Collection: Challenges, Achievements and New Directions.
- Kruskal, W., and Mosteller, F. (1979), Representative sampling III: Current statistical literature, International Statistical Review, Vol. 47, pp. 245-265.
- Ouwehand, P., and Schouten, B., (2014), Measuring Representativeness of Short-Term Business Statistics, Journal of official Statistics, Vol. 30, No. 4, pp. 623-649.
- Schouten, B., Bethlehem, J., Beullens, K., Kleven, O., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., and Skinner, C. (2012), Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. International Statistical Review, Vol. 80, No. 3, pp. 382-399.
- Schouten, B., Calinescu, M., and Luiten, A. (2013), Optimizing quality of response through adaptive survey designs, Survey Methodology, Vol. 39, No. 1, pp. 29-58.
- Schouten, B., Cobben, F., and Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, Vol. 35, No. 1 pp 101-113.
- Schouten, B., Shlomo, N., and Skinner, C. (2011), Indicators for monitoring and improving survey response, Journal of Official Statistics, Vol. 27, pp. 231-253.
- She, Y. and Owen, A. B. (2011), Outlier detection using nonconvex penalized regression, Journal of the American Statistical Association, 106, 626-639

빅데이터의 경제통계 활용 현황 및 시사점

정보화 시대에 빅데이터에 대한 경제적·사회적 관심이 높아지고 통계 환경이 변화됨에 따라 빅데이터의 공식통계 활용 방안에 관한 논의가 국제기구 및 주요국 통계기관을 중심으로 시작되었다. 현재 한국은행에서도 주요 경제통계 작성에 빅데이터를 활용하는 방안을 모색하고 있다. 이와 관련하여 빅데이터가 경제통계에 활용된 해외 사례를 살펴보고 빅데이터의 통계적 활용 가능성을 평가해보고자 한다.

빅데이터의 활용 사례를 살펴본 결과 스캐너 데이터, 웹 스크래핑 데이터, 지급결제 데이터 등 비교적 정형화된 빅데이터가 물가, 소비지출 등 경제통계 작성에 부분적으로 이용되고 있다. 소셜미디어, 뉴스 등 텍스트 데이터와 인터넷 검색 데이터도 경제 및 금융 관련 지표 분석에 널리 활용되고 있다. 이러한 사례들은 빅데이터의 특성에 기인한 기초자료로서의 한계에도 불구하고 향후 빅데이터가 경제통계 작성 및 분석에 유용하게 활용될 가능성이 있음을 시사한다.

따라서 빅데이터가 쉽게 적용 가능한 경제통계 영역을 발굴하여 시험편제를 실시해보고, 기존 활용 사례들을 벤치마킹하여 유용성을 검증할 필요가 있다. 또한 빅데이터의 정제·처리·분석을 위한 전문적 지식을 습득하고 빅데이터에 적합한 통계 작성기법에 대한 장기적인 연구도 필요하다.

I. 검토배경

II. 빅데이터의 경제통계 활용 현황

1. 물가지표
2. 심리지표
3. GDP 관련 지표
4. 기타

III. 시사점 및 향후 과제

I. 검토배경

최근 인터넷 및 모바일 기기의 확산으로 빅데이터(big data)에 대한 사회적 관심이 높다. 빅데이터는 데이터의 규모(volume), 생성속도(velocity), 형태(variety) 측면에서 기존의 데이터와 상당한 차이가 있다. 따라서 빅데이터를 처리·분석할 수 있는 관련 기술 개발, 의미 있는 정보 추출 및 새로운 가치 창출에 대한 기대가 확산되고 있다.

이처럼 빅데이터의 출현으로 “데이터”에 대한 중요성과 기대가 높아진 반면 기존 통계 작성 환경은 점점 열악해지고 있다. 1인 가구, 맞벌이 가구 등의 증가로 가구 대상 서베이 조사의 애로사항이 많아지고, 정보보호 요구 및 업무상 비협조 등으로 사업체 대상 조사의 응답거부율도 높아지고 있다. 반면 디지털 경제 하에 신규로 포착해야 할 통계의 수요는 커지고 있다. 이러한 한계를 극복하기 위해 통계 작성기관들이 행정자료, 온라인 자료 등 새로운 형태의 데이터에 주목하게 되었다.

빅데이터의 공식통계 활용 방안에 대한 본격적인 논의는 UN, Eurostat, OECD 등 국제기구를 중심으로 진행되어 왔다. UN 통계위원회(UN Statistical Commission)는 2013년 제45차 회의에서 빅데이터를 정식의제로 다루었고 글로벌 워킹그룹(Global Working Group, GWG)을 구성¹⁾하여 빅데이터 프로젝트 등 관련 활동을 지원하기 시작하였다. 한편 UN 유럽경제위원회(UN Economic Commission for Europe, UNECE)는 고위 전문가 그룹(High-Level Group for the Moderation of statistical production and services, HLG)이 주축이 되어 빅데이터의 개념 및 과제 등에 관한 가이드라인을 제시하고 국제협력 프로젝트에 참여²⁾하고 있다. Eurostat, OECD 등도 주요국의 통계 작성기관과 함께 빅데이터 활용을 위한 국제적 논의와 프로젝트에 동참하고 있다.

이와 같이 빅데이터의 통계적 활용에 대한 관심이 높아짐에 따라 한국은행에서도 주요 경제통계 작성에 빅데이터를 활용하는 방안을 모색하고 있다. 빅데이터는 빠르게 변화하는 정보화 사회를 잘 반영하는 유용한 자료이지만 전통적 방식의 통계작성에 실제로 활용되려면 극복해야 할 한계와 과제들이 많다. 따라서 빅데이터의 경제통계 활용 사례 연구를 통해 빅데이터의 통계적 활용 가능성과 잠재가치를 점검해 볼 필요가 있다. II장에서는 빅데이터가 경제통계 작성 및 분석에 활용된 구체적 사례를, III장에서는 빅데이터의 통계적 활용 가치와 시사점을 살펴보고자 한다.

1) <http://unstats.un.org/bigdata>

2) UNECE Big Data Inventory 참고

II . 빅데이터의 경제통계 활용 현황

빅데이터는 자료의 형태에 따라 일정한 규칙을 갖고 체계적으로 정리된 수치형태의 정형적(structured) 데이터와 텍스트, 사진, 동영상 등의 비정형적(unstructured) 데이터로 구분된다. 정형적 데이터는 그 자체로 의미 해석이 가능하고 통계로 활용하기 용이하나 비정형적 데이터는 데이터의 구조가 복잡하여 전문적인 처리 및 분석 과정을 거친 후에 의미 있는 통계로 활용 가능하다.

UNECE는 <표 1>과 같이 빅데이터를 정보소스의 유형에 따라 사회관계망(social networks), 거래내역(traditional business systems) 및 사물인터넷(internet of things) 자료로 구분하였다. 이 중 공공기관이나 기업의 사업과정을 매개로 하는(process-mediated) 거래내역 자료와 센서, 컴퓨터 등 기계장치에 의해 생성되는(machine-generated) 사물인터넷 자료는 정형적 데이터에 가깝다. 반면 사회관계망 자료는 인간의 사회활동 과정에 생성되는 정보(human-sourced information)로 주로 비정형적인 형태를 가진다. 이하에서는 이러한 다양한 유형의 빅데이터가 주요 경제통계에 활용된 사례를 살펴보도록 한다.

<표 1>

빅데이터의 분류

사회관계망 자료	거래내역 자료	사물인터넷 자료
<ul style="list-style-type: none"> - 소셜미디어(페이스북, 트위터 등) - 블로그, 코멘트 - 개인문서 - 사진 - 동영상(유튜브 등) - 인터넷 검색 - 모바일 데이터 - 사용자 생성 지도 - 이메일 	<ul style="list-style-type: none"> • 공공기관 생성자료 <ul style="list-style-type: none"> - 의료기록 • 기업 생성자료 <ul style="list-style-type: none"> - 상업적 거래 - 은행/증권 기록 - 전자상거래 - 신용카드 	<ul style="list-style-type: none"> • 센서 데이터 (고정센서) <ul style="list-style-type: none"> - 홈자동화 - 기후/오염 센서 - 교통센서/웹캠 - 과학센서 - 보안/감시용 비디오/사진 (이동센서) - 휴대폰 위치 - 자동차 - 위성사진 • 컴퓨터 시스템 데이터 <ul style="list-style-type: none"> - 로그 - 웹로그

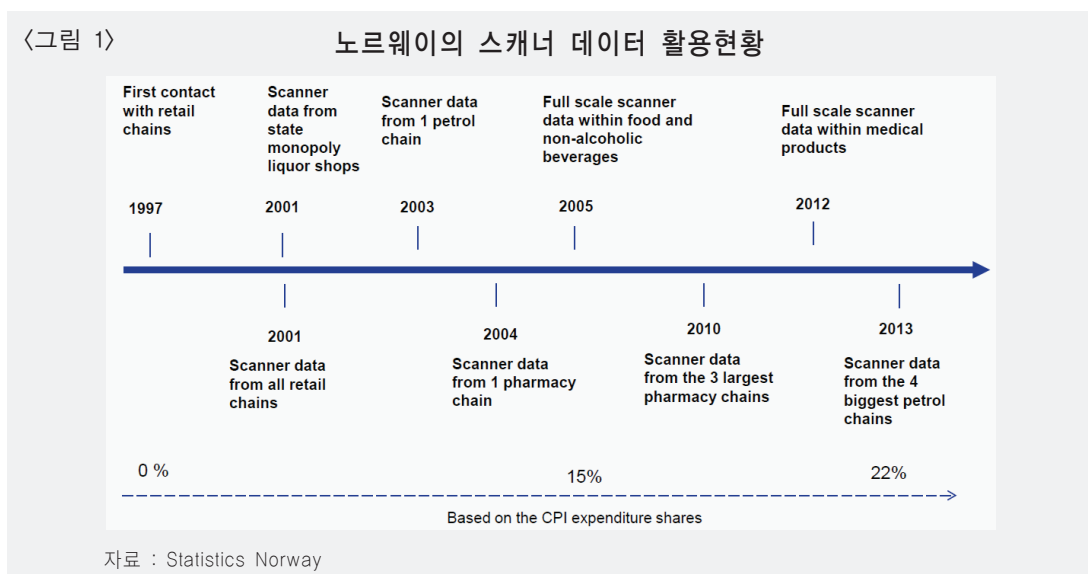
출처 : UNECE (2013) : Classification of Types of Big Data

1. 물가지표

소비자, 생산자, 수출입 물가지수 등 대표적인 공식 물가지수는 주로 거래 규모가 큰 대표 품목을 표본조사하여 작성한다. 그러나 품목의 종류 및 거래 채널이 다양해지고 제품의 생성주기 및 수명도 단축됨에 따라 서베이 기반 물가지수를 보완하는 데 빅데이터를 이용하는 방법들이 제안되고 있다.

가. 스캐너 데이터(scanner data)

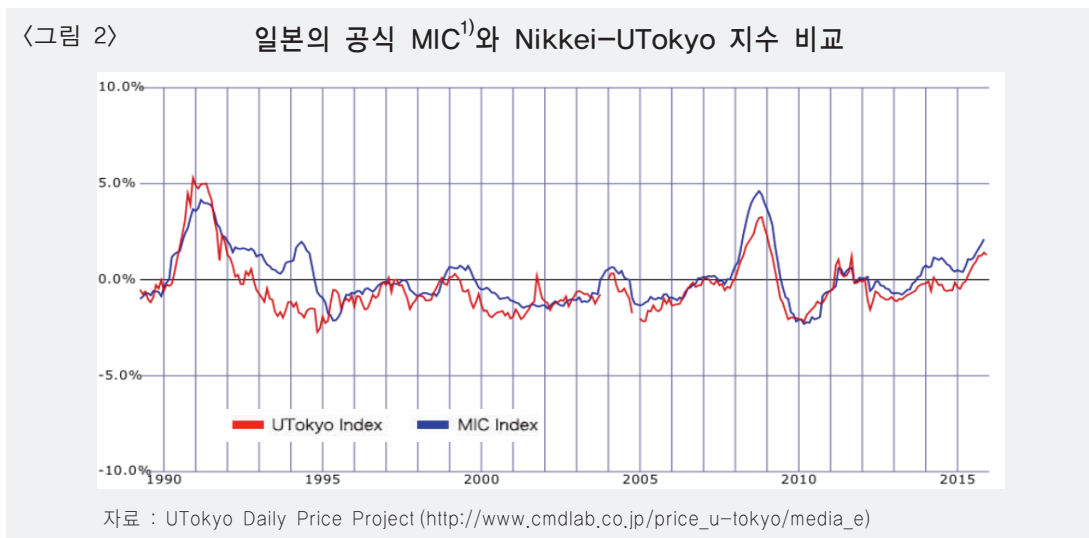
스캐너 데이터는 제품에 부착된 바코드(bar code)가 판매시점에 스캐너에 의해 읽혀지면서 입력되는 데이터로서 POS(Point of Sale) 데이터라고도 한다. 최근 소규모 상점에도 판매 시스템이 구비됨에 따라 스캐너 데이터에 포착되는 품목의 범위가 규모가 확대되고 있다. 스캐너 데이터는 실시간으로 수집되는 실제 거래내역 자료이므로 속보성과 신뢰성이 높고, 판매 가격뿐만 아니라 수량 및 기타 세부 정보도 확보 가능하다. 이러한 장점으로 노르웨이, 네덜란드, 스위스 등 유럽 국가들은 일찍부터 소비자물가지수 편제에 스캐너 데이터를 사용³⁾하고 있다. <그림 1>에서 보는 바와 같이 노르웨이 통계청은 이미 2000년대 초반부터 스캐너 데이터를 활용하기 시작해서 적용 품목을 점점 확대하고 있다.



3) 이들 국가의 통계청에서 스캐너 데이터를 소비자물가지수 편제에 활용하는 구체적인 방법은 Rodriguez and Haraldsen (2006), Randi (2016), Müller(2010), Van der Grient and Hann (2010) 등을 참고하기 바란다.

한편 Eurostat은 통합소비자물가지수(Harmonized Index of Consumer Prices, HICP)⁴⁾의 정확성을 유지하기 위하여 스캐너 데이터의 입수, 처리, 이용 등 일련과정에 관한 가이드라인을 제시하고 회원국에도 스캐너 자료의 활용을 권고하고 있다.

도쿄 대학의 Tsutomu Watanabe 교수는 Nikkei Inc.의 일별 POS 데이터를 이용하여 일본의 Nikkei-UTokyo 일별지수를 개발⁵⁾하였다. <그림 2>에서 2015년까지 Nikkei-UTokyo 지수⁶⁾의 움직임을 보면 일본 총무성(Ministry of Internal Affairs and Communications)의 공식 식료품물가지수(MIC Index for grocery component)⁷⁾와 비슷한 추이를 보이고 있다.



이외에도 Watanabe 교수는 전체 일본인의 40% 이상이 보유하고 있는 T 포인트 카드의 제휴사(슈퍼마켓, 약국, 편의점, 외식 체인 등)를 통해 수집되는 구매가격 데이터를 이용하여 T-point Index(TPI)를 개발하였다. TPI는 식품, 생활용품 이외에도 패션, 주거, 영상, 음악 등 소비의 다양한 측면을 포괄할 수 있고, 구매자의 속성을 파악하여 성별·연령별 물가지수 산출이 가능한 장점⁷⁾이 있다.

한편 관세청의 수출입 통관자료에도 수출입되는 모든 제품들의 가격과 수량 정보가 기록되므로 스캐너 데이터의 일종으로 볼 수 있다. 현재 일부 품목의 통관자료가 수출입 물가지수 편제에 활용되고 있으나 그 활용범위를 확대하기 위해서는 통관자료의 동일 HS 코

4) EU 회원국들의 소비자물가지수를 가중평균하여 산출한 물가지수이다.

5) Watanabe and Watanabe (2014) 참고

6) 2016년 1월부터는 Nikkei CPINow로 명칭을 변경하여 Nikkei.Inc와 Nowcast Inc.가 공동으로 제공한다.

7) TPI의 대상 품목 수는 총무성 CPI 품목의 약 23.4%에 해당한다.(www.cccmk.co.jp/tpi/)

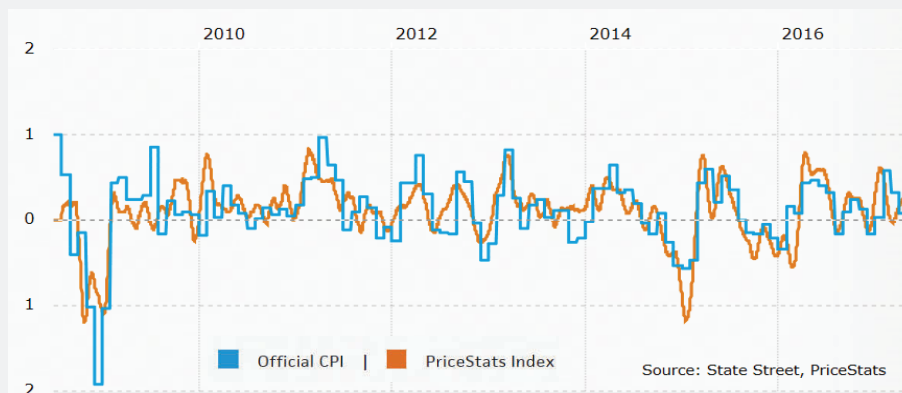
드내에 혼재되어 있는 다양한 품목들에 대한 정확한 품질 식별 방안을 강구할 필요가 있다. 한국은행은 외부연구용역 사업(강규호 외 (2015), 장영재 외 (2017) 참고)을 통해 수출입 물가지수 작성 관련 연구를 실시하였으며 향후 통관자료의 정제·처리 및 활용 방법을 구체적인 모색할 필요가 있다.

나. 웹 스크래핑 데이터(web scraping data)

MIT의 BPP(Billion Price Project, Cavallo and Rigobon (2016) 참고)는 온라인 가격정보를 이용하여 물가지수를 개발한 대표적인 사례이다. 동 방식에서는 자동화된 웹 스크래핑 소프트웨어가 온라인 소매업체 웹의 HTML 코드를 분석하여 수많은 상품의 가격정보를 수집한다. 이러한 온라인 가격자료는 서베이 자료보다 저렴하게 수집할 수 있고, 상품의 브랜드, 사이즈 등에 관한 부가 정보 및 신제품에 대한 정보를 신속히 입수할 수 있는 장점이 있다. 또한 일별 물가지수로 산출 가능하므로 시의성 높은 정보를 제공할 수 있다. 반면 온라인에 존재하지 않는 품목의 가격정보를 수집할 수 없고 온라인 가격과 오프라인 가격 차이 또는 실거래 가격과의 괴리가 발생할 수 있는 단점이 있다.

