

제3장 집세지수 작성방안 개선방법



정미옥 · 이익규

제1절 서론

소비자물가지수는 도시가구가 일정한 소비생활을 영위할 때 소요되는 비용이 어떻게 변화되는가에 대한 객관적 측정지표이다. 통계청에서는 소비자물가지수를 작성하기 위해 도시가계가 구매하는 489개¹⁾의 대표품목을 결정하고, 매월 상품가격과 서비스요금을 조사하여 그 변동을 측정한다. 이때 소비자물가지수 품목은 도시가구의 소비구조를 반영할 수 있도록 기준년도의 도시가계 품목별 소비지출액 비중을 기초로 선정한다. 이렇게 선정된 품목들 중에서 월세와 전세의 두 품목을 포괄하는 집세지수가 <표 3-1>에서 보는 바와 같이 9.78%로 가장 큰 비중을 차지하고 있어 소비자물가지수 내 상대적 중요도가 가장 높은 품목이다.

집세지수를 작성하기 위해 매월 표본가구에 조사원이 방문조사 하는 집세조사가 수행된다. 집세조사의 조사항목은 집세와 주택유형, 방수가 있으며, 아파트에 대해서는 전용면적과 건축년도의 항목이 2009년부터 추가되었다. 표본가구는 통계청의 가구대상 표본조사에 활용될 수 있도록 경제활동인구조사가 중심이 되어 구축된 다목적표본에서 임차가구만을 대상으로 선정하고, 표본 수가 부족한 일부 도시에 대해서는 보완하여 줌으로써 전체 표본 수는 약 10,000여 가구를 유지하고 있다. 또한 소비자물가지수 대상 도시는 38개로 이들 도시에 대해서는 전·월세를 구분하지 않는 집세지수를 공표하고, 16개 시도에 대해서는 전세와 월세지수를 각각 공표한다. 2011년에는 소비자물가지수 개편 작업이 진행될 예정으로 대표품목의 선정, 지수 작성방법과 품목과 도시별 가중치 작성 등이 개편 시 주요하게 검토되어야 할 과제이다.

다목적표본을 이용하고 있는 집세조사와 같이 설계목적이 다른 표본을 차용하고 있거나, 표본설계 시 주요 관심변수를 고려할 수 없는 경우 선정된 표본에 모집단의 세부

1) 2005년 기준 소비자물가지수 개편 결과

적인 특성 반영이 미흡할 수밖에 없다. 집세조사에서는 10,000여 표본가구에서 수집한 집세정보로 38개 도시별 지수를 작성하고 있다. 또한 지수작성에 기초 자료로 쓰이는 도시별 평균 전·월세 추정 시 가중치를 적용하지 않는 단순 산술평균의 방식을 취하고 있다. 이와 같은 집세조사의 다목적표본과 추정방법에 대한 논의가 통계개발원의 연구진에 의해 진행되었다. 전백근(2006)은 표본현황과 문제점을 살펴보았는데, 집세 표본조사구가 경제활동인구조사, 가계조사 등 다목적으로 설계된 표본으로 정확한 집세 가격변동 파악에 어려움이 있으므로 집세조사를 위한 독립표본이 필요하다고 하였다. 그러나 예산과 인력, 그리고 시스템을 마련하는 것이 현실적으로 어렵기 때문에 이러한 한계를 통계적 추정방법으로 극복하고자 하는 개선안을 제시하였다. 전·월세 금액을 추정함에 있어 연면적, 사용방수, 가구원수 등 다양한 보조변수를 이용하는 것이 집세지수의 정확성과 신뢰성을 높이는 방안이라고 제안하였다. 또한 집세조사 규격이 주택유형별로 세분화되어 있지 않아 유형별 집세의 동향을 파악하기 어렵다고 지적하였다.

다음으로 정미옥(2009)은 집세표본이 집세변동 측정을 목적으로 설계된 표본이 아니라 한계점을 언급하고, 주요 특성들에 대해 집세표본이 모집단의 구조를 얼마만큼 반영하고 있는지를 분석하였다. 이 연구에서는 이러한 구조의 차이 등을 근거로 하여 집세조사의 전용표본 도입이 필요하다고 제안하였으며, 사후층화 가중값을 작성하는 등의 연구가 필요하다고 하였다. 이 연구를 기초로 정미옥 외(2009)는 집세조사의 독립표본설계 방법에 대해서 연구를 진행하였으나, 현실적으로 해결하기 어려운 예산·인력 등의 난관에 부딪쳐 추진되지는 못하였으며, 구조의 차이가 심한 도시에 대해서는 일부 표본을 추가로 보완해주는 방식으로 전환되었다. 마지막으로 소비자물가지수의 지수산식은 박현정(2010)에 의해 연구되었으나, 전세와 월세의 도시별 기본지수 작성은 상품 및 그 외 서비스 가격지수 작성방법과 다른 체계를 갖고 있어 해당 연구의 범위에서 제외되었다. 이처럼 선행연구들에서 거론되어 왔던 독립표본의 전환은 실현되기 어려운 측면이 있었다. 그러나 표본 선정방법 이외에도 통계적인 추정방법의 도입이나 지수 작성체계에 대한 타당성 검토 등과 같이 집세지수 작성과 관련하여 개선되어야 할 부분은 여전히 남아 있다.

집세는 임차가구가 거주하는 주택의 규격이 매우 다양하고, 대상처가 아닌 표본가구에서 가격자료를 수집하는 등 그 특성과 조사방법이 소비자물가 내 다른 품목들과는 상이하다. 본 연구에서는 그동안 논의되어 왔던 집세지수 작성방법을 전반적으로 검토하고, 현실을 좀 더 반영할 수 있는 개선 방안을 찾아보고자 한다. 먼저, 국제적으로 비교 가능한 통계작성을 위해 국제노동기구(ILO)에서 권고하고 있는 소비자물가지수 작성 매뉴얼과 해외사례를 살펴보고, 지수 작성이론과 실무적 고려사항들을 검토한다. 그리고 다목적표



본에 의해 나타나는 모집단과의 구조차이를 보정해주기 위해 사후층화(post-stratification)방법에 의해 사후가중치를 작성해보고자 한다. 사후층화란 표본 선택 후의 표본조사 결과에 알려져 있는 모집단 정보를 이용하여 가중치를 부여하는 일종의 가중치 조정기법이다. 임차가구의 모집단 자료로는 2005년 인구주택총조사 결과를 사용하고, 임차가구가 거주하는 주택의 유형과 특성들에 의하여 사후층을 구성한 후 전·월세 평균을 추정하여 사후가중치 적용 전과 후를 비교해 본다. 다음으로 현행 지수산식 검토를 위해 일반적으로 많이 사용되는 세 가지 주요 방법에 대해서 살펴본다. 대표적인 하위 품목별 지수 작성 방법으로는 가격비의 산술평균(Carli)방법, 산술평균의 가격비(Dutot)방법, 가격비의 기하평균(Jevons)방법이 있다. 현행 집세지수는 듀토티방법을 이용하여 작성하고 있는데, 이 방법은 집세와 같이 가격수준이 다양한 경우 가격수준이 더 높은 주택에 더 큰 내재적 가중치를 줌으로써 지수에 편이가 발생할 수 있다고 알려져 있다. 이러한 지수 작성방법의 장·단점 및 특징을 알아보고, 과거 자료를 이용하여 집세지수를 시산·비교해 봄으로써 집세지수 작성방법의 개선방향을 제안하고자 한다. 또한 주택유형에 의한 지수와 같은 현행보다는 좀 더 세분화된 지수를 추가적으로 작성할 수 있는지에 대한 가능성을 검토하고 모의지수를 작성해본다.

본 연구는 사후가중치의 적용을 통해 집세표본에 모집단 대표정도를 부여함으로써 다목적표본에 의해 발생할 수 있는 편향(bias)이 조정되고 정도(precision)가 향상되기를 기대한다. 또한, 지수 작성방법에 대한 타당성을 확보하면서 이용자에게 좀 더 유용한 정보를 제공할 수 있도록 현행 집세지수를 세분화시켜 작성하는 방안에 대해 다시 한번 고민해보는 계기가 되었으면 한다.

제2절 소비자물가 집세지수 작성방법

1. 집세지수 작성 현황

소비자물가지수의 작성은 도시가구가 일정한 소비생활을 영위하기 위해서 구입하는 상품과 서비스의 가격변동을 종합적으로 측정하기 위한 것으로 생활수준의 향상이나 가구원수의 변동, 자녀의 성장 등에 따른 지출규모의 변동에 의한 영향은 제외된다. 소비자물가지수는 도시가구의 소비생활에 영향을 미치는 물가수준의 변동을 파악하는데 쓰이고, 임금, 국민소득 등의 실질수준 측정을 위한 디플레이터(deflator)로 활용되며, 물가정책, 연금지급 등 국민생활과 관련된 정책자료로도 이용된다.

〈표 3-1〉 소비자물가지수 기본분류 내역

분류내역	가중치	품목수	분류내역	가중치	품목수
총 지 수	1000.0	489			
(식료품·비주류음료)	140.4	131	(보건 의료)	51.6	28
곡 물	28.0	14	의 약 품	16.8	16
육 류	21.0	7	의 료 용 품	2.7	5
어 개	16.6	19	의료 서비스	30.6	4
낙 농 품	11.2	5	기타 의료서비스	1.5	3
유 지	1.1	2	(교 통)	109.0	31
과 실	17.3	15	차량 구입비	21.7	6
채 소·해 초	19.2	32	차 량 연 료	47.0	3
과 자·당 류	8.1	9	차량부품·수리비	6.8	5
조미료·기타식품	10.9	18	기타차량서비스	3.6	5
차	1.7	4	철도 이용료	5.3	2
음 료	5.3	6	육상 이용료	20.6	4
(주 류 · 담 배)	14.6	9	항공 이용료	2.4	4
주 류	3.8	7	기타차량이용서비스	1.6	2
담 배	10.8	2	(통 신)	60.2	13
(의 복 · 신 발)	58.4	42	우편 서비스	0.1	1
남 자 의 류	23.4	15	전 화 기	3.6	2
여 자 의 류	19.5	14	전화·정보이용료	56.5	10
아동복·유아복	4.3	2	(교 양 · 오 락)	56.3	64
기 타 의 류	1.9	4	영상·음향기기	4.0	5
의 류 서비스	2.1	2	정보처리기기·소모품	5.3	7
신 발	7.2	5	악기·오락용품	5.8	9
(주거 및 수도·광열)	170.4	19	오락·운동서비스	9.7	15
집 세	97.5	2	문화 서비스	12.6	11
주택설비수리	6.8	5	도 서	7.7	7
상하 수도료	7.8	2	신 문·잡 지	1.4	2
기 타 주 거	14.6	3	문 방 구	1.8	6
전 기 료	19.0	1	단 체 여 행	8.0	2
가 스 료	17.6	3	(교 육)	110.9	23
연 료	7.1	3	유치원·초등교육	36.3	6
(가구집기·가사용품)	41.7	52	중 등 교 육	40.0	6
가 구	7.9	8	고 등 교 육	26.6	6
침구·직물제품	2.4	3	기 타 교 육	8.0	5
가정용 기구	12.7	15	(의 식 · 숙 박)	132.7	43
가정용기구수리서비스	1.2	3	의 식	130.6	39
주 방 용 품	2.5	10	숙 박	2.1	4
가사 소모품	3.7	11	(기 타 잡 비)	53.8	34
가사 서비스	11.3	2	이미용서비스	17.3	4
			이 미 용 품	17.5	20
			개 인 용 품	8.4	5
			기타 서비스	10.6	5



2005년 기준 조사대상 품목은 489개로 도시가구 월평균 소비지출 총액에 대한 항목별 지출액이 1/10,000(월 평균 185원) 이상인 품목이 선정되었다. 가중치는 소비지출액 구성비로서 천분비로 나타내며 <표 3-1>에서 보는 바와 같다. 이러한 상품과 서비스의 가격은 약 19,000개 소매점포에서 조사하고 집세는 대상처가 아닌 10,000여 임차가구에 서 조사하며, 소비자물가조사 지역은 전국 38개 주요도시에서 이루어지고 있다. 서울을 포함한 7대 특·광역시와 9개의 도(道)에서 인구와 지역분포를 고려하여 2~7개 시(市)를 선정하여 전국적으로 대상지역 표본이 골고루 반영되도록 하고 있다. 조사대상 도시를 살펴보면, 서울, 부산, 대구, 인천, 광주, 대전, 울산(이상 특별시 및 광역시), 수원, 성남, 의정부, 안양, 부천, 고양, 안산(이상 경기도), 춘천, 원주, 강릉(이상 강원도), 청주, 충주(이상 충청북도), 천안, 보령, 서산(이상 충청남도), 전주, 군산, 남원(이상 전라북도), 목포, 여수, 순천(이상 전라남도), 포항, 경주, 안동, 구미(이상 경상북도), 창원, 김해, 마산, 전주(이상 경상남도), 제주, 서귀포(이상 제주도)이다.

지수작성 주기는 매월이며, 지수계산방식은 라스파이레스산식(기준시점 고정 가중산술평균법)을 따르고 품목지수는 아래와 같이 기준 시의 가격을 100으로 했을 때 비교 시의 가격수준으로 구한다. 또한 총 지수 등 상위분류지수는 품목지수를 품목별 가중치로 가중 평균하여 계산한다.

$$\text{품목별지수} = \frac{P_t}{P_0} \times 100, \quad P_t : \text{기준 시 가격}, \quad P_0 : \text{비교 시 가격},$$

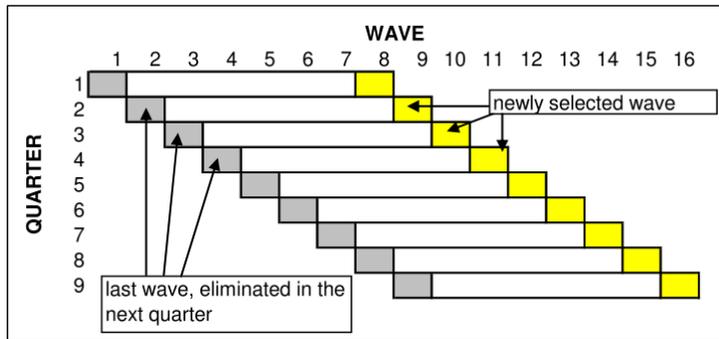
$$\text{전도시 품목별 지수} = \frac{\sum \frac{P_t}{P_0} W_e}{\sum W_e} \times 100,$$

$$\text{총지수} = \frac{\sum \left(\frac{\sum \frac{P_t}{P_0} W_e}{\sum W_e} \right) W_a}{\sum W_a} \times 100,$$

여기서 P_0 : 도시별 품목별 기준 시 가격
 P_t : 도시별 품목별 비교 시 가격
 W_e : 품목별 도시 간 가중치
 W_a : 전도시의 품목별 가중치

2. 해외사례

스위스²⁾는 소비자물가지수 내 집세지수의 가중치가 약 20%로 다른 품목지수들에 비해 상대적 중요도가 매우 높다. 집세를 조사하기 위해 단순임의추출(Simple Random Sampling)에 의해 5,000여 아파트 표본가구를 선정하여 자료를 수집하고 있다. 분기별 1/8 연동표본제(rotation sample system)를 따르고 있으며, 조사는 매분기(2월, 5월, 8월, 11월)마다 수행된다. 집세조사에서는 집세뿐만 아니라 방의 개수, 면적, 건축년도, 개보수 사항 등의 항목을 조사하고, 필요 시 등록자료(Register of Building and Dwellings)에 의해 내용을 확인한다.



[그림 3-1] 스위스 집세조사 1/8 연동표본제

현재 사용 중인 표본추출틀은 주로 가구 조사들을 위한 것으로 주거 목적으로 아파트를 임차하는 가구가 기준인 집세지수의 모집단과는 정확하게 일치하지 않는다. 또한 추출틀은 응급전화 데이터베이스로부터 가져 온 유선 전화번호 명부로 구성되어 있다. 이 추출틀의 가구 커버리지는 90~95% 정도라고 추정하고 있다. 이러한 추출틀로부터 표본을 베르누이 방법에 의해 단순임의추출하고, 자가, 무상, 상업 혹은 휴일용 아파트 여부를 파악하여 적격여부를 사전에 심사한다. 전화번호부 추출틀의 문제점을 보완하기 위해 사후층화를 수행하는 데, [그림 3-2]와 같이 방수(6그룹)와 아파트연령(4그룹)별로 총 24개 층으로 사후층을 구성한다. 또한 각 층의 가중치는 임차구조조사(rent structure survey)로부터 임차인의 지출액(renter's expenditure)에 기초하여 계산한다. 스위스의 집세 추정은 각 셀 내에서는 기하평균 방법에 의해 집세평균을 구하며, 이때 설계가중치(design weight)에 기초하여 작성된 가중치를 적용하여 계산한다. 그리고 이 평균집세로 전월과의 비를 구한 다음 <표 3-2>와 같이 각 셀의 상대적인 중요도인 사후층화가중치에 의해 가중 평균함으로써 지수를 작성한다.

2) Vermeulen 외(2009) 인용

		Age of apartment				
		0-5 years	5-10 years	10-20 years	20+ years	Total
Number of rooms	1	0.0442	0.1332	0.3159	3.7241	4.2174
	2	0.3363	0.8432	1.6650	10.8420	13.6865
	3	1.0825	2.0882	3.5370	25.8936	32.6013
	4	1.9577	2.7432	4.8034	23.1534	32.6577
	5	1.1870	1.2383	2.0177	8.3954	12.8384
	6	0.2569	0.2607	0.5829	2.8982	3.9987
	Total	4.8646	7.3068	12.9219	74.9067	100.0000

[그림 3-2] 스위스 집세조사 사후층 및 각 셀 내 가중치(%)

다음으로 캐나다³⁾에서는 소비자물가지수의 바꾸니 품목들 중 집세지수가 5.5%를 대표하고 있다. 캐나다도 우리나라와 마찬가지로 집세조사만을 위한 독립표본이 아닌 노동력조사(Labor Force Survey)의 표본을 이용하여 자료를 수집하고 있다. 노동력조사의 매월 표본의 크기가 약 54,000개 거처이고 이 중 임차가구는 매월 평균적으로 약 14,000개 거처이며, 16개 도시에 대해 집세지수가 공표되고 있다. 노동력조사는 6개월 연동표본제를 채택하고 있고, 표본으로 선정된 가구는 6개월 동안 연속적으로 조사에 참여한다. 우리나라도 경제활동인구조사의 표본가구 중 임차가구를 집세조사의 표본가구로 선정하고, 경제활동인구조사는 36개월 연동표본제를 따르고 있는 등 표본의 형태가 캐나다와 유사하다.

또한 캐나다의 경우 집세지수 작성을 위한 평균가격을 추정할 때 표집가중치(sampling weight)를 사용하지 않는다. 집세평균과 전월 대비 집세평균의 비(month-to month price index ratios)는 전국(nation), 주(province), 도시(city) 수준에서 계산한다. 각각의 주 내에서는 두 개의 층으로 나뉘는데, h_1 층은 인구 30,000보다 적은 도시이고 h_2 층은 인구 30,000보다 큰 도시로 구분한다. 이렇게 구분된 각각의 층 내에서는 모든 도시들에 대해 가중치를 적용하지 않은 단순 산술평균으로 평균 집세를 구한다. 그리고 주의 수준을 구하기 위해서는 <표 3-2>의 수식처럼 층별로 구한 추정치를 CPI 가중치에 의해 가중 평균한다. 여기서 CPI 가중치는 주에서 차지하는 h_1 층과 h_2 층의 상대적 중요도를 의미하며, 1991년 센서스 결과로부터 구한다. 이와 같이 추정된 금월의 집세평균을 전월추정치와 비교하는 산술평균의 비의 방식으로 집세지수를 작성한다.

3) Claveau 외(2009) 인용

<표 3-2> 각국의 집세지수 산식

	방법	지수산식
스위스	기하평균의 비 방식(Jevons) 층 내 기하평균(\tilde{x}_j^t) 전월과 금월 간의 상대변동비(L_j^t) *사후가중치로 상대변동비 가중평균	$L^T = \frac{\sum_j w_j L_j^t}{\sum_j w_j}, \quad L_j^t = \frac{\tilde{x}_j^t}{x_j^{t-1}} \cdot 100$ $\tilde{x}_j^t = \left(\prod_{i=1}^{n_j^t} (x_{ij}^t)^{p_i^t} \right)^{1/\sum_i p_i^t} = \exp \left[\frac{\sum_{i=1}^{n_j^t} (p_i^t \log x_{ij}^t)}{\sum_{i=1}^{n_j^t} p_i^t} \right]$
캐나다	산술평균의 비 방식(Dutot) 층 내 산술평균, 층(도시) 간 가중평균 *도시 간 상대적 중요도인 가중치는 센서스로부터 계산	rent index $= \frac{\sum_{i \in h1j \in sr} \sum \frac{r_{h1ij}^t}{n_{h1i}} * w_{h1} + \sum_{i \in h2j \in sr} \sum \frac{r_{h2ij}^t}{n_{h2i}} * w_{h2}}{\sum_{i \in h1j \in sr} \sum \frac{r_{h1ij}^{t-1}}{n_{h1i}} * w_{h1} + \sum_{i \in h2j \in sr} \sum \frac{r_{h2ij}^{t-1}}{n_{h2i}} * w_{h2}}$
미국	산술평균의 비 방식(Dutot)의 변형 집세지수($IX_{A,t}^{Rent}$)는 전월 지수와 금월 상대변동비($REL_{A,t-6,t}^{Rent}$) 6승근의 곱 금월 집세총합과 6개월 전 집세총합의 비 *집세총합은 표집가중치에 의한 가중 합	$IX_{A,t}^{Rent} = IX_{A,t-1}^{Rent} \times \sqrt[6]{\frac{REL_{A,t-6,t}^{Rent}}{\sum_{i \in A} RW_s * ER_{i,t}}}$ $REL_{A,t-6,t}^{rent} = \frac{\sum_{i \in A} RW_s * ER_{i,t-6}}{\sum_{i \in A} RW_s * ER_{i,t}}$
일본	산술평균의 비 방식(Dutot) 기준년도(2005년도 1월~12월) 집세와 금월 집세의 비 *금월 집세는 3개월 총 집세 수준	house rent per $3.3m^2$ $= \frac{\sum_{n=0}^2 (t-n) \text{월 조사된 집세총합}}{\sum_{n=0}^2 (t-n) \text{월 조사된 총면적}} \times 3.3$

Claveau 외(2009)의 최근 연구에 의하면, 가격지수에 대해서는 표집가중치를 사용하지 않는 것이 일반적이고, 캐나다에서도 현재 집세지수 작성 시 표집가중치를 사용하지 않는다. 그러나 노동력조사의 표집가중치를 집세와 관련이 높은 변수에 대해 모집단에 보정시킨 가중치(calibrated weight)를 쓰게 된다면 다음과 같은 장점이 있다고 하였다. 첫째, 모집단의 외부 추정치와의 일치성을 확보하고, 둘째, 과소포함(undercoverage)에 대한 설명을 할 수 있으며, 셋째, 추정치의 효율성을 증대할 수 있다. 이러한 복합보정추정량



(composite calibration estimator)은 특히 변동의 추정치에 대한 효율성을 높인다. 이 연구에서는 노동력조사의 표집가중치나 집세지수와 관련된 변수로 보정한 가중치의 사용이 집세지수의 품질향상에 보다 유용할 것이라는 개선안을 제시하였다.

미국과 일본의 집세조사 방법과 표본설계에 대한 내용은 정미옥(2009)을 참고하길 바라고, 여기서는 지수산식에 대해서만 언급하고자 한다. <표 3-2>를 보면, 미국⁴⁾의 지수는 연쇄적인 방법으로 전월 지수와 금월 상대변동비의 6승근을 곱하여 구한다. 여기서 상대변동비는 금월의 집세총합(total)과 6개월 전의 집세총합의 비(ratio)로 구하며, 집세총합은 표집가중치에 의해 세그먼트별로 가중합을 한다. 미국은 표본가구를 6개의 그룹으로 구분하여 6개월에 한 번 표본가구를 조사하는 방식을 채택하고 있기 때문에 조사시점의 변동을 6개월로 배분하는 방법의 지수산식을 사용하고 있다. 일본⁵⁾에서는 기준년도의 집세평균과 금월의 집세평균의 비의 방식으로 지수를 작성한다. 2005년 기준 소비자물가지수 개편결과에 의하면 기준년도의 집세평균은 2005년 1월부터 12월까지 조사된 결과를 모두 이용하여 평균수준을 결정한다. 그리고 금월의 집세평균을 구할 때는 금월을 포함하여 지난 2개월 간 조사된 결과를 모두 이용하여 단순 산술평균을 함으로써 3개월 동안의 평균수준을 구한다. 이때 집세평균을 면적(3.3m²)당 집세로 구함으로써 추정치에 주택크기에 대해서는 규격화를 하고 있는 셈이다. 또한 일본은 가중치를 적용하지 않는 산술평균의 비의 방식을 따르고 있다.

3. 집세지수 작성방법 실무적 지침⁶⁾

임차가구들은 자가와는 달리 다른 사람이 소유하고 있는 주거공간을 점유할 권리를 산다. 이로써 시장거래가 발생하게 되므로 자가주거비에 비하여 시장에서 집세를 관측하기는 상대적으로 쉽다. 또한 소비자물가지조사 바구니에서 임차가구에 대한 가중치를 구하는 것은 복잡하지 않은데, 일반적으로 가계조사(Household Budget Surveys, HBS)와 같은 표본조사에서의 집세지출액 비중을 이용하여 가중치를 산출하게 된다. 가계조사에서는 가구로부터 그들의 거주지에 대한 집세 정보를 수집하며, 집세의 연간지출액을 추정한다. 이때 그 가구가 주된 주거지 이외에 직장근처나 휴일 여가생활을 위해 다른 주택을 추가적으로 임차하고 있는지에 대해서도 질문을 해야 주거를 위해 지불하는 정확한 지출액을 파악할 수 있다. 집세는 세금은 포함하고 보조금은 제외한 가구에서 실제로 지불한 금액으로 정의한다. 집세조사의 응답자는 주거의 점유자가 되거나, 집주인, 또는 집주인의 대리인이 될 수도 있다. 그러나 표본가구에 대한 지식이 있는 응답자로부터 수집한 집세정보라고

4) Bureau of Labor Statistics(2001) 참고

5) Japan Statistics Bureau(2006) 참고

6) United Nations(2009) 인용

할지라도 만약 세금이나 보조금이 있다면 그 가구가 지불하는 금액이 주인이 받는 금액과는 같지 않게 될 것이기 때문에 주의가 필요하다.

소비자물가조사의 고정바구니 원칙을 존중하기 위해서 집세조사는 일정 기간 동안 동일한 거주단위로부터 가격을 수집하도록 노력해야 한다. 이로 인하여 발생할 수 있는 여러 어려움이 있음에도 불구하고, 임차가구의 종적(longitudinal)조사 방법이 집세를 적절히 수집할 수 있는 최선의 해결책이라 할 수 있다. 그러나 차선택으로 집세를 노동력조사(우리의 경제활동인구조사)와 같이 이미 존재하는 경상 가구조사로부터 수집하는 것도 가능하다. 이러한 경우 주된 조사인 노동력조사에 참여하는 임차인에게 집세조사를 위한 부가조사표를 배부할 수 있다. 이때 중요하게 고려해야 할 점은 집세는 여러 시점에 걸쳐 제공해야 하지만 표본가구의 전출 등의 이유로 주택단위의 고정바구니 원칙을 어쩔 수 없이 따를 수 없게 된다. 따라서 사전에 정의된 주택 특성 또는 고정바구니 주택유형의 평균집세에 대한 증가가 결정될 수 있도록 집세 데이터는 집세 결정요인에 의해 증화되어야 한다.

만약 집세가 자주 변동하지 않는다면, 소비자물가지수가 매월 계산되지 않을지라도 매월 정보를 수집하지 않으면서 상대적으로 더 많은 표본을 사용하는 것이 좀 더 효율적이다. 집세조사는 지역의 거주단위를 포함하고 있는 표본들로부터 표본이 추출될 수 있는데 인구주택총조사나, 우편번호, 거리 주소록이 될 수도 있다. 만약 인구주택총조사가 평균집세 또는 지역별 주거가치 등의 정보를 제공할 수 있다면, 집세나 주거가치에 대한 확률비례법에 의해 지역을 추출하는 것이 최종표본의 대표성을 향상시킬 수 있을 것이다. 그러나 표본추출틀이 이용 가능하도록 준비되지 않았다면, 지역의 지도에 격자무늬를 그리고, 격자 내의 셀이 표본으로 선택될 수 있다. 일단 넓은 지역의 집세나 가치에 대한 이용 가능한 정보를 사용하여 넓은 지역을 1차로 선택한 후에 표본추출의 두 번째 단계에서 이러한 격자방법을 활용하면 된다. 개발도상국을 포함한 여러 국가에서 목표지역이나 셀 안의 주택을 정확하게 확인하기 위해 위성사진을 이용해왔다. 이상적으로는 선택된 셀의 모든 거처를 열거하고 랜덤하게 표본을 선정한 후 임차거주여부를 확인하여 적격인 경우 표본에 포함시키고 집세를 조사한다. 조사 대상 지역의 초기선정은 무응답과 자가 거주 가구에 대응할 수 있을 만큼 충분히 크면서도, 관리가능하고 상대적으로 동질한 특성을 지닐 수 있을 만큼의 작은 지역을 선정하도록 한다. 조사비용이나 여러 문제들로 인해 조사가 불가능한 경우 판단(judgmental) 또는 유의(purposive) 표집방법이 적용될 수도 있지만, 이러한 경우 대표성이 없는 표본이 선정될 가능성을 최소화하기 위해 격자 내 개별 셀들이 거처의 종류, 크기, 품질에 대해 상대적으로 동질해야 한다는 점을 각별히 유의해야 한다.

집세의 경우 매월 변동이 일어나는 것이 아니라 한 달 간격보다 덜 자주 일어나는 데,



이러한 경우의 집세지수 계산에 대해 살펴본다. 앞서 언급한 바와 같이, 특정한 거처에 대해 매월보다 덜 자주 가격을 매김으로써(즉, 매월 부표본에 대한 가격을 매김) 가격관측치를 퍼뜨리는 것은 자료수집 활동을 위해 더 많은 자원을 배정하지 않고도 표본을 증가시키는 데 사용될 수 있는 하나의 전략이다. 어떤 경우에는 이러한 접근법을 적용함으로써 집세변동 추정치의 통계적 품질을 심지어 향상시킬 수도 있다. 지수의 지연(lag)발생과 계절효과를 없애는 것은 상충(trade-off) 관계에 있다. 만약 대부분의 집세가 매년 같은 월, 예를 들어 1월에 전통적으로 변동한다면, 이러한 방법의 사용은 1월의 집세변동을 이후 12개월에 분산시켜 주게 되어 지수에는 지연이 발생하게 되지만 계절효과는 없앨 수 있다.

매월 전체 표본에 대한 집세변동을 얻을 수 없는 경우, t 월과 $t-1$ 월 사이의 집세변동률($\Delta_{Rent}^{t-1 \rightarrow t}$)은 해당 월의 부표본으로부터 구해질 수 있다. 장기간에 걸친 집세변동의 전체 추정치를 얻기 위해 부표본의 서로 다른 월별 값들이 연속적으로 묶이게 된다. 하나의 부표본의 월별 집세지수를 계산하기 위해, 동일한 집세거처가 m 개월 간의 간격으로 값이 매겨진다. t 시점에서 월별변동률은 t 시점의 표본집세들의 가중합과 $t-m$ 시점의 표본집세들의 가중합의 비에 대한 m 제곱근으로 구해지며 수식으로 표현하면 아래와 같다.

$$\Delta_{Rent}^{t-1 \rightarrow t} = \sqrt[m]{\frac{\sum_i w_i \times r_i^t}{\sum_i w_i \times r_i^{t-m}}}$$

여기서 r_i^t 는 i 번째 표본거처의 t 시점의 집세, r_i^{t-m} 은 i 번째 표본거처의 $t-m$ 시점의 집세(수집된 거처에 대한 지난 시점의 집세), w_i 는 i 번째 표본 거처의 가중치이다.

표본거처가 표본추출틀의 임차거처의 수와 대략 같은 수를 대표한다면, 간편성을 위해 표본가중치가 무시될 수도 있다. 이는 각각의 표본에 대해 수식 상에서 가중치가 1로 주어진 경우라고 할 수 있는데, 이때엔 무응답과 이례적인 표본추출을 적당하게 조정해주는 메커니즘을 적용해야 할 것이다. 추가로 언급하면, 듀퍃지수의 결점을 피하기 위해 위의 식에 산술평균대신 기하평균이 채택될 수 있다. 여기서 결점이란 거처의 측정단위의 변화에 불변량(invariant)하지 않다는 것과 매우 비싼 거처가 큰 내재적 가중치를 갖게 된다는 것이다. 이에 대한 대안으로 값비싼 거처는 이상점으로 취급하여 계산에서 제외시킬 수 있다.

다음으로 결측치 처리에 대해서 살펴보자. 응답자와의 접촉불능 등의 사유로 셀 내에서 일시적으로 정보수집이 불가능하여 어느 한 거처가 무응답이 되고, 그 셀 내에 다른 거처들의 정보가 수집된 경우에는 정보가 수집된 거처에 결측된 거처의 가중치를 주어야 한다. 예를 들어, 한 셀 내에 3개의 거처가 배정되고 각 거처의 가중치가 1(셀 내 가중치가 동일하기 때문에)이며 두 거처의 자료는 수집되고, 한 가구가 수집되지 않은 경우에 무응

답 거처의 집세는 다른 두 거처의 값으로부터 다음과 같이 대체될 수 있다.

$$\hat{r}_{missing}^t = \hat{r}_{missing}^{t-m} \times \frac{\hat{r}_{collected1}^t + \hat{r}_{collected2}^t}{\hat{r}_{collected1}^{t-m} + \hat{r}_{collected2}^{t-m}}$$

만일 자료수집에 문제가 발생해서 셀 내의 모든 관측치가 결측일 경우, 해당 월의 계산에서 제외시킬 수 있다. 그러나 이러한 거처들의 집세는 다음 번 지수작성에 필요하므로 반드시 추정되어야 하는데, 그 거처의 집세는 다음의 수식에 의해 값을 대체해준다.

$$\hat{r}_{missing}^t = \hat{r}_{missing}^{t-m} \times (\Delta_{Rent}^{t-1 \rightarrow t})^m$$

이러한 집세 대체값은 $t+m$ 시점까지 사용되지 않는다.

소비자물가지수에 포함하기 부적합하게 된 거처(예를 들어, 화재 등으로 사람이 거주할 수 없게 된 경우)가 회복(수리)되지 않을 것으로 보인다면 표본에서 탈락시킬 수 있다. 그러나 만일 복구되어 다시 자료를 활용할 수 있을 것으로 보인다면 무응답으로 처리할 수도 있다. 또한 자가나 공실로 상태가 변경되는 경우처럼 임차의 상태가 중단되는 거처는 표본에서 탈락되지만, 가능하다면 근처에서 대체거처를 찾아 주는 것이 좋다. 그리고 대체거처가 지수에 사용될 수 있을 때까지 기존거처는 무응답으로 처리되어야 한다. 만일 하위지수에 Jevons 방법을 적용했다면, Jevons 산식과 일치성이 있는 기하평균 대체법을 사용하는 것이 좀 더 바람직하다.

집세표본은 모든 소비자물가지조사와 마찬가지로, 최신의 상태를 유지해야 한다. 이것은 만약 새로운 임차주택이 건설된 경우에는 특히 더 중요하다. 이를 위해 완전하게 새로운 표본을 한 번에 뽑아 기존의 표본과 병행하여 운영하다가 기존표본을 탈락시킬 수 있다. 그 새로운 표본은 새로운 표본추출틀을 기반으로 하여야 한다. 새 표본을 한 번에 채택하는 것은 매우 비용이 높을 수가 있기 때문에 매년 표본의 일부를 교체하는 방법이 대안이 되기도 한다. 이는 집세표본을 여러 개의 부표본들로 구성하여 자연스럽게 매년마다 하나의 부표본을 연동하는 방법이다. 예를 들면, 한해에 1월과 7월의 부표본을 교체하고, 그 다음 해에 4월과 10월을 교체하는 방식이다. 새 지역 셀들을 선택하고 그들을 부표본으로 모두 한 번에 할당하지만 자료수집은 한 번에 하나의 부표본을 수집하도록 제한한다. 기존과 새로운 부표본들에 대한 집세는 동시에 수집될 필요가 있는데 소비자물가지수는 새로운 부표본에 대한 초기집세를 수집하는 동안 마지막에 대한 기존 부표본을 사용한다.

마지막으로 금월의 지수는 금월의 집세변동에 전월의 지수를 곱하는 연쇄방법으로 아래와 같이 계산될 수 있다.

$$I_{Rent}^t = I_{Rent}^{t-1} \times \Delta_{Rent}^{t-1 \rightarrow t}$$

제3절 사후총화

1. 사후총화방법 개요

표본조사에서는 표본설계 시 구해지는 추출확률에 따른 설계가중치와 표본조사를 수행하면서 발생할 수 있는 무응답에 대한 조정가중치를 고려한 후, 사후총화가중치를 적용하여 최종가중치를 작성하는 것이 일반적이다. 여기서 사후총화란 어떤 특성에 대한 가중 표본분포를 모집단의 분포와 일치하게 조정하는 방법이다. 사후총화의 아주 간단한 예를 들어보면, 임차가구가 거주하는 주택유형의 구성비가 모집단에서 단독주택 50%, 아파트 40%, 연립다세대주택 10%이고, 아파트의 집세가 단독주택과 연립다세대주택보다 일반적으로 높다고 하자. 그런데 표본조사에서는 단독주택이 70%, 아파트가 10%, 연립다세대주택이 20%와 같은 가중 표본분포를 보이고 있다고 가정해 보자. 이러한 경우 집세가 높은 아파트가 모집단에 비해 표본비중이 낮기 때문에 이 결과를 그대로 이용하여 추정된 평균집세는 모집단의 평균집세보다 낮게 추정될 것이다. 그러나 사후총화를 통해 모집단 주택유형의 비율과 유사할 수 있도록 사후총화가중치를 작성하여 조정해 주면, 아파트의 가중치는 기존보다 높아지고, 단독주택의 가중치는 다소 낮아지게 된다. 이렇게 사후조정을 거친 추정치는 그렇지 않은 경우보다 모집단의 평균에 근사하게 될 것이다. 이와 같이 사후총화방법은 가중 표본분포가 모집단의 주요 구성 비율을 제대로 반영하지 못할 때 적합하며(김영원 외, 1998, p187), 사후총화의 올바른 적용은 표본선정편향(sample selection bias)을 조정하고, 정도(precision)를 향상시키는 데 도움이 된다.

실무에서 표본을 설계하다 보면, 표본설계 시 주요변수에 대하여 총화를 하고 싶지만 표본이 뽑히기 전까지는 해당 변수를 원하는 층에 배치할 수 없는 경우가 있다. 또는 예산과 인력 등의 현실적인 여건에 의해 다목적표본제를 채택해야 하는 상황에서 여러 조사의 공통관심 변수들 위주로 표본설계를 수행하다보니 특정조사에서의 주요변수를 미처 고려할 수 없는 경우도 있다. 이처럼 표본추출틀에 원하는 총화변수의 정보가 없거나 표본설계의 정교함이 충분하지 못하여 이를 보완하고자 할 때 사후총화는 유용하게 활용될 수 있다.

집세조사에서는 소비자물가조사 대상 38개 도시에 거주하고 있는 일반 임차가구를 모집단으로 정의한다. 또한 다목적표본제를 채택하고 있으며, 표본을 추출하기 위한 모집단 자료는 2005년 인구주택총조사 10% 자료를 이용한다. 인구주택총조사의 10% 자료에는 가구의 점유형태에 대한 정보가 있지만 이는 매우 유동적인 정보이고, 표본설계와 조사 시점, 인구주택총조사 통계작성 시점 간에는 시차가 존재한다. 따라서 표본을 설계한 후 해당 가구와 접촉하기 전까지는 점유형태를 미리 알 수가 없다. 또한 표본설계를 할 때 다목적표본제의 주된 조사인 경제활동인구조사와 상관성이 높은 변수들을 우선적으로 활용해야 하기 때문에 집세에 영향을 미치는 특성변수들을 고려하지 못하고 있는 것이 현실이다.



이와 같이 불가피하게 표본선정에 제약을 받고 있는 상황에서 집계조사에서는 도시별 평균집세를 추정할 때 표집가중치를 고려하지 않고 있다. 물론 캐나다에서도 표집가중치를 사용하지 않으며, 2절에서 살펴본 바와 같이 UN(2009)의 집계지수 작성방법 실무적 지침에서도 간편성을 위해 표집가중치가 생략될 수도 있다고 언급하고 있다. 그러나 이는 표본거처가 표본추출틀의 임차 거처의 수와 대략 같은 수를 대표하고 있다는 전제조건이 필요하다. 즉, 표집가중치를 고려하지 않는다는 것은 선정된 표본의 구조가 모집단과 유사할 경우, 다시 말해 표본이 대표성 있게 추출되었다는 것이 보장되어야 가능한 일이다. 앞에서 살펴본 선행연구에서도 집계조사에서는 표본선정의 한계에 의해 38개 도시 수준에서 모집단과 표본 간 주택특성들의 분포에 차이가 발생하고 있다고 지적하고 있다. 이러한 상황에서 가중치를 무시하고 추정하게 되면 심각한 편향이 발생할 수 있다는 점을 유념해야 한다. 따라서 집세의 수준과 변동에 관련성이 높은 층화변수를 선정하고, 표본가구에 가중치를 사후적으로나마 부여함으로써 모집단의 주요 특성분포에 대한 조정이 필요할 것으로 보인다.

본 연구에서는 표본조사의 분산을 줄이고, 목표모집단에서 불충분한 특정그룹 표본의 커버리지를 조정하기 위해 사용되는 일반적인 기법인 사후층화를 통해 사후가중치를 작성해 보고자 한다. 여기서 가중치는 목적에 따라 주택재고나 지출액을 기준으로 작성할 수 있다. 또한 가중치는 기준시점을 참고하여 작성되는데, 일반적으로 일 년이 되며, 몇 년 동안 고정가중치를 사용할 수 있다. 이러한 사후층화는 수행하고자 하는 표본조사에서의 주요 관심사항에 대하여 상관성이 높은 특성을 층화변수로 선정하는 것이 매우 중요하다. 이와 더불어 선정된 층화변수로 사후층의 경계를 어느 수준까지 구분해 줄 것인지에 따라 사후층화의 효과도 크게 다르게 나타난다. 매우 세분화된 사후층의 구성은 각 층 내에서 동질성은 향상시킬 수는 있지만, 층별 평균관측치의 수를 줄이게 되어 결국 지수에 표준 오차를 증가시킬 수 있다. 또한 모든 표본에서 사후층의 구분에 사용되는 정보를 모두 확인할 수 있어야 하며, 세부그룹별로 가중치를 부여하기 위해 필요한 정확한 데이터를 찾기도 현실적으로 어렵기 때문에 사후층화의 세분화의 수준을 적절하게 설정해야 한다.

만일 사용자가 서로 다른 시장별로 가격지수를 원한다면 이러한 층화방법을 사용하여 별도의 지수를 작성할 수도 있다. 여러 개의 하위그룹이나 층으로 전체 표본가구를 구분하고, 각 층에 대한 중심을 파악할 수 있는 측도(평균)를 구한 후, 층별지수를 가중평균함으로써 구할 수 있다. 소비자물가지수에서 집계지수는 전세와 월세지수로 구분이 가능한데, 단독주택과 아파트 등과 같이 시도별 주택유형별 지수를 작성할 수 있는지도 다음 절에서 검토한다.

2. 작성대상 자료현황

집세조사의 조사모집단은 소비자물가조사 대상 38개 도시에서 주택을 임차하여 거주



하는 모든 일반가구⁷⁾로 정의할 수 있다. 집세조사에서 요구하는 임차가구에 대한 모집단 정보는 인구주택총조사 자료가 유일하다. 인구주택총조사는 5년마다 수행되고 있으며, 본 연구가 수행되고 있는 2010년 현재에 가장 최근 자료는 2005년 자료이다. 2005년 11월에 인구주택총조사가 실시된 이래로 모집단의 구조는 계속해서 변하고 있다. 따라서 2010년의 임차가구의 모집단 분포는 2005년과는 분명한 차이가 있을 것이다. 그러나 2005년 이후 임차가구에 대한 모집단 구조의 변동을 확인할 수 있는 자료가 없다. 따라서 본 연구에서는 2005년 인구주택총조사 자료를 모집단 자료로 이용하였으며, 이는 본 연구의 근본적인 한계라 할 수 있다. 그리고 집세조사 자료는 2006년 1월부터 2010년 4월까지의 52개월 자료를 이용하였다.

앞서 언급한 것처럼 사후층의 구성에 사용할 층화변수를 선택하기 위해서는 모집단에서의 정보도 파악할 수 있고, 표본조사에서도 해당변수가 조사되어야 한다. 두 조사에서의 집세관련 공통 항목들을 <표 3-3>에 비교해 보았다. 점유형태, 사용 방 수, 주택유형, 연건평(전용면적), 건축년도에 대한 항목이 두 조사에서 공통적으로 조사되고 있다. 위의 다섯 개 항목에 대해 두 조사 간 정의의 차이와 층화변수로서의 활용가능성을 살펴보면, 먼저 점유형태는 집세조사에서는 전세와 월세를 구분하여 지수를 작성하고 있어 층화변수로 사용할 수 없다. 다음으로 인구주택총조사의 사용 방 수는 침실, 침실 이외의 방, 거실, 식당과 같이 방의 용도별로 구분하여 조사한다. 그러나 집세조사에서는 용도의 구분 없이 사용 방 수의 합계가 ‘1개’, ‘2개’, ‘3개’, ‘4개 이상’과 같이 4개 카테고리로 구분되어 있다. 주택유형의 경우 주택 이외의 거처는 집세조사의 대상이 아니라 제외하면 단독주택과 아파트의 구분은 같다. 그러나 인구주택총조사는 연립주택과 다세대주택을 구분하지만 집세조사는 이를 구분하지 않고 연립주택으로만 표시하고 있다. 또한 인구주택총조사의 비거주용 건물 내 주택은 집세조사에서는 기타에 해당한다.

인구주택총조사에서 연건평은 가구가 거주하는 면적이 아닌 전체 주택에 대한 면적을 의미하며, 아파트에 대해서는 전용면적을 조사한다. 즉, 단독주택이나 연립·다세대주택처럼 여러 가구가 하나의 주택에 거주하는 경우엔 모든 가구의 거주면적이 합산되어 있어 개별가구의 주거면적은 파악할 수 없다. 집세조사에서는 최근에 아파트에 대한 전용면적이 조사항목으로 추가되었으며, 2009년 10월부터 결측치 없이 완전하게 조사되고 있다. 1주택 1가구가 일반적인 아파트에 대해서는 전용면적 항목이 향후 유용하게 활용될 수 있을 것으로 보인다. 마지막으로 건축년도도 집세조사에 최근에 추가된 항목으로 아파트에 대해서만 조사하며, 2009년 10월부터 완전한 자료가 이용 가능하다. 그러나 인구주택총조사 자료에는 2005년 이후 건축된 주택에 대한 정보가 없고, 신축아파트에 대한 행정자료

7) 가족으로 이루어진 가구, 가족과 가족 이외의 사람이 함께 사는 가구, 1인 가구, 가족이 아닌 남남끼리 함께 사는 5인 이하의 가구

를 이용하여 총조사 자료를 보완하더라도 신축아파트의 임차가구 비율을 확인할 수 없다. 따라서 건축년도의 층화변수 선정은 다소 신중하게 접근할 필요가 있다고 판단된다.