현재 BPP는 약 70개국 1,000개 이상의 온라인 소매업체로부터 수집한 가격정보를 토대로 22개국⁸⁾의 일별 물가지수를 PriceStats을 통해 공표하고 있다. <그림 3>에서 미국의 BPP 지수를 보면 공식 CPI에 약간 선행하는 움직임을 보이고 있다.

<그림 3> 미국의 공식 CPI와 BPP 지수 비교¹⁾



자료 : PriceStats (www.pricestats.com)

주 : 1) food & beverages; furnishing & household products; recreation & culture; clothing & footwear; housing, electricity & fuel; health 부분의 합성지수

8) 아르헨티나, 호주, 브라질, 캐나다, 칠레, 중국, 콜롬비아, 프랑스, 독일, 그리스, 아일랜드, 이탈리아, 일본, 한국, 네덜란드, 러시아, 남아프리카공화국, 스페인, 터키, 영국, 미국, 우루과이 등

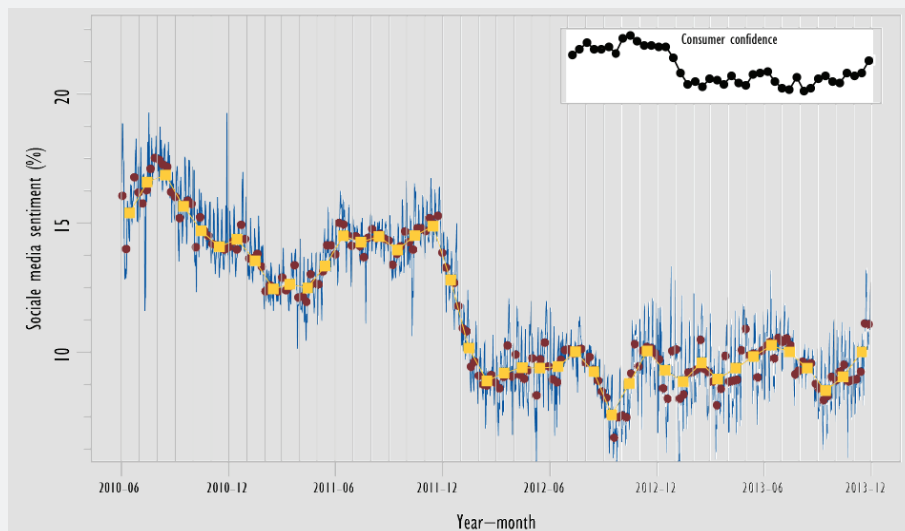
료가 경제주체들의 심리를 추출하거나 경제 및 금융 상황 모니터링을 위한 새로운 정보소스로 활용되고 있다.

가. 소셜미디어 데이터

네덜란드 통계청에서는 페이스북, 트위터 등 소셜미디어 메시지에 나타난 감성(sentiment)을 이용하여 소셜미디어지수(social media index, SMI)를 작성하였다. 네덜란드의 소셜미디어 모니터링 전문업체 Coosto는 각종 소셜미디어에 나타난 메시지들을 수집하여 문장단위로 분석한 후 해당 메시지의 전반적인 의견을 긍정, 중립, 부정으로 분류¹⁰⁾하는 서비스를 제공한다. 네덜란드 통계청은 Coosto에서 구입한 소셜미디어 메시지 감성분류 결과의 긍정의견 비중에서 부정의견 비중을 차감하여 SMI를 산출하였다. <그림 5>에서 SMI는 네덜란드의 공식 심리지표인 소비자신뢰지수(consumer confidence index, CCI)와 유사한 움직임을 보이며 0.78의 상관관계를 가진 것으로 나타났다. 자료 취득원별로 보면 페이스북과 트위터의 SMI가 언론기사, 블로그 등의 SMI보다 CCI와 더 유의미한 관계를 보이는 것으로 나타났다. 보다 자세한 연구결과는 Daas and Puts (2014), Daas et al. (2015), Van den Brakel et al. (2016) 등을 참고하기 바란다.

<그림 5>

네덜란드의 소셜미디어지수와 소비자신뢰지수 비교



자료 : Daas and Puts (2014)

10) 예를 들어 “Despite the high unemployment, the economy is doing well.”는 부정적 측면(high unemployment)도 포함하고 있으나 전반적인 경기인식은 긍정적(the economy is doing well)이므로 긍정 의견으로 분류한다.

네덜란드의 SMI를 벤치마킹하여 한국은행이 우리나라의 소셜미디어 데이터를 이용한 경제심리지표를 내부적으로 시산해 본 결과 공식 심리지표인 한국은행의 소비자심리지수와 유사한 움직임을 보이고 대표 실물경제지표인 GDP에 선행하는 것으로 나타났다. 이와 같이 소셜미디어 데이터를 이용한 심리지표는 속보성 및 시의성이 높은 심리 정보를 제공하고, 경제주체의 판단, 의도, 기대 등의 심리뿐만 아니라 그 원인에 대한 정보도 제공 가능하여 기존 경제심리지표의 보완지표 또는 정책결정을 위한 속보성 있는 모니터링 지표로 활용 가능하다. 그러나 소셜미디어를 적극적으로 활용하는 특정 집단의 의견이 더 많이 반영될 우려가 있고, 텍스트 마이닝(text mining)¹¹⁾의 기술적 제약으로 인해 경제주체의 심리를 완벽하게 추출해 내기 어려운 한계도 있다.

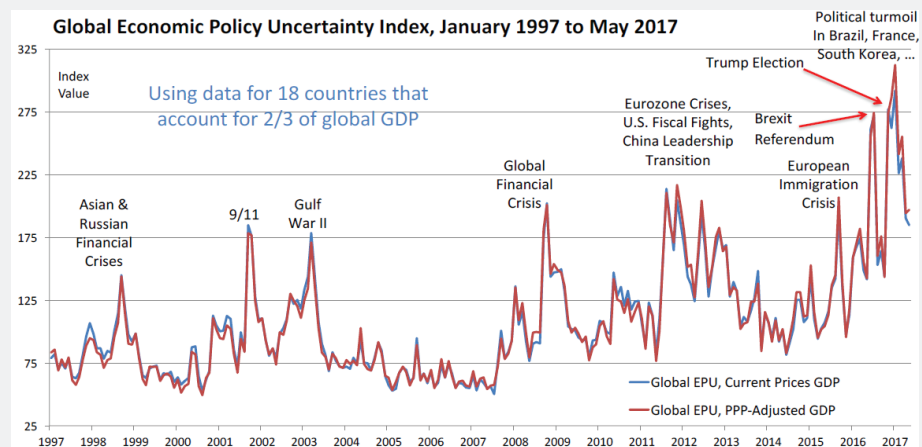
나. 뉴스 및 신문기사

Baker et al. (2015)는 주요 신문사의 기사를 이용하여 경제정책 불확실성 지수(economic policy uncertainty index, EPU)를 개발하였다. 구체적으로 경제(economic 또는 economy), 불확실(uncertain 또는 uncertainty), 정책관련 용어(congress, deficit, Federal Reserve, legislation, regulation, 또는 White House) 등 세 가지 범주의 단어를 모두 포함하는 기사의 빈도수를 전체 기사의 개수로 나누어 EPU 지수를 산출하였다. 현재 미국 등 18개국의 개별 EPU 지수와 이들을 합성한 글로벌 EPU 지수¹²⁾가 웹사이트에 공개되고 있다. <그림 6>을 보면 2016년말 한국의 정치적 불안이 글로벌 EPU 지수에 반영된 것을 확인할 수 있다. 이와 같이 EPU 지수는 정보의 신뢰성, 정확성, 일관성 측면에서 잠재적 한계가 있으나 경제정책 관련 동향을 신속하게 포착하고 국제적 비교가 용이한 장점이 있다.

11) 텍스트 데이터를 계량화하여 분석하기 위해 사용되는 통계적 방법들과 정보처리 방법을 통칭한다.

12) 1997년~2015년 기간을 표준화구간으로 하여 각국의 EPU 지수를 표준화한 다음 IMF의 World Economic Outlook의 GDP를 가중치로 하여 합성지수를 산출한다.

〈그림 6〉 글로벌 경제정책 불확실성 지수



자료 : www.policyuncertainty.com

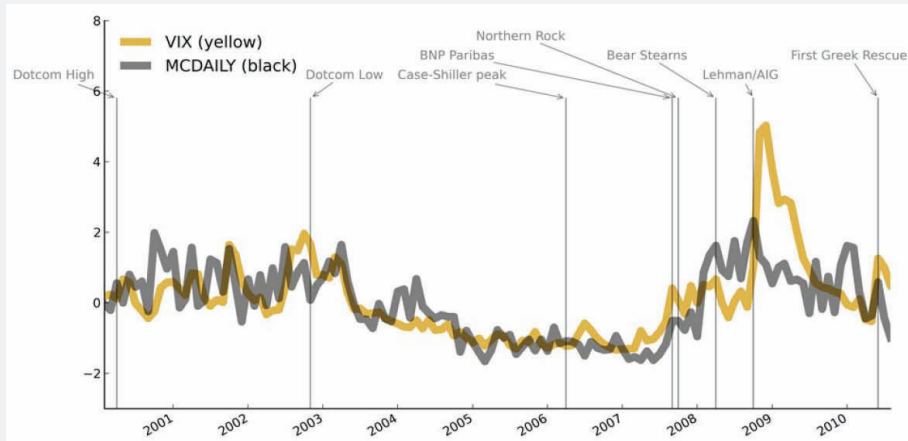
한편 샌프란시스코 연방준비은행의 Economic Letter(2017.4월)¹³⁾는 경제에 관한 뉴스 기사나 사실이 기사나 인터뷰 응답자들에 대한 감정을 전달하여 공식 자료의 단순제공 이상의 역할을 한다고 평가하고 경제·금융 관련 뉴스 기사를 이용하여 심리지표를 작성한 Shapiro et al. (2017)의 연구를 소개하였다. 동 심리지표는 현재의 경제상황을 잘 반영하는 것으로 나타났고, 미래 경제상황 예측력 측면에서 미시간 대학의 소비자심리지수(consumer sentiment index), Conference Board의 소비자신뢰지수(consumer confidence index) 등 전통적 서베이 방식의 심리지표보다 우월한 것으로 평가되었다.

영란은행(Bank of England)은 금융시장 심리지표를 산출하는 공동연구를 런던대학교와 함께 실시하였다. Nyman et al. (2016)은 영란은행 내부 보고서(internal market commentary), 브로커 보고서(broker reports), 로이터통신 뉴스(Reuters news archive) 등의 텍스트를 분석하여 금융시장 관련 심리를 기대(excitement)와 불안(anxiety)의 두 가지 범주로 구분하였다. <그림 7>은 영란은행 내부 보고서에서 추출한 불안지수(MCDAILY)가 금융시장이 불안했던 시기에 높았음을 보여준다. 미국 월가의 ‘공포지수’로 불리는 시카고 옵션거래소(CBOE)의 변동성 지수(Volatility Index, VIX)와는 0.65의 상관관계를 보이는 것으로 나타났다.

13) What's in the News? A New Economic Indicator.

〈그림 7〉

금융시장 불안지수(MCDAILY)와 VIX 비교



자료 : 영란은행, 시카고옵션거래소

이외에도 금융 및 경제 뉴스에 나타난 텍스트를 대상으로 투자자들의 심리를 분석하여 기업의 주가를 예측하거나, 기업의 부실위험을 예측하는 등 뉴스 정보를 이용한 다양한 연구 사례가 많다. 이는 뉴스에서 추출한 심리가 경제 및 금융 상황에 대한 유의미한 정보를 제공할 수 있음을 시사한다.

3. GDP 관련 지표

물가지표 편제에 대표적으로 활용되고 있는 스캐너 데이터는 품목별 소매동향 등 소비 지출 관련 정보로도 활용 가능하다. 실제로 미국 경제분석국(BEA)은 가계소비지출 통계 작성에 스캐너 데이터를 참고한다. 한국은행은 현재 신한카드와의 공동연구를 통해 가계소비 지출, 서비스업 생산 등 국민소득 구성항목 가운데 신용카드 빅데이터로 추정할 수 있는 항목을 발굴해 GDP 추계에 활용하는 방안을 모색하고 있다. Galbraith and Tkacz (2015)는 신용카드뿐만 아니라 직불카드, 수표 거래 등을 통해 발생하는 전자결제(electronic payment) 빅데이터를 이용하여 캐나다의 GDP를 추정(nowcast) 하였다.

한편 국제금융 및 외환거래를 위한 국제금융통신망 SWIFT(Society for Worldwide Interbank Financial Telecommunication)에도 결제 관련 빅데이터가 축적된다. SWIFT 전문 양식(FIN message)에 금융기관간 거래정보 확인, 지급지시, 이체 확인 등 각 메시지 종류에 따라 거래유형, 결제일, 계좌정보 등의 다양한 로그정보가 기록된다. SWIFT는 고객자금 이체와 관련된 메시지(MI 1031⁴⁾) 중 실제 경제활동(real economic activity)과 관련된 메시지들만 추출,

월별 규모를 측정하여 SWIFT Index를 작성¹⁵⁾한다. <표 2>는 SWIFT 통신 메시지 분석 빅데이터를 이용한 GDP 추정(nowcast) 결과인데 공식 GDP 성장률과 다소 차이가 있다. 그러나 SWIFT 빅데이터는 실제 금융거래에 기반한 속도성 있는 경제 모니터링 지표로 활용 가능해 보인다.

<표 2> SWIFT 빅데이터를 이용한 GDP 성장률 추정

Region/ Country	Q3-2013 vs. Q3-2012 (Year-on-Year %)	Q4-2013 vs. Q4-2012 (Year-on-Year %)	Q1-2014 vs. Q1-2013 (Year-on-Year %)	Forecast Q1-2014 Trend	
	GDP Actual ⁽¹⁾ (published by OECD)	GDP Nowcast	GDP Forecast	Direction ⁽²⁾	Rate of change ⁽³⁾
OECD	1.1%	1.4%	1.5%	Growing	Faster
EU27	0.1%	0.7%	0.9%	Growing	Faster
US	1.6%	2.1%	2.2%	Growing	Faster
UK	1.5%	2.3%	2.6%	Growing	Faster
Germany	0.6%	1.4%	1.6%	Growing	Faster

자료 : OECD(2013.11.12일), "SWIFT Index anticipates a strong start to 2014 for the UK and US economies."

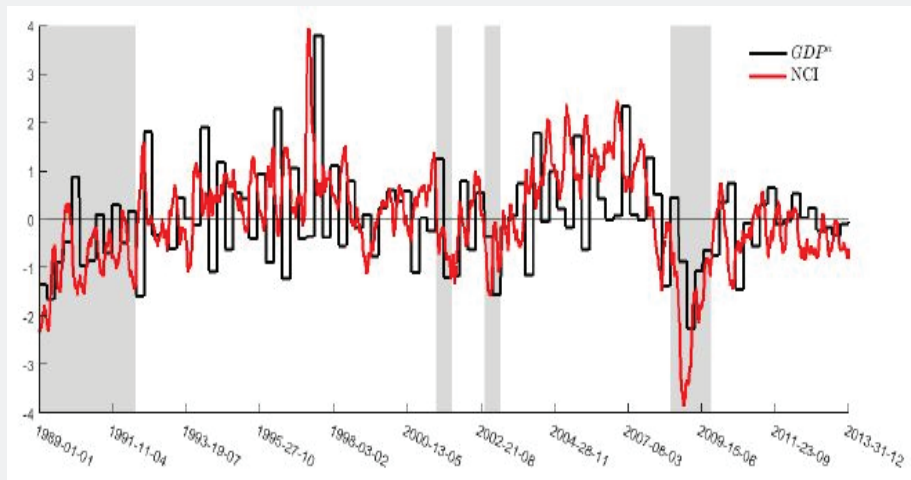
이외에도 신문기사로부터 추출한 텍스트 정보도 경기순환지수 및 분기 GDP 성장률을 추정하는 데 이용 가능하다. 예를 들어 Thorsrud는 노르웨이의 주요 일간 경제신문인 Dagen Naringsliv(DN)의 뉴스 기사를 Latent Dirichlet Allocation(LDA) 모형을 이용하여 기사에 나타나는 단어들을 바탕으로 신문 기사를 여러 주제(topic)들로 분해하고 각 주제가 기사에 언급된 빈도수를 계산하였다. 그리고 하버드 IV-4 심리사전에 정의된 긍정/부정 단어 목록을 이용하여 기사에서 각 주제에 대해 사용된 단어들을 긍정단어와 부정단어로 분류하고 그 차이를 계산하였다. 이러한 과정을 통해 신문기사의 텍스트 자료는 최종적으로 어떤 주제에 대한 기사가 많았는지, 각 주제에 대한 기사의 전반적인 분위기는 어떠했는지 반영된 시계열 자료로 변환되었다.

14) SWIFT는 전문 양식마다 "MT+세 자리 숫자" 형태의 명칭을 부여하는데 이 중 MT 103 메시지는 고객자금 이체에 이용되는 전문이다.

15) 현재 200개 이상의 국가 11,000여개 금융기관이 SWIFT의 서비스를 이용하고 있으며 SWIFT Index는 매일 작성되어 SWIFT 기관회원에게 제공되고 있다.

Thorsrud (2016a)는 이와 같이 시계열 자료로 표현된 텍스트 정보를 동적인자모형(Dynamic Factor Model, DFM)에 적용하여 일별 경기동행지수(Newly Coincident Index of Business Cycles, NCI)를 추정하였다. 그 결과 NCI는 일반적으로 사용되는 경기순환지수 대비 정확성과 시의성이 우수하고 특히 2000년대 초반의 경기침체를 주식이나 채권과 같은 시장지표들보다 더 잘 예측한 것으로 나타났다.

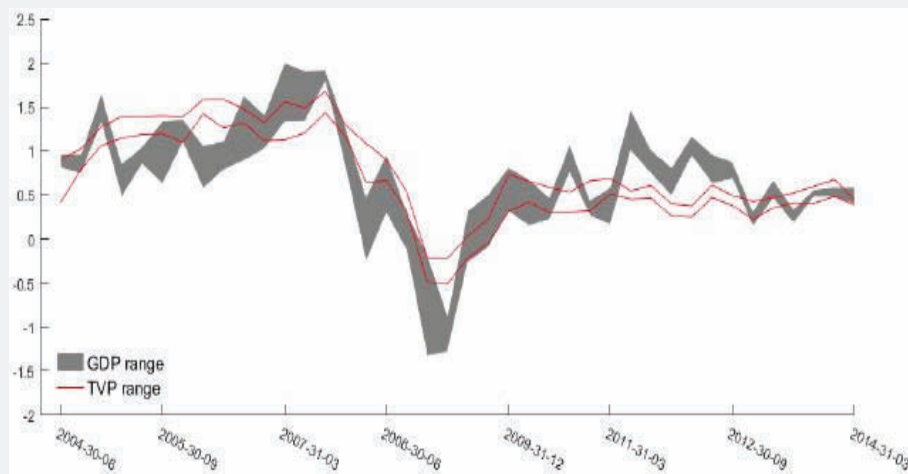
<그림 8> NCI와 GDP 비교¹⁾



자료 : Thorsrud (2016a)
 주 : 1) 회색 음영은 경기침체기

한편 Thorsrud (2016b)는 신문기사에서 추출한 텍스트 정보 시계열과 추세제거와 표준화를 거친 보정된 실시간 GDP 성장률을 DFM, 시변모수(Time-Varying Parameter, TVP) 리스케일링(rescaling) 모형을 이용하여 분기 GDP 성장률을 추정(nowcast)하였다. <그림 9>에서 공식 GDP 성장률(속보치 및 잠정치)의 범위(최대치 및 최소치)와 TVP 리스케일링 모형을 통한 GDP 성장률 예측치 범위를 비교한 결과를 보면 신문기사를 활용한 모형이 전반적으로 경제상황을 잘 예측한 것으로 나타났다. 또한 동 모형을 통한 GDP 예측결과는 노르웨이 중앙은행의 나우캐스팅 시스템이나 여타 벤치마크 모형들과 비슷한 수준의 예측오차를 가지나 경기 전환점 포착 측면에서 더 우수한 것으로 평가되었다.

〈그림 9〉 GDP 성장률과 TVP 모형의 GDP 성장률 예측결과 비교



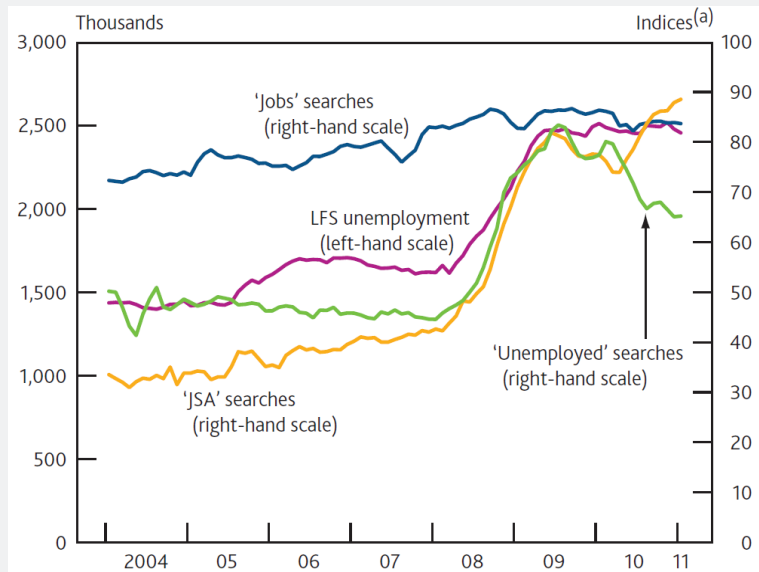
자료 : Thorsrud (2016b)

4. 기타

마지막으로 소개할 인터넷 검색 관련 빅데이터도 고용 시장 연구에 많이 활용되어 있다. Google의 실업 관련 검색 데이터를 이용하여 Choi and Varian (2009)은 미국의 신규 실업수당 청구건수를, Askitas and Zimmerman (2009)과 D'Amuri (2009)는 각각 독일과 이탈리아의 실업률을 예측하였다. <그림 10>은 영란은행의 Quality Bulletin에 게재된 영국 통계청의 노동 서베이(Labor Force Survey, LFS)에 의한 실업자 수(좌축)와 실업 관련 인터넷 검색 데이터 지수(우축)의 비교 자료이다. McLaren and Shanbhogue (2011)는 이러한 실업 관련 검색 빅데이터가 영국의 실업자 수 변화를 예측하는 데 유용하다고 평가하였다. 한국은행에서도 외부연구용역 사업(이궁희 외 (2016))을 통해 네이버 고용 검색지표를 개발하여 고용 관련 공식통계와 비교분석해 보았다.

〈그림 10〉

실업 관련 서베이 데이터와 인터넷 검색 데이터 비교



자료 : Google, ONS, Bank of England

주 : (a) 주별 검색 데이터를 표준화한 지수를 평균하여 산출한 월별 지수

이외에도 인터넷 검색 데이터는 주택가격 등 부동산 동향을 예측하는 데에 활용되기도 하였다. 이와 같이 Google Trends¹⁶⁾ 등을 통한 검색 데이터는 특정 관심 주제에 대한 시간·지역별 추이 관찰이 용이하고, 고정된 질문을 사용하는 서베이에 비해 특정 사건 전후의 변화를 신속·유연하게 분석할 수 있는 장점이 있다. 그러나 검색 데이터도 소셜미디어 데이터와 같이 모집단 대표성이 부족하고 검색 단어의 선택에 민감하며 잡음(noise)이 많다는 단점이 있으므로 의미 있는 정보를 추출하기 위한 정교한 작업이 필요하다.

16) www.google.co.kr/trends/

III. 시사점 및 향후 과제

빅데이터는 디지털 경제 하에서 새로운 경제현상을 포착하고 변화하는 통계환경에 대응하기 위한 수단으로 제안되고 있다. 무작위로 수집된 빅데이터를 사후적으로 의미 있는 통계로 가공하는 방식은 사전적 설계와 조사과정을 거쳐야 하는 전통적 서베이 방식에 비해 저렴하고 응답부담이 없으며 다양한 부가정보 확보가 가능한 장점이 있다. 또한 빅데이터는 실시간으로 생성·축적되므로 자동화된 데이터 처리과정을 통해 시의성 있는 정보를 신속하게 제공할 수 있다.

반면 빅데이터는 관심의 대상이 되는 목표 모집단에 대한 대표성을 보장할 수 없고 지속적이고 안정적인 자료 확보가 어려운 단점이 있다. 또한 빅데이터의 상당수가 민감한 개인정보를 포함하고 데이터의 소유권도 논란의 여지가 많아 빅데이터의 공유 및 활용에 대한 원칙과 제도적 장치 마련이 필요한 상황이다. 한편 빅데이터는 자료의 양이 방대하여 통계분석 수행시 엄청나게 많은 양의 계산이 요구될 뿐만 아니라 작업시간도 매우 길다. 따라서 빅데이터 처리·분석 관련 전문성을 확보하고 IT 시스템을 구축할 필요가 있다.

이와 같은 현실로 인해 단기간에 빅데이터가 새로운 공식통계를 개발하거나 기존 공식통계를 대체하는 데에 활용되는 것을 기대하기는 쉽지 않다. 그러나 현재 스캐너 데이터, 웹 스크래핑 데이터, 지급결제 데이터 등 비교적 정형화된 빅데이터가 물가, 소비지출 등 경제통계 작성에 부분적으로 이용되고 있다. 또한 소셜미디어, 뉴스 등 텍스트 데이터와 인터넷 검색 데이터도 경제 및 금융 관련 지표 분석에 널리 활용되고 있다. 이러한 사례들은 빅데이터의 단점과 한계에도 불구하고 향후 빅데이터가 경제통계 작성 및 분석에 유용하게 활용될 가능성이 있음을 시사한다.

따라서 빅데이터가 쉽게 적용 가능한 경제통계 영역을 발굴하여 시험편제를 실시해보고 기존 활용 사례들을 벤치마킹하여 유용성을 검증해 볼 필요가 있다. 또한 빅데이터의 정제·처리·분석을 위한 전문적 지식을 습득하고 빅데이터에 적합한 통계 작성기법에 대한 조사 연구도 필요하다.

<경제통계국 국민소득총괄팀 과장 문혜정, 빅데이터통계연구반 과장 이해영>

참고문헌

- 강규호·김민수·김성은 (2015), “스캐너타입 자료를 활용한 물가지수 작성방법 연구,” 국민계정리뷰, 2015년 제4호.
- 이궁희·김용대·황희진 (2016), “빅데이터를 이용한 고용지표 개발,” 국민계정리뷰, 2016년 제1호.
- 장영재·박종문·김민수 (2017), “통관자료를 이용한 수출입물가지수 표본추출기법 연구,” 국민계정리뷰, 2017년 제2호.
- Armah, N. (2013), “Big Data Analysis: The Next Frontier,” Bank of Canada Review, Summer, pp. 32-39.
- Askitas, N. and K. Zimmermann (2009), “Google Econometrics and Unemployment Forecasting,” *Applied Economics Quarterly*, Vol. 55(2), pp. 107-120.
- Baker, S., N. Bloom and S. Davis (2015), “Measuring Economic Policy Uncertainty,” NBER Working Paper, No. 21633.
- Cavallo, A. and R. Rigobon (2016), “The Billion Prices Project: Using Online Prices for Measurement and Research,” *Journal of Economic Perspectives*, Vol. 30(2), pp. 151-178.
- Choi, H. and H. Varian, (2009), “Predicting Initial Claims for Unemployment Benefits,” available at: <http://research.google.com/archive/papers/initialclaimsUS.pdf>.
- Dass, P. and M. Puts (2014), “Social Media Sentiment and Consumer Confidence,” European Central Bank Statistics Paper Series, No. 5.
- Daas, P., M. Puts, B. Buelens and P. van den Hurk (2015), “Big Data as a Source for Official Statistics,” *Journal of Official Statistics*, Vol. 31(2), pp. 249-262.
- D’Amuri, F. (2009), “Predicting unemployment in short samples with internet job search query data,” Bank of Italy Research Department, available at: http://mpra.ub.uni-muenchen.de/18403/1/MPRA_paper_18403.pdf.
- Galbraith, J. and G. Tkacz (2015), “Nowcasting GDP with electronic payments data,” European Central Bank Statistics Paper Series, No. 10.
- McLaren, N. and R. Shanbhogue (2011), “Using internet search data as economic indicators,” Bank of England Quarterly Bulletin (Q2), pp. 134-140.
- Müller R. (2010), “Scanner data in the Swiss CPI: An alternative to price collection in the field,” Swiss Federal Statistical Office, available at: <http://www.unece.org/fileadmin/DAM/>

stats/documents/ece/ces/ge.22/2010/zip.10.e.pdf.

- Nyman, R., D. Gregory, S. Kapadia, P. Ormerod, D. Tuckett and R. Smith (2016), “News and narratives in financial systems: Exploiting big data for systemic risk assessment,” September 2016, available at: <http://www.norges-bank.no/contentassets/49b4dce839a7410b9a7f66578da8cf74/papers/smith.pdf>.
- Office for National Statistics (2016), “Research indices using web scraped price data: clustering large datasets into price indices (CLIP),” available at: <http://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/clusteringlargedatasetsintopriceindicesclip>.
- Randi, J. (2016), “Scanner data in CPI/HICP,” Statistics Norway, available at: http://ec.europa.eu/eurostat/cros/content/use-scanner-data-norwegian-cpi_en.
- Rodriguez, J. and F. Haraldsen (2006), “The use of scanner data in the Norwegian CPI: The ‘new’ index for food and non-alcoholic beverages,” *Economic Survey*, Vol. 4, pp. 21-28.
- Shapiro, A. and D. Wilson (2017), “What’s in the News? A New Economic Indicator,” Federal Reserve Bank of San Francisco Economic Letter, 2017-10.
- Shapiro, A., M. Sudhof and D. Wilson (2017), “Measuring News Sentiment,” Federal Reserve Bank of San Francisco Working Paper, 2017-01.
- SWIFT (2016), “SWIFT Index Service Description,” October, available at: http://www2.swift.com/uhbonline/books/a2z/swift_index.htm.
- Thorsrud, L. A. (2016a), “Words are the new numbers: A newsy coincident index of business cycles,” Working Paper 44, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.
- Thorsrud, L. A. (2016b), “Nowcasting using news topics. Big Data versus big bank,” Working Paper 46, Centre for Applied Macro- and Petroleum economics(CAMP), BI Norwegian Business School.
- Van den Brakel, J., E. Söhler, P. Daas and B. Buelens (2016), “Social media as a data source for official statistics; the Dutch Consumer Confidence Index,” Statistics Netherlands Discussion paper, 2016-01.
- Van der Grient, H. and J. Hann (2010), “The use of supermarket scanner data in the Dutch CPI,” Statistics Netherlands.
- Watanabe, K. and T. Watanabe (2014), “Estimating Daily Inflation Using Scanner Data: A Progress Report,” CARF Working Paper, No. 37.

세계산업연관표(WIOT)의 구조와 이해

세계산업연관표(WIOT, World Input-Output Tables)는, 21세기 들어 글로벌 가치 사슬(global value chains)의 중요성이 커지고 있음에도 불구하고 현행 통계 시스템으로는 이를 정확히 분석하기 어렵다는 문제 제기에 따라 유럽연합 집행위원회(European Commission)의 후원 아래 11개 기관 소속 학자 및 전문가들이 결성한 WIOD(World Input-Output Database) 프로젝트의 결과물이다.

기존에도 글로벌 무역 분석 프로젝트(Global Trade Analysis Project), 아시아 국제산업연관표(Asian International Input-Output Tables), 아시아개발은행(ADB)의 다지역 산업연관표(Multi-Regional Input-Output Tables) 등 국가 간 산업연관표에 대한 연구가 다수 있었지만, 본 프로젝트는 국가 간 무역(bilateral trade) 통계를 적극적으로 활용하여 새로운 가정을 적용한 것이 기존 연구와의 차별점이라고 할 수 있다. 즉, 모든 경제 주체가 특정 재화 수요를 동일한 비율의 수입품으로 충당한다는 표준 수입 비율 가정에서 벗어나 상대 국가별로 동일한 수입품이라도 자국의 이용 목적별 비율에 차이가 있다는 가정을 통해 거래 국가 간 별도의 수입거래표를 작성한 것이다.

본 자료에서는 Dietzenbacher et al.(2013)를 바탕으로 세계산업연관표 작성 과정을 설명하고자 하며 나아가 우리나라의 지역 간 산업연관표에 적용할 수 있는 시사점을 살펴보고자 한다.

I. 머리말

II. 세계산업연관표 작성 과정

1. 개요
2. 국가별 공급사용표 시계열 작성
3. 세계공급사용표 시계열 작성
4. 세계공급사용표로부터 세계투입산출표 변환

III. 맺음말

I. 머리말

세계산업연관표(World Input-Output Tables)는 재화 및 서비스의 생산이 글로벌 가치 사슬(global value chains)을 통해 생산되는 경우가 점점 확대되자 이에 따른 경제적 영향을 분석하기 위한 통계가 필요하다는 인식에서 출발한 WIOD(World Input-Output Database) 프로젝트의 결과물이다. 국가(혹은 경제권) 간 재화 및 서비스 이동을 집계한 기존 통계로는 국제수지통계가 있으나, 국가를 단일한 경제 주체로 가정한 ‘산출물 단위’ 무역 기준 통계라는 특징 때문에 글로벌 가치 사슬 및 그 영향을 분석하는 데는 한계가 있었다. 이에 Grossman and Rossi-Hansberg(2008)는 기업의 오프쇼링 선택 모형을 통해 오프쇼링 기회 확대가 총요소집약적 기술변화(factor-augmenting technological change)와 비견됨을 보이면서 ‘교역 가능한 생산공정(tradeable tasks)’이라는 개념을 도입했는데, 결론에서 ‘생산과정의 글로벌화는 무역 데이터를 집계하는 데 있어서 새로운 접근법을 필요로 하는데, 이는 다년간에 걸친 국내 거래 데이터를 기록하는 것과 상당히 유사한 방식으로 국제 무역을 기록하는 것을 의미한다.’¹⁾ 라고 언급하며 기존과 다른 틀로 구축된 통계의 필요성을 제기하였다.

본 프로젝트는 유럽연합 집행위원회(European Commission)의 후원 하에 11개 기관²⁾의 참가로 2009년 5월 출범하여 2012년 4월 종료되었다. 종료 시점에는 27개의 유럽 연합(European Union) 회원국 및 13개 주요 국가의 개별 투입산출표 및 기타 국민계정통계를 토대로 명목 기초가격(current basic prices) 및 전년 기준 기초가격(basic prices of the previous year) 기준으로 작성된 최초 세계산업연관표를 발표하였다. 발표된 시계열은 1995년부터 2009년까지였으며, 2013년 11월에는 기존의 시계열을 2011년까지 확장한 추가 세계산업연관

1) Grossman, G. M. and Rossi-Hansberg, E., “Trading Tasks: A Simple Theory of Offshoring” (American Economic Review 2008, Vol. 98 No. 5), p. 1996

2) 참가기관은 다음과 같다.

1. Groningen 대학(University of Groningen, The Netherlands)
2. Institute for Prospective Technological Studies, Seville, Spain
3. The Vienna Institute for International Economics Studies, Vienna, Austria
4. Zentrum für Europäische Wirtschaftsforschung, Mannheim, Germany
5. Österreichisches Institut für Wirtschaftsforschung, Vienna, Austria
6. Hochschule Konstanz, Germany
7. The Conference Board Europe, Brussels, Belgium
8. CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands
9. Institute of Communication and Computer Systems, Athens, Greece
10. Central Recherche S.A., Paris
11. OECD

표가 발표되었다. 이후 2016년 11월에는 EU 회원국 1개국 및 주요 국가 2개국을 추가하여 2000년부터 2014년까지의 시계열을 갖춘 개정판이 발표되었다.