이와 같이 두 조사의 5개의 공통 항목 중에서는 모든 주택유형에 대해서 임차가구가 거주하는 주택유형과 사용 방 수의 정보를 활용하여 층을 구성해 볼 수 있다. 또한 아파트에 대해서는 전용면적, 건축년도의 정보도 부분적으로 활용하여 좀 더 세분화된 층을 구성할 수 있을 것이다. 그러나 본 연구에서 시험작성해볼 수 있는 시계열이 충분히 구축되어 있지 않은 관계로 본 연구에서는 해당변수에 대한 검토는 하지 않았다. 그러나 전용면적과 건축년도는 집세의 수준과 변동에 매우 밀접한 관련이 있고, 품질조정을 위한 유용한 정보이므로 향후에도 지속적으로 자료를 축적할 필요가 있다.

〈표 3-3〉 작성대상 자료 항목비교

항목		2005 인구주택총조사	집세조사
가구 특성	점유형태	자기 집	※(조사대상 아님)
		전세(월세 없음)	전세
		보증금 있는 월세	보증부월세
		보증금 없는 월세	순수월세
		사글세	사글세 보증사글세
		무상	※(조사대상 아님)
거주 주택 특성	사용 방 수	침실 수	'총 사용 방 수' 1개/2개/3개/4개 이상
		침실 이외의 방 수	
		거실 수	
		식당 수	
	주택유형	단독주택	단독주택
		아파트	아파트
		연립주택	연립주택
		다세대주택	
		비거주용 건물 내 주택	
	주택 이외의 거처	기타 ※(조사대상 아님)	
	연건평	주택 전체의 전용면적(평 또는 m ²)	'아파트만 전용면적(m ²) 조사'
	건축년도	2000/2001/2002/2003/2004/2005	'아파트만 조사'
		1995~1999	
		1990~1994	
1985~1989			
1980~1984			
1070~1979			
1960~1969			
~1959			



3. 사후층의 구성

사후층화의 첫 번째 단계는 사후층을 명확하게 정의하는 것이다. 사후층을 구성할 때 어떠한 변수를 사용하는지에 따라 사후층화의 효과는 매우 달라지므로 층화변수의 선정은 매우 중요하다. 층화변수는 표본조사의 주요 관심 변수와 높은 상관관계를 갖고 있으면서, 모집단의 구조가 비교적 안정적이라고 인정될 만한 변수를 선택하는 것이 바람직하다. 집세조사의 최종 관심사항은 집세의 변동이지만 이를 파악하기 위해서는 전월과 금월의 집세수준을 추정해야 한다. 따라서 집세수준추정치의 정확도가 집세변동을 결정하는 중요한 요인이 되므로, 집세수준의 정확한 추정이 선행되어야 한다. 정미옥 외(2009)에 의하면 앞서 살펴보았던 변수들 중 주택유형, 사용 방 수, 전용면적, 건축년도는 집세에 영향을 미치는 요인으로 분석되었다. 모집단에서의 주택의 재고와 특성의 구조변화는 신축과 멸실주택을 제외하면 시간에 따라 크게 변하지 않는다. 그러므로 이와 같은 네 항목들을 사후층화 변수로 검토해보는 것은 타당하다고 판단된다.

층화변수 선정 못지않게 층의 세분화정도를 결정하는 문제도 중요하다. 사후층을 세밀하게 나누게 되면 각 층 내 동질성은 향상되어 보다 추정의 정도를 높일 수 있다. 그러나 현실적으로 층의 세부정보를 파악할 수 없어 너무 세분화할 수 없을 수가 있다. 뿐만 아니라 층을 너무 세분화하게 되면 층에 배정되는 평균관측치의 수를 줄이게 되어 어떤 층에 대해서는 관측치가 거의 없거나 하나도 없는 빈 셀이 발생할 수 있다. 이와 같이 관측치가 거의 없는 층에는 가중치가 매우 크거나 작아지는 극단치가 발생할 수도 있다. 또한 관측치가 없는 층에 대해서는 특성이 유사한 인접의 다른 층과 병합을 해줘야 하는 등의 복잡한 후속작업을 수반한다. 따라서 사후층의 경계를 합리적이고 적절한 수준으로 결정해야 하는데, 사후층화는 표본 수가 크고, 각 사후층 내의 표본 수도 상대적으로 큰 값을 가질 때 좋은 결과를 가져다준다고 알려져 있다.

사후층화변수를 선정하기 위해 인구주택총조사 결과를 분석하여 집세수준과 각 주택 특성 간의 상관관계를 확인해 보았다. 분석대상 자료는 2005년 인구주택총조사의 10% 표본자료 중에서 소비자물가조사 대상 38개 도시에 거주하는 일반 임차가구의 임차료 항목을 이용하였다. 사용 방 수를 5개 그룹(1개/2개/3개/4개/5개 이상)으로 범주화하고, 분석의 편의성을 위해 월세를 전세평가액으로 환산⁸⁾하였다. 점유형태와 주택유형의 사용방수별 집세수준에 대한 분산분석을 수행한 <표 3-4>의 결과를 보면, 모든 유형에서 유의확률(p-value)이 <0.0001로 매우 작아 모든 주택유형과 점유형태에서 사용 방 수에 따라 평균 집세의 차이가 존재함을 확인할 수 있었다.

8) 적용이자율 1%로 가정

<표 3-4> 사용방수별 집세수준에 대한 분산분석 결과

점유형태	주택유형	DF	F Value	Pr > F
전세	단독주택	4	5252.51	<.0001
	아파트	4	9871.42	<.0001
	연립·다세대·기타	4	920.88	<.0001
월세 (사글세포함)	단독주택	4	2272.44	<.0001
	아파트	4	7208.13	<.0001
	연립·다세대·기타	4	236.94	<.0001

분석자료 : 2005 인구주택총조사 10%표본 38개 도시거주 일반 임차가구

점유형태와 주택유형별 임차가구의 사용 방 수에 대한 모집단 백분위수를 <표 3-5>와 같이 분석하였다. 전세가구의 사용 방 수의 33백분위수는 세 주택유형 모두 3개이며, 66백분위수는 단독주택은 3개, 아파트와 연립다세대기타는 4개이다. 월세가구의 사용 방 수의 33백분위수는 단독주택과 연립다세대기타가 2개 아파트가 3개이고, 66백분위수는 세 주택유형 모두 3개이다. 전세 거주 가구의 사용 방 수가 월세에 비해서 더 많은 것을 알 수 있으며, 같은 점유형태 내에서도 단독주택보다는 아파트의 사용 방 수가 더 많음을 확인할 수 있었다. 이와 같이 점유형태와 주택유형에 의해 다소 차이가 있지만 임차가구의 사용 방 수는 대체적으로 2~4개 사이에 집중되어 있음을 알 수 있다.

<표 3-5> 임차가구 사용 방 수의 모집단 백분위수

점유형태	주택유형	25 백분위수	33 백분위수	50 백분위수	66 백분위수	75 백분위수
전세	단독주택	2	3	3	3	4
	아파트	3	3	4	4	4
	연립·다세대·기타	3	3	4	4	4
월세 (사글세포함)	일반주택	1	2	2	3	3
	아파트	3	3	3	3	4
	연립·다세대·기타	2	2	3	3	4

분석자료 : 2005 인구주택총조사 10%표본 38개 도시거주 일반 임차가구

앞의 작성대상 자료현황에서 살펴보았지만, 사후층을 구성하기 위해 이용가능한 변수가 많지 않다. 또한 집세평균 추정 단위인 38개 도시별 전세그룹과 월세그룹 내에 표본의 크기가 크지 않아 각 그룹 내에서 층을 세세하게 구성하기에는 어려움이 있었다. 이에 본 연구에서는 모집단과 표본의 분포에 대해 살펴본 개략적인 내용과 현실적인 측면을 고려하여 다음과 같이 층 구성안을 간단하게 정리하였다. 먼저 단독주택, 아파트, 연립·다세



대·기타주택의 3개 층으로 구분하여 주택유형에 의한 층1을 구성하였다. 다음으로 주택 유형과 사용 방 수를 동시에 고려하여 주택유형의 3개 층과 사용 방 수의 2개 이하, 3개, 4개 이상으로 구분된 3개 층을 조합하여 총 9개의 그룹으로 층2를 구성하였다.

다음으로 층의 경계의 적절성을 확인해보고자 표본과 모집단의 층별 분포의 특성을 살펴보았다. 집세조사 2009년 결과를 분석해보니 38개 도시 중 주택유형이 연립·다세대·기타주택에 거주하는 가구가 매월 하나도 조사되지 않는 도시도 있고, 10가구 미만이 9개 도시, 10~20가구 미만이 7개 도시인 분포를 보이고 있었다. 또한 모집단에서도 연립·다세대·기타주택은 서울, 경기, 인천, 제주를 제외하고 나머지 도시에서 임차가구가 거주하는 비중이 매우 낮아 별도의 층으로 만들었을 때 빈 셀 발생 가능성이 매우 높았다. 사후층화는 표본의 수(n)가 크고 각 층별 표본의 수(n_i) 역시 상대적으로 큰 값을 가질 때 좋은 결과를 주며, 각 층에 대하여 $n_i \geq 20$ 이라면 표본선택 후의 층화방법은 비례배분을 갖는 층화임의표본 만큼이나 정확하다고 하였다(김영원 외, 1998, p187~189). 이에 따라 연립·다세대·기타주택을 집세의 수준이나 특성이 유사한 단독주택으로 병합을 하고, 주택유형을 ‘아파트’와 아파트 이외의 ‘일반주택’의 두 개 그룹으로 만들어 층1-1이라 하였다.

<표 3-6> 16개 시도 주택유형별 임차가구 사용 방 수의 모집단 분포

시 도	(단위 : %)					
	일반주택 방 2개 이하	일반주택 방 3개	일반주택 방 4개 이상	아파트 방 2개 이하	아파트 방 3개	아파트 방 4개 이상
서울	24.57	31.66	18.72	2.62	10.03	12.42
부산	26.64	27.53	18.29	5.87	8.72	12.96
대구	29.49	23.68	20.6	0.93	11.56	13.74
인천	23.08	25.25	20.09	1.88	13.58	16.11
광주	26.06	9.73	19.81	4.37	12.69	27.34
대전	31.11	19.74	14.64	2.64	14.17	17.71
울산	32.49	24.23	14.93	1.57	11.63	15.15
경기	24.18	31.5	14.9	1.25	11.23	16.93
강원	33.71	12.21	10.13	4.2	17.3	22.44
충북	34.01	15.15	10.4	7.88	14.39	18.17
충남	32.24	10.03	10.15	4.89	18.01	24.68
전북	25.19	11.82	14.37	3.67	19.28	25.68
전남	14.23	11.07	12.86	0.34	18.16	43.34
경북	33.14	16.42	14.11	2.19	16.06	18.08
경남	30.15	20.74	16.63	1.07	12.01	19.41
제주	22.15	22.27	31.92	1.46	11.09	11.13

또한 9개의 그룹이었던 층2도 층2-1로 통합하였는데, 먼저 주택유형은 일반주택과 아파트로 구분하였다. <표 3-6>의 16개 시도 주택유형별 임차가구 사용 방 수의 모집단 분

포를 살펴보면, 서울은 일반주택 방 3개에 거주하는 임차가구가 31.66%로 가장 높았으며 아파트 방 2개 이하에 거주하는 임차가구는 2.62%로 가장 낮았다. 다른 광역시나 시도 수준에서도 대체로 일반주택은 방 2개 이하의 비중이 가장 높으며 아파트 방 2개 이하의 극히 낮은 비율을 차지하고 있었다. 이러한 모집단 분포를 감안하여 사용 방 수를 일반주택의 경우에는 방 2개 이하와 방 3개 이상으로 구분하고, 아파트는 방 3개 이하와 4개 이상으로 구분하여 총 4개의 층으로 구성하였다.

4. 사후가중치 작성

사후층화를 통한 가중치 조정 시 유의해야 할 점은 분산의 크기를 유지하면서 편향을 줄이는 효과를 통제하기 위해서 극단가중치가 발생하는 것을 억제해야 한다는 것이다. 특정 셀의 가중치 크기가 사전에 정해놓은 수준을 벗어나 극단가중치가 발생한다면, 최초 사후층화 셀을 인근의 다른 셀과 병합하거나 가중치의 크기를 제한하는 등의 후속 조치를 해야 한다. 이러한 후속 조치의 실행 여부를 결정하기 위해 우리는 다음과 같은 두 가지 기준을 고려해 볼 수 있다. 첫 번째는 해당 셀의 표본 수와 목표모집단에서의 수의 비율로 정의되는 초기조정인자(IAF: Initial Adjustment Factor)를 기준으로 삼을 수가 있다. 즉, 사후층화를 통해 초기에 계산된 가중치의 크기가 정해놓은 임계치를 넘으면 병합을 해주게 된다. 두 번째로는 표본 수를 기준으로 활용할 수 있는데, 어느 셀의 표본 수가 매우 적은 경우에도 초기조정인자가 불안정하다는 근거에 의해 셀을 병합할 수 있다(Kim 외, 2007). 극단가중치의 범위에 대해 공식적으로 명시된 기준은 없으며 각 조사의 특성에 맞춰 개별적으로 설정해야 하지만, 실무에서는 경험적으로 사후가중치의 크기가 0.3보다 작거나 3.0보다 큰 경우를 극단가중치로 정의하고 있다. 본 연구에서도 이러한 정의를 따라 사후가중치의 크기가 0.3~3.0의 범위를 벗어나는 경우를 극단가중치로 설정한다.

앞에서 설정해 놓은 층 구분에 의해 사후층을 구성한 후 2005년 인구주택총조사 전수조사 결과를 이용하여 사후가중치를 작성해보았다. 사후가중치는 각 층화방법별로 구성해 놓은 층의 개수만큼 작성된다. <표 3-7>에서 보면, 층1의 경우 사후가중치 수는 총 228개(38개 도시*2개 점유형태*3개 주택유형)이다. 분석대상 기간인 52개월 동안 매월 발생할 수 있는 극단가중치의 평균 개수를 구해보았는데, 매월 약 15개의 극단가중치가 평균적으로 발생하여 전체 중 약 6%의 비중을 차지하고 있었다. 또한 층의 개수가 152개로 가장 적은 층화1-1에서의 극단가중치 비율이 3%로 가장 낮음을 알 수 있으며, 층의 개수가 가장 많은 방법인 층화2에서 극단가중치의 비율이 16%로 매우 높다는 것을 알 수 있다.

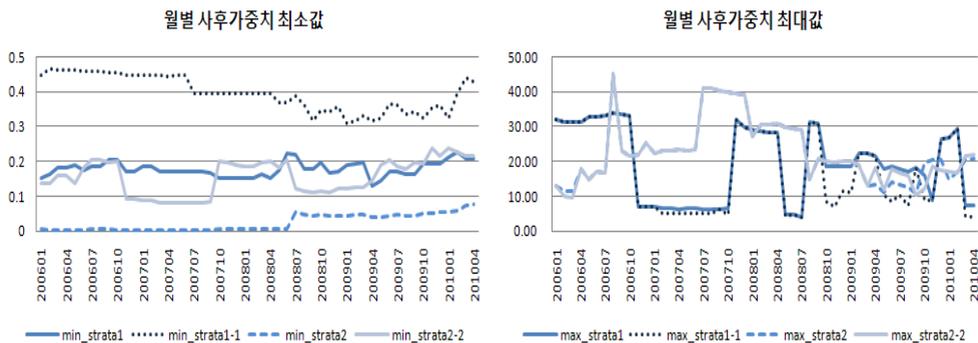
〈표 3-7〉 층화방법별 사후가중치 작성결과

	층화1	층화1-1	층화2	층화2-1
사후가중치 수	228	152	684	304
극단치 수	15	5	113	35
극단치 비율	0.06	0.03	0.16	0.11

* 극단치 : 가중치<0.3 or >3.0

** 52개월 평균

[그림 3-3]은 2005년 인구주택총조사의 모집단 자료를 기준으로 2006년 1월부터 2010년 4월까지의 집세조사 52개월 자료에 각 층화방법을 적용하여 작성된 사후가중치의 월별 최소값과 최대값에 대한 그림이다. 층화1-1(strata1-1)방법의 경우 최소값은 전체 기간 동안 0.3보다는 항상 큰 값을 유지하고 있지만, 최대값은 3.0보다 큰 극단가중치가 거의 매월 발생하며 때에 따라서는 30보다 큰 극단치도 종종 발생하고 있음을 알 수 있다. 층화1방법과 층화2-2방법에서도 매월 0.3보다 작은 사후가중치의 극단치가 발생하고 있으며, 3.0보다 큰 극단가중치도 매월 발생한다. 또한 층화2(strata2)방법은 최소값이 0.3보다 훨씬 작은 0에 가까운 극단치가 매월 발생하고 있으며, 사후가중치의 최대값이 40보다 크게 나타나는 월도 있는 등 극단가중치가 매월 10보다 큰 수준을 유지하고 있다.

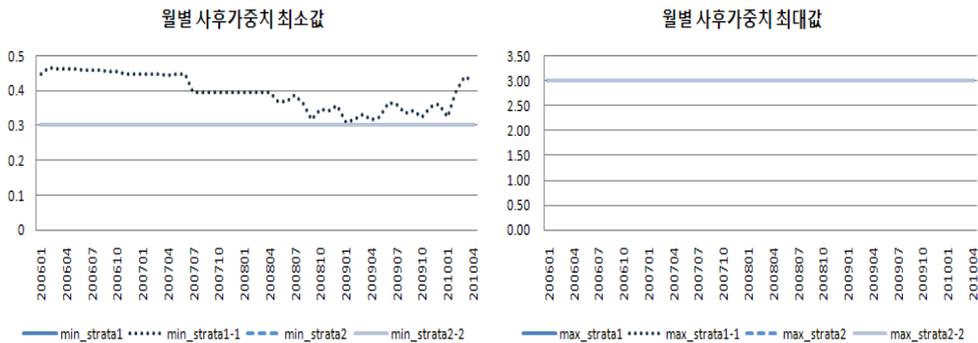


[그림 3-3] 가중치 제한 전의 사후가중치 최소값과 최대값

이와 같이 극단가중치의 발생 빈도가 높거나 그 크기가 일정범위를 벗어나 아주 크거나 작은 값을 갖게 된다면, 이러한 가중치를 기초로 계산한 추정치의 분산도 매우 커지게 된다. 극단가중치의 발생은 추정의 정도를 저하시키므로 사후가중치의 작성과 적용은 신중해야 한다. 만일 극단가중치가 발생한 층이 있으면 일차적으로 특성이 유사한

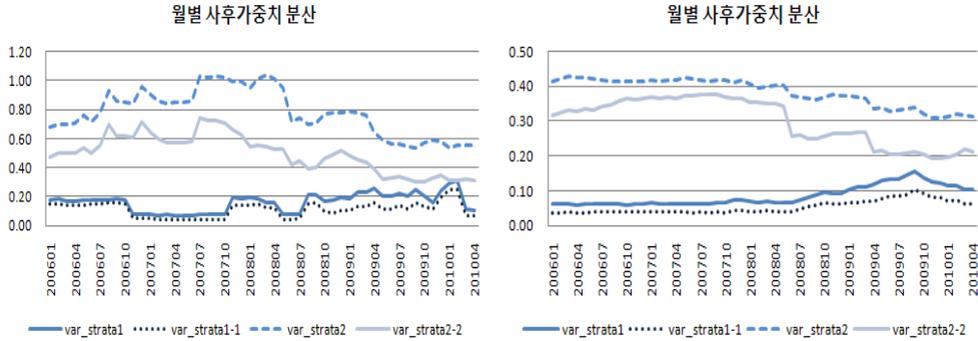
인접 층으로 병합하는 방법을 고려해볼 수 있다. 그러나 집세조사의 추정단위가 매우 협소하여 38개 도시별 전·월세 가구에 대해 각각의 추정치를 작성해야 한다. 또한 층화변수인 주택유형이나 사용 방 수에 의한 층의 개수도 많지가 않기 때문에 인접 층과 병합을 하는 데에도 한계가 있다. 따라서 극단가중치를 제어하기 위한 층의 병합보다는 이 단계에서는 가중치를 일정한 범위를 두고 제한하는 방법을 적용하였다.

앞서 정의해 놓은 것처럼 극단가중치의 임계치를 0.3과 3.0으로 설정하고, 사후가중치가 0.3보다 작으면 0.3으로, 3.0보다 크면 3.0으로 값을 제한함으로써 극단가중치의 발생을 억제하였다. 이와 같이 가중치의 크기를 제한한 후 사후가중치의 월별 최소값과 최대값을 구해보았다. [그림 3-4]에서 보면, 최소값이 모두 0.3보다 커서 가중치 제한이 필요 없었던 층화1-1방법을 제외하고는 다른 층화방법들 모두 가중치의 최소값이 0.3으로 제한되었으며, 가중치의 최대값도 3.0으로 제한되었음을 확인할 수 있다.



[그림 3-4] 가중치 제한 후의 사후가중치 최소값과 최대값

다음으로 사후가중치 퍼짐의 정도를 측정해보고자 사후가중치들의 분산을 월별로 구해보았다. [그림 3-5]에서 보는 바와 같이 층화1-1방법의 분산이 0.2보다 작은 값으로 다른 층화방법들에 비해 가장 낮은 수준을 보이고 있으며, 층화2방법은 0.6~1.0 수준을 유지하고 있다. 또한 이렇게 가중치의 크기가 제한된 후에는 사후가중치들의 분산이 층화1방법과 층화1-1방법은 월별로 0.1보다 낮은 수준을 유지하고, 층화2-2방법은 0.2~0.4, 층화2방법은 0.3~0.4의 수준을 보이고 있다. 가중치 제한 전과 후의 수준을 비교하였을 때 가중치를 제한한 후에 분산이 거의 절반 수준으로 줄어드는 것은 어쩌면 당연한 결과이겠지만, 사후가중치의 크기를 제한해주는 것은 필요하다고 판단된다.



[그림 3-5] 가중치 제한 전(좌측)과 후(우측)의 사후가중치 분산

5. 평균 집세의 추정

집세지수는 각 도시에서 전세지수와 월세지수를 작성하고, 도시 내 전월세가중치를 이용하여 도시의 집세지수가 작성된다. 따라서 기본지수인 도시별 전세지수와 월세지수를 작성하기 위해서 도시별로 금월과 전월의 전세평균값과 월세평균값을 추정한 후 이들을 비교하여 전세와 월세의 변동을 계산한다. 즉, 지수작성을 위해 추정해야 하는 기본단위는 금월과 전월의 전세와 월세 평균이므로 각 도시 내에서 전세가구의 평균전세와 월세가구의 평균월세를 각각 추정해야 한다.

사후층화에 의한 평균 추정은 다음과 같다.

$$\bar{y}_p = \sum_{h=1}^H \frac{N_h}{N} y_h = \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} w_h y_{hi}}{N}$$

여기서 첨자 p 는 사후층화, h 는 사후층을 의미하고, w_h 는 층별 사후가중치이다. 사후가중치는 사후층별로 모집단가구 수를 표본가구 수로 나눠주면 된다. 또한 사후층화추정량 \bar{y}_{st} 의 분산은 층화 표본평균의 분산과 같지 않은데, 그 이유는 층화가 표집단계에서 설계되지 않았기 때문이다. 사후층화추정량 \bar{y}_{st} 의 근사 분산은 다음과 같이 구한다(김영원 외, 1998, p.188)

$$\widehat{V}_p(\bar{y}_{st}) = \frac{N-n}{Nn} \sum_{h=1}^H \left(\frac{N_h}{N} \right) s_h^2 + \frac{1}{n^2} \sum_{h=1}^H \left(1 - \frac{N_h}{N} \right) s_h^2$$

마찬가지로 집세조사의 사후가중치는 아래와 같은 수식으로 표현할 수 있다.