〈표 1〉

세계산업연관표 분석 대상 국가

EU 회원국		비유로존	북미 (NAFTA 회원국)	동아시아	기타
유로존					
오스트리아	이탈리아	불가리아	캐나다	중국	브라질
벨기에	룩셈부르크	체코	멕시코	일본	러시아
키프로스	몰타	덴마크	미국	한국	인도
에스토니아	네덜란드	헝가리		대만	인도네시아
핀란드	포르투갈	라트비아			호주
프랑스	슬로바키아	리투아니아			터키
독일	슬로베니아	폴란드			스위스 ¹⁾
그리스	스페인	루마니아			노르웨이 ¹⁾
아일랜드		스웨덴			
		영국			
		크로아티아 ¹⁾			

주 : 1) 2016년 개정판 발표 당시 추가된 국가

II . 세계산업연관표 작성 과정

1. 개요

프로젝트에서는 개별 국가의 공급사용표를 기초자료로 이용하였다. 공급표 및 사용표는 「상품(재화 및 서비스)×산업」 행렬로 작성되는데, 공급표에서는 개별 산업의 개별 상품 생산액 및 상품별 수입액을 알 수 있으며 사용표에서는 수입된 상품 및 개별 국가 내부에서 생산된 상품이 개별 산업과 최종수요 제도단위에서 얼마나 사용되었는지 파악할 수 있다. 이러한 특징 덕분에 국가별 공급사용표를 연결할 때는 상품 기준으로 작성된 국제무역 데이터와 산업 기준으로 작성된 사회경제적 혹은 환경 관련 데이터를 이중으로 이용할 수 있으며, 이를 통해 국가별 투입산출표를 직접 연결하는 것보다 더 직관적이고 정교한 통계를 만들게 된다.

두 번째로는 국민소득통계 등 여타 국민계정통계를 비교 자료로 이용하였다. 공급사용표는 기준년으로부터 일정 기간까지 연장하여 추계하기 때문에 장기 시계열이 단절되며, 한번 발표된 시계열은 개편하기가 용이하지 않다. 반면, 국민계정통계는 시계열 단절 문제가 없고 과거 시계열에 대한 개편이 상대적으로 자주 이루어지며 이 과정에서 과거 상황에 대해 갱신된 정보가 반영되기 때문에³⁾ 과거에 작성된 공급사용표가 반영하지 못한 정보를 보충해줄 수 있다. 구체적으로는 국민계정통계로부터 총생산과 총부가가치, 총수입 및 총수출, 최종소비지출의 시계열을 도출한 다음 Temurshoev and Timmer(2011)가 개발한 SUT-RAS 방법⁴⁾을 이용하여 최신 기준년도 공급사용표로부터 공급사용표 시계열을 연장 추계한다.

국가별 공급사용표 시계열이 준비되면, 다음에는 국제무역통계(International Trade Statistics)를 이용해서 세계공급사용표 시계열을 작성하고, 여기에 작성 대상 국가가 아닌 나머지 국가와의 무역을 반영한 세계산업연관표를 도출한다. 이때, 모든 경제 주체가 총 소비 대비 동일한 비율로 수입품을 이용한다는 표준 수입 비율 가정을 완화하여 이용 목적 항목별로 적용하였다. 즉, 무역통계를 만들기 위한 원시데이터에 부여된 BEC(Broad Economic Categories) 코드⁵⁾ 및 거래 상대 국가 정보를 토대로 원수출 국가별 수입품을 이용 목적에 따라 각각

3) 국민소득통계의 경우 기준년 개편 시 기준년보다 과거에 있는 시계열도 최신 정보 및 통계편제기법을 반영하여 다시 작성된다.

4) 실측 편제된 기준년 통계로부터 연장표를 추계 및 작성할 때 정사각행렬인 투입산출표에 적용되는 RAS 방법을 대부분 직사각행렬인 공급사용표에 응용하는 방법

중간소비, 최종소비, 총고정자본형성 등으로 분류하여 각 거래상대국가 및 이용 목적별로 별도의 수입 비율을 적용하였다. 한편, 서비스 무역의 경우 개별 국가간 무역량을 확인할 수 있는 표준화된 통계가 없어⁶⁾ 여러 원천 데이터를 통합하여 데이터베이스를 구축하였다.

세계산업연관표를 작성하기 위한 데이터는 검증의 용이성 및 향후 파생 통계 작성 가능성을 고려하여 각국 공인 통계 기관에서 공개적으로 발표한 통계에서 인용하였으며 전년 기준 기초가격 기준 통계 작성시에는 디플레이터로 국민계정통계에서 도출한 총산출 디플레이터(gross output deflators)를 이용하였다. 총산출 디플레이터를 이용할 경우, 수출자 가격을 조정하게 되는데 이는 암묵적으로 수입자 가격도 조정하는 효과를 갖는다.

2. 국가별 공급사용표 시계열 작성

세계산업연관표의 기초가 되는 국가별 공급사용표는 비록 각국이 UN에서 배포한 국민계정체계 매뉴얼에 기초하여 작성하지만 각국의 사정에 따라 실측년도 및 연장 추계 기간에 차이가 있으며, 분석 대상이 되는 산업 및 상품의 종류도 다르다. 이에 따라 일차적으로 국가별 공급사용표를 표준화시키는 작업이 필요하다.

프로젝트에서는 우선 각국의 공급사용표를 59개의 상품과 35개의 산업으로 정리하였는데, 상품 및 산업 분류는 각각 EU에서 정립한 분류 체계인 CPA(Classification of Products by Activity) 및 NACE(Nomenclature statistique des activités économiques dans la Communauté européenne)⁷⁾를 기준으로 하였다. 이외에 일부 국가에서 총공급과 총수요가 맞지 않거나 보안 문제로 일부를 공개하지 않는 문제 등이 있어 총공급과 총수요의 차이를 최종소비에 안분하거나 공개되지 않는 부분을 추정하여 해결하였다.

한편, 수출입은 현행 적용 기준인 2008 SNA가 아닌 1993 SNA에 의거하였다. 수출입의 경우 1993 SNA와 2008 SNA 모두 상품의 소유자 변경을 기준으로 기록하는 것을 원칙으로 삼고 있으나 1993 SNA에서는 2008 SNA와 달리 가공무역에 예외를 두었다. 즉, 한 국가의 가공무역 위탁업자가 본인이 소유권을 유지하는 재화를 수탁업자에게 수출하고 가공이 끝

5) SITC(Standard International Trade Classification) 체계의 하나로 정의되어 있는 코드로, 무역대상 재화를 최종 이용 목적별로 분류하였다. 현재 3차 개정판(rev.3)이 발표되었으며 UN 통계국 홈페이지(<https://unstats.un.org/unsd/class>)에 기존 SITC 및 HS(Harmonized System) 코드와의 연계표가 게시되어 있다.

6) 재화(goods)의 경우 관세 협정에 따라 각국의 관세 당국이 수출입되는 재화마다 국제적으로 HS 코드를 부여하고 행선국가를 확인하지만 서비스(service)는 그러한 절차를 거치지 않고 거래되기 때문에 개별 거래에 유형 및 국가별로 표준화된 코드를 부여하여 식별하기 어렵다.

7) 각각 UN에서 제시한 CPC(Central Product Classification) 및 ISIC(International Standard Industrial Classification)를 기반으로 EU에서 세부 분류를 세밀하게 다듬은 체계. 세계산업연관표 작성 당시에 NACE는 1차 개정판(revision 1)을 인용하였다. 이는 ISIC 3차 개정판(revision 3)과 대응한다.

난 재화는 재수입하는 경우 2008 SNA에 따르면 위탁업자가 수탁업자로부터 가공무역 서비스를 수입한 것으로 기록하고 가공무역 대상 재화의 수출입은 기록에서 제외하지만 1993 SNA에서는 가공 전 재화의 수출(수입)액과 가공 후 부가가치가 더해진 재화의 수입(수출)액을 모두 재화의 수출입으로서 기록하는 것이다. 이는 가공무역 대상이 되는 재화를 중간 투입으로 계상하기 위한 것으로, 프로젝트에서는 이러한 조치가 생산 기술을 가장 잘 반영할 수 있다고 보았다⁸⁾. 중계무역에 따른 재수출도 가공무역과 동일한 기준을 적용하였다.

이에 따라 국민계정통계를 반영한 시계열을 작성하기 전 준비된 국가별 공급사용표는 다음과 같다. 이때 공급사용표는 실측된 공표치를 채용하며 시계열 확장을 위한 추계 시 기준년도 지표로 삼게 된다. 금액 단위는 공표 국가별 법정통화 기준이다.

〈표 2〉 표준화된 국가별 공급사용표

	상품	산업	최종소비 (고정자본형성 포함)	재고증감	수출	총계
상품		\mathbf{U}	\mathbf{Y}	\mathbf{s}	\mathbf{e}	\mathbf{q}
산업	\mathbf{V}_b					\mathbf{x}_b
수입	\mathbf{m}'					m
부가가치		\mathbf{w}'_b				w_b
마진	\mathbf{t}'_m					t_m
순생산물세	\mathbf{t}'_n					t_n
총계	\mathbf{q}'	\mathbf{x}'_b	\mathbf{y}'	s	e	

\mathbf{U} : 사용표(구매자가격 기준, 59(상품)×35(산업) 행렬)
 \mathbf{Y} : 최종소비(구매자가격 기준, 59×4 행렬)
 — 항목별 4개 벡터로 구성(가계소비지출, 가계봉사비영리단체 소비지출, 정부소비지출, 총고정자본형성)
 — 벡터별 총합은 \mathbf{y}' (1×4 벡터)
 \mathbf{s} : 재고증감(구매자가격 기준, 총재고증감은 s , 59×1 벡터)
 \mathbf{e} : 수출(FOB 기준, 총수출은 e , 59×1 벡터)
 \mathbf{q} : 총수요(구매자가격 기준, 59×1 벡터)
 — \mathbf{q}' : 총공급(구매자가격 기준, 1×59 벡터)
 \mathbf{V}_b : 공급표(기초가격 기준, 35(산업)×59(상품) 행렬)
 \mathbf{x}_b : 총산출(기초가격 기준, 35×1 벡터)
 \mathbf{m}' : 수입(CIF 기준, 총수입은 m , 1×59 벡터)
 \mathbf{w}'_b : 부가가치(총부가가치는 w_b , 1×35 벡터)
 $\mathbf{t}'_m, \mathbf{t}'_n$: 마진(도소매마진 및 화물운임 포함), 순생산물세(원소의 총합은 각각 t_m 및 t_n , 1×59 벡터)
 $'$: 전치행렬 및 전치벡터, b (아래첨자) : 기초가격 기준 표시

8) 그러나 가공무역을 세계공급사용표에 반영하는 방법에 대해서는 추가적인 연구가 필요하다고 하였다 (Dietzenbacher, E. et al., “The Construction of World Input-Output Tables in the WIOD Project” (Economic Systems Research 2013, Vol. 25, No. 1), p. 78).

<표 2>에 국민계정통계를 반영하고 기준년도 지표로부터 연장 추계하여 국가별 공급사용표 시계열을 만드는 과정은 다음과 같다.

가. 국민계정통계 반영 및 기초가격 기준 공급사용표 작성

1) 국민계정통계를 이용한 기준년 공급사용표의 연장

국민계정통계는 총수출 및 총수입(\bar{e} , \bar{m}), 총산출 및 부가가치(\bar{x}_b , \bar{w}_b), 최종소비(\bar{y}), 총재고증감(\bar{s}), 총마진 및 총순생산물세(\bar{t}_m , \bar{t}_n)를 연장하는 데 이용된다. 여기에서 \bar{x} 는 해당 수치는 국민계정통계에서 인용하였음을 나타낸다.

수출 및 수입의 경우 벡터 \mathbf{m} 및 \mathbf{e} 의 추정치 $\hat{\mathbf{m}}$ 및 $\hat{\mathbf{e}}$ 는 국민계정통계를 반영하여 다음과 같이 추정할 수 있다.

$$\hat{m}_i = m_i(\bar{m}/m), \quad \hat{e}_i = e_i(\bar{e}/e)$$

\hat{x} 는 x 의 최종 추정치를 의미한다.

2) 사용표를 기초가격 기준으로 변환

<표 2>에서 밝혔듯이, 표준화된 공급사용표에서 공급표는 기초가격 기준, 사용표는 구매자가격 기준으로 작성되는 것으로 하였다. 이는 프로젝트 진행과정에서 세계산업연관표 작성 대상 국가의 공급사용표를 수집할 때 기초가격 기준 사용표를 공표하는 국가가 많지 않은 반면, 공급표는 기초가격 기준 데이터에 마진과 순생산물세 정보를 추가하여 구매자가격 기준 데이터도 같이 제공하는 경우가 많은 데 따른 것이다. 프로젝트에서는 모든 데이터를 기초가격 기준으로 조정하기로 하였으며 이를 위해서 사용표를 기초가격 기준으로 재작성하였다. 이때 각국에서 공급된 상품 i 의 총공급액(총수요액) 구매자가격 평가 금액과 기초가격 평가 금액사이에는 다음의 관계가 성립한다.

$$q_b^i = q_i - (t_m^i + t_n^i)$$

t_m 및 t_n 은 상품별 데이터만 담고 있지만, 해당 데이터를 산업까지 확장해서 「상품×산업」의 형태로 마진표 및 순생산물세표를 만들 수 있다. 이를 가격전환행렬(Valuation

Matrices)⁹⁾이라고 하는데, 사용표를 기초가격 기준표로 전환하기 위해서는 반드시 필요하나 대부분의 국가에서는 마진표 및 순생산물세표를 공표하지 않는다. 이에 따라 상품별 마진율 및 순생산물세율을 별도로 도출하여 사용표에 적용함으로써 가격전환행렬을 추정하였으며 추정 과정에서 국민계정통계가 일부 인용되었다. 동 과정을 개략적으로 살펴보면 다음과 같다.

첫 번째로, 상품 i 의 마진과 순생산물세의 합 t_i 는 다음과 같이 추정된다. \tilde{x} 는 x 의 일차 추정치이다.

$$\hat{t}_i = \frac{(t_m^i + t_n^i)(\bar{t}_m + \bar{t}_n)}{t_m + t_n}$$

상품은 마진이 있는 것과 없는 것(주로 서비스)으로 나뉘는데, 마진이 없는 경우 순생산물세의 일차 추정치 \tilde{t}_n^i 는 \hat{t}_i 와 동일하며 마진이 있는 경우 상품 i 의 순생산물세율 τ_i 및 구매자가격 기준 총공급 q_i 를 적용하여 $\tilde{t}_n^i = \tau_i q_i$ 로 추정할 수 있다. 이때 $\tau_i = \tilde{t}_n^i / q_i$ 는 공급표에서 도출된다. 그러나 국민계정통계 \bar{t}_n 에는 $t_n = \sum_i \tilde{t}_n^i = \sum_i \tau_i q_i$ 보다 최신 정보가 포함되어 있으므로 두 수치는 같을 수 없다. 따라서 마진이 있는 상품의 최종 추정 순생산물세 \hat{t}_n^i 은 다음과 같이 표준화하여야 한다.

$$\hat{t}_n^i = \tilde{t}_n^i \frac{\bar{t}_n - \sum_{i \in \{\text{무마진상품}\}} \tilde{t}_n^i}{\sum_{i \in \{\text{마진상품}\}} \tilde{t}_n^i}$$

결국 마진이 없는 상품은 $\hat{t}_n^i = \tilde{t}_n^i$, $\hat{t}_m^i = 0$ 이 되며 마진이 있는 상품의 \hat{t}_n^i , \hat{t}_m^i 는 다음과 같이 계산된다.

$$\hat{t}_n^i = \tilde{t}_n^i \frac{\bar{t}_n - \sum_{i \in \{\text{무마진상품}\}} \tilde{t}_n^i}{\sum_{i \in \{\text{마진상품}\}} \tilde{t}_n^i} = \tau_i q_i \frac{\bar{t}_n - \sum_{i \in \{\text{무마진상품}\}} \tilde{t}_n^i}{\sum_{i \in \{\text{마진상품}\}} \tilde{t}_n^i}$$

9) Eurostat, *Eurostat Manual of Supply, Use and Input-Output Tables*(Luxembourg, 2008), p. 161-188.

$$\hat{t}_m^i = \hat{t}_i - \hat{t}_n^i = \frac{(t_m^i + t_n^i)(\bar{t}_m + \bar{t}_n)}{t_m + t_n} - \tau_i q_i \frac{\bar{t}_n - \sum_{i \in \{\text{무마진상품}\}} \tilde{t}_n^i}{\sum_{i \in \{\text{마진상품}\}} \tilde{t}_n^i}$$

최종 추정 마진과 순생산물세를 토대로 마진을 및 순생산물세율을 추정하기 위해서는 한 가지 가정이 더 필요한데, 그것은 마진 및 순생산물세는 수출품에 적용되지 않는다는 것이다. 이를 반영하면 상품 i 의 순생산물세율 $\hat{\tau}_n^i$ 및 마진을 $\hat{\tau}_m^i$ 는 다음과 같이 추정된다.

$$\hat{\tau}_n^i = \hat{t}_n^i / (\hat{q}_i - \hat{e}_i), \quad \hat{\tau}_m^i = \hat{t}_m^i / (\hat{q}_i - \hat{e}_i)$$

순생산물세율 벡터 $\hat{\tau}_n$ 및 마진을 벡터 $\hat{\tau}_m$ (모두 59×1 벡터)는 사용표 \mathbf{U} 및 최종소비 \mathbf{Y} 의 산업별 각 원소에 일괄적으로 적용할 수 있으며, 최종적으로 사용표에 대한 순생산물세표 \mathbf{N}_U 및 마진표 \mathbf{M}_U , 최종소비에 대한 순생산물세표 \mathbf{N}_Y 및 마진표 \mathbf{M}_Y 를 도출할 수 있다. 즉, \mathbf{N}_U 및 \mathbf{N}_Y 의 도출 과정은 다음과 같다.

$$\mathbf{N}_U = \begin{bmatrix} u_{1,1} \hat{\tau}_n^1 & \cdots & u_{1,35} \hat{\tau}_n^1 \\ \vdots & \ddots & \vdots \\ u_{59,1} \hat{\tau}_n^{59} & \cdots & u_{59,35} \hat{\tau}_n^{59} \end{bmatrix}, \quad \mathbf{N}_Y = \begin{bmatrix} y_{1,1} \hat{\tau}_n^1 & \cdots & y_{1,4} \hat{\tau}_n^1 \\ \vdots & \ddots & \vdots \\ y_{59,1} \hat{\tau}_n^{59} & \cdots & y_{59,4} \hat{\tau}_n^{59} \end{bmatrix}$$

이제 사용표 \mathbf{U} 에서 순생산물세표 \mathbf{N}_U 및 마진표 \mathbf{M}_U 를 빼면 기초생산가격 기준 사용표 ($\hat{\mathbf{U}}_b^0$)가 되며 최종소비 \mathbf{Y} 에 대해서도 마찬가지로 기초생산가격 기준 최종소비($\hat{\mathbf{Y}}_b^0$)를 도출할 수 있다.