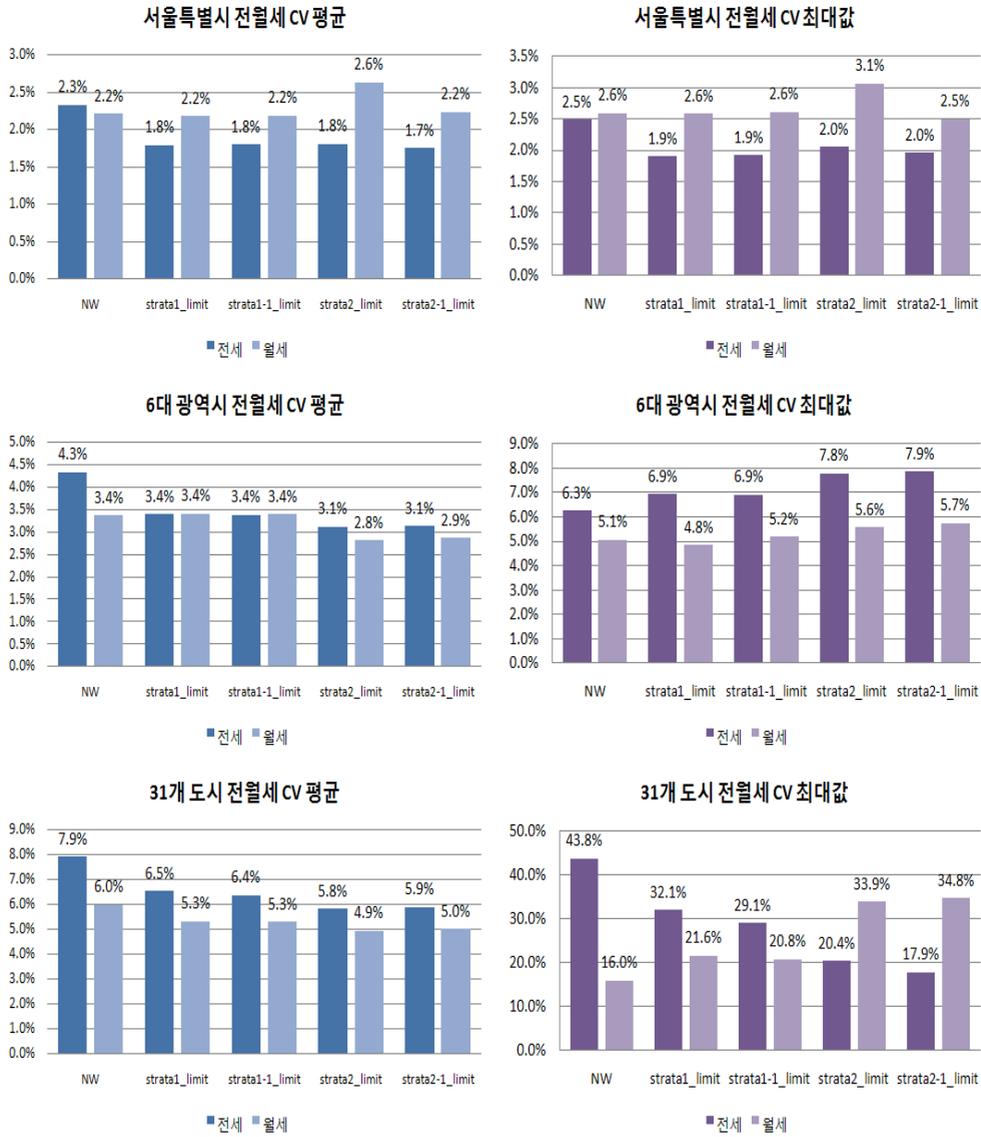


$$w_{cdh(ym)} = \frac{N_{cdh}}{n_{cdh(ym)}}$$

여기서 $c = 1, 2$ (전세, 월세), $d = 1, \dots, 38$ (도시), $ym = 200601, \dots, 201004$ (년월)이다.

다음은 사후층화를 수행한 후, 도시별 전·월세 평균을 추정하고 근사적으로 상대표준오차(CV : Coefficient of Variation)를 52개월의 자료에 대해서 구해보았다. 월별 도시별로 구한 전·월세 평균 추정치의 CV 그래프는 <부록>에서 확인할 수 있다. 서울특별시, 6대 광역시, 31개 도시와 같이 도시와 표본의 규모별로 그룹을 만들어 52개월 동안의 CV 평균과 최대값을 요약해보았다. [그림 3-6]의 NW는 사후가중치를 사용하지 않고 현행 집계조사의 추정방법처럼 단순 산술평균한 경우이고, 나머지는 가중치의 크기 제한 후의 사후가중치를 층화1방법부터 2-1방법까지를 적용하였을 때의 결과이다. 먼저 서울은 사후조정을 하지 않은 NW의 전세 CV의 평균 수준은 약 2.3%인데, 층화1방법과 1-1방법을 적용하게 되면 약 1.8%로 줄어들고 층화2-1방법에서 약 1.7%로 가장 낮게 나타난다. 또한 월세 CV의 평균은 NW와 층화1, 1-1, 2-1방법에서 모두 2.2% 수준을 보이고 있다. 이와 같이 서울에서는 사후 조정 방법에 의한 CV 수준에 현격한 차이가 나타나지 않는 것은 서울은 표본의 수가 많고 사후층별 비율도 모집단과 가장 유사한 편이기 때문인 것으로 보인다. 그러나 9개의 그룹으로 세분화한 층화2방법에서 CV가 2.6% 수준까지 오히려 높아지는 것은 층의 구분을 지나치게 세분한 결과라고 판단된다. 6대 광역시와 31개 도시에서는 사후가중치를 적용했을 때 전·월세 모두 CV의 평균 수준이 낮아진다. 그러나 CV의 최대값은 6대 광역시의 전·월세와 31개 도시의 월세에서 사후층화를 한 경우가 더 크게 나타난다.

지나치게 세분화한 층화방법은 추정치의 정도를 오히려 저하시킬 우려가 있다. 또한 표본이 충분히 크고, 모집단의 구조를 잘 반영하고 있는 경우 사후층화의 효과가 크게 나타나지 않는다. 이와는 반대로 사후층 내에 표본의 크기가 크지 않거나 모집단과 표본 간 구조차이가 심하여 층별 사후가중치 크기의 차이가 클 때에는 가중치 조정으로 인해 CV가 단순히 줄어들지만은 않는다는 것을 알 수 있다. 따라서 추정치의 정도를 향상시켜 줄 수 있는 적절한 층화방법을 선택하는 것이 바람직하다. 현행 표본의 크기나 추정치의 CV를 고려한다면 서울과 6대 광역시는 4개의 층으로 구성된 층화2-1방법, 31개 도시에 대해서는 2개의 층으로 구성된 층화1-1방법이 적절한 층화의 수준이라고 보인다.



[그림 3-6] 52개월 동안의 전·월세 평균 추정에 대한 CV 평균과 최대값 요약

제4절 하위지수 작성방법

1. 하위지수 작성방법 이론

소비자물가지수 작성은 일반적으로 크게 두 단계를 거쳐 진행된다. 첫 번째 단계에서는 하위가격지수가 하위지출을 집계하여 추정된다. 두 번째 단계에서는 하위지출의 상대적인 가치를 가중치로 사용하여 하위가격지수들의 평균을 구함으로써 상위가격지수가 계산된다. 즉, 하위가격지수는 수집된 가격자료에 의존하는 좀 더 근본적인 개념이라 할 수 있으며, CPI를 구성하는 기초 요소가 된다. 이 단계에서 부적절한 산식을 사용하게 된다면 소비자물가지수 전반에 걸쳐 심각하게 좋지 않은 영향을 미칠 수 있다.

즉, 38개 도시별 집세에 대한 지수를 구하는 것이 첫 번째 단계이고, 이렇게 구해진 38개 도시별 집세지수를 도시별 소비자물가지수, 16개 시도와 전국의 집세지수로 종합하고, 나아가서는 전국 소비자물가지수로 종합하는 과정이 두 번째 단계이다.

지난 수세기 동안 지수에 대한 많은 종류의 수학적 식이 제안되어 왔지만, 모든 상황에 적용될 수 있는 이상적인 하나의 식은 존재하지 않을 수도 있다. 다시 말해 각자의 목적과 현실에 맞는 방법을 채택하는 것이 최선이라 할 수 있다. 본 연구에서는 하위지수인 집세지수를 작성하는 방법에 대해 중점적으로 알아보려고 하므로 하위지수 작성방법론에 대해서만 살펴보고자 한다. 개별 가격관측치로부터 하위지수를 작성하기 위해 일반적으로 많이 사용되는 세 가지 주요 방법에 대해서 살펴보고, 이들 방법의 장·단점을 비교해 본다.

가. 듀퐁(Dutot)방법

첫 번째로는 프랑스의 경제학자 듀퐁(1738)에 의해 알려진 하위지수식은 다음과 같다.

$$P_D^{0:t} = \frac{\sum_i p_i^t / n}{\sum_i p_i^0 / n}$$

여기서 하위품목에는 n 개의 상품가격관측치가 있다고 가정을 하면, p_i^t 는 t 시점에서 i 번째 상품의 가격을 나타낸다. 위의 식에서 보듯이 듀퐁방법은 비교시점 t 의 n 개 가격의 산술평균을 기준시점 0의 n 개 가격의 산술평균으로 나뉜 것이다. 이 방법은 평균의 비 방법(RA: Ratio of Averages)으로도 불린다.



나. 칼리(Carli)방법

두 번째로는 이탈리아 경제학자에 칼리(1764)에 의한 하위지수식이다.

$$P_C^{0:t} = \frac{1}{n} \sum_i \frac{p_i^t}{p_i^0}$$

이 지수식은 n 개 상품의 가격비를 산술평균한 것이다. 이 방법은 가격비의 평균방법(AR: Average of Relatives)이라고도 불린다. 이 방법은 만일 n 개 상품가격이 모두 동등하게 중요한 경우에 적용하는 것이 가장 적절하다. 즉, 개별가격이 동일한 지출비중을 차지하고 있거나, 이러한 지출비중이 시간이 흐름에 따라 변하지 않을 경우에 적합하다.

다. 제본스(Jevons)방법

세 번째로 영국 경제학자 제본스(1863)에 의해 소개된 하위지수식은 아래와 같다.

$$P_J^{0:t} = \left(\prod_i \frac{p_i^t}{p_i^0} \right)^{1/n} = \frac{\left(\prod_i p_i^t \right)^{1/n}}{\left(\prod_i p_i^0 \right)^{1/n}}$$

이 식은 M 개의 가격비를 기하평균하는 것으로, 기하평균의 비를 구하거나 상대가격에 기하평균을 취하거나 동일한 결과를 수행하게 된다. 이 방법은 기하평균방법(GM: Geometric Mean)으로도 불린다. 이 방법을 사용할 때 주의해야 할 점은 하나의 관측치가 0의 값을 가질 경우, 다른 관측치들의 값에 상관없이 기하평균 값이 0이 된다는 사실이다. 제본스 방법은 극단치에 매우 민감하므로 개별가격의 비에 대해 10과 0.1과 같이 각각 상한과 하한을 부여할 필요가 있다. 물론 극단적인 관측치는 어떠한 다른 오류로부터 기인할 수 있지만, 어쨌든 극단적인 가격의 움직임은 신중하게 확인해야 한다.

라. 각 방법의 장단점

<표 3-8>에는 예제데이터를 이용하여 각 방법들의 지수를 산출하여 비교해 보았다. 표에서 보는 바와 같이 서로 다른 세 가지 지수식은 매우 다른 결과를 가져오게 된다. 하위지수에 대해 가장 주요하게 고려되고 있는 세 가지 성질인 측정단위의 변경 테스트(commensurability test), 시간역전성 테스트(time reversal test), 이행성 테스트(transitivity test)에 의해 각 방법들의 장·단점을 비교해 보고자 한다. 듀퐁방법은 가격수준에 의해 영향을 받는다. 듀퐁방법에서는 가격변동이 기준시점의 가격에 의해 내재적인 가중

치를 갖게 된다. 즉, 동일한 가격변동이 있더라도 가격수준이 높은 제품의 가격변동이 낮은 제품에 비해 더 높은 가중치를 갖게 된다. 하위지수를 구성하는 상품들은 가능하면 동질적이어야 하는데, 동질적이지 못하다면 듀퍃방법에서는 측정단위를 변경해도 지수가 변하지 않아야 하는 측정단위의 변경 테스트를 통과하지 못하여 심각한 문제가 발생한다.

〈표 3-8〉 예제데이터를 이용한 작성방법별 지수

	1월	2월	3월	4월	5월	6월	7월
품목A	6.00	6.00	7.00	6.00	6.00	6.00	6.60
품목B	7.00	7.00	6.00	7.00	7.00	7.20	7.70
품목C	2.00	3.00	4.00	5.00	2.00	3.00	2.20
품목D	5.00	5.00	5.00	4.00	5.00	5.00	5.50
산술평균	5.00	5.25	5.50	5.50	5.00	5.30	5.50
기하평균	4.53	5.01	5.38	5.38	4.53	5.05	4.98

칼리	m2m	100.00	112.50	108.93	101.85	91.25	113.21	100.07
	연쇄	100.00	112.50	122.54	124.81	113.89	128.93	129.02
	직접(Jan.)	100.00	112.50	125.60	132.50	100.00	113.21	110.00
듀퍃	m2m	100.00	105.00	104.76	100.00	90.91	106.00	103.77
	연쇄	100.00	105.00	110.00	110.00	100.00	106.00	110.00
	직접(Jan.)	100.00	105.00	110.00	110.00	100.00	106.00	110.00
제본스	m2m	100.00	110.67	107.46	100.00	84.09	111.45	98.70
	연쇄	100.00	110.67	118.92	118.92	100.00	111.45	110.00
	직접(Jan.)	100.00	110.67	118.92	118.92	100.00	111.45	110.00

따라서 일반적으로 하위품목 내에 서로 다른 종류의 상품이 들어 있다면, 듀퍃방법은 적합하지 않다. 또한 상품들이 동질적이고, 같은 단위로 측정이 되었다고 할지라도, 듀퍃방법의 내재적가중치는 여전히 만족스럽지 못하다. 왜냐하면, 총 지출에서 차지하는 비중이 작더라도 가격이 비싼 상품의 가격변동에 더 많은 가중치가 주어지기 때문이다. 반면, 칼리와 제본스 방법은 가격의 비에 의존하기 때문에, 가격수준에 의해서는 영향을 받지 않게 되어 측정단위 변경 테스트를 만족시킨다.

다음으로 모든 데이터의 두 시점이 서로 바뀌게 된다면, 이때의 지수는 원 지수의 역수와 같아야 한다는 것이 시간역전성이다. 칼리방법은 이 테스트를 통과하지 못하지만, 듀퍃과 제본스 방법은 이 성질을 만족한다. 또한 다른 중요한 특성은 듀퍃과 제본스 방



법은 이행성(transitivity)이 있는데 비하여 칼리는 그렇지 않다. 여기서 이행성이란 연속적인 곱에 의해 월별변동을 연결시킴으로써 연쇄적으로 지수를 구한 것과 각각의 연속되는 월들에 대해서 기준시점과 직접 비교하여 지수를 낸 것이 동일하다는 것이다. 실질적으로 많은 하위지수들이 전월과 금월을 비교하는 방법과 함께 연쇄적 지수방법에 의해 계산되기 때문에, 이와 같은 특성은 실제적으로 매우 중요하다.

여기서 직접적지수는 현재시점의 가격을 기준시점의 가격과 직접적으로 비교하는 것이고, 연쇄적지수는 각 시점별 가격들을 바로 그 직전 기간의 가격들과 비교하는 것이다. 연쇄적지수는 단기지수를 함께 연결시킴으로써 장기지수를 구하게 되는 결과를 낳게 된다. 듀툏과 제본스 방법의 연쇄적지수는 이행성을 만족하지만, 연쇄적 칼리방법은 이행성이 없어서 상향된 편향을 갖게 된다.

세 가지 하위지수식을 평가해보면, 칼리와 듀툏 방법은 아직까지 각국의 통계청에서 널리 쓰이고 있지만 심각한 단점을 가지고 있다. 칼리방법은 시간역전성과 이행성 테스트를 통과하지 못한다. 듀툏방법은 동질적인 상품집합에 대해 의미가 있지만, 다양한 상품에 대해서는 지수의 임의성이 커지게 된다. 반면 제본스방법은 위에서 언급한 모든 테스트를 통과하므로, 공리적 접근법에 의하면 제본스방법이 아직까지는 많이 사용되지는 않으나 가장 좋은 방법이라 할 수 있다. 이러한 이유로 칼리나 듀툏 방법을 사용하던 통계청들이 제본스방법으로 변경하는 추세가 증가하고 있다.

이와 같이 지수들의 특징을 이해한다고 해도, 하위지수를 작성하기 위해 어떤 방법을 사용해야 하는지에 대한 충분한 근거는 될 수 없을 것이다. 각 지수식들의 특징을 이해함과 동시에 하위지수를 작성하고자 하는 품목과 상황에 맞는 방법을 선택하는 것이 바람직하다.

〈표 3-9〉 공리적 접근법에 의한 작성방법 비교

	측정단위 변경 테스트 (commensurability test)	시간역전성 테스트 (time reversal test)	이행성 테스트 (transitivity test)
듀툏	fail	satisfy	satisfy
칼리	satisfy	fail	fail
제본스	satisfy	satisfy	satisfy

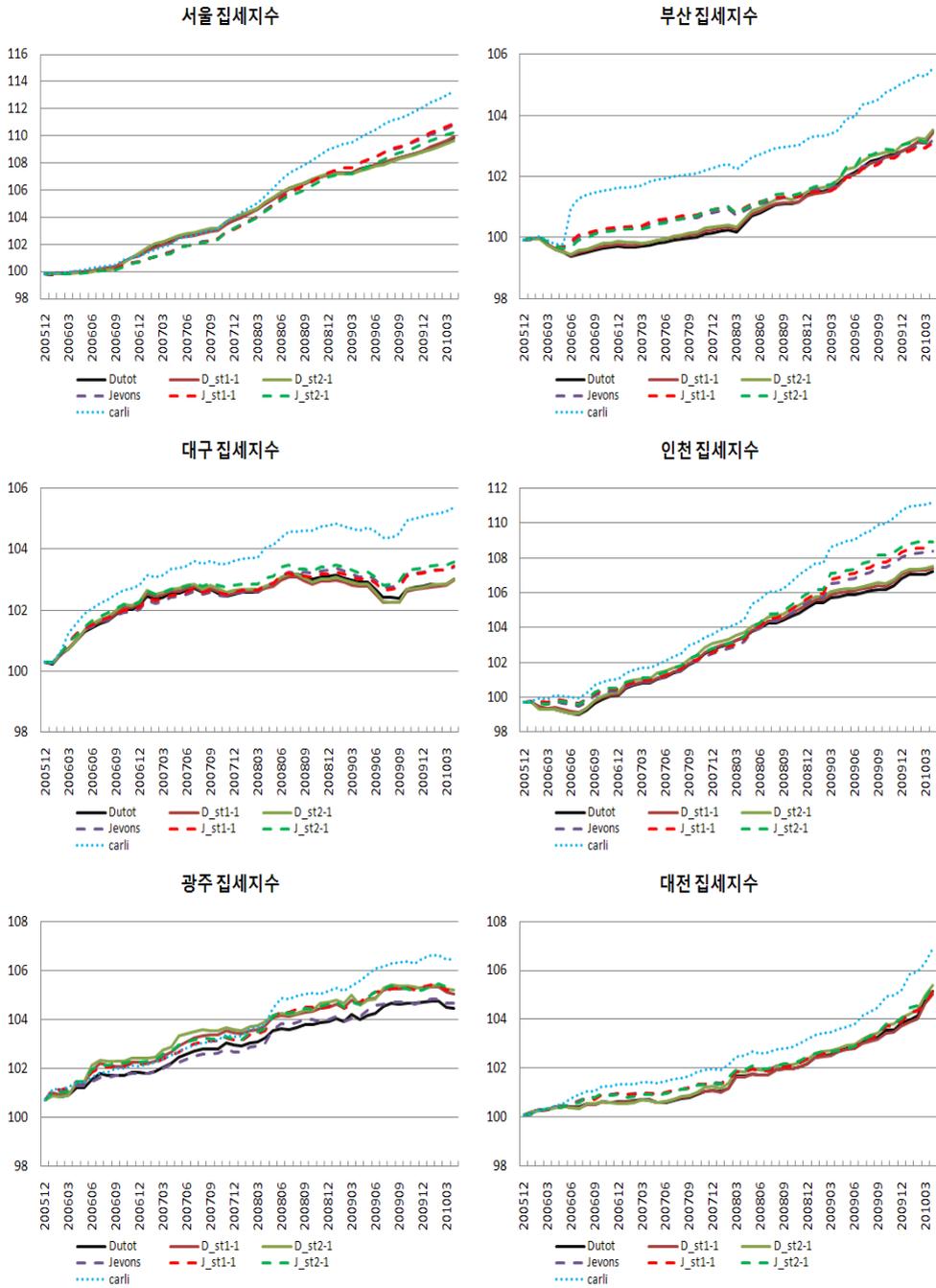
2. 모의지수 작성 및 결과 비교

이 절에서는 앞에서 살펴본 세 가지 하위지수 작성방법인 칼리, 듀툏, 제본스 방법에 대해서 사후가중치를 적용했을 때와 적용하지 않았을 때의 지수를 각각 구해보고, 총 7가지의 방법을 비교해보고자 한다. 먼저 방법 1은 현행 집세지수 작성방법과 동일하게 가중치

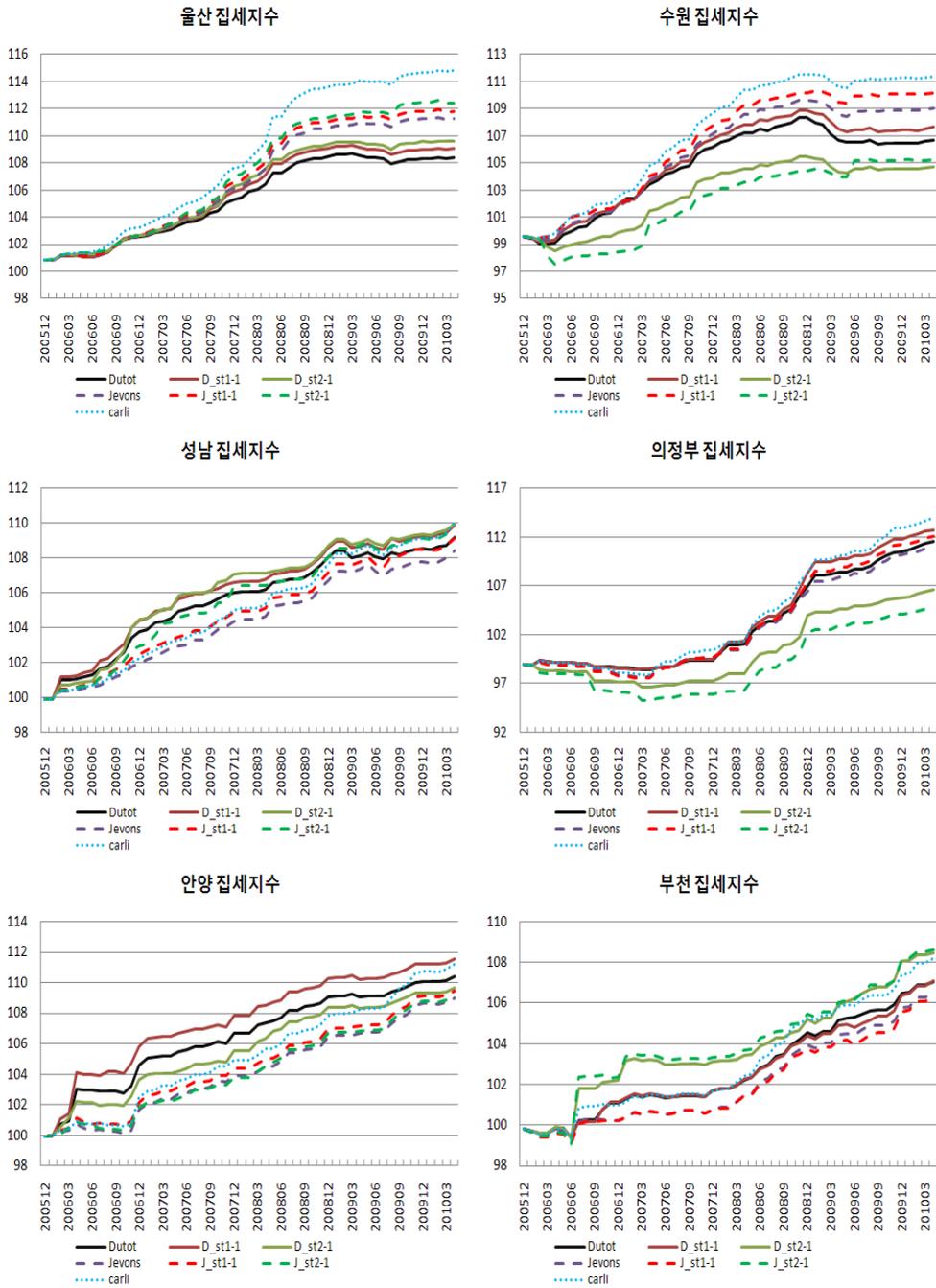
를 고려하지 않은 듀퐁방법을 적용한 것이다. 방법 2는 증화1-1에 의해 주택유형에 대한 사후가중치를 사용한 듀퐁방법이고, 방법 3은 주택유형과 사용 방 수에 의한 사후가중치인 증화 2-1를 적용한 듀퐁방법으로 지수를 작성한다. 방법 4는 가중치가 없는 제본스방법이고, 방법 5는 증화 1-1에 의한 가중치로 제본스방법을 적용한 경우이며, 방법 6은 증화 2-1에 의해 가중치가 있는 제본스방법으로 지수를 작성하였다. 마지막으로 방법 7은 가중치를 고려하지 않은 칼리방법이다. 수식으로 표현하면 각각 <표 3-10>과 같다.

<표 3-10> 하위지수 작성방법

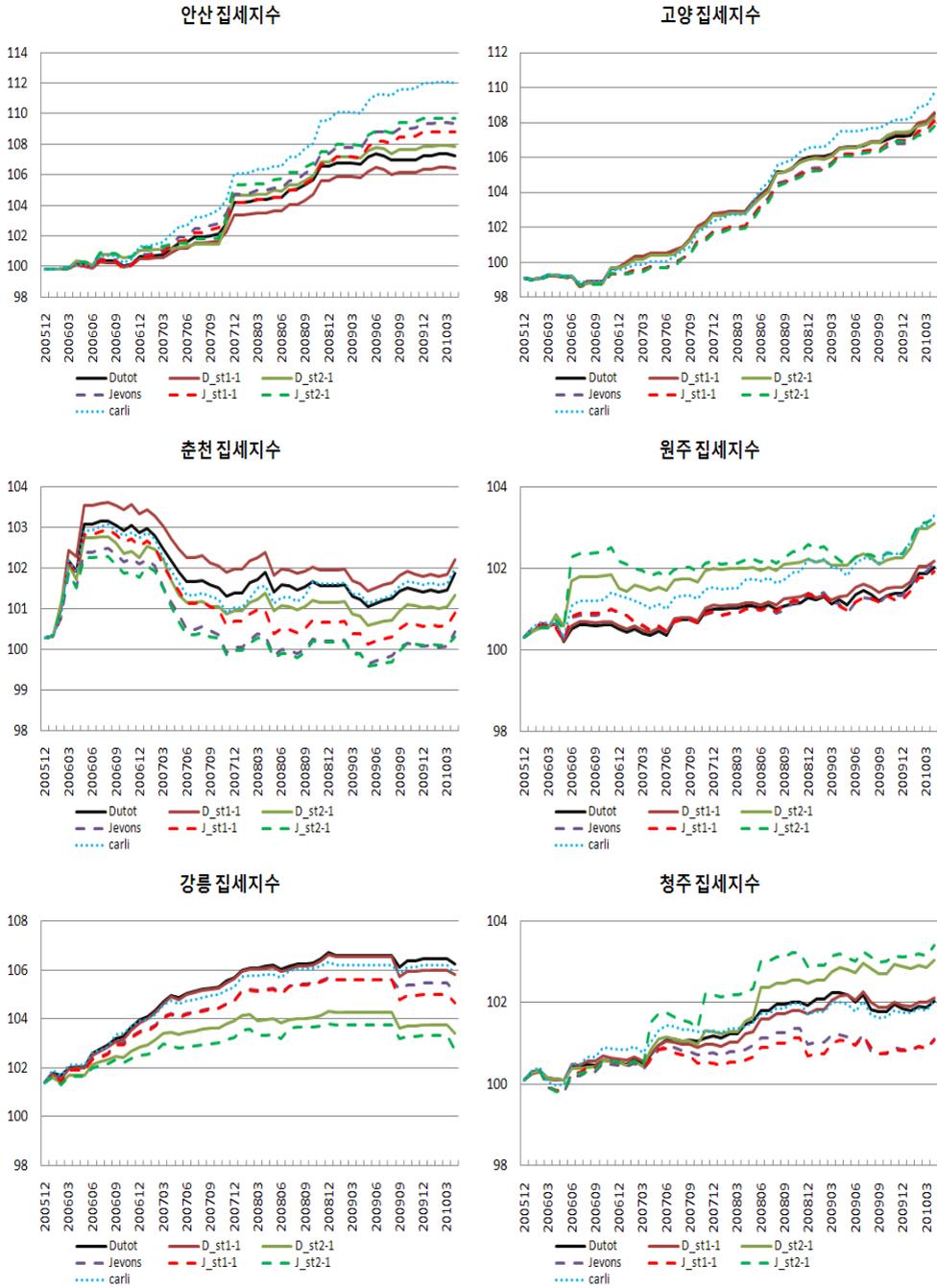
	하위지수 작성방법	적용 가중치	산식
방법 1 (Dutot)	듀퐁	없음	$P_{Dutot}^{t-1:t} = \frac{\frac{1}{n} \sum_{i=1}^n p_i^t}{\frac{1}{n} \sum_{i=1}^n p_i^{t-1}}$
방법 2 (D_st1-1)	듀퐁	사후증화1-1	$P_{D_{st1-1}}^{t-1:t} = \frac{\sum_{h=1}^2 \sum_{i=1}^{n_h} w_h p_{hi}^t}{\sum_{h=1}^2 \sum_{i=1}^{n_h} w_h p_{hi}^{t-1}}$
방법 3 (D_st2-1)	듀퐁	사후증화2-1	$P_{D_{st2-1}}^{t-1:t} = \frac{\sum_{h=1}^4 \sum_{i=1}^{n_h} w_h p_{hi}^t}{\sum_{h=1}^4 \sum_{i=1}^{n_h} w_h p_{hi}^{t-1}}$
방법 4 (Jevons)	제본스	없음	$P_{Jevons}^{t-1:t} = \left(\prod_{i=1}^n \frac{p_i^t}{p_i^{t-1}} \right)^{1/n}$
방법 5 (J_st1-1)	제본스	사후증화1-1	$P_{J_{st1-1}}^{t-1:t} = \sum_{h=1}^2 w_h \left(\prod_{i=1}^{n_h} \frac{p_{hi}^t}{p_{hi}^{t-1}} \right)^{1/n_h}, \quad \sum_{h=1}^2 w_h = 1$
방법 6 (J_st2-1)	제본스	사후증화2-1	$P_{J_{st2-1}}^{t-1:t} = \sum_{h=1}^4 w_h \left(\prod_{i=1}^{n_h} \frac{p_{hi}^t}{p_{hi}^{t-1}} \right)^{1/n_h}, \quad \sum_{h=1}^4 w_h = 1$
방법 7 (Carli)	칼리	없음	$P_{Carli}^{t-1:t} = \frac{1}{n} \sum_{i=1}^n \frac{p_i^t}{p_i^{t-1}}$



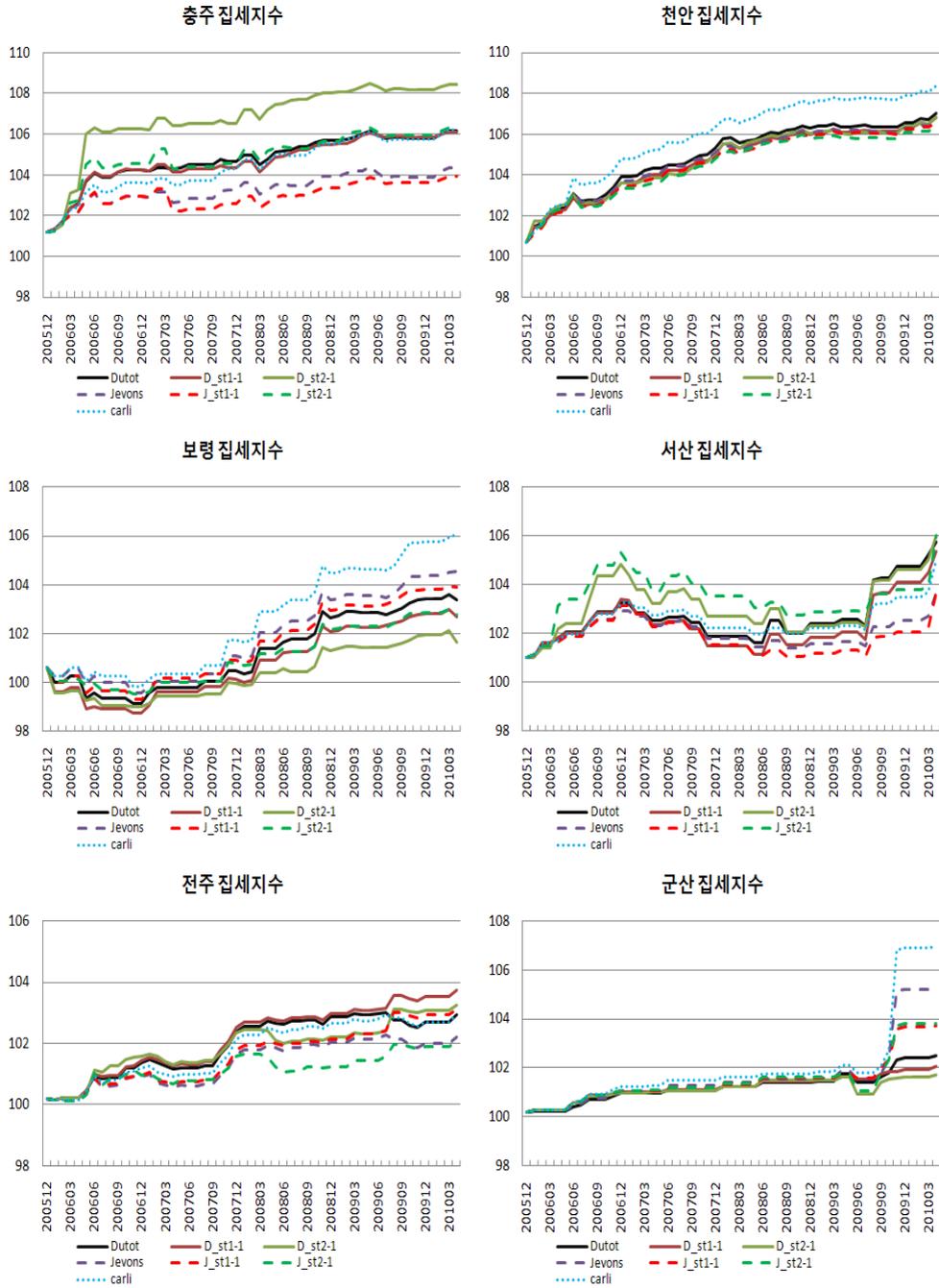
[그림 3-7] 작성방법별 모의지수



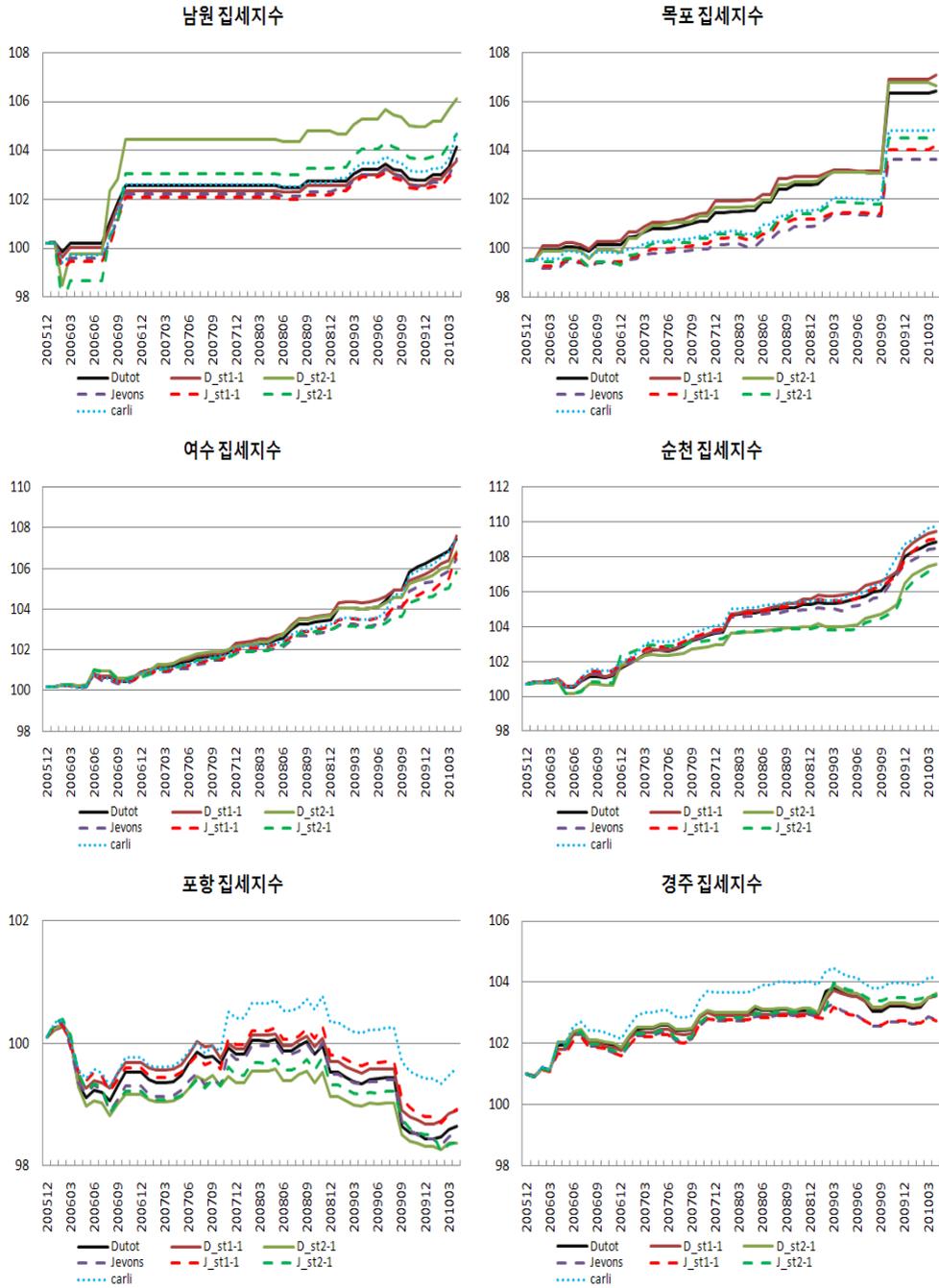
[그림 3-7] 작성방법별 모의지수(계속)



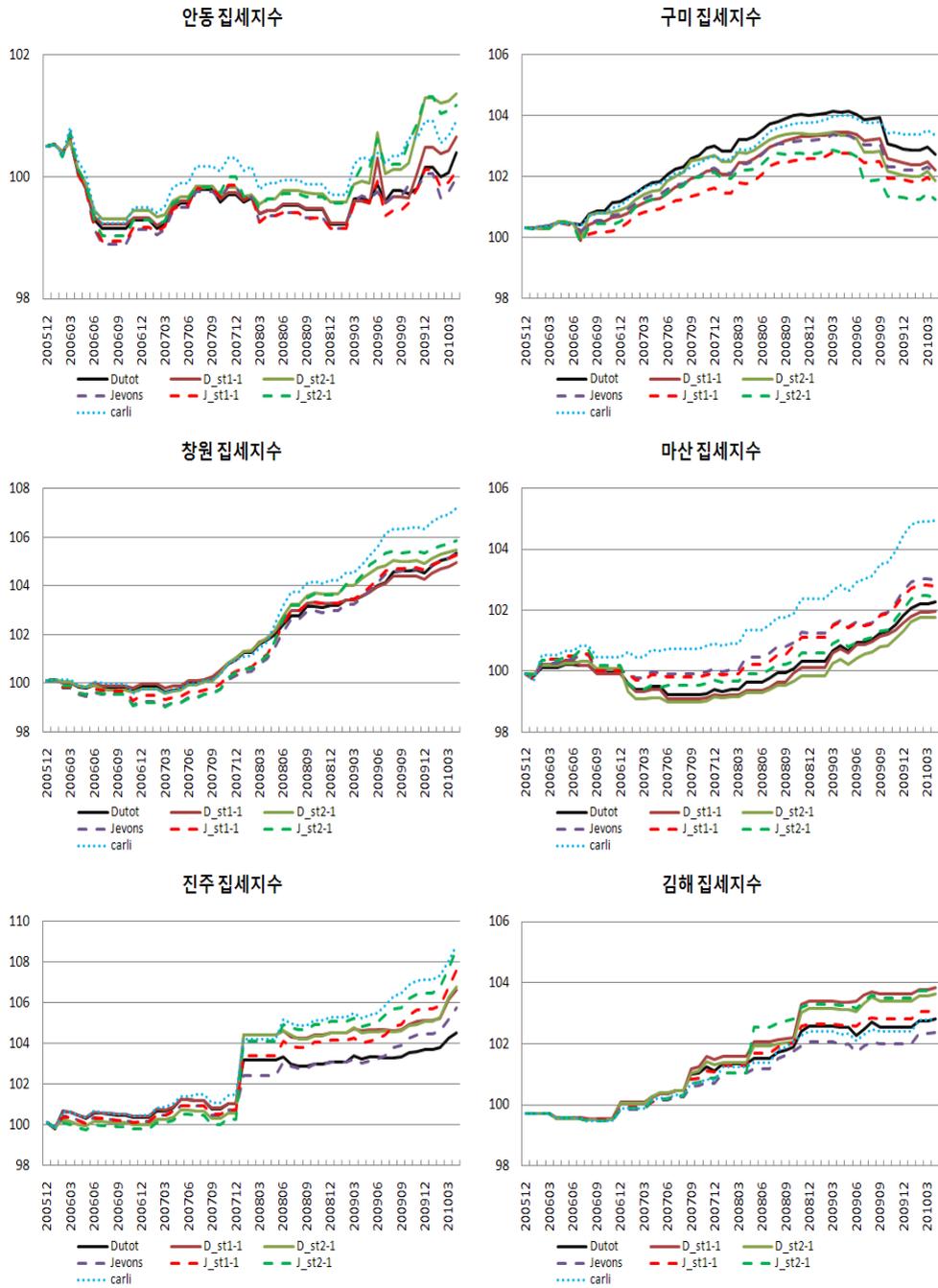
[그림 3-7] 작성방법별 모의지수(계속)



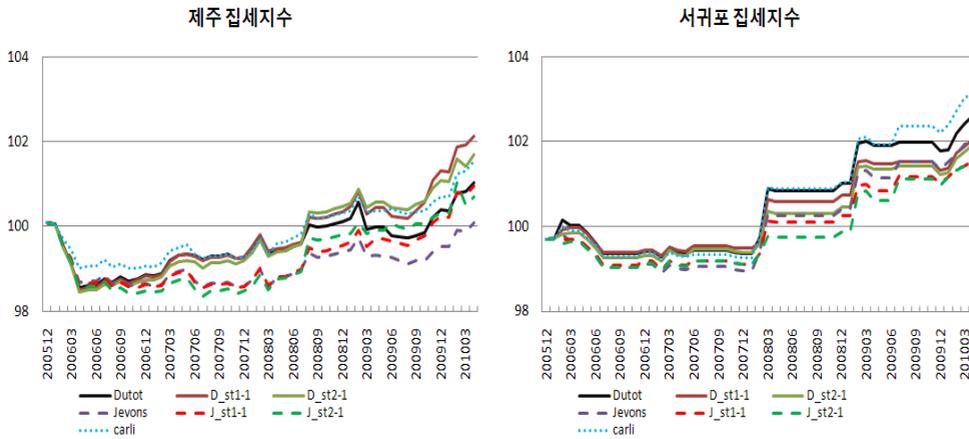
[그림 3-7] 작성방법별 모의지수(계속)



[그림 3-7] 작성방법별 모의지수(계속)



[그림 3-7] 작성방법별 모의지수(계속)



[그림 3-7] 작성방법별 모의지수(계속)

[그림 3-7]은 앞에서 설명한 7가지 방법에 의해 시험적으로 작성된 38개 도시별 집세지수결과이다. 먼저 서울을 비롯한 6대 광역시에서는 칼리방법에 의한 지수는 다른 방법에 비해 상향된 방향성을 보이고 있다. 칼리방법을 연쇄적으로 사용하였을 때 상향편향을 갖게 된다는 이론을 실증적으로 확인할 수 있는 부분이다. 서울의 경우 주택유형 층(층화1-1)이나 주택유형과 사용 방 수를 동시에 고려한 층(층화2-1)의 모집단 분포와 표본분포 간의 차이가 크지 않기 때문에 사후가중치의 적용 효과가 미미하게 나타나고 있다. 또한 표본 수도 약 2,000가구로 많기 때문에 듀토틀방법에 의해 나타날 수 있는 가격대가 높은 주택의 영향이 일정 부분 상쇄될 수 있는 여지가 있는 것으로 보인다. 다음으로 광주에서는 기존의 방식에 비해 나머지 6개의 방법들의 지수가 조금 더 높게 나타나며, 대구와 인천은 듀토틀방법에 의한 지수에 비해 제본스방법에 의한 지수들이 다소 높게 나타나고 있다. 이처럼 서울과 같이 모집단에 대한 대표성이 있는 표본을 확보하고 있는 대도시의 경우에는 사후가중치나 지수산식에 의한 차이가 기존의 방식과는 크게 차이가 나타나지 않았다.

그러나 7개 특·광역시와는 다르게 일반 도시에서는 사후가중치와 작성방법에 의한 지수의 차이가 크게 나타나고 있다. 몇몇 도시의 예를 살펴보면, 먼저 수원시는 집세조사 표본의 주택유형 구조가 모집단과 크게 다르지 않지만 사용 방 수에서는 차이가 나는 지역이다. 사용 방 수를 고려한 층화2-1의 사후가중치를 적용했을 때의 지수가 기존 지수에 비해 매우 낮게 나타나며, 층화1-1은 기존의 지수에 비해 다소 높아진 경향을 확인할 수 있었다. 또한 사후가중치를 적용하지 않았을 경우나 층화1-1 가중치를 적용하였을 때의 제본스방법이 듀토틀방법보다 다소 지수가 커짐을 알

수 있다.

다음으로 부천의 집세지수는 2006년 6월에 들어 층화2-1에 의한 사후가중치를 적용했을 때와 층화1-1에 의해 적용했을 때의 지수에 차이가 발생한다. 이 시기의 자료의 특성을 확인해본 결과, 일반주택이면서 사용 방 수가 3개 이상인 층의 사후가중치가 3의 값을 가지고 있었다. 이는 해당 층에 표본 수가 적어 사후가중치에 극단치가 발생되었기 때문에 사후가중치가 임계치에 의해 제한되었다. 또한 해당 층에 있는 어느 한 가구가 전 시세로 오랫동안 거주하고 있다가 전출을 가고 전입되어 온 가구는 현 시세를 적용받아 해당 가구의 전세가가 2.29배 올랐던 것이다. 즉, 가중치를 크게 받고 있는 사후층 내에 있던 한 가구의 집세가 크게 변동했던 것이다. 이와 같이 어느 특정한 한 가구의 특수한 상황에 의해 도시의 지수가 크게 변동된다면, 이는 안정적인 지수라고 할 수 없을 것이다. 반면 주택유형에 의해서만 층을 구분한 층화1-1에서는 사후가중치에 극단치 발생이 덜하게 되어 이러한 가구의 영향이 지수에 상대적으로 덜 반영될 수 있는 것으로 볼 수 있다.