3) 재고증감을 기초가격 기준으로 변환 후 국민계정통계를 반영

재고증감도 앞서 도출한 $\hat{\tau}_n$ 을 활용하여 다음과 같이 기초가격 기준 통계로 변환할 수 있다.

$$\tilde{s}_i = s_i \cdot \tau_i$$

국민계정통계는 앞선 작업을 거친 뒤 반영한다. 그런데 재고증감은 각 통계 작성기관에

서 실제로 측정하기보다는 공급과 수요를 일치시키는 과정에서 잔차로 처리하는 경향이 있다. 따라서 재고증감 수치는 국민계정통계와 공급사용표상 수치의 차를 상품별 재고증감에 안분하는 방식으로, 국민계정통계를 간접적으로 반영하여 추정한다.

$$\hat{s}_i = \tilde{s}_i + \frac{|\tilde{s}_i|}{\sum_j |\tilde{s}_j| (\bar{s} - \tilde{s})}$$

나. 국가별 공급사용표 시계열 작성

1) 기준년도 기초가격 기준 공급사용표 재작성

공급사용표에 국민계정통계를 반영하고 가격전환행렬을 별도 추정하여 모두 기초가격 기준으로 도출한 공급사용표는 다음과 같다.

<표 3> 국가별 공급사용표(기초가격 기준, 국민계정통계 반영)

	상품	산업	최종소비 (고정자본형성 포함)	재고증감	수출	총계
상품		\hat{U}_b^0	\hat{Y}_b^0	\hat{s}	\hat{e}	...
산업	V_b					\bar{x}_b
수입	\hat{m}'					\bar{m}
순생산물세 부가가치		$\hat{t}'_{\text{중간수요}}$	$\hat{t}'_{\text{최종소비}}$	$t_{\text{재고증감}}$		\bar{t}_n
		\bar{w}_b				\bar{w}_b
총계	...	\bar{x}'_b	\bar{y}	\bar{s}	\bar{e}	

\hat{U}_b^0, \hat{Y}_b^0 : 사용표 및 최종소비(기초가격 기준, 59×35(사용표), 59×4(최종소비) 행렬)
 $\hat{t}'_{\text{중간수요}}, \hat{t}'_{\text{최종소비}}$: 순생산물세표(N_U, N_Y)의 열합벡터(각각 1×35, 1×4 벡터)
 $t_{\text{재고증감}}$: $\hat{\tau}_n' s$ (스칼라)
 $-\sum_i \hat{t}_{\text{ind}}^i + \sum_i \hat{t}_{\text{fin}}^i + t_{\text{inv}} = \bar{t}_n$

<표 3>에서 음영 처리된 부분은 국민계정통계를 직간접적으로 반영한 부분이며 이들은 고정된 수치로 간주된다. 음영 밖에 있는 공급표 및 사용표, 최종소비 행렬($V_b, \hat{U}_b^0, \hat{Y}_b^0$)에는 국민계정통계가 반영되지 않았기 때문에 <표 3>의 공급사용표는 행렬 간 합이 맞지 않

는 등 정합성 문제가 있어 바로 이용할 수 없다. 이때 음영 부분을 제약조건, $\mathbf{V}_b, \hat{\mathbf{U}}_b^0, \hat{\mathbf{Y}}_b^0$ 를 초기값으로 하여 앞서 언급한 SUT-RAS 방법을 적용하면 별도의 외생적 정보 없이도 정합성을 갖춘 기초가격 기준 공급사용표 $\hat{\mathbf{V}}_b, \hat{\mathbf{U}}_b, \hat{\mathbf{Y}}_b$ 를 만들 수 있다. 최종적으로는 기초 가격 기준 총공급(총수요) 추정치 $\hat{\mathbf{q}}_b$ 를 도출하게 된다.

2) 국가별 비교년도 공급사용표 추계

앞서 설명한 과정과 같이 기준년도 기초가격 기준 공급사용표가 준비되었으면 비교년도 기초가격 기준 공급사용표도 국민계정통계가 적용된 수치를 제약조건으로, 기준년도 공급 표 및 사용표, 최종소비 행렬을 초기값으로 하여 SUT-RAS 방법을 통해 작성할 수 있다.

두 개의 기준년도 BY_1 과 BY_2 사이에 있는 비교년도 $t(BY_1 < t < BY_2)$ 의 공급사용표를 추계할 때 국민계정통계에서 직접 인용 가능한 부분은 <표 3>에서 $\bar{\mathbf{x}}_b, \bar{m}, \bar{t}_n, \bar{\mathbf{w}}_b, \bar{w}_b, \bar{\mathbf{y}}, \bar{s}, \bar{e}$ 이다. 그러나 $\hat{\mathbf{s}}, \hat{\mathbf{e}}, \hat{\mathbf{m}}'$, 그리고 $\hat{\mathbf{t}}'_{\text{중간수요}}, \hat{\mathbf{t}}'_{\text{최종소비}}$ 는 별도의 추정을 통해 추정치를 확정하여야 한다. 이중 순생산물세 및 마진($\hat{\mathbf{t}}'_{\text{중간수요}}, \hat{\mathbf{t}}'_{\text{최종소비}}$)은 기준년도의 가격 전환행렬을 작성할 때와 동일한 과정으로 추정할 수 있으나 수출입($\hat{\mathbf{e}}, \hat{\mathbf{m}}'$) 및 재고증감($\hat{\mathbf{s}}$)은 별도의 추정 과정이 필요하다.

재고증감($\hat{\mathbf{s}}$)은 비교년도 국민계정통계에서의 재고증감과 시간상 앞서 있는 기준년도 BY_1 의 재고증감의 차를 상품별 재고증감 절대값으로 안분한 추정치($\hat{\mathbf{s}}^{t, BY_1}$)와 마찬가지로 기준년도 BY_2 의 재고증감과 비교년도 재고증감의 차를 안분한 추정치($\hat{\mathbf{s}}^{t, BY_2}$)를 평균하여 추정한다.

$$\hat{s}_i^{t, BY_1} = s_i^{BY_1} + \frac{|s_i^{BY_1}|}{\sum_i |s_i^{BY_1}|} (\bar{s}^t - s^{BY_1})$$

$$\hat{s}_i^{t, BY_2} = s_i^{BY_2} + \frac{|s_i^{BY_2}|}{\sum_i |s_i^{BY_2}|} (\bar{s}^t - s^{BY_2})$$

$$\hat{s}_i = (\hat{s}_i^{t, BY_1} + \hat{s}_i^{t, BY_2}) / 2$$

수출입의 경우 기준년도 수치는 국민계정통계를 반영하는 방법을 그대로 적용한다. 수출을 예로 들면, 기준년도 BY_1 과 BY_2 의 상품 i 의 수출 $\hat{e}_i^{BY_1}, \hat{e}_i^{BY_2}$ 는 각각 다음과 같다.

$$\hat{e}_i^{BY_1} = e_i^{BY_1} (e^{-BY_1} / e^{BY_1}), \quad \hat{e}_i^{BY_2} = e_i^{BY_2} (e^{-BY_2} / e^{BY_2})$$

하지만 국민계정통계에서는 산업연관표를 제외하고는 상품별 수출 데이터를 공표하지 않기 때문에 국제무역통계를 인용하여야 하나 국제무역통계상 상품별 수출의 총합과 국민계정통계의 수출이 일치하지 않을 수 있다. 그래서 비교년도 t 의 상품 i 의 수출 \hat{e}_i^t 는 전년도 수출 추정치 \hat{e}_i^{t-1} 에 국제무역통계를 이용한 전년대비 수출 증감률과 기준년도 추정치와의 정합성을 유지하기 위한 보정치를 곱하여 추정한다. 여기에서 국제무역통계에서의 비교년도 t 의 상품 i 의 수출은 국민계정통계상 수치와 구분하기 위해 ew_i^t 라 한다.

$$\hat{e}_i^t = \hat{e}_i^{t-1} \cdot \frac{ew_i^t}{ew_i^{t-1}} \cdot \left(\frac{\hat{e}_i^{BY_2} / \hat{e}_i^{BY_1}}{ew_i^{BY_2} / ew_i^{BY_1}} \right)^{1/(BY_2 - BY_1)}$$

해당 추정치 \hat{e}_i^t 는 국민계정통계상 비교년도 t 의 수출 e^t 와의 정합성을 위해 표준화 절차를 거친다.

$$\hat{e}_i^t = e^t \frac{\hat{e}_i^t}{\sum_i \hat{e}_i^t}$$

그리고 비교년도 t 의 상품별 수입도 수출과 같은 과정을 거쳐 추정한다.

한편, 비교년도 t 의 공급표 및 사용표, 최종소비 행렬 초기값($\hat{\mathbf{V}}_b^{t,0}$, $\hat{\mathbf{U}}_b^{t,0}$, $\hat{\mathbf{Y}}_b^{t,0}$)은 기준년도 추정치의 산술가중평균으로 추정할 수 있다.

$$\hat{\mathbf{V}}_b^{t,0} = \frac{(BY_2 - t) \mathbf{V}_b^{BY_1} + (t - BY_1) \mathbf{V}_b^{BY_2}}{BY_2 - BY_1}$$

$$\hat{\mathbf{U}}_b^{t,0} = \frac{(BY_2 - t) \mathbf{U}_b^{BY_1} + (t - BY_1) \mathbf{U}_b^{BY_2}}{BY_2 - BY_1}$$

$$\hat{\mathbf{Y}}_b^{t,0} = \frac{(BY_2 - t) \mathbf{Y}_b^{BY_1} + (t - BY_1) \mathbf{Y}_b^{BY_2}}{BY_2 - BY_1}$$

이후 SUT-RAS 방법을 통해 최종적으로 기초가격 기준 비교년도 t 의 공급사용표를 추계할 수 있다. 이를 다년간에 걸쳐 추계하고, 특히 비교년도 이전 혹은 이후에 기준년도가 없는 경우에는 한쪽 기준년도만을 기준으로 추정치 및 초기값을 추정한 후 추계하면 국가별 공급사용표 시계열이 작성된다.

3. 세계공급사용표 시계열 작성

국가별 공급사용표를 연결하여 세계공급사용표를 작성하기 위해서는 국가 간 무역 데이터베이스(bilateral trade database)가 포함된 국제무역통계가 필요하다. 재화의 경우에는 각국의 관세당국이 통관 업무를 수행하는 과정에서 수출입 실적에 HS(Harmonized System) 코드를 부여하여 관리하므로 신뢰성 있는 데이터를 얻을 수 있다. 프로젝트에서는 UN Comtrade 데이터베이스¹⁰⁾로부터 1995년부터 2010년까지의 시계열을 확보하였다. 그러나 몇몇 국가의 경우 일부 연도의 시계열이 없고, 일부 국가의 무역 통계는 정치적인 이유로 공개하지 않거나 편제하지 않는 사례가 있었다. 이러한 문제로 누락된 시계열은 각국 통계기관에 직접 문의하고, 특정 국가의 비공표 통계는 특별분류 혹은 기타국가로 분류하는 관행에 따라 상대 국가를 유추하여 해결하였다¹¹⁾. 또, 여섯 자리 숫자 기준의 HS 코드로 식별되는 국가 간 무역 데이터베이스 내 수출입품목은 대략 5,000여 개에 달한다. 이를 이용 목적별로 구분하기 위해 BEC 코드(3차 개정판 기준)를 적용하여 중간소비, 최종소비 및 자본재로 분류하였으며 여러 목적으로 이용될 수 있는 품목은 HS 코드별로 목적별 가중치를 부여하였다. 최종적으로는 모든 품목을 프로젝트에서 삼는 기준에 맞게 CPA 및 NACE 기준으로 재분류하였다. 이에 따라 5,000여 개의 HS 코드별 재화 수출입품은 59개 상품(재화 포함) 그룹으로 정리되었으며 거래 상대국가별·이용 목적별 가중치 및 HS 코드별 재화 수출입품과 세계산업연관표 기준 재화 그룹과의 관계는 수입품 기준으로 다음과 같다.

10) 홈페이지(<http://comtrade.un.org>)에서 국가 간 무역 데이터를 포함한 다양한 데이터를 확인할 수 있다.

11) 중국은 홍콩과 마카오에 대해서는 별도 무역 통계를 편제하였지만, 대만과의 무역 통계는 명시적으로 공표하지 않았다. 하지만 프로젝트에서는 홍콩과 마카오도 중국 경제권으로 간주하였으며, 대만은 별도 국가로 분리하였다. 따라서 홍콩과 마카오의 무역 통계는 상계하고 두 지역의 타 국가간 무역 통계는 중국 본토의 통계에 포함시켰으며, 대만은 중국 입장에서 무역 상대방이 '기타 아시아'로 분류된 통계를 대만 무역 데이터로 간주하였다. 대만의 국가 간 무역 통계는 OECD에서 인용하였다.

〈표 4〉 수입품의 원수출 국가별·이용 목적별 가중치 및 재화 분류기준별 관계

<p>HS 코드별 원수출 국가별 수입품</p>	<p>중간재 : $m_{jo}^{\text{중간재}} = \alpha_{jo}^{\text{중간재}} m_{jo}^{\text{ITS}}$ 최종소비재 : $m_{jo}^{\text{최종소비재}} = \alpha_{jo}^{\text{최종소비재}} m_{jo}^{\text{ITS}}$ 총고정자본형성 : $m_{jo}^{\text{총고정자본형성}} = \alpha_{jo}^{\text{총고정자본형성}} m_{jo}^{\text{ITS}}$</p>
<p>59개 상품 (프로젝트 편제) 기준 원수출 국가별 수입품</p>	<p>중간재 : $m_{io}^{\text{중간재}} = \sum_{j=1}^{n_i} m_{jo}^{\text{중간재}}$ 최종소비재 : $m_{io}^{\text{최종소비재}} = \sum_{j=1}^{n_i} m_{jo}^{\text{최종소비재}}$ 총고정자본형성 : $m_{io}^{\text{총고정자본형성}} = \sum_{j=1}^{n_i} m_{jo}^{\text{총고정자본형성}}$</p>
<p>59개 상품 기준 이용 목적 가중치</p>	<p>중간재 : $\alpha_{io}^{\text{중간재}} = m_{io}^{\text{중간재}} / m_i^{\text{ITS}}$ 최종소비재 : $\alpha_{io}^{\text{최종재}} = m_{io}^{\text{최종재}} / m_i^{\text{ITS}}$ 총고정자본형성 : $\alpha_{io}^{\text{총고정자본형성}} = m_{io}^{\text{총고정자본형성}} / m_i^{\text{ITS}}$</p>

여기에서 아래첨자 o 는 수입품의 원수출 국가이며 2016년 공표 기준 43개 국가 및 기타 지역(Rest of World) 1개¹²⁾ 중 자국을 뺀 43개 국가를 나타낸다. j 는 HS 코드 기준 수입품을 나타낸다. CPA 기준 상품 i 중 재화는 각각 HS 코드 기준 재화 집합 $\{1, 2, \dots, n_i\}$ 에 대응되며 $m_i^S = \sum_{j=1}^{n_i} m_j^S$ 가 된다. α 는 가중치이며 $\sum_o (\alpha_{jo}^{\text{중간재}} + \alpha_{jo}^{\text{최종소비재}} + \alpha_{jo}^{\text{총고정자본형성}}) = 1$ 이 된다. 위첨자 ITS는 국민계정통계를 반영한 공급사용표상 수입(m_i^{SUT})과 구분하기 위하여 국제무역통계에서 인용한 수입 통계임을 나타내는 표시이다.

서비스 무역에도 재화와 동일하게 국가 간 거래 개념 및 거래 상대국가별·이용 목적별 수입 비율 가중치를 적용하였다. 그러나 무형이며 저장 불가능한 서비스는 신뢰성 있는 데이터 확보가 어려워 UN을 포함하여 Eurostat, OECD에서 데이터를 인용한 후 원천별로 데이터를 교차검증하고 최종적으로 정리된 데이터를 조정하여 이용하였다. 이때 서비스 무역의 기준은 WTO 협정 중 일부인 서비스 무역에 관한 일반 협정(General Agreement on Trade in Services)의 모드 1만을 따랐다¹³⁾. 가중치는 Eurostat에서 공표한 수입품 기준 투입

12) 향후 세계공급사용표 및 세계투입산출표 작성 시 작성 대상 국가 이외의 지역은 국가 간 산업연관표의 의 잔차로 처리할 수 있으나, 이 경우 양수가 되어야 하는 항목이 음수가 되는 문제가 발생할 수 있어 별도의 단일 지역으로 처리하였다.

13) 협정에 따르면, 서비스 무역은 네 가지 종류로 정의된다.

1. 모드 1 : 국경 통과 - 특정 국가의 영토에서 타국 영토로 이전되어 제공
2. 모드 2 : 해외 소비 - 특정 국가의 영토에서 타국의 소비자에게 제공

산출표를 기반으로 상품 i 의 전체 국가 및 연도별 평균을 계산하여 각 연도별로 모든 국가 간 서비스 무역에 동일하게 적용하였다⁴⁾.

수출입 데이터가 준비되었으면, 모든 국가의 공급사용표 단위를 미국달러(USD) 기준으로 변환한다. 이때 이용된 환율은 연도별 평균 미국달러 환율이며, IMF IFS(International Financial Statistics) 데이터베이스에서 인용하였다.