그리고 목포의 경우에는 2009년 9월에 지수가 큰 폭으로 상승하였음을 확인할 수 있다. 이때의 상승폭을 각 방법별로 비교해보면, 방법1인 기존의 방식은 전월에 비해 3.3p 상승하고, 층화1-1과 층화2-1을 듀툃방법에 적용한 방법 2와 3은 각각 3.7p가 상승하였다. 이에 비해 제본스 산식을 적용한 방법 3은 전월에 비해 지수가 2.3p, 방법 4, 5는 각각 2.6p, 2.7p씩 상승하였다. 즉, 듀툃방법의 상승폭이 제본스방법에 비해 약 1p정도 크게 나타나고 있음을 알 수가 있다. 이러한 차이의 원인을 조사된 데이터에서 찾아보았다. 아파트에 방 4개 이상인 층에서 다른 전세가구들에 비해 집세의 수준이 높은 가구가 재계약을 하면서 전세 보증금이 1.6배 올랐기 때문이었다. 즉, 가격이 높은 특정한 가구의 변동이 도시의 지수를 끌어올리는 데 큰 영향을 주었고, 그 영향력은 제본스방법보다 듀툃방법에서 더 크게 나타나고 있음을 확인할 수 있었다. 비싼 주택에 더 큰 내재적 가중치를 주게 되는 듀툃방법의 결정적인 단점이 표본규모가 크지 않은 도시에서 보다 심각하게 나타날 수 있다는 것이다.

이상으로 살펴보았던 도시별 상황을 정리해보면, 7개 특·광역시를 제외한 나머지 31개 도시에서는 층별 표본분포가 모집단과 차이가 많이 나타날수록 사후가중치에 대한 효과가 더 크게 나타나고 있음을 확인할 수 있다. 그러나 그 차이가 너무 커서 층별로 빈 셀이 나타나거나 매우 적은 표본이 배정되어 극단치가 발생하여 가중치 제한이 된 경우에는 사후가중치를 적용한다고 하더라도 그에 합당한 효과를 기대하기는 어렵다. 사후가중치란 모집단의 분포와 유사할 수 있도록 선정된 표본의 분포를 조정해주는 기능을 하는 것이다. 만일 선정된 표본의 대표성이 현저하게 떨어진다면 사후가중치를 적용할지라도 추정의 정도를 높이는 데에는 한계가 있다.



지수작성 방법이나 사후가중치를 적용했을 때의 모의지수 결과는 기존 도시 표본과 모집단의 구조차이에 매우 의존하게 된다. 7개 특·광역시와 같이 층화2-1에 의해 4개의 층으로 구분해도 표본의 크기가 충분하고, 사후가중치의 극단치 발생 우려가 적은 도시에서는 좀 더 세분화된 층인 사용 방 수와 주택유형에 의한 사후가중치를 적용하는 것이 타당하다. 그러나 나머지 31개 도시에서는 표본의 규모도 크지 않고, 층의 구분을 너무 세분할 경우 앞에서 살펴보았던 것과 같이 어떤 특수한 상황이 도시의 지수를 결정짓게 될까 우려된다. 따라서 이러한 소규모 도시에서는 주택유형에 의해서만 층을 구분한 사후가중치를 적용하는 것이 보다 바람직할 것이라 본다.

하위지수 작성방법 역시 앞에서 살펴보았던 각 방법의 이론에서 설명하고 있던 장·단점들이 여실히 드러나고 있음을 실증적으로 확인해볼 수 있었다. 칼리방법의 경우 연쇄적으로 지수를 작성할 때 나타날 수 있다는 상향된 편향을 발견하였다. 듀퍃방법은 집세가 비싼 가구의 변동에 민감하여 집세수준이 크게 차이가 나는 도시에서는 부적합한 것으로 보인다. 또한 제본스의 경우는 큰 폭의 변동에 민감하여 정책의 결정이나 급변하는 시장상황 등에 의해 급등락이 발생한 도시나 집세가 일시에 대폭 조정된 가구에 대해서는 좋지 않은 결과를 가져다준다. 어느 한 방법이 절대적으로 우월하다고는 말하기 어렵기 때문에 각각의 장·단점을 잘 이해하고 적용해야 하며, 이들 방법 모두 지수작성 시 이상치에 대한 주의 깊은 관찰과 제어가 필요하다.

3. 추가 지수 작성 검토

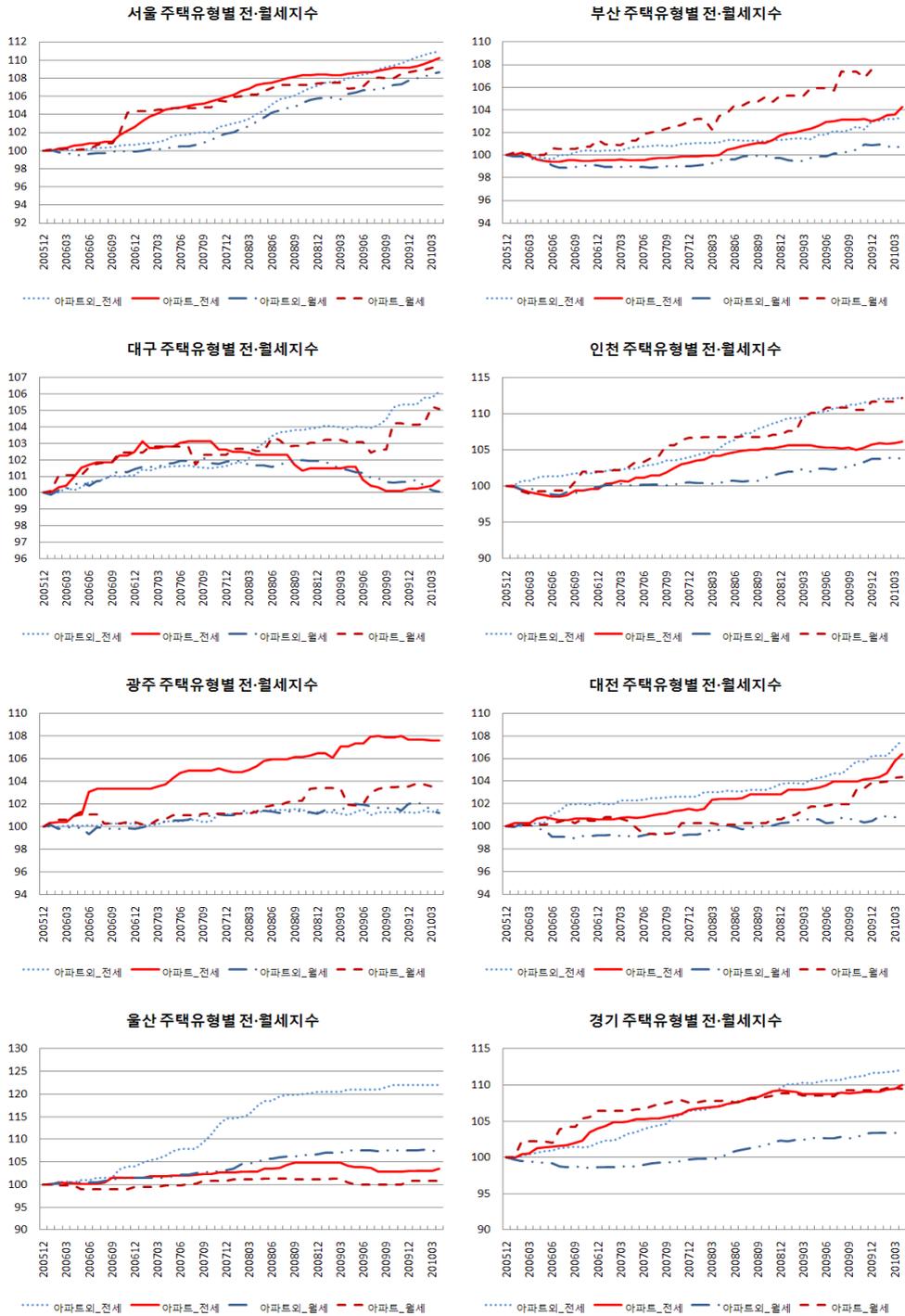
다음으로 현행 지수 작성범위를 좀 더 세분화하여 지수를 작성할 수 있는지를 검토하기 위해 16개 시도별 주택유형별 지수를 구해보고자 한다. 2005 인구주택총조사 전수조사 자료를 이용하여 시도별 각 도시의 점유형태별 주택유형 재고비율을 기준으로 가중치를 작성하였다. 특·광역시의 경우는 9개 시도와 달리 하위도시가 없기 때문에 특·광역시 내 점유형태별 주택유형 재고비율을 기준으로 하여 가중치를 작성하였다.

[그림 3-8]에서 보는 바와 같이 서울의 아파트의 경우 전세와 월세 지수가 유사한 흐름을 보이고 있으며, 아파트 외 주택의 전·월세 지수는 아파트에 비해 낮게 나타나다가 2009년 하반기에 들어서 거의 아파트의 상승세를 따라 잡게 되는 모습을 확인할 수 있다. 대구에서는 아파트 전세와 아파트 외 월세 지수는 소폭 상승을 하다가 2008년 하반기부터 하향세를 보이는 반면, 아파트의 전세와 아파트 월세 지수는 지속적으로 상승하는 형태를 띠고 있다. 경기도의 경우에는 아파트 외 주택의 월세의 상승폭이 적은 반면,

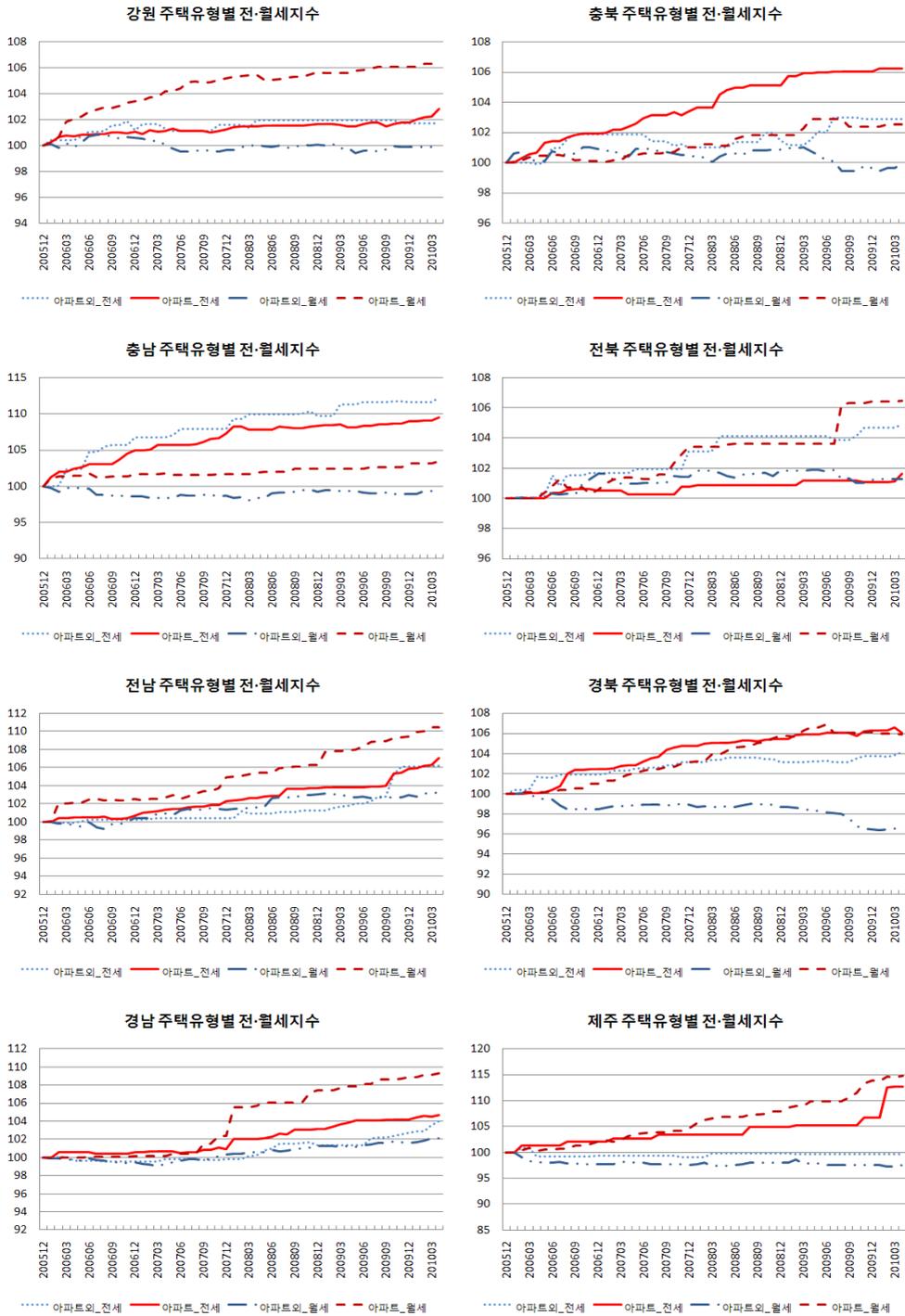
다른 유형의 지수들은 모두 크게 오르고 있다. 충남에서는 전세지수가 지속적으로 상승하는 곡선을 보이지만, 월세는 크게 변동이 없이 보합의 상태를 보이고 있음을 확인할 수 있다. 같은 점유형태 내에서도 주택유형에 의해 지수의 방향성과 변동폭이 차별적인 형태를 보이고 있다. 이와 같이 개별적인 주택유형에 대한 지수의 특성을 살펴볼 수 있게 된다면 집세지수가 보다 이용자에게 유용한 정보를 제공하게 되고 통계의 활용성 또한 증대할 것이다.

본 연구에서 시험적으로 작성해 본 주택유형별 지수의 타당성에 대해서는 통계적인 측면과 현실 반영적인 측면으로 면밀한 검토가 필요하다. 주택유형별 지수작성 시 기초자료로 필요한 집세 추정치에 대한 상대표준오차(CV)를 구해봄으로써 통계로서의 신뢰성을 간접적으로나마 파악하고자 하였다. [그림 3-9]에서 시도별 CV를 살펴보면, 서울의 경우 아파트 외의 주택에 대한 CV는 2% 수준으로 가장 낮으며 아파트 전세의 CV는 3%, 아파트 월세는 7~10% 수준을 보이고 있다. 6개 광역시에서도 아파트 월세의 CV가 약 10% 수준으로 높게 나타나고, 나머지는 5% 수준이거나 그보다 낮게 나타나고 있다. 그리고 9개 시도에서는 대체적으로 CV가 10%보다는 낮은 수준을 유지하고 있다. 경기도, 전라북도, 제주도와 같이 CV가 20% 전후로 대폭 크게 나타난 일부 기간에 대해서는 이상치가 존재하기 때문인 것으로 보이는데, 지수작성에 앞서 관측된 자료의 이상치에 대한 철저한 내용 검토가 요구된다.

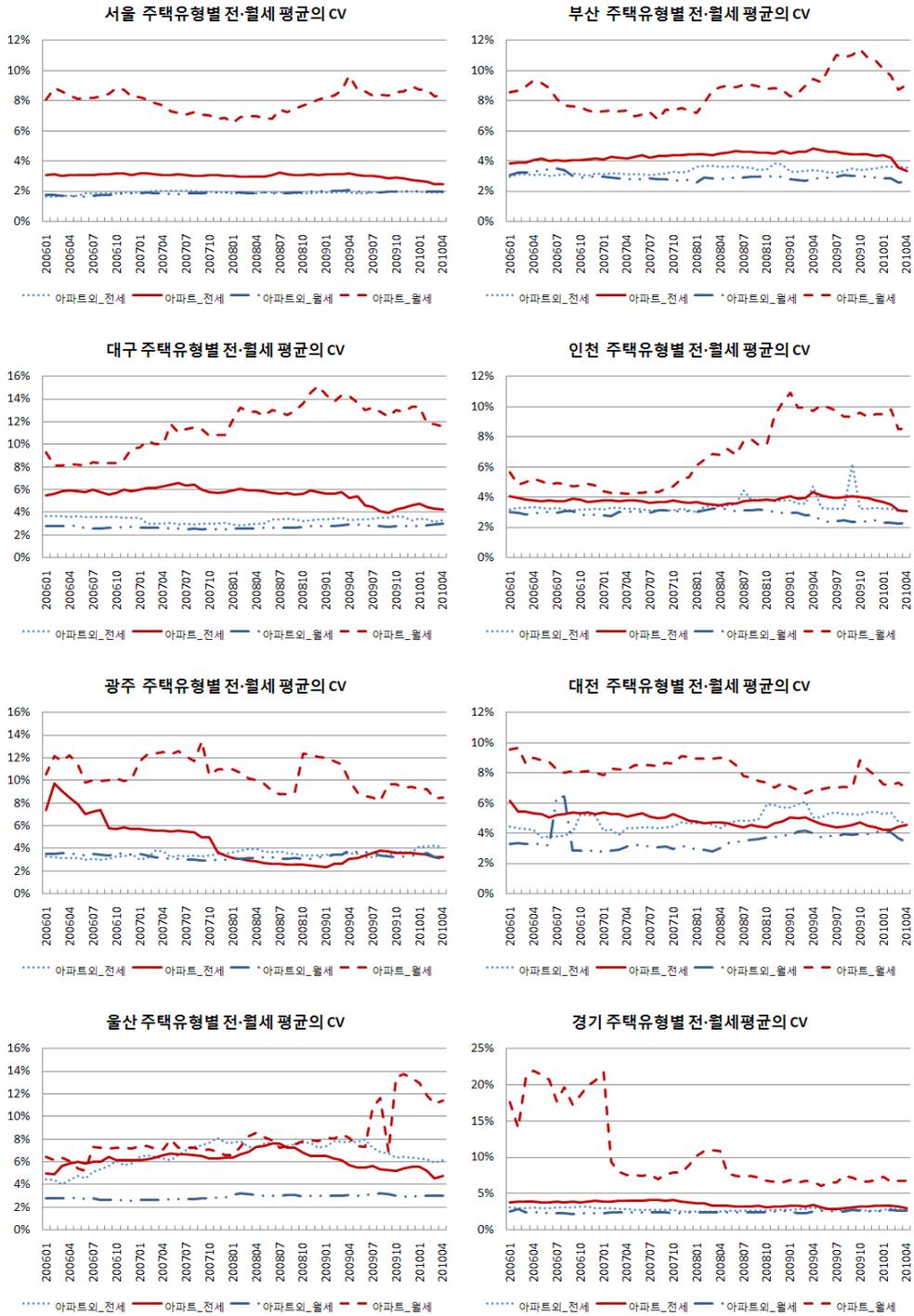
소비자물가에서는 집세의 평균 추정치를 전월 추정치와 비교한 변동에 관심이 있기 때문에 일반적인 통계청의 가구 표본조사에서 지역별로 기대하는 CV의 수준보다는 목표기준을 다소 높게 잡아도 무방하다. 만일 목표 CV를 10%로 설정한다면 16개 시도별 주택유형 전·월세 추정치는 허용 가능한 범위에서 추정되고 있다고 볼 수 있다. 그러나 이는 집세 추정에 근거한 간접적인 평가일 뿐이며, 변동에 대한 평가는 조금 더 다른 측면을 고려할 필요가 있다. 주택유형별 지수는 추정치에 대한 평가뿐만 아니라 각 지역 사정을 잘 아는 전문가의 의견을 바탕으로 현실 반영도에 대한 타당성을 평가 받은 후 정책적 의지에 의해 공표가 가능할 것이라 본다.



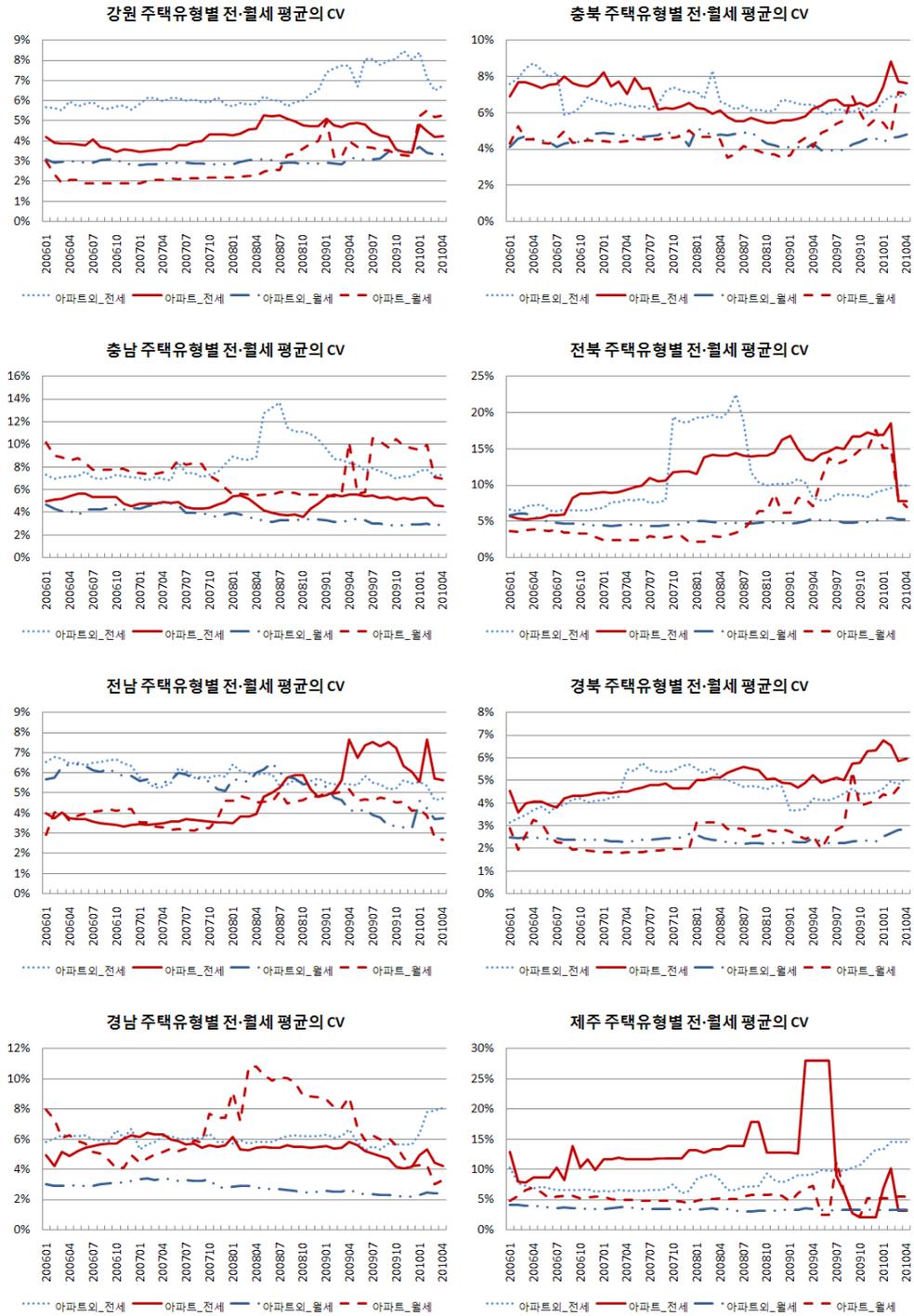
[그림 3-8] 시도별 주택유형별 전·월세 지수



[그림 3-8] 시도별 주택유형별 전·월세 지수(계속)



[그림 3-9] 시도별 주택유형별 전·월세 평균 추정의 CV



[그림 3-9] 시도별 주택유형별 전·월세 평균 추정 CV(계속)



제5절 결 론

본 연구는 2011년 소비자물가지수 개편에 맞춰 현행 집세지수 작성방법에 대해 검토하고자 하였다. 집세는 규격이 다양하고, 표본조사에 의해 자료를 수집하는 등 소비자물가 내 다른 품목들과는 조사방법과 특성이 매우 상이하다. 또한 집세표본이 다목적표본을 이용함에 따라 나타나는 모집단과 표본의 분포 차이를 조정해줄 수 있는 방법의 적용이 필요하다.

이를 위해 집세지수 작성방법 이론 및 실무적 내용을 검토하고 세 가지 주요 기본지수산식에 대한 이론적 장·단점 및 특징 검토하였다. 또한 집세지수 작성방법에 대한 UN·ILO의 권고사항과 해외사례를 파악하였다. 캐나다의 집세지수가 우리와 유사한 표본체계와 추정방법을 적용하고 있었는데, 현재는 설계가중치를 적용하지 않은 듀트방법을 사용하고 있다. 그러나 최근 발표된 연구결과에 의하면 설계가중치나 사후가중치 등의 적용이 추정의 정도 제고를 위해 필요하다고 하였다. 또한 스위스에서는 전화번호부에서 표본가구를 단순임의추출함으로써 발생할 수 있는 모집단 구조와의 차이를 조정하고자 사후가중치를 적용한 제본스방법을 사용하고 있었다.

사후층화를 통한 사후가중치의 시험작성을 위해 임차가구에 대한 모집단 자료로는 2005년 인구주택총조사 전수결과를 활용하였다. 그리고 집세조사 자료는 2006년 1월부터 2010년 4월까지의 52개월 자료를 이용하였다. 이러한 자료에서 즉시 활용이 가능한 공통항목인 주택유형과 사용 방 수를 사후층화 변수로 선택하였다. 아파트에 대해서는 전용면적과 건축년도가 최근 들어 조사항목으로 추가되었기 때문에 향후에는 이들 항목에 대해서도 사후층 구성방안을 검토해볼 수 있을 것이다. 사후층화 변수 중 주택유형으로 층화1방법을 구성하였는데, 단독주택, 아파트, 연립·다세대·기타주택과 같이 3개의 층으로 구분하였다. 다음으로 주택유형과 사용 방 수를 동시에 고려하여, 방수는 2개 이하, 3개, 4개 이상으로 구분하고 주택유형은 층화1방법과 동일하게 구분하여 총 9개의 층 조합으로 층화2방법을 구성해보았다.

그러나 층이 너무 세분화되어 표본가구의 수가 적게 배정되는 층이 발생하였고, 이에 따라 층을 다시 통합하였다. 층화1방법은 연립·다세대·기타주택을 단독주택과 통합함으로써 일반주택(아파트 이외)과 아파트의 두 개의 그룹을 만들어 층화1-1방법으로 구성하였다. 또한 층화2방법은 주택유형을 일반주택과 아파트로 구분하고, 사용 방 수는 일반주택인 경우에는 2개 이하와 3개 이상으로, 아파트는 방 3개 이하와 4개 이상으로 구분하여 총 4개의 층인 층화2-1방법을 구성하였다. 이와 같이 설정된 층 구분에 의해 사후층을 구성하고, 사후가중치를 작성하였다. 사후가중치의 크기가 0.3보다 작거나 3.0보다 큰 경우를 실무에서는 일반적으로 극단가중치로 정의한다. 이러한 극단가중치는

사후층의 수가 많을수록 더 많이 발생하는 결과를 확인하였다. 만일 극단가중치의 빈도가 매우 잦거나 그 크기가 매우 크거나 작은 값을 가지면 추정치의 분산이 커지게 되고 결과적으로 통계의 신뢰성을 저하시킨다. 이에 따라 작성되는 사후가중치에 임계치를 두고 제한하는 방법을 사용하였다. 사후가중치가 0.3보다 작게 되면 0.3으로, 3.0보다 크게 되면 3.0으로 값에 제한을 두어 극단가중치의 발생을 억제할 수 있도록 하였다. 그 결과로 가중치 제한 전의 사후가중치들의 분산의 수준이 가중치를 제한하고 나서는 거의 절반 가까이 줄어들게 됨을 알 수 있었고, 사후가중치의 크기 제한은 필요하다고 판단되었다. 이와 같이 가중치를 제한한 후의 사후가중치를 이용하여 집세의 평균을 추정하였다.

다음으로는 집세지수 작성방법을 검토하면서 하위지수를 작성하기 위해 일반적으로 많이 사용되는 세 가지 주요 방법에 대해서 살펴보았다. 첫 번째 방법은 칼리방법으로 개별가격비의 산술평균으로 계산되는데, 이 방법을 연쇄적으로 이용하면 상향된 편향을 갖게 된다고 알려져 있다. 이러한 이유로 칼리방법은 국제사회에서는 이제는 더 이상 권고되지 않는 방법이다. 두 번째 방법은 듀트방법으로 산술평균의 비의 방식으로 계산된다. 이 방법은 동질적인 상품 집합에 대해 의미가 있지만, 다양한 상품에 대해서는 지수의 임의성이 커지게 된다. 다시 말하면, 가격변동이 기준시점의 가격에 의해 내재적인 가중치를 갖게 되어 가격수준이 높은 제품의 가격변동이 낮은 제품에 비해 지수에 더 큰 영향을 주게 된다. 마지막으로 제본스방법은 개별가격비의 기하평균방법으로 구할 수 있으며, 기하평균의 비의 방식으로도 같은 결과를 얻을 수 있다. 따라서 제본스방법은 가격수준에 의해서는 영향을 받지 않게 되는 장점이 있다. 그러나 개별 가격비의 극단치에 매우 민감하다는 단점이 있어 극단적인 가격의 움직임은 신중하게 확인해야 한다.

이러한 세 가지 주요 산식에 사후가중치의 적용 여부에 따라 총 7가지 방법으로 집세지수를 시험작성해보고, 그 결과를 비교해보았다. 우선 서울을 비롯한 6대 광역시에서는 칼리방법에 의한 지수가 다른 방법들에 비하여 높게 작성되어 상향된 편향가능성을 보이고 있었다. 또한 모집단에 대한 구조와 표본의 구조가 크게 차이가 나지 않았던 특·광역시 위주의 대도시에서는 사후가중치나 지수산식에 의한 차이가 기존의 방식과 크게 차이가 나타나지 않는 것으로 나타났다. 그러나 이러한 7개 대도시와는 다르게 나머지 31개 도시에서는 층별 표본분포가 모집단과 차이가 많이 나타날수록 사후가중치에 대한 효과가 더 크게 나타났다. 그러나 그 차이가 극심하여 층별로 빈 셀이 나타나거나 극단가중치의 발생으로 가중치 제한이 적용된 도시에서는 사후가중치를 적용한다고 하더라도 그에 합당한 효과를 기대하기 어려웠다. 사후가중치는 체계적으로 선정된 표본이 모집단의 분포와 유사한 구조를 가질 수 있도록 일부 조정하여 추정의 정도를 높일



수 있도록 하는 보조적인 기능이 있는 것이지, 대표성이 현저히 떨어지는 표본구조를 근본적으로 해결할 수 있는 수단으로 활용될 수는 없다.

이와 같이 지수 작성방법이나 사후가중치를 적용했을 때의 모의지수 결과는 기존 도시표본과 모집단의 구조 차이에 매우 의존하게 됨을 알 수 있다. 7개 특·광역시와 같이 층화2-1에 의해 4개의 층으로 구분해도 표본의 크기가 충분하고, 사후가중치의 극단치 발생 우려가 적은 도시에서는 좀 더 세분화된 층인 사용 방 수와 주택유형에 의한 사후가중치를 적용하여도 충분할 것으로 보인다. 그러나 나머지 31개 도시에서는 표본의 규모도 크지 않아서 층의 구분을 너무 세분할 경우 앞에서 살펴보았던 것과 같은 특수한 상황에 기인하여 지수가 작성될 우려가 있다. 따라서 이러한 도시에서는 주택유형에 의해서만 층을 구분한 사후가중치를 적용하는 것이 보다 바람직하다고 판단된다.

하위지수 작성방법 역시 앞에서 살펴보았던 각 방법의 이론에서 설명하고 있던 특징들이 여실히 드러나고 있음을 실증적으로 확인해볼 수 있었다. 듀트방법의 경우 집세가 비싼 가구의 변동에 민감하여 집세수준이 크게 차이가 나는 도시에서는 부적합하다. 또한 제본스의 경우는 큰 폭의 변동에 민감하여 시장상황에 의해 급등락이 발생한 도시나 집세가 일시에 대폭조정된 가구에서는 불리하다. 다시 말하면, 어느 한 방법이 절대적으로 우월하다고는 말하기 어렵기 때문에 각각의 장·단점을 잘 이해하고 적용해야 한다. 듀트방법이 근본적으로 가질 수밖에 없는 집세가 비싼 가구의 변동에 민감한 반응을 보이는 지수식의 단점을 근본적으로 해결해줄 수 있는 제본스방법으로의 전환도 고려해볼 만하다. 그러나 지수작성 시 이 두 방법 모두 이상치에 대한 주의 깊은 관찰과 제어가 반드시 필요하다고 하겠다.

다음으로 현행 지수 이외에 좀 더 세분화된 지수가 작성될 수 있는지를 검토하기 위해 16개 시도별 주택유형별 지수를 구해보았다. 이때 각 도시의 점유형태 가구별 거주주택의 재고비율을 기준으로 가중치를 작성하였다. 같은 점유형태 내에서도 주택유형에 의해 지수의 방향성과 변동폭이 차별적인 형태를 보이고 있다. 이와 같이 개별적인 주택유형에 대한 지수의 특성을 살펴볼 수 있게 된다면 집세지수가 보다 이용자에게 유용한 정보를 제공하게 되고 통계의 활용성 또한 증대할 것으로 보인다. 이러한 지수가 의미를 갖기 위해서는 추후 현실 반영적인 측면에 대해 각 지역사정을 잘 아는 전문가의 의견을 바탕으로 지역별 주택유형지수에 대한 객관적인 평가가 필요하다.

이와 같이 사후층을 구성하여 사후가중치를 적용하고 지수 작성방법을 비교하여 현행 방법들에 대한 개선안을 제시해보고자 하였다. 그러나 현재로서는 사후가중치를 적용하더라도 정도 높은 추정치를 제공하는 데는 다소 한계가 있다고 보이며, 근본적인 해결책은 아니라고 생각한다. 집세조사가 다목적표본을 이용하면서 나타나는 모집단과의 구조차이뿐만 아니라 다음과 같은 문제점도 제기해볼 필요가 있다. 예를 들어, 주어진

표본가구들 중에서 매월 일정수준 이상의 계약갱신이 발생하지 않는다면 그러한 표본에서 구한 지수에 우연성이 개입할 가능성이 매우 높다. 다시 말해, 도시 전체의 움직임과는 전혀 다르게 어느 특정한 가구의 특수한 상황이 도시의 집세변동을 결정짓게 될 수도 있다는 것이다. 소비자물가지수에서는 38개 도시에 대해 집세지수를 공표하고 있다. 그러나 지방 소도시는 수도권이나 대도시에 비해 전체 가구 중 임차가구의 비중이 크지 않아 표본규모가 약 200여 가구가 되지 않는 곳도 있다. 또한 우리나라의 전세제도는 통상적으로 2년마다 재계약을 하게 되는데, 200여 가구를 24개월로 나누게 되면 약 8가구가 매월 계약갱신이 일어날 수 있을 것이라고 추측해볼 수 있을 것이다. 이러한 8가구의 변동이 도시 전체를 설명한다고 가정해보면, 지수의 임의성이 없으리라는 것을 보장할 수 없는 상황이 된다. 또한 표본의 크기가 크지 않아 현재 이용 가능한 변수로도 층을 세분하여 구성할 수가 없으며, 개별 도시의 평균 전·월세 CV를 보면, 10%수준을 훨씬 넘어서는 도시들도 존재하고 있음을 확인해볼 수 있었다.

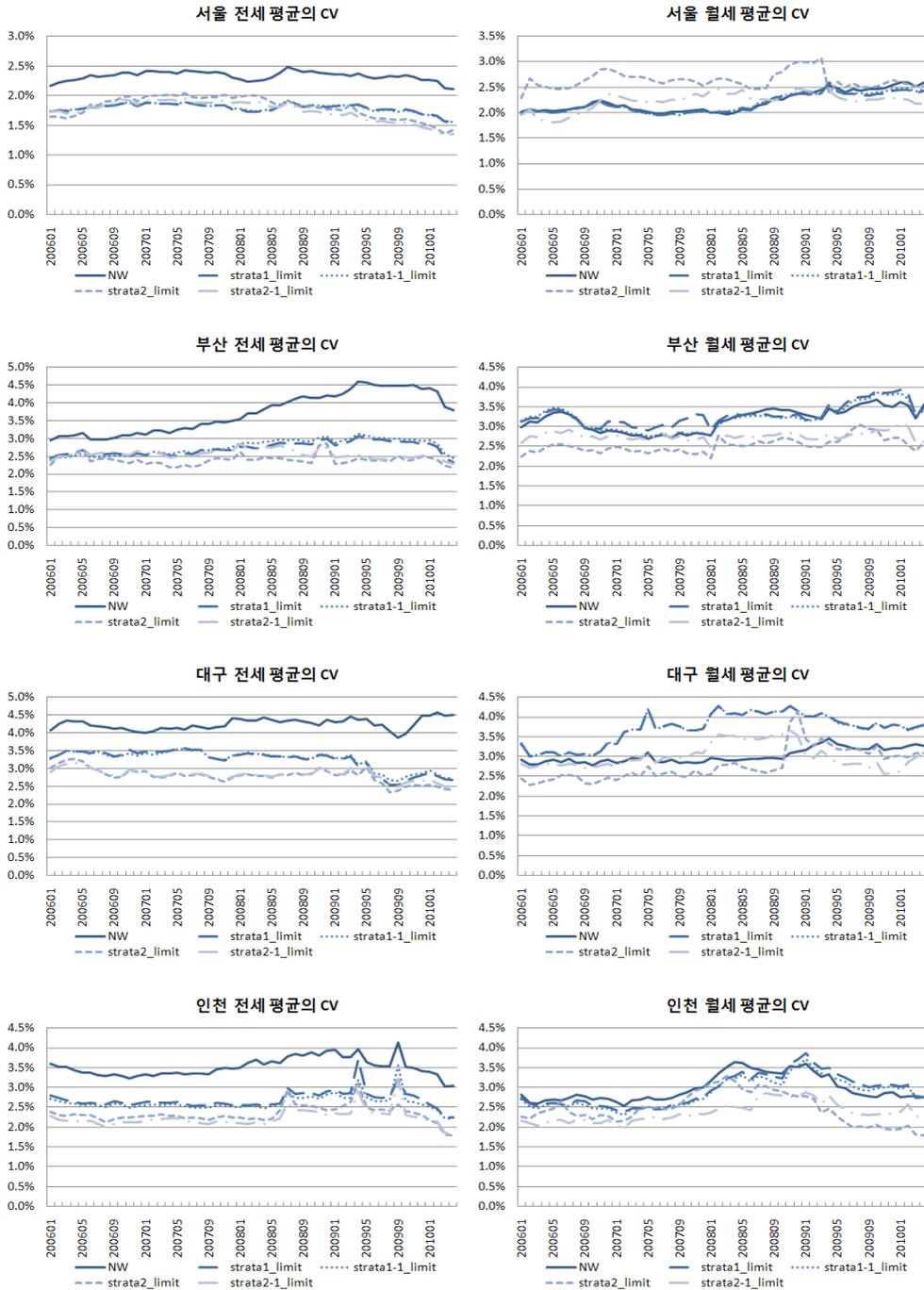
다목적표본에 의해서는 집세조사에서 기대하는 수준의 표본구성을 기대하기는 어렵기 때문에, 집세지수를 위한 별도의 표본을 도입하거나 표본의 크기를 늘려줌으로써 표본의 대표성을 확보할 수 있는 방안을 마련해야 한다. 그러나 표본을 추가적으로 더 확보하거나 독립표본을 설계하는 것이 예산과 인력의 문제로 인하여 실현되기가 매우 어려운 것이 현실이다. 이와 같이 제약된 여건에서 독립표본의 도입이 불가능하다면 차선책으로 공표범위에 대한 재검토를 제안하고자 한다. 38개 도시별로 일괄적으로 집세지수를 작성하여 제공하기보다는 수도권과 비수도권, 주택유형에 의한 혹은 임차주택의 규모 그룹에 의한 지수 등과 같은 다각적인 지표를 설정하여 보다 현실을 반영하고 이용자들에게 필요한 다양한 정보를 제공하는 것이 필요하지 않을까 기대한다.

참고문헌

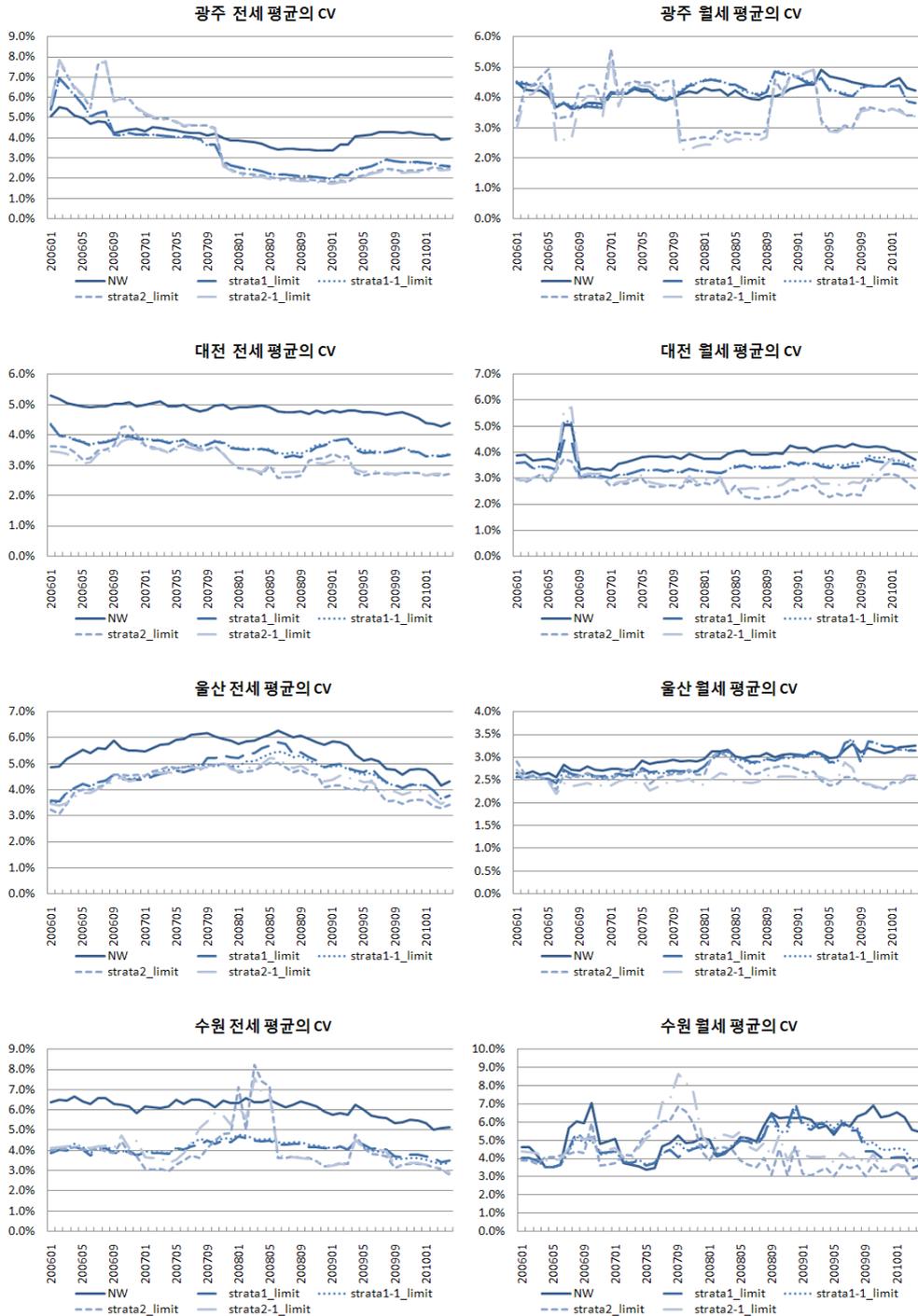
- 김영원, 류제복, 박진우, 홍기학(1998), 「표본조사의 이해와 활용」, 자유아카데미.
- 박현정(2010), “소비자물가지수 현실반영도 제고를 위한 지수산식 연구”, 「통계개발원 2009년 하반기 연구보고서 II」, pp.1-46, 통계개발원.
- 전백근(2006), “집세조사 방법론 - 추정법 중심으로 고찰”, 「2006년도 연구결과 모음집」, pp.275-321, 통계개발원.
- 정미옥(2009), “집세조사의 표본 구조분석”, 「국가 통계의 분류·수집·공표체계 품질향상을 위한 연구」, 통계개발원.
- 정미옥, 박소현(2009), “집세조사의 표본설계방안”, 「통계개발원 2009년 상반기 연구보고서 I」, pp.167-221, 통계개발원.
- 통계교육원(2008), 「지수론」.
- 통계청(2005), 2005 인구주택총조사 조사지침서.
- 통계청(2010), 2010 집세조사 지침서.
- 통계청 물가통계과(2007), 「2005년 기준 소비자물가지수 개편 보고서」.
- Bureau of Labor Statistics(2001), “Revision of the CPI housing sample and estimators”.
- C. B. Vermeulen, H. M. Herren(2009), “Rents in Switzerland: sampling and quality adjustment”, 11th Meeting-Ottawa Group.
- Japan Statistics Bureau(2006), “2005-Base Explanation of the Consumer Price Index”, <http://www.stat.go.jp/english/data/cpi/1586.htm>.
- Jay J. Kim, Jianzhu Li and Richard Valliant(2007), “Cell collapsing in the poststratification”, Survey Methodology, Vol 33. No.2. pp.139-150.
- Jeannine Claveau, Jack Lothian, Sylvie Gauthier(2009), “The Rent Component of The Consumer Price Index”, SSC Annual Meeting.
- United Nation(2004), 「CPI Manual - Theory and Practice」.
- United Nation(2009), 「Practical Guide - To producing Consumer Price Indices」, Chater 9. Special Cases - Housing.
- RRPI Handbook the 1st Draft(2010), Joint UNECE/ILO meeting on Consumer Price Indices, <http://www.unece.org/stats/documents/2010.05.cpi.htm>.