가. 원수출 국가별 수입거래표 추정

<표 3>에서 수입거래표, 즉 \hat{U}_b, \hat{Y}_b 의 수입 행렬을 $\hat{U}_b^{\text{수입}}, \hat{Y}_b^{\text{수입}}$ 라 하면 각 행렬별 원소는 $u_{ik}^{\text{수입}}, y_{ik}^{\text{수입}}$ 라 할 수 있다. 여기에서 k 는 산업 및 최종수요 항목별 열을 의미한다.

$$\hat{U}_b = \hat{U}_b^{\text{수입}} + \hat{U}_b^{\text{국내}}, \quad \hat{Y}_b = \hat{Y}_b^{\text{수입}} + \hat{Y}_b^{\text{국내}}$$

$u_{ik}^{\text{수입}}, y_{ik}^{\text{수입}}$ 는 각각 중간소비, 최종소비 및 총고정자본형성에 대응되며 <표 3>의 \hat{m}' ($= \{m_1^{\text{SUT}}, \dots, m_i^{\text{SUT}}\}$)에 이용 목적별 가중치를 곱하여 추정한다.

$$\text{중간소비} : u_{ik}^{\text{수입}} = \frac{u_{ik}}{\sum_{l=1}^{35} u_{il}} \left(\sum_o \alpha_{io}^{\text{중간재}} \right) m_i^{\text{SUT}}$$

$$\text{최종소비} : y_{ik}^{\text{수입}} = \frac{y_{ik}}{\sum_{l=1}^3 y_{il}} \left(\sum_o \alpha_{io}^{\text{최종소비재}} \right) m_i^{\text{SUT}}$$

$$\text{총고정자본형성} : y_{i4}^{\text{수입}} = \left(\sum_o \alpha_{io}^{\text{총고정자본형성}} \right) m_i^{\text{SUT}} \quad 15)$$

특정 상품 i 의 재고증감을 제외한 총 국내 수요는 $\sum_k u_{ik} + \sum_k y_{ik}$ 와 같은데 이것이 m_i^{SUT} 보다 작을 경우, 즉 수입물량이 국내 수요를 상회할 경우 초과분은 국내 생산과정에

-
- 3. 모드 3 : 상업적 실체 - 타국 영토에서 특정 국가의 사업장을 통해 제공되는 서비스(직접투자 등)
 - 4. 모드 4 : 자연인적 실체 - 타국 영토에서 특정 국가의 자연인을 통해 제공되는 서비스
 - 14) Dietzenbacher et al.(2013)은 국제수지통계에서의 상품분류와 CPA와의 차이, 특정 상품 부문에서는 제한적인 국가 간 데이터, 그리고 이용 목적별 데이터를 확인할 수 없는 한계로 인해 Eurostat의 수입품 기준 투입 산출표를 인용하였고, 그나마 통계를 바로 사용한 것이 아닌 거친 추정(a rough estimate)을 위해 통계를 면밀하게 실사하였다고 밝혔다.
 - 15) 총고정자본형성은 \hat{Y}_b 행렬(59×4)에서 가장 마지막 열을 의미한다.

서 이용되지 않는 잉여 상품으로 간주하여 재수출(re-export)¹⁶⁾분으로 취급한다.

$$\text{재수출} : e_i^{\text{재수출}} = m_i^{\text{SUT}} - \left(\sum_k u_{ik} + \sum_k y_{ik} \right)$$

상품 i 에서 재수출이 일어날 경우 모든 산업 및 최종수요 항목에 대해 $u_{ik}^{\text{수입}} = u_{ik}$, $u_{ik}^{\text{국내}} = 0$, $y_{ik}^{\text{수입}} = y_{ik}$, $y_{ik}^{\text{국내}} = 0$ 이 성립하며 상품 i 의 수입 m_i 는 $m_i' = m_i - e_i^{\text{재수출}}$ 로 조정된다.

국가 간 수입거래표를 구성하는 수입 행렬은 $\hat{U}_b^{\text{수입}}$, $\hat{Y}_b^{\text{수입}}$ 로부터 도출할 수 있는데 이를 $\hat{U}_b^{\text{수입},o}$, $\hat{Y}_b^{\text{수입},o}$ 라 한다. 행렬별 원소는 원수출 국가별·이용 목적별 가중치의 가중평균을 통해 추정한다.

$$\text{국가 간 중간소비} : u_{ik}^{\text{수입},o} = \frac{\alpha_{io}^{\text{중간재}}}{\sum_o \alpha_{io}^{\text{중간재}}} u_{ik}^{\text{수입}}$$

$$\text{국가 간 최종소비} : y_{ik}^{\text{수입},o} = \frac{\alpha_{io}^{\text{최종소비재}}}{\sum_o \alpha_{io}^{\text{최종소비재}}} y_{ik}^{\text{수입}}$$

$$\text{국가 간 총고정자본형성} : y_{i4}^{\text{수입},o} = \frac{\alpha_{io}^{\text{최종소비재}}}{\sum_o \alpha_{io}^{\text{최종소비재}}} y_{i4}^{\text{수입}}$$

이와 같은 추정 과정이 본 프로젝트가 기존의 글로벌 가치 사슬 분석 및 지역 간 산업 연관분석에 대해 가지는 핵심적인 차별점이라고 할 수 있다. 기존의 방법은 대체로 모든 경제 주체가 특정 재화 수요를 동일한 비율의 수입품으로 충당한다는 표준 수입 비율 가정을 적용하였지만¹⁷⁾ 본 프로젝트에서는 같은 상품이라도 수입품 충당 비율이 이용 목적 뿐만 아니라 이용 목적 안에서도 원수출 국가별로 달라진다는 Puzzello(2012)의 주장을 수용하여 상품과 산업 및 최종수요 항목별로 각기 다른 수입 비율이 적용되는 수입거래표 및 국가 간 수입거래표를 도출하였다.

16) 가공무역에서의 재수출 혹은 중계무역에서의 재수출로서 II장 2절에서 언급한 재수출과는 의미가 다르다.
 17) Dietzenbacher et al.(2013)에서는 대표적인 예로 Peters et al.(2011) 및 Johnson and Noguera (2012)를 들고 있다. 전자는 글로벌 무역 분석 프로젝트(Global Trade Analysis Project) 데이터베이스를 활용한 다지역 산업연관모형(Multi-Regional Input-Output)에 관한 논문이며, 후자에서는 국가 간 무역 데이터와 투입산출표를 결합하여 총수출 중 국가간 부가가치 비율(ratio of Value Added to Gross Exports)을 다루고 있다.

한편, <표 2>와 <표 3>에서 수출은 FOB(Free on Board) 기준인 것과 달리 수입은 CIF (Cost, Insurance and Freight) 기준으로 계상되었다. 향후 세계공급사용표를 작성하기 위해서는 자국 입장에서의 상대국으로부터의 수입과 상대국 입장에서의 자국으로의 수출이 일치해야 하며, 1993 SNA에서는 수출입 금액을 모두 FOB 기준으로 산출하도록 하고 있기 때문에 수출에 비해 정보량이 더 많은 수입을 FOB 기준에 따라 조정해야 한다¹⁸⁾. 본 프로젝트에서는 두 기준의 차액인 화물운임 및 보험료를 국가 간 무역 마진¹⁹⁾으로 간주하여 이를 추정하였다. 우선 UN Comtrade 데이터베이스를 거리와 내륙국가 여부, 그리고 동일 대륙 더미변수를 설명변수로 하는 중력모형에 대입하여 수출단가 대비 수입단가 비율을 추정하였고, 여기에서 상품별 CIF-FOB 마진을 도출하였다²⁰⁾. 계산의 편의상 상품별 마진을 상품 대응 벡터 \mathbf{tm}^o (59×1 벡터)로 표현하기로 한다.

나. 국가 간 공급사용표 및 세계공급사용표 작성

앞서 도출한 CIF-FOB 마진을 벡터 \mathbf{tm}^o 의 원소(상품별 마진율)를 순생산물세표 및 마진표 작성 방법과 마찬가지로 국가 간 수입 행렬 $\hat{\mathbf{U}}_b^{\text{수입},o}$, $\hat{\mathbf{Y}}_b^{\text{수입},o}$ 의 산업별 각 대응 상품 원소에 곱하면 국가별 CIF-FOB 마진 행렬 \mathbf{TM}_U^o , \mathbf{TM}_Y^o 을 도출할 수 있으며 수입 행렬에서 마진 행렬을 차감하면 FOB 기준 수입 행렬 $\hat{\mathbf{U}}_b^{\text{수입},o,\text{FOB}}$, $\hat{\mathbf{Y}}_b^{\text{수입},o,\text{FOB}}$ 이 도출된다.

$$\mathbf{TM}_U = \begin{bmatrix} \hat{u}_{1,1}^{\text{수입},o} & tm_1 & \cdots & \hat{u}_{1,35}^{\text{수입},o} & tm_1 \\ \vdots & \ddots & & \vdots & \\ \hat{u}_{59,1}^{\text{수입},o} & tm_{59} & \cdots & \hat{u}_{59,35}^{\text{수입},o} & tm_{59} \end{bmatrix}, \quad \mathbf{TM}_Y = \begin{bmatrix} \hat{y}_{1,1}^{\text{수입},o} & tm_1 & \cdots & \hat{y}_{1,4}^{\text{수입},o} & tm_1 \\ \vdots & \ddots & & \vdots & \\ \hat{y}_{59,1}^{\text{수입},o} & tm_{59} & \cdots & \hat{y}_{59,4}^{\text{수입},o} & tm_{59} \end{bmatrix}$$

$$\hat{\mathbf{U}}_b^{\text{수입},o,\text{FOB}} = \hat{\mathbf{U}}_b^{\text{수입},o} - \mathbf{TM}_U^o, \quad \hat{\mathbf{Y}}_b^{\text{수입},o,\text{FOB}} = \hat{\mathbf{Y}}_b^{\text{수입},o} - \mathbf{TM}_Y^o$$

18) 1993 SNA에서는 기본적으로 총수입 및 총수출은 FOB 기준으로 계상하되 상품별 수출입액 평가시 FOB 기준 데이터를 구할 수 없을 경우 CIF 기준 평가금액과 FOB 기준 조정금액을 같이 산출하도록 했다 (System of National Accounts(1993), §3.85, p.96).

19) 엄밀한 의미에서 도소매마진 및 화물운임, 보험료 등은 구분되어야 하나 본 프로젝트에서는 <표 2>의 도출 과정에서부터 이러한 항목을 '마진'이라는 단일한 분류로 두고 있다.

20) 프로젝트에서는 Streicher and Stehrer(2012)가 제시한 방법론에 따라 세계산업연관표 분석 대상 국가의 마진 부문 경상수지를 기준으로 도출된 마진을 검증하였고, 그 결과 수입액 대비 평균 5~7%가 총 운송 비용으로 투입되는 것으로 추정되었다.

이를 토대로 FOB 기준 국가 간 수입거래표를 작성하는데, RAS 방법을 통해 자국의 수출입 총액과의 정합성을 갖춘 수입거래표를 추정하면 이것이 국가 간 사용표가 된다. 한 예시로, A국과 B국, 그리고 기타지역(Rest of World)이 존재하는 3국 모형에서 A국의 국가 간 공급사용표는 다음과 같다.

<표 5> A국 기준 국가 간 공급사용표(기초가격 기준)

	상품	산업	최종 소비	재고 증감	수출	총계
상품(A국)		U^{AA}	Y^{AA}	s^A	e^A	$\sim^A q$
B국으로부터 A국 수입 (FOB 기준)		U^{BA}	Y^{BA}			$\sim^{BA} m$
기타지역으로부터 A국 수입 (FOB 기준)		U^{RA}	Y^{RA}			$\sim^{RA} m$
산업(A국)	V^A					x^A
수입(A국 전체, CIF 기준)	$(m^{BA} + m^{RA})'$					m^A
부가가치		$(w^A)'$				w^A
국제 운송마진(운송비용)		$(tm_U^A)'$	$(tm_Y^A)'$			tm^A
총계	$(q^A)'$	$(x^A)'$	$(y^A)'$	s^A	e^A	

$(tm_U^A)'$, $(tm_Y^A)'$: A국의 상대국가별 CIF-FOB 마진 행렬 TM_U^O , TM_Y^O 의 열합벡터를 모두 합한 벡터(각각 1×35 , 1×4 벡터)

음영 : CIF 기준 가격으로 산출된 부분
 총고정자변형성 : 최종소비 항목에 포함

위 표에서 U^{BA} , Y^{BA} , U^{RA} , Y^{RA} 는 A국의 전체 수출입 금액(FOB 기준)을 제약조건으로, 앞서 도출한 $\hat{U}_b^{수입,o,FOB}$, $\hat{Y}_b^{수입,o,FOB}$ 중 B국 및 기타지역에 해당하는 수입 행렬을 초기값으로 하여 RAS 방법으로 추정한 행렬이다.

4. 세계공급사용표로부터 세계투입산출표 변환

세계공급사용표는 작성 대상 국가의 각 국가 간 공급사용표(기타지역 포함 2016년 공표 기준 총 1,892개)를 모두 연결한 것이다. 예를 들어 <표 5>의 기타지역 포함 3국 모형의 각 국가 간 공급사용표를 연결하면 3국 간 세계공급사용표를 만들 수 있다.

〈표 6〉 3국 세계공급사용표(기초가격 기준)

	상품 (A국)	상품 (B국)	산업 (A국)	산업 (B국)	최종 소비 (A국)	최종 소비 (B국)	재고 증감	수출 (기타 지역행)	총계
상품(A국)			U^{AA}	U^{AB}	Y^{AA}	Y^{AB}	s^A	e^{AR}	$\sim^A q$
상품(B국)			U^{BA}	U^{BB}	Y^{BA}	Y^{BB}	s^B	e^{BR}	$\sim^B q$
상품(기타지역)			U^{RA}	U^{RB}	Y^{RA}	Y^{RB}			
산업(A국)	V^A								x^A
산업(B국)		V^B							x^B
수입(CIF)	$(m^{BA} + m^{AB} + m^{RA})'$ $(m^{RB})'$								
부가가치			$(w^A)'$	$(w^B)'$					
국제 운송마진 (운송비용)			$(tm_U^A)'$	$(tm_U^B)'$	$(tm_Y^A)'$	$(tm_Y^B)'$			tm
총계	$(q^A)'$	$(q^B)'$	$(x^A)'$	$(x^B)'$					

음영 : CIF 기준 가격으로 산출된 부분

프로젝트에서는 투입산출표를 「산업×산업」 행렬²¹⁾ 구조로 작성하였다²²⁾. 이에 따라 공급사용표를 투입산출표로 전환하기 위하여 고정 상품 판매구조 가정(fixed product sales structure assumption)을 채택하였다. 고정 상품 판매구조 가정은 동일한 상품으로서 특정 산업의 주생산물과 다른 산업의 부차적 생산물이 시장에 판매될 때 두 경우 모두 동일한 판매 구조를 보인다는 것이다. 다른 가정으로는 각 산업은 주생산물과 부차적 생산물 조합에 상관없이 고유의 판매 구조에 따라 판매된다는 고정 산업 판매구조 가정(fixed industry sales structure assumption)이 있으나, 가정 자체가 현실과 부합하지 않으며 전환된 투입산출표에 음(-)인 수치가 있을 수 있다는 문제가 있어 채택되지 않았다.

고정 상품 판매구조 가정에 따르면 공급사용표를 투입산출표로 변환하기 위해서는 사용표 행렬에 공급표 행렬을 이용한 변환 행렬을 곱해주면 된다. 이때 변환 행렬의 각 원소는 특정 상품의 총수요 중 산업별 국산품 공급 비율을 의미한다. <표 6>의 3국간 모형에서 A국과 B국의 변환 행렬은 다음과 같다²³⁾.

21) Dietzenbacher et al.(2013), p. 90.

22) Eurostat은 「상품×상품」 및 「산업×산업」 행렬 투입산출표를 모두 제공하고 있으며, OECD에서는 「산업×산업」 행렬 투입산출표만을 제공하고 있다.

23) 해당 가정을 토대로 한 수식 전개에 관한 자세한 설명은 Eurostat(2008), p. 349~352 참고.

$$\mathbb{B}^A = \mathbb{V}^A (\text{diag}(\tilde{\mathbf{q}}^A)^{-1}), \quad \mathbb{B}^B = \mathbb{V}^B (\text{diag}(\tilde{\mathbf{q}}^B)^{-1}) \quad 24)$$

$\text{diag}(\mathbf{q})$: 벡터 \mathbf{q} 의 대각행렬

이를 <표 6>의 항목별 사용표 행렬(중간수요, 총고정자본형성 포함 최종소비, 재고증감 및 수출)에 적용하면 투입산출표의 각 행렬을 도출할 수 있다.

<표 7> 세계사용표 각 행렬의 세계투입산출표로의 변환

중간수요	$\mathbb{Z}^{AA} = \mathbb{B}^A \mathbf{U}^{AA}, \quad \mathbb{Z}^{AB} = \mathbb{B}^A \mathbf{U}^{AB}$ $\mathbb{Z}^{BA} = \mathbb{B}^B \mathbf{U}^{BA}, \quad \mathbb{Z}^{BB} = \mathbb{B}^B \mathbf{U}^{BB}$
최종소비	$\mathbb{F}^{AA} = \mathbb{B}^A \mathbf{Y}^{AA}, \quad \mathbb{F}^{AB} = \mathbb{B}^A \mathbf{Y}^{AB}$ $\mathbb{F}^{BA} = \mathbb{B}^B \mathbf{Y}^{BA}, \quad \mathbb{F}^{BB} = \mathbb{B}^B \mathbf{Y}^{BB}$
재고증감	$\tilde{\mathbf{s}}^A = \mathbb{B}^A \mathbf{s}^A, \quad \tilde{\mathbf{s}}^B = \mathbb{B}^B \mathbf{s}^B$
수출(기타지역행)	$\tilde{\mathbf{e}}^{AR} = \mathbb{B}^A \mathbf{e}^{AR}, \quad \tilde{\mathbf{e}}^{BR} = \mathbb{B}^B \mathbf{e}^{BR}$

$\mathbb{Z}^{AA}, \mathbb{Z}^{AB}, \mathbb{Z}^{BA}, \mathbb{Z}^{BB}$: 중간수요 행렬(35(산업)×35(산업) 행렬)

$\mathbb{F}^{AA}, \mathbb{F}^{AB}, \mathbb{F}^{BA}, \mathbb{F}^{BB}$: 총고정자본형성 포함 최종수요 행렬(35×4(항목) 행렬)

한편 프로젝트에서는 잔차 문제를 제거하기 위해 기타지역을 별도의 국가로 간주하였기 때문에, 여타 분석 대상 국가와 마찬가지로 투입산출표 변환 행렬 및 공급사용표 작성 시 별도로 산출하지 않았던 기타지역 내 투입산출행렬을 별도로 추정하였다. 우선 기타지역의 변환 행렬은 분석 대상 국가 43개국의 공급표를 모두 더한 \mathbb{V}^R 에 \mathbb{V}^R 의 열합벡터 $\tilde{\mathbf{q}}^R$ 을 토대로 도출되었다. 이는 기타지역의 산업 구조가 분석 대상 국가와 유사하다는 가정 하에 분석 대상 국가의 평균 총수요 중 국산 조달 비율을 기타지역에 적용한다는 의미를 갖는다.

$$\mathbb{B}^R = \mathbb{V}^R (\text{diag}(\tilde{\mathbf{q}}^R)^{-1})$$

기타지역의 세계투입산출표를 작성하기 위해서는 <표 6>에서 분석 대상 국가의 기타지역으로부터의 수입 $\mathbf{U}^{RA}, \mathbf{U}^{RB}, \mathbf{Y}^{RA}, \mathbf{Y}^{RB}$ 이외에도 분석 대상 국가의 기타지역으로의 수

24) 대각행렬의 역행렬은 원대각행렬 각 원소에 역수가 적용된 행렬이 된다.

출, 기타지역 내 투입산출표가 필요하다. 이 중 기타지역으로부터의 수입(기타지역에서의 수출)은 \mathbb{B}^R 을 적용하여 투입산출표로 변환할 수 있다.

$$\begin{aligned} \mathbf{Z}^{RA} &= \mathbb{B}^R \mathbf{U}^{RA}, & \mathbf{Z}^{RB} &= \mathbb{B}^R \mathbf{U}^{RB} \\ \mathbf{F}^{RA} &= \mathbb{B}^R \mathbf{Y}^{RA}, & \mathbf{F}^{RB} &= \mathbb{B}^R \mathbf{Y}^{RB} \end{aligned}$$

기타지역으로의 수출(기타지역에서의 수입)은 국가 간 수출금액 중 평균 산업별·이용 목적별 비율을 적용하여 수출 벡터 $\tilde{\mathbf{e}}^{AR}$, $\tilde{\mathbf{e}}^{BR}$ 을 중간수요 행렬(\mathbf{Z}^{AR} , \mathbf{Z}^{BR}) 및 최종소비 행렬(\mathbf{F}^{AR} , \mathbf{F}^{BR})로 분리하였다. 평균은 기타지역이 대부분 개발도상국가임을 감안하여²⁵⁾ 선진국(개발도상국 지정 제외 국가)으로부터 지정된 개발도상국(브라질, 러시아, 인도, 중국, 인도네시아, 멕시코)으로 수출되는 물품 중 산업별·이용 목적별 비율로부터 계산하였다.

기타지역 내 투입산출표(\mathbf{Z}^{RR} , \mathbf{F}^{RR})는 UN 국민계정통계와 UNIDO 산업 통계로부터 도출하였다. 우선 프로젝트 분석 대상 국가를 제외한 모든 국가의 경제활동별·최종수요별 GDP를 더함으로써 부가가치 데이터를 산출하고, 항목별 총투입(총산출)은 지정된 개발도상국의 항목별 부가가치율을 앞서 산출한 GDP 데이터에 적용하여 추정하였다. 이후 프로젝트의 산업 분류별 총투입을 도출하기 위해 분석 대상 국가 제외 UNIDO 산업 통계의 평균 국가별 총산출 산업별 비율을 산출하여 앞서 추정한 항목별 총투입에 적용하였다. 마지막으로, 세부 항목별 총투입을 투입산업별로 분리하기 위해 지정된 개발도상국의 투입산업별 대 수요산업 투입액 가중평균을 산출하여 적용하고 이에 따라 도출된 기타지역 내 투입산출표 행렬을 초기값, 나머지 부분을 제약조건으로 하는 RAS 방법을 이용하여 최종적으로 기타지역 내 투입산출표를 편제하였다. 이와 같은 과정을 거쳐 「산업×산업」 형태의 세계투입산출표를 작성하였으며, 이해의 편의를 위하여 <표 6>의 3국간 모형을 적용한 세계투입산출표는 다음과 같다.

25) 분석 대상 국가는 모든 G7 국가 및 대부분의 OECD 회원국(아이슬란드, 노르웨이, 뉴질랜드, 칠레, 이스라엘 제외)을 포함하고 있어 사실상 분석 대상에서 제외된 국가는 대부분 개발도상국으로 볼 수 있다.

〈표 8〉 3국 세계투입산출표(기초가격 기준)

	중간수요			최종수요			재고증감	총계
	A국	B국	기타	A국	B국	기타		
산업(A국)	Z^{AA}	Z^{AB}	Z^{AR}	F^{AA}	F^{AB}	F^{AR}	\tilde{s}^A	x^A
산업(B국)	Z^{BA}	Z^{BB}	Z^{BR}	F^{BA}	F^{BB}	F^{BR}	\tilde{s}^B	x^B
산업(기타지역)	Z^{RA}	Z^{RB}	Z^{RR}	F^{RA}	F^{RB}	F^{RR}		x^R
부가가치	$(w^A)'$	$(w^B)'$	$(w^R)'$					
국제운송마진	$(tm_U^A)'$	$(tm_U^B)'$	0	$(tm_Y^A)'$	$(tm_Y^B)'$	0	0	tm
총계(총산출)	$(x^A)'$	$(x^B)'$	$(x^R)'$					

III. 맺음말

세계산업연관표는 국가별 산업연관표와 달리 전 세계를 하나의 단일한 폐쇄경제로 간주하고 작성되었다. 즉, 개방경제를 전제로 하는 국가별 산업연관표에서는 수입 및 국내생산품을 중간수요 및 소비와 투자, 재고 축적에 이용하고 해외 수요는 수출이라는 단일 항목으로 계상되지만 세계산업연관표에서는 해당 수출은 모두 다른 국가의 중간수요 및 소비와 투자를 충족시키는 데 이용되기 때문에 세계 전체에서는 수출 항목이 별도로 존재하지 않는다. 이는 <표 8>의 세계투입산출표 중 A국의 투입산출표를 따로 분리하여 보면 알 수 있다.

<표 9> 3국 중 A국의 투입산출표(기초가격 기준)

	중간수요 산업	최종수요	재고 증감	수출	총계
산업	$\mathbf{Z}^{AA} + (\mathbf{Z}^{BA} + \mathbf{Z}^{RA})$	$\mathbf{F}^{AA} + (\mathbf{F}^{BA} + \mathbf{F}^{RA})$	$\tilde{\mathbf{s}}^A$	$\tilde{\mathbf{e}}^A$	\mathbf{x}^A
부가가치	$(\mathbf{w}^A)'$				
국제운송마진	$(\mathbf{tm}_U^A)'$	$(\mathbf{tm}_Y^A)'$			tm^A
총계(총산출)	$(\mathbf{x}^A)'$				

$$\tilde{\mathbf{e}}^A : \tilde{\mathbf{e}}^{AR} + (\mathbf{Z}^{AB} + \mathbf{F}^{AB})_l$$

- l : 행합 벡터(35×1 벡터, 각 원소는 모두 1)

<표 9>의 국가별 투입산출표에서는 수입품의 원수출국이나 수출품의 행선국을 별도로 파악할 필요가 없으며, 다만 개방경제인 관계로 수입거래표 및 수출입 벡터 작성이 필요하다. 그러나 <표 8>에서는 전 세계 이외 영역과의 교류가 없는 것을 전제하기 때문에 수출입 항목은 존재하지 않으며 국가 간 수출입에 있어서도 잔차를 남길 수 없기 때문에 편제 과정에서 수출입 행선지 파악 및 특정 국가 간 수출 및 수입 액수 동기화가 반드시 필요하다. 이는 세계산업연관표를 작성하는 데 있어서 국가 간 무역(bilateral trade) 통계가 매우 중요한 역할을 하고 있음을 의미한다. 이러한 통계가 부실하였다면, 프로젝트에서 특정 국가에서 여러 다른 국가로부터 동일한 상품을 수입하더라도 거래상대 국가에 따라 상품의 이용 목적별 수요 비율이 달라진다는 전제에 따라 거래국가별·상품별로 별도의 수입충당비

을 추정하는 것이 매우 어려웠을 것이다. 특히 재화 무역과 달리 서비스 무역은 신뢰성 있는 국가 간 무역 통계가 적어 수입충당비율 추정이 어려웠는데, Dietzenbacher et al.(2013)은 결론에서 UN 통계국이 서비스 무역 통계 분과를 무역 통계 위원회에 설치한 것과 여행수지 의존 국가의 통계 품질 향상 가능성 등의 노력을 현 상황을 개선시키는 기회로 보고 있다고 밝혔다.

세계산업연관표의 접근 방법은 우리나라의 지역 간 산업연관표에도 적용 가능하다. 이는 신뢰성 있는 통계를 작성하기 위해서는 지역 내 총생산 등 지역 간 경제 통계가 세밀하게 구축되어 있어야 하되 서비스 무역의 경우처럼 정확한 통계를 찾기 어려우면 여러 통계를 교차검증하여 이용하는 등 대안을 염두에 두고 있어야 함을 의미한다. 따라서 지역 간 산업연관표 편제 시 기반 통계를 면밀히 점검하고, 통계 정합성에 대한 대안을 연구해 나갈 필요가 있겠다.

<경제통계국 투입산출팀 이재민 조사역, 정영호 과장>

참고문헌

- 이종호, 양시환, “글로벌 가치사슬이 산업별 생산성에 미치는 영향”, 『조사통계월보』, 제 2017-7호, 한국은행, 2017, pp. 55-87.
- 한국은행, 『산업연관분석해설』, 2014.
- Dietzenbacher, Erik, et al. “The construction of world input-output tables in the WIOD project.” *Economic Systems Research* 25.1 (2013): 71-98.
- Eurostat. *Eurostat Manual of Supply, Use and Input-Output Tables*, 2008.
- Grossman, Gene M., and Esteban Rossi-Hansberg. “Trading tasks: A simple theory of offshoring.” *The American Economic Review* 98.5 (2008): 1978-1997.
- Johnson, Robert C., and Guillermo Noguera. “Accounting for intermediates: Production sharing and trade in value added.” *Journal of International Economics* 86.2 (2012): 224-236.
- Miller, Ronald E., and Peter D. Blair. *Input-output analysis: foundations and extensions*. 2nd ed. Cambridge University Press, 2009.
- O'Mahony, Mary, and Marcel P. Timmer. “Output, input and productivity measures at the industry level: the EU KLEMS database.” *The Economic Journal* 119.538 (2009).
- Peters, Glen P., Robbie Andrew, and James Lennox. “Constructing an environmentally-extended multi-regional input-output table using the GTAP database.” *Economic Systems Research* 23.2 (2011): 131-152.
- Puzzello, Laura. “A proportionality assumption and measurement biases in the factor content of trade.” *Journal of International Economics* 87.1 (2012): 105-111.
- Temurshoev, Umed, and Marcel P. Timmer. “Joint estimation of supply and use tables.” *Papers in Regional Science* 90.4 (2011): 863-882.
- Timmer, Marcel P., et al. *An anatomy of the global trade slowdown based on the WIOD 2016 release*. No. GD-162. Groningen Growth and Development Centre, University of Groningen, 2016

한국은행 경제통계 안내

Information on Statistics of the Bank of Korea

경제통계시스템(ECOS) 수록 통계

No	통계명칭	내 용
1	통화 및 유동성지표	
	주요 통화금융지표	주요 통화금융지표
	본원통화	본원통화 구성내역(말잔)
		본원통화 구성내역(평잔)
	M2(광의통화)	M2 상품별구성내역(말잔)
		M2 상품별구성내역(평잔)
		M2 경제주체별 보유현황(말잔)
		M2 경제주체별 보유현황(평잔)
		M2 기관별구성내역(말잔)
		M2 기관별구성내역(평잔)
	예금취급기관 개관표 (종합표, 말잔)	중앙은행 개관표(말잔)
		기타예금취급기관 개관표(말잔)
	Lf(금융기관 유동성) 관련통계	Lf 상품별 구성내역(말잔)
		Lf 상품별 구성내역(평잔)
		Lf 기관별 구성내역(말잔)
Lf 기관별 구성내역(평잔)		
L(광의유동성) 구성내역(말잔)	L(광의유동성) 구성내역(말잔, 계절조정계열)	
	L(광의유동성) 구성내역(말잔, 원계열)	
2	한국은행 주요계정 및 기준금리	
	한국은행 주요계정(말잔)	한국은행 주요계정(말잔)
	화폐발행잔액	화폐발행잔액
	예금은행 지급준비액(평잔)	예금은행 지급준비액(평잔)
	예금은행 지급준비액(평잔, 신, 2008.2월 상반월)	예금은행 지급준비액(평잔, 신, 2008.2월 상반월)
	한국은행 원화대출금(말잔)	한국은행 대출금(말잔)
	한국은행 기준금리 및 여수신금리	한국은행 기준금리 및 여수신금리
3	예금/대출금/기타금융	
	예금은행 수신	예금은행 종별예금(말잔)
		예금은행 종별예금(평잔)
		예금은행 예금주별예금(말잔)
		예금은행 기간별 정기예금(말잔)
		예금은행 총수신(말잔)
		예금은행 총수신(평잔)
	예금은행 자금별 대출금(말잔)	예금은행 자금별 대출금(말잔)
	산업별 대출금	예금취급기관
		예금은행
		예금은행 지역별
		비은행예금취급기관
비은행금융기관 수신 및 여신	비은행기관 수신(말잔)	

No	통계명칭	내 용
	비은행금융기관 수신 및 여신	비은행기관 수신(평균) 비은행기관 여신(말잔)
	지역별 여수신	예금은행 수신 및 여신
		비은행예금취급기관 수신 및 여신
		생명보험회사 수신 및 여신
	가계신용	가계신용(분기별)
		예금취급기관의 가계대출(월별)
		예금취급기관의 가계대출 지역별(월별)
	기타금융통계	예금은행 예금회전율
		금융기관 점포 및 인원현황
		은행대출금 연체율
		보험계약실적
		예금규모별 계좌 및 금액
	4	금리
시장금리	시장금리(일별)	
	시장금리(월, 분기, 년)	
예금은행 가중평균금리	수신금리	
	대출금리	
예금은행 금리수준별 여수신 비중	예금은행 금리수준별 여수신 비중	
예금은행 고정 및 특정 금리연동대출 비중	신규취급액기준	
	잔액기준	
비은행금융기관 가중평균금리	수신금리(신규취급액기준)	
	대출금리(신규취급액기준)	
5	지급결제	
어음교환 및 부도	어음교환 및 부도	
	지역별 부도율	
한은금융망	한은금융망	
지로시스템	지로시스템	
은행공동망	은행공동망	
BIS기준통계	결제시스템별 통계	
	지급수단별 통계	
전자금융통계	인터넷뱅킹	
	전자어음	
	전자화폐	
회계시스템통계	한은금융망결제통계표	
	오전/오후 차액결제표	
	일일지급결제동향	
신용카드	신용카드	
	지역별 소비유형별 개인 신용카드	
금융정보화 통계	은행	
	증권	
	보험	

No	통계명칭	내 용
	금융정보화 통계	기타
6	증권/재정	
	증시자금 조달상황	증시자금 조달상황
	주식거래 및 주가지수	주식시장(일별) 주식거래 및 주가지수 투자자별 주식거래 증시주변자금동향 소유자별 주식소유현황 주가지수선물거래 주가지수옵션거래
	채권거래	채권거래
	주요 국공사채 발행 및 잔액	주요 국공사채 발행 및 잔액
	재정	통합재정수지 조세징수액
7	물가	
	생산자물가지수(2010=100)	생산자물가지수(기본분류)(2010=100) 생산자물가지수(특수분류)(2010=100) 생산자물가지수(품목별)(2010=100) 생산자물가지수 - 총지수(1910~1964)(2010=100)
	국내공급물가지수(2010=100)	국내공급물가지수(2010=100)
	총산출물가지수(2010=100)	총산출물가지수(2010=100)
	소비자물가지수(2010=100)	소비자물가지수(2010=100)(전국) 소비자물가지수(2010=100)(전국, 특수분류)
	수출물가지수(2010=100)	수출물가지수(기본분류)(2010=100) 수출물가지수(특수분류)(2010=100) 수출물가지수(품목별)(2010=100)
	수입물가지수(2010=100)	수입물가지수(기본분류)(2010=100) 수입물가지수(특수분류)(2010=100) 수입물가지수(품목별)(2010=100) 수입물가지수(용도별)(2010=100)
	기타가격지수	주택매매가격지수(KB) 주택전세가격지수(KB) 지역별 지가변동률 유형별 주택매매가격지수 유형별 주택전세가격지수 유형별 주택월세통합가격지수 아파트실거래가격지수 상권별 오피스임대가격지수 상권별 중대형매장용부동산임대가격지수
8	국제수지/외채/환율	
	국제수지	국제수지 계절변동조정 경상수지(경상수지 S.A.)

No	통계명칭	내 용
	국제수지	서비스무역세분류통계
		지식재산권 무역수지
		수출결제통화
		수입결제통화
	지역별 국제수지	지역별 경상수지
		지역별 자본수지·금융계정
	통관기준 수출입	수출
		수입
	무역지수 및 교역조건지수	수출금액지수
		수출물량지수
		수입금액지수
		수입물량지수
		교역조건지수
	외환보유액	외환보유액
	국제투자대조표(IIP)	국제투자대조표(IIP)
		지역별 국제투자대조표
통화별 국제투자대조표		
대외채무/채권	대외채무	
	대외채권	
	순대외채권	
환율	일일환율	
	평균환율, 기말환율	
해외투자/차관	해외직접투자(신고기준)	
	외국인투자(신고기준)	
9	기업경기 / 소비자동향 / 경제심리	
	기업경기실사지수	기업경기실사지수(한국은행, 전국, 전기간)
		업종별 기업경기실사지수(한국은행, 전국)(~2009.8)
		업종별 기업경기실사지수(한국은행, 전국)(2009.8~)
		기업경기실사지수(지역)
소비자동향조사	소비자동향조사(한국은행, 전국)(월)(2008년 9월~)	
	소비자동향조사(한국은행, 전국)(분기) (1995년 3분기~2008년 2분기)	
	소비자동향조사(기대인플레이션율)(전국)	
	소비자동향조사(한국은행, 지역)(월)	
경제심리지수	경제심리지수	
10	국민계정	
	주요지표	연간지표
		분기 지표
경제활동별, 지출항목별 규모	경제활동별 GDP 및 GNI	
	국내총생산에 대한 지출	
	GDP 디플레이터	

No	통계명칭	내 용
	생산, 지출, 자본, 국외거래(종합계정)	국내총생산과 지출
		국민처분가능소득과 처분(명목, 연간)
		자본거래(명목, 연간)
		국외거래(명목, 연간)
	제도부문별 생산, 소득, 자본계정	제도부문별 생산계정(명목, 연간)
		제도부문별 소득계정(명목, 연간)
		제도부문별 자본계정(명목, 연간)
	GDP에 대한 성장기여도	경제활동별 성장기여도(계절조정, 실질, 분기)
		경제활동별 성장기여도(원계열, 실질, 분기 및 연간)
		지출항목별 성장기여도(계절조정, 실질, 분기)
		지출항목별 성장기여도(원계열, 실질, 분기 및 연간)
	총자본형성	자본재형태별 총자본형성
		주체별 총자본형성
		경제활동별 총자본형성
		경제활동별 설비투자
	최종소비지출	가계의 최종소비지출
		가계에 봉사하는 비영리단체의 최종소비지출 및 투입, 산출
	일반정부 및 공공부문계정	공공부문의 부문별 총수입, 총지출, 저축투자차액(명목, 연간)
		일반정부의 부문별 총수입, 총지출, 저축투자차액(명목, 연간)
		일반정부의 부문별 기능별 최종소비지출(명목, 연간)
일반정부의 부문별 기능별 총지출(명목, 연간)		
총저축과 총투자(명목)	총저축과 총투자(계절조정, 명목, 분기)	
	총저축과 총투자(원계열, 명목, 분기 및 연간)	
정보통신산업(실질)	정보통신산업(계절조정, 실질, 분기)	
	정보통신산업(원계열, 실질, 분기 및 연간)	
국민대차대조표	자산별 자본스톡	
	제도부문별 대차대조표	
	제도부문별 순자본스톡	
	제도부문별 비금융자산별 보유손익(명목)	
	제도부문별 생산자본스톡	
	제도부문별 자본서비스물량지수	
	경제활동별 순자본스톡	
	경제활동별 생산자본스톡	
	경제활동별 자본서비스물량지수	
	주택 시가총액(명목, 연말기준)	
	시도별 토지자산 순자본스톡(명목, 연말기준)	
11	자금순환	
	2008 SNA기준	금융거래표(2009~)

No	통계명칭	내 용
	2008 SNA기준	금융자산부채잔액표(2008~)
	1993 SNA기준(2002~2013)	금융거래표(2003~2013)
		금융자산부채잔액표(2002~)
	1968 SNA기준(1975~2005)	금융거래표
		금융자산부채잔액표
	1968 SNA기준(1969~1974, 연간)	금융거래표(1970~1974)
		금융자산부채잔액표(1969~1974)
	1953 SNA기준(1962~1970, 연간)	금융거래표(1963~1970)
		금융자산부채잔액표(1962~1970)
12	기업경영분석	
	기업경영분석지표	기업경영분석지표(2009~, 전수조사)
		기업경영분석지표(2007~2010)
		기업경영분석지표(~2007)
	성장성에 관한 지표	성장성에관한지표(중소기업기본법, 2015~, 전수조사)
		성장성에관한지표(조세특례제한법, 2019~2014, 전수조사)
		성장성에관한지표(2007~2010, 표본조사)
		성장성에관한지표(1998~2007, 표본조사)
		성장성에관한지표(1990~1997, 표본조사)
	손익의 관계비율	손익의관계비율(중소기업기본법, 2015~, 전수조사)
		손익의관계비율(조세특례제한법, 2009~2014, 전수조사)
		손익의관계비율(2007~2010, 표본조사)
		손익의관계비율(1998~2007, 표본조사)
		손익의관계비율(1990~1997, 표본조사)
	자산·자본의 관계비율	자산·자본의관계비율(중소기업기본법, 2015~, 전수조사)
		자산·자본의관계비율(조세특례제한법, 2009~2014, 전수조사)
		자산·자본의관계비율(2007~2010, 표본조사)
		자산·자본의관계비율(1998~2007, 표본조사)
		자산·자본의관계비율(1990~1997, 표본조사)
	자산·자본의 회전을	자산·자본의회전을(중소기업기본법, 2015~, 전수조사)
		자산·자본의회전을(조세특례제한법, 2009~2014, 전수조사)
		자산·자본의회전을(2007~2010, 표본조사)
		자산·자본의회전을(1998~2007, 표본조사)
		자산·자본의회전을(1990~1997, 표본조사)
	생산성에 관한 지표	생산성에관한지표(중소기업기본법, 2015~, 전수조사)
		생산성에관한지표(조세특례제한법, 2009~2014, 전수조사)
		생산성에관한지표(2007~2010, 표본조사)
		생산성에관한지표(2002~2007, 표본조사)
		생산성에관한지표(1998~2006, 표본조사)
		생산성에관한지표(1990~1997, 표본조사)
	부가가치의 구성	부가가치의구성(중소기업기본법, 2015~, 전수조사)
		부가가치의구성(조세특례제한법, 2009~2014, 전수조사)

No	통계명칭	내 용
	부가가치의 구성	부가가치의구성(2007~2010, 표본조사)
		부가가치의구성(2002~2007, 표본조사)
		부가가치의구성(1998~2006, 표본조사)
		부가가치의구성(1990~1997, 표본조사)
	재무상태표(대차대조표)	재무상태표(대차대조표)(중소기업기본법, 2015~, 전수조사)
		재무상태표(대차대조표)(조세특례제한법, 2009~2014, 전수조사)
		재무상태표(대차대조표)(2008~2010, 표본조사)
		재무상태표(대차대조표)(2000~2007, 표본조사)
		재무상태표(대차대조표)(1999, 표본조사)
		재무상태표(대차대조표)(1997~1998, 표본조사)
		재무상태표(대차대조표)(1990~1996, 표본조사)
	손익계산서	손익계산서(중소기업기본법, 2015~, 전수조사)
		손익계산서(조세특례제한법, 2009~2014, 전수조사)
		손익계산서(2008~2010, 표본조사)
		손익계산서(2000~2007, 표본조사)
		손익계산서(1999, 표본조사)
		손익계산서(1997~1998, 표본조사)
		손익계산서(1990~1996, 표본조사)
	제조원가명세서	제조원가명세서(중소기업기본법, 2015~, 전수조사)
		제조원가명세서(조세특례제한법, 2009~2014, 전수조사)
		제조원가명세서(2008~2010, 표본조사)
		제조원가명세서(2000~2007, 표본조사)
		제조원가명세서(1990~1999, 표본조사)
	이익잉여금처분계산서	이익잉여금처분계산서(중소기업기본법, 2015~, 전수조사)
		이익잉여금처분계산서(조세특례제한법, 2009~2014, 전수조사)
		이익잉여금처분계산서(2007~2010, 표본조사)
		이익잉여금처분계산서(2000~2007, 표본조사)
		이익잉여금처분계산서(1990~1999, 표본조사)
	현금흐름표	현금흐름표(2012~)
		현금흐름표(2008~2011)
		현금흐름표(1995~2007)
	현금흐름관련 기타주요재무항목	현금흐름관련 기타주요재무항목(2012~)
		현금흐름관련 기타주요재무항목(2008~2011)
		현금흐름관련 기타주요재무항목(1995~2007)
	현금흐름비율	현금흐름비율(2012~)
		현금흐름비율(2008~2011)
		현금흐름비율(1995~2007)
	기업경영분석(분기)	성장성에 관한 지표
		손익의 관계비율

No	통계명칭	내 용
	기업경영분석(분기)	자산자본의 관계비율
		자산의 회전을
	기업경영분석 분위수	1분위
		2분위(중위수)
		3분위
13	산업연관표	
14	대출행태서베이	
	대출태도	대출태도
	신용위험	신용위험
	대출수요	대출수요
15	산업 및 고용	
	산업일반	산업별 사업체수 및 종사자수
		전산업 생산지수(농림어업제외)
		설비투자지수
		경기종합지수
	농림어업	농림생산지수
		어획량
		농어가인구 및 가구수
		경지면적
		농어가 자산 및 부채
	광업 및 제조업	산업별 생산·출하·재고지수
		재별 제조업생산지수
		제조업생산능력 및 가동률지수
		제조업 평균가동률
		설비용 기계류 생산지수
		기계수주액
		품목별 광공업 생산·출하·재고·내수·수출량
		제조업 재고율지수
	전기, 가스, 수도사업, 건설업	부문별 전력사용량
		발전전력량
		에너지소비량
		석유제품수급
		국내건설수주액
		건축허가현황
		건축착공현황
		건설기성액
		주택건설인허가실적
		주택보유수 및 주택보급률
		미분양주택현황
	서비스업	산업별 서비스업생산지수
		소매업태별 판매액지수

No	통계명칭	내 용	
	서비스업	재별 및 상품군별 판매액지수	
		도소매업 재고액지수	
		소비재 수입액	
		여객수송실적	
		자동차등록대수	
		정보통신서비스 이용실적	
	고용, 임금, 가계, 사회	경제활동인구	
		추계인구	
		고용보험가입현황	
		농가가계수지	
		실업급여수급실적	
		산업별 근로일수 및 근로시간, 임금총액(5인 이상)	
		직종별평균임금	
		노동생산성지수	
		단위노동비용지수	
시간당 명목임금지수			
가구당 월평균 가계수지(전국, 2인 이상)			
지니계수			
16		거시경제분석 지표	
17		해외/북한	
	주요국제통계	본원통화	
		협의통화	
		광의통화	
		중앙은행할인율	
		주요국제금리	
		주가지수	
		생산자물가지수	
		소비자물가지수	
		국제상품가격	
		경상수지	
		외환보유액	
		환율	
		산업생산지수(계절변동조정)	
		실업률(계절변동조정)	
		경제성장률	
		국내총생산	
		1인당 GNI	
		국내총투자율	
		총저축률	
		국내총생산(PPP적용)	
PPP환율			

No	통계명칭	내 용
	남북한 주요지표 비교	남북한 주요지표비교(총량)
		남북한 주요지표비교(남한/북한 비율)
	북한GDP 관련통계	북한GDP 관련통계

2017년 4/4분기 통계공표일정

* 통신/방송/인터넷 매체에서 취급 가능한 일자 및 시각

공표일	시각	대상통계
2017.10.17(화)	12:00	2017년 9월 수출입물가지수
2017.10.18(수)	12:00	2017년 8월중 통화 및 유동성
2017.10.18(수)	12:00	2017년 8월중 예금취급기관 가계대출
2017.10.20(금)	12:00	2017년 9월 생산자물가지수
2017.10.23(월)	12:00	2017년 9월 무역지수 및 교역조건
2017.10.24(화)	12:00	2016년 기업경영분석
2017.10.26(목)	08:00	2017년 3/4분기 실질 국내총생산(속보)
2017.10.27(금)	06:00	2017년 10월 소비자동향조사
2017.10.27(금)	12:00	2017년 9월중 금융기관 가중평균금리
2017.10.31(화)	06:00	2017년 10월 기업경기실사지수(BSI) 및 경제심리지수(ESI)
2017.11.03(금)	08:00	2017년 9월 국제수지(잠정)
2017.11.08(수)	12:00	2017년 9월중 통화 및 유동성
2017.11.13(월)	12:00	2017년 10월 수출입물가지수
2017.11.20(월)	12:00	2017년 10월 생산자물가지수
2017.11.22(수)	12:00	2017년 3/4분기중 가계신용(잠정)
2017.11.23(목)	12:00	2017년 3/4분기중 예금취급기관 산업별 대출금
2017.11.23(목)	12:00	2017년 9월말 국제투자대조표(잠정)
2017.11.24(금)	06:00	2017년 11월 소비자동향조사
2017.11.24(금)	12:00	2017년 10월 무역지수 및 교역조건
2017.11.27(월)	12:00	2017년 10월중 금융기관 가중평균금리
2017.11.28(화)	06:00	2017년 11월 기업경기실사지수(BSI) 및 경제심리지수(ESI)
2017.12.01(금)	08:00	2017년 3/4분기 국민소득(잠정)
2017.12.05(화)	08:00	2017년 10월 국제수지(잠정)
2017.12.12(화)	12:00	2017년 11월 수출입물가지수
2017.12.13(수)	12:00	2017년 10월중 통화 및 유동성
2017.12.13(수)	12:00	2017년 10월중 예금취급기관 가계대출
2017.12.15(금)	12:00	2017년 3/4분기 기업경영분석(속보)
2017.12.19(화)	12:00	2017년 11월 생산자물가지수
2017.12.26(화)	12:00	2017년 11월 무역지수 및 교역조건
2017.12.27(수)	06:00	2017년 12월 소비자동향조사
2017.12.27(수)	12:00	2017년 3/4분기중 자금순환(잠정)
2017.12.28(목)	06:00	2017년 12월 기업경기실사지수(BSI) 및 경제심리지수(ESI)
2017.12.28(목)	12:00	2017년 11월중 금융기관 가중평균금리

『국민계정리뷰』 투고 안내

한국은행 경제통계국은 국민계정 통계에 대한 이용자의 편의를 제고하고 이해를 돕기 위하여 국민계정과 관련된 동향과 분석자료, 새로운 통계 및 분석기법 등에 대한 연구자료를 종합·수록하여 『국민계정리뷰』를 발간하고 있습니다. 동 책자에 게재되는 원고는 한국은행 직원은 물론 모든 사람에게 개방되어 있으며, 원고 게재를 희망 하는 분은 다음의 요령에 의거 원고를 작성·송부하여 주시기 바랍니다.

〈논문주제〉

- 국민소득, 산업연관, 국제수지, 자금순환, 계절변동조정, 초단기예측, 지수이론, 표본 설계 등 국민계정의 편제 및 분석에 관한 연구

〈원고 투고요령 및 게재〉

- 본지의 발간기준일(3월, 6월, 9월, 12월 말일)을 참고하여 발간 2개월 전까지 원고 및 파일을 제출
- 원고 제출시 저자의 이력서 및 영문성명, 소속기관 및 그 주소와 전화번호, 그리고 자택주소와 전화번호를 별지에 기재하여 송부
- 제출된 원고는 행내외 전문가 2인 이상의 심사를 거친 후 『국민계정리뷰』 편집위원회가 수록여부를 최종 결정하며 수록된 원고에 대하여는 소정의 원고료를 지급

〈원고작성 요령〉

- 원고는 A4용지에 한글 워드프로세서로 작성. 글씨 크기는 본문 11, 각주 9의 크기로 하여 행간은 160으로 설정하고 분량은 A4용지 30매 내외로 작성
- 원고의 처음은 제목, 저자명, 요약, 목차 등을 포함하고 요약은 첫 장을 넘지 않도록 작성
- 본문의 장·절 구분은 I, 1, 가, (1), (가)의 순으로 하며, 원고 작성은 한글을 원칙으로 하되 꼭 필요한 경우에 한하여 한자 사용
- <표> <그림>의 일련번호는 장절 구분하지 않으며 글씨체는 돋움(표 제목 10, 표 내용 10, 표 주석 9)으로 설정

- 참고문헌은 동양문헌의 경우 가나다순, 서양문헌의 경우 알파벳순으로 배열하되 기재순서는 저자명, 논문명, 서명, 호, 발행처, 연도, 인용쪽의 순으로 기재하며 이중 논문명은 “ ”로, 서명은 동양문헌의 경우 『 』, 서양문헌의 경우 이탤릭체로 표기

〈문의처〉

(04531) 서울특별시 중구 남대문로 39 (남대문로3가 110)

한국은행 경제통계국 국민계정부 국민소득총괄팀 『국민계정리뷰』 담당자 앞

전화 : (02)759-5265, 4413 팩스 : (02)759-4387 E-mail : bokstat3@bok.or.kr

『국민계정리뷰』 수록 원고 목록 안내

『국민계정리뷰』 각 호에 수록된 원고의 목록은 한국은행 홈페이지(www.bok.or.kr)의 ① 조사·연구-② 종류별자료-③ 계간-④ 국민계정리뷰에 게시되어 있으며 매분기마다 갱신됩니다.

The screenshot shows the '계간' (Quarterly Bulletin) section of the Bank of Korea's website. The page title is '계간' and the subtitle is '경제분석 | 국민계정리뷰 | Quarterly Bulletin'. A search bar is present with the text '경제분석 | 국민계정리뷰 | Quarterly Bulletin'. Below the search bar, there is a table listing recent publications.

전체	번호	제목	첨부	날짜	조회
<input type="checkbox"/>	335	[2011년 제4호] 한국은행 경제통계 안내		2012.01.19	40
<input type="checkbox"/>	334	[2011년 제4호] 경제주체별 통화지표 편제결과		2012.01.19	39
<input type="checkbox"/>	333	[2011년 제4호] 2011년 국제 경제통계 세미나 개최 결과		2012.01.19	9
<input type="checkbox"/>	332	[2011년 제4호] 우리나라의 최연구조 및 노동역과후과		2012.01.19	68

국 민 계 정 리 뷰

QUARTERLY NATIONAL ACCOUNTS REVIEW

발행일 | 2017. 9.

발행인 | 이 주 열

편집인 | 정 규 일

발행처 | 한국은행

서울특별시 중구 남대문로 39 (남대문로3가 110)

인쇄처 | (주)동화인쇄공사

- ▣ 본 책자의 내용을 인용할 때에는 반드시 “한국은행 『국민계정리뷰』
○년 ○호 ○페이지에서 인용” 하였음을 표시하여야 합니다.
- ▣ 본 책자에 대한 질의 또는 제안은 한국은행 국민소득총괄팀으로 연락
주시기 바랍니다. (☎ 02-759-5265, 4413)
- ▣ 국민계정리뷰는 한국은행 화폐박물관 서적 / 기념품코너(759-4805,
www.seowonbok.co.kr) 또는 정부간행물센터(734-6818, www.gpcbooks.
co.kr) 등에서 매권당 5,000원으로 판매하고 있습니다.