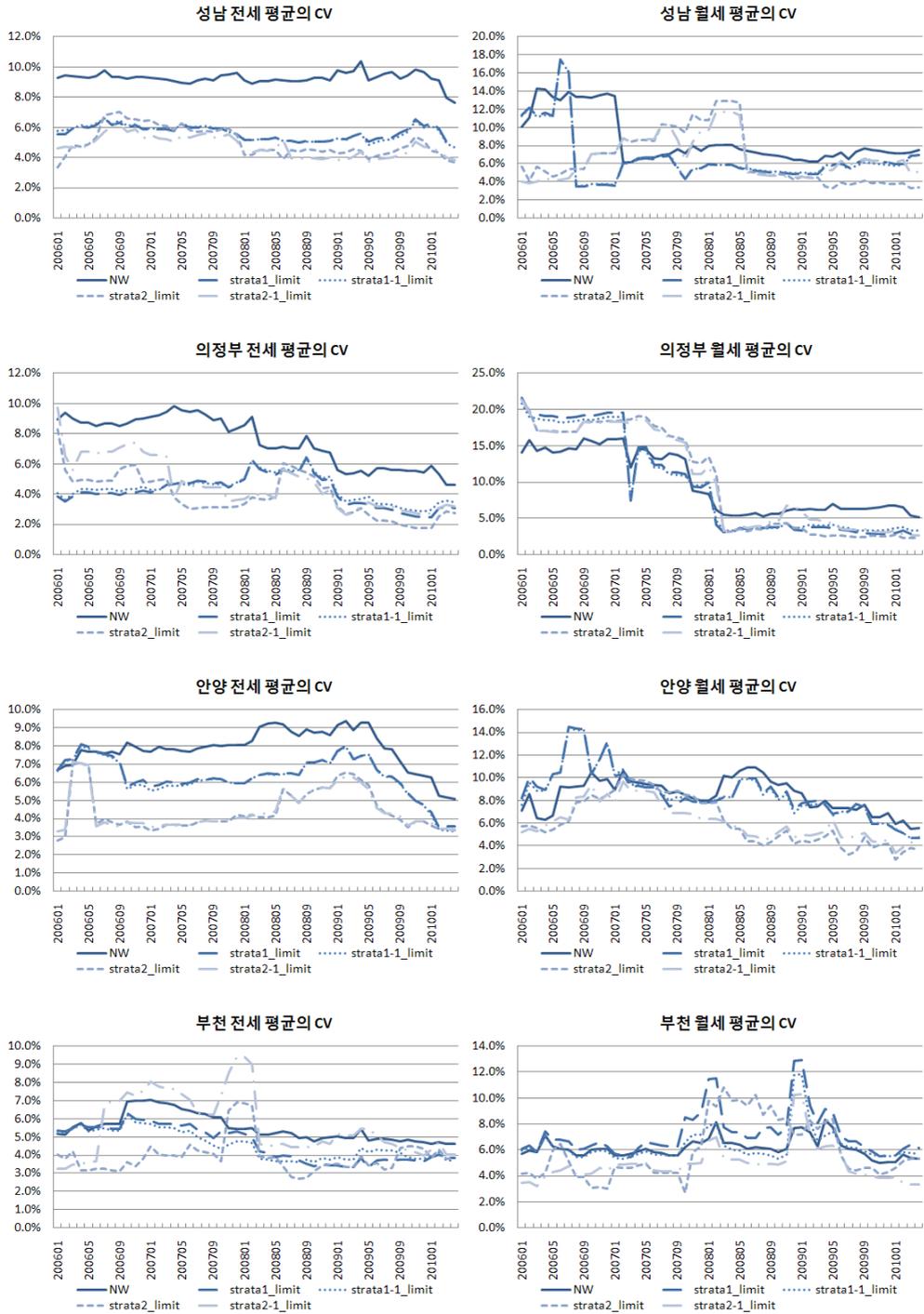
<부 록>



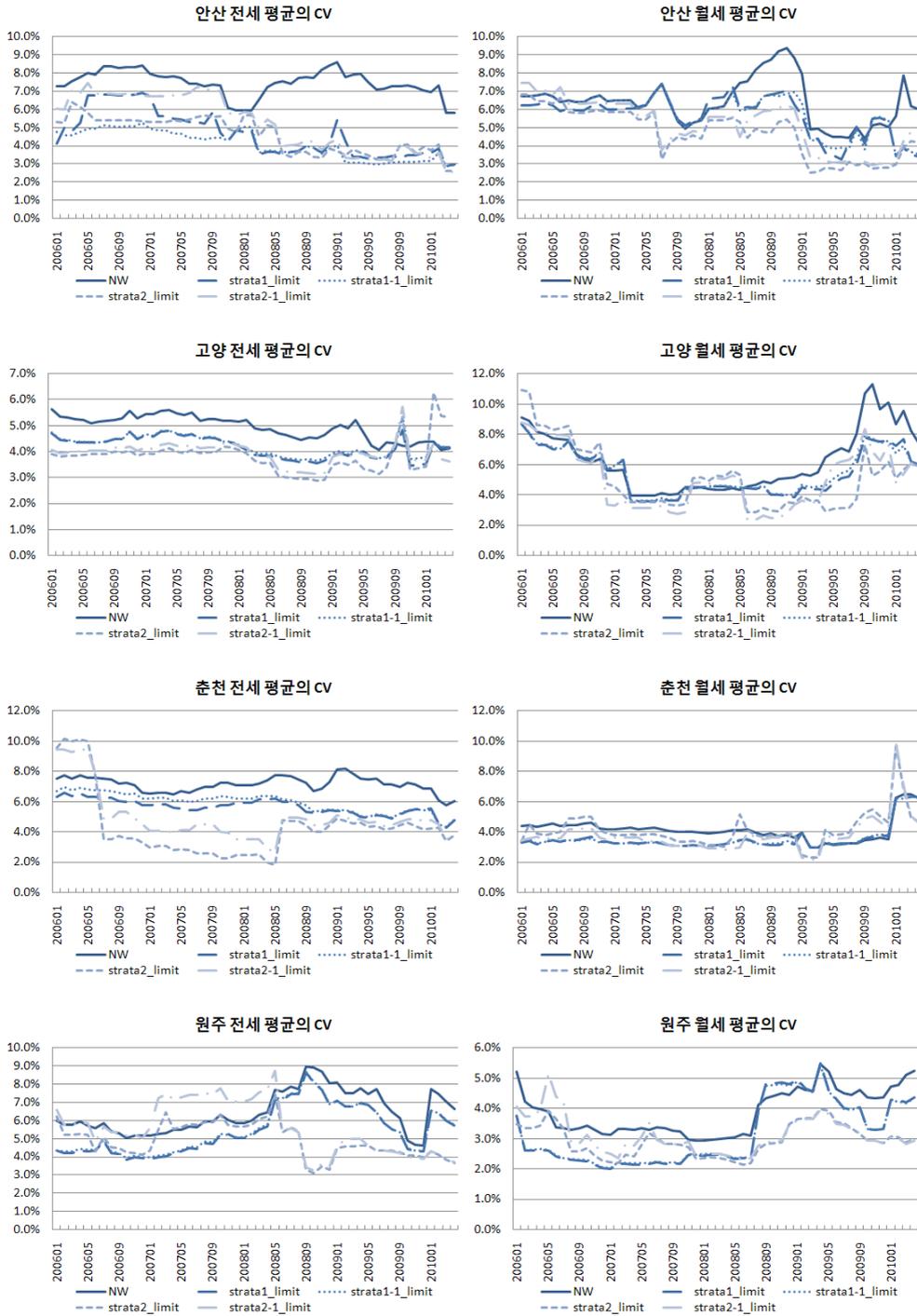
[부도 3-1] 도시별 전월세 평균 추정치의 CV



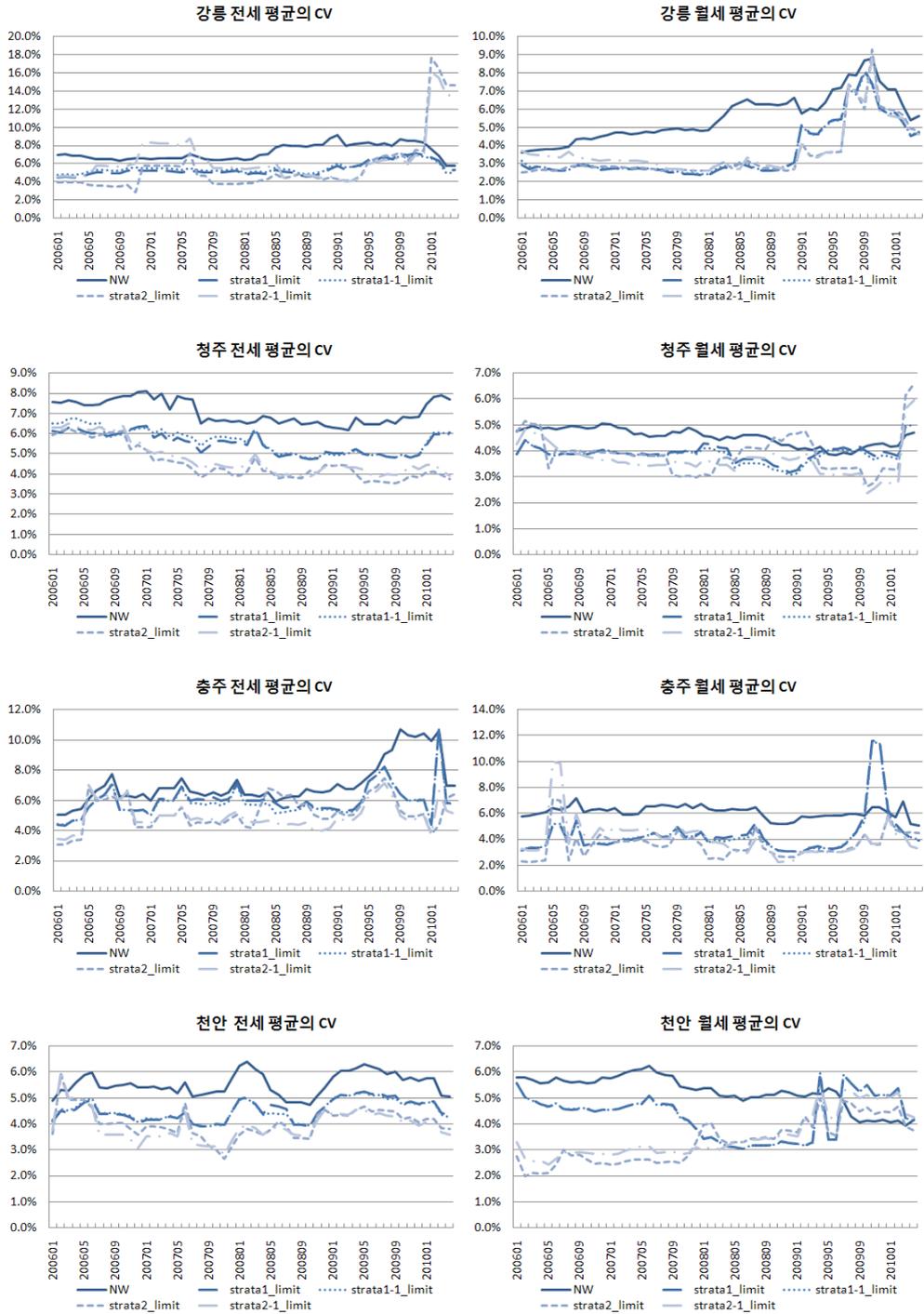
[부도 3-1] 도시별 전월세 평균 추정의 CV(계속)



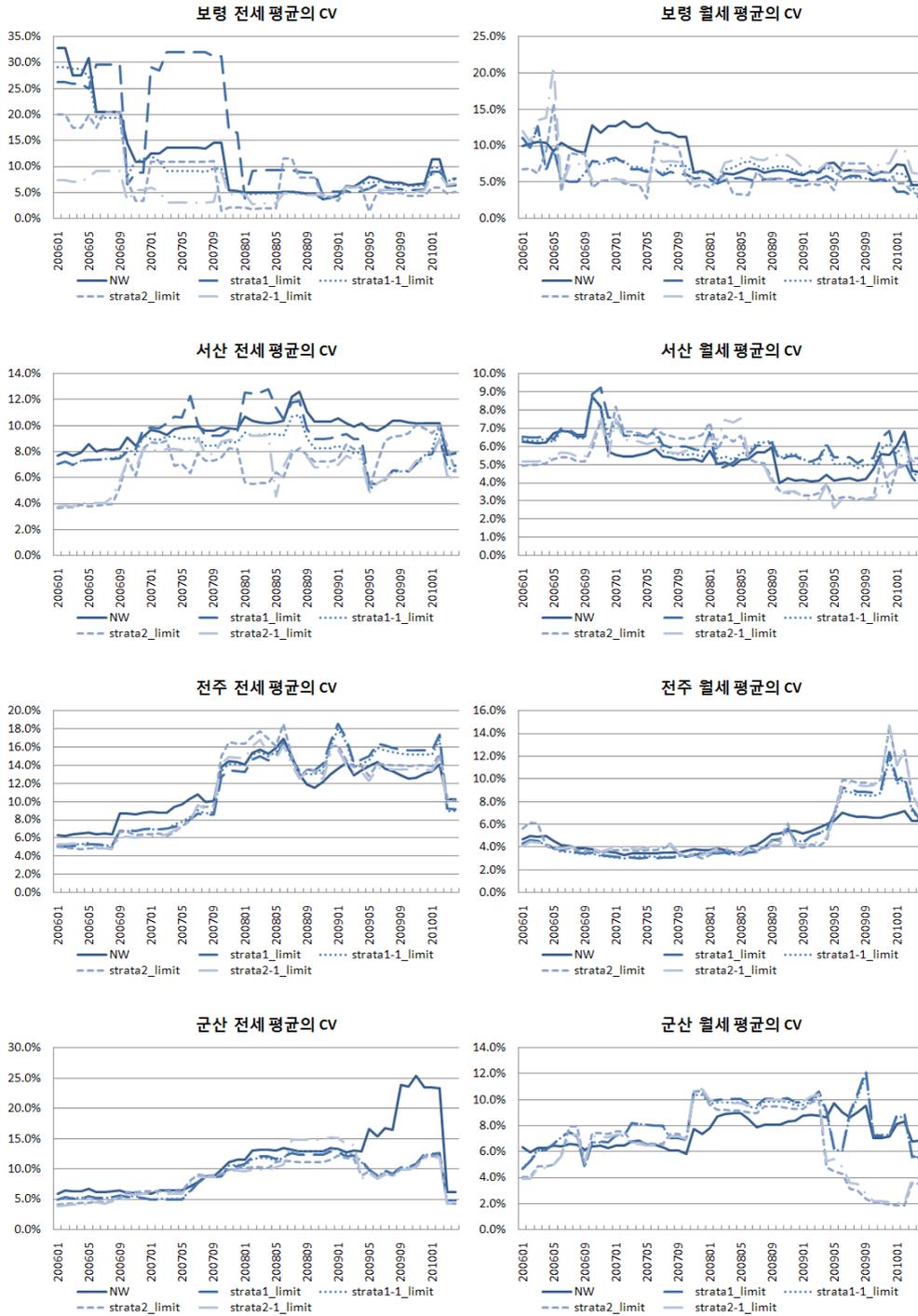
[부도 3-1] 도시별 전월세 평균 추정의 CV(계속)



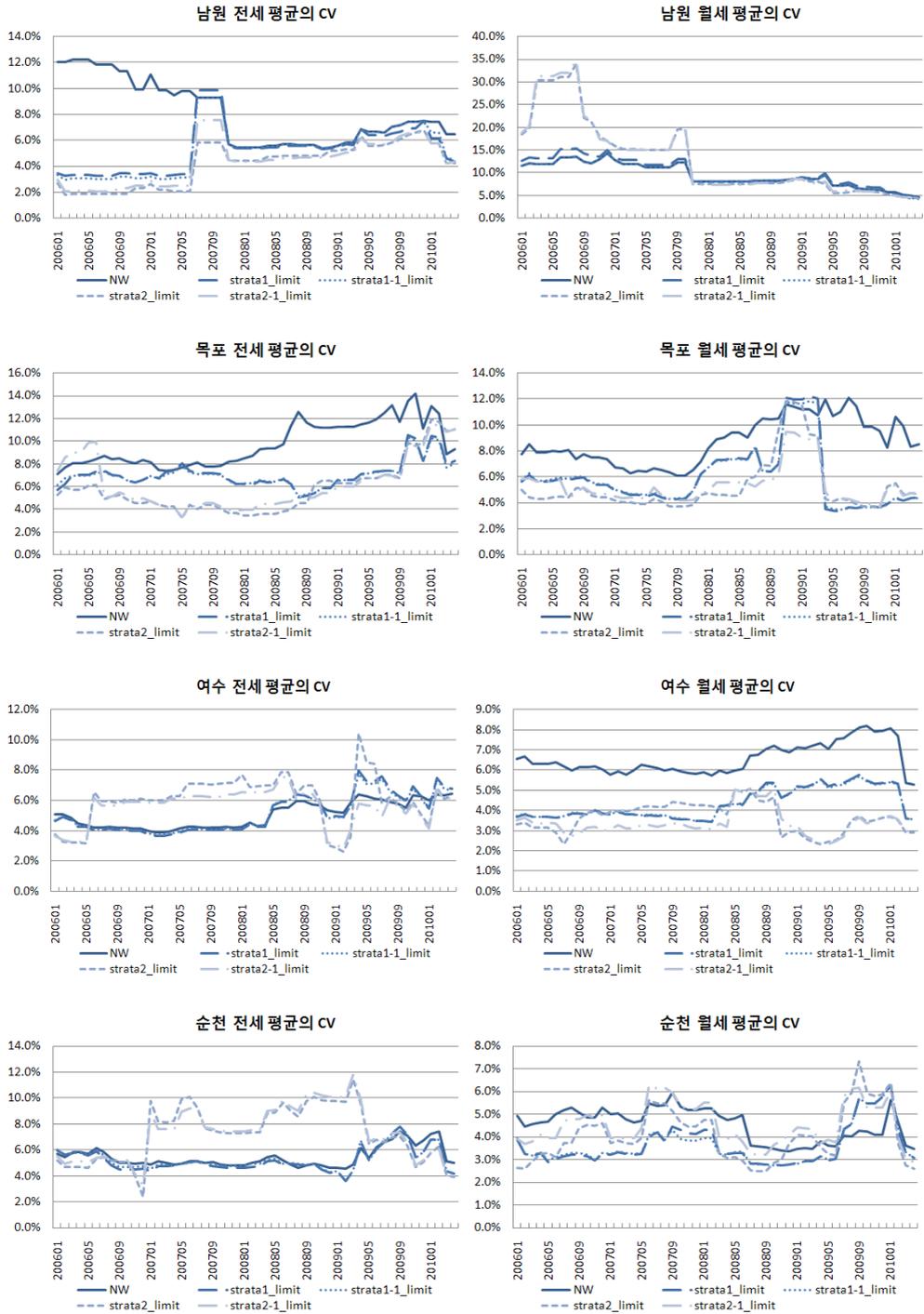
[부도 3-1] 도시별 전월세 평균 추정의 CV(계속)



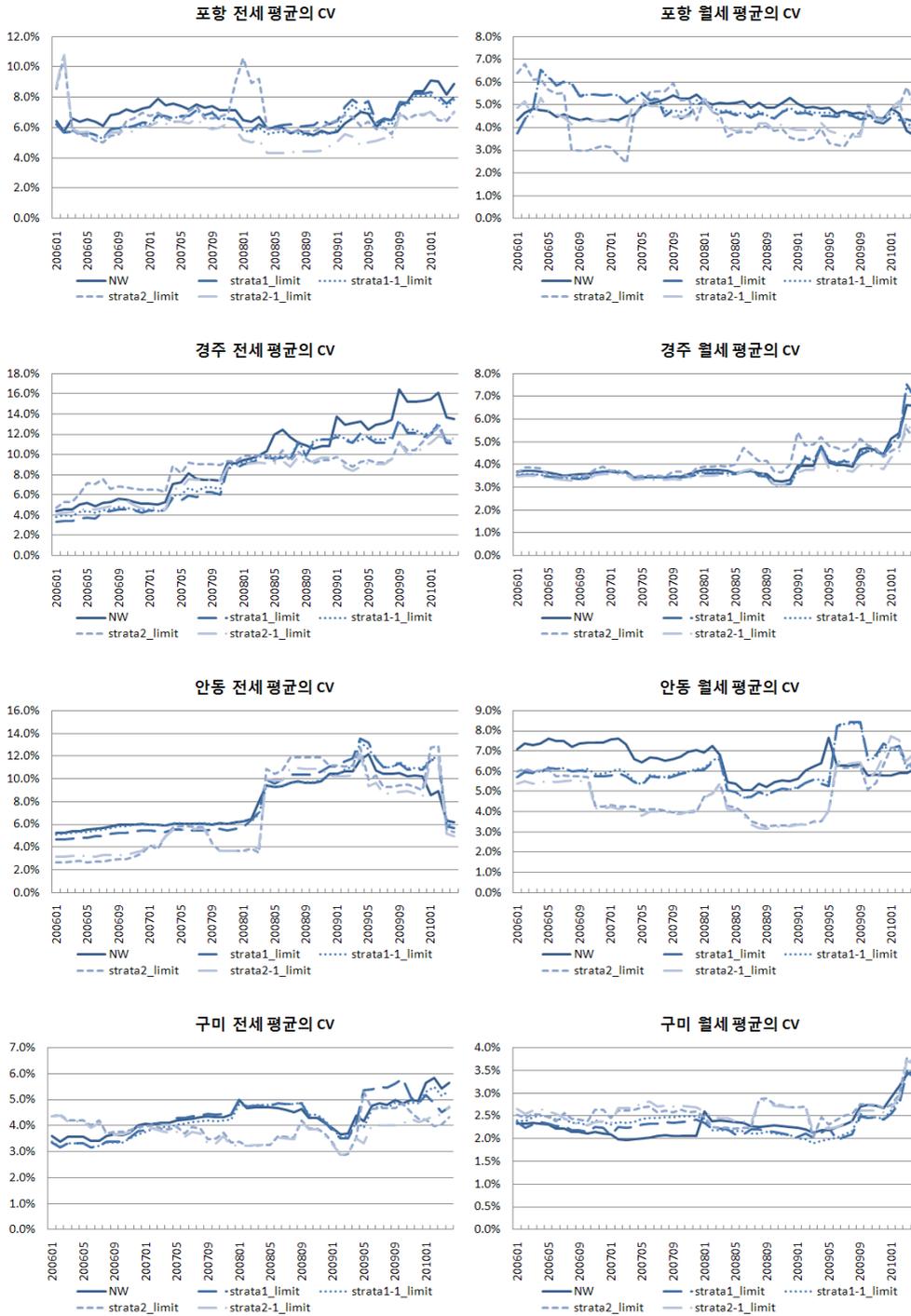
[부도 3-1] 도시별 전월세 평균 추정치의 CV(계속)



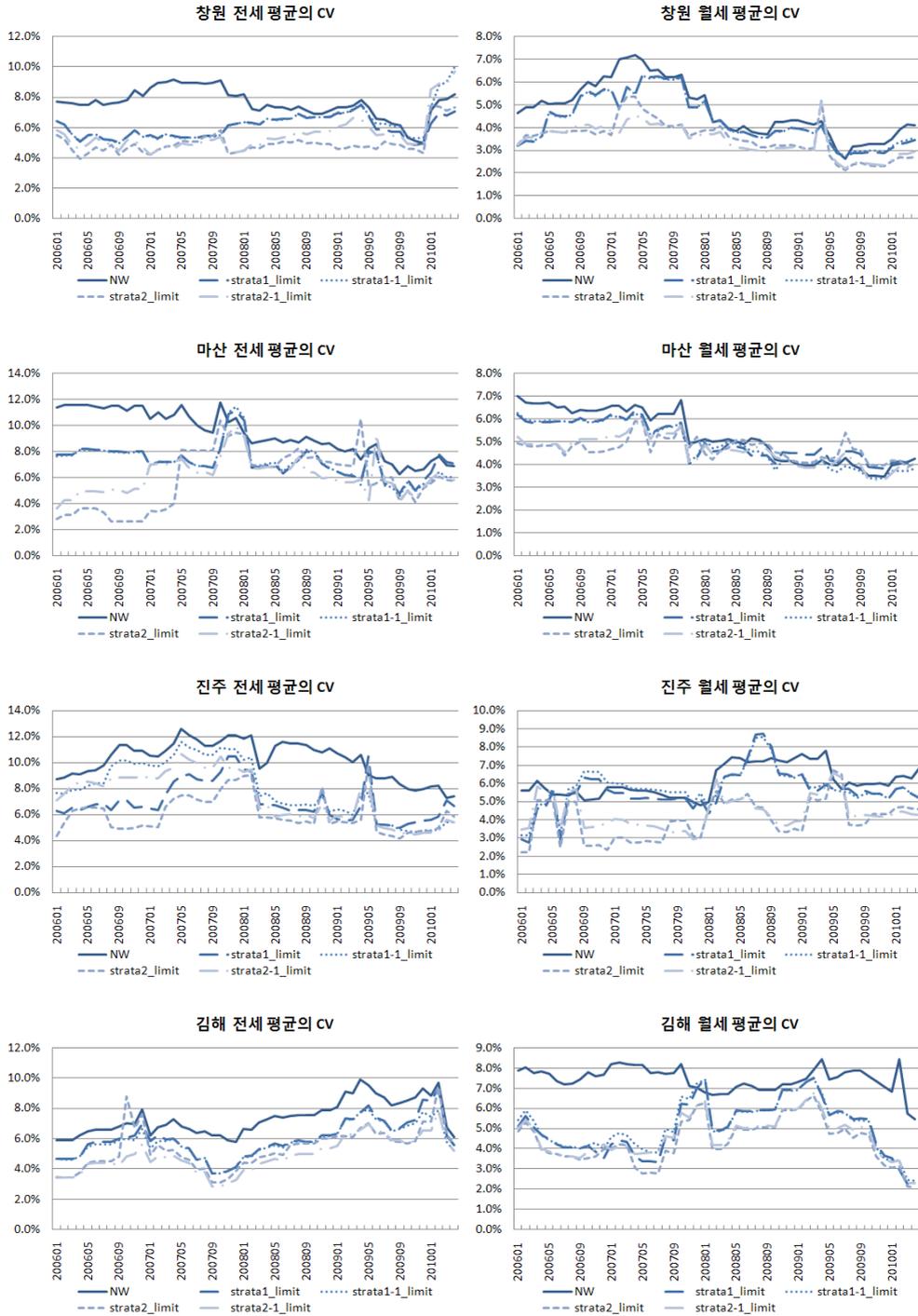
[부도 3-1] 도시별 전월세 평균 추정의 CV(계속)



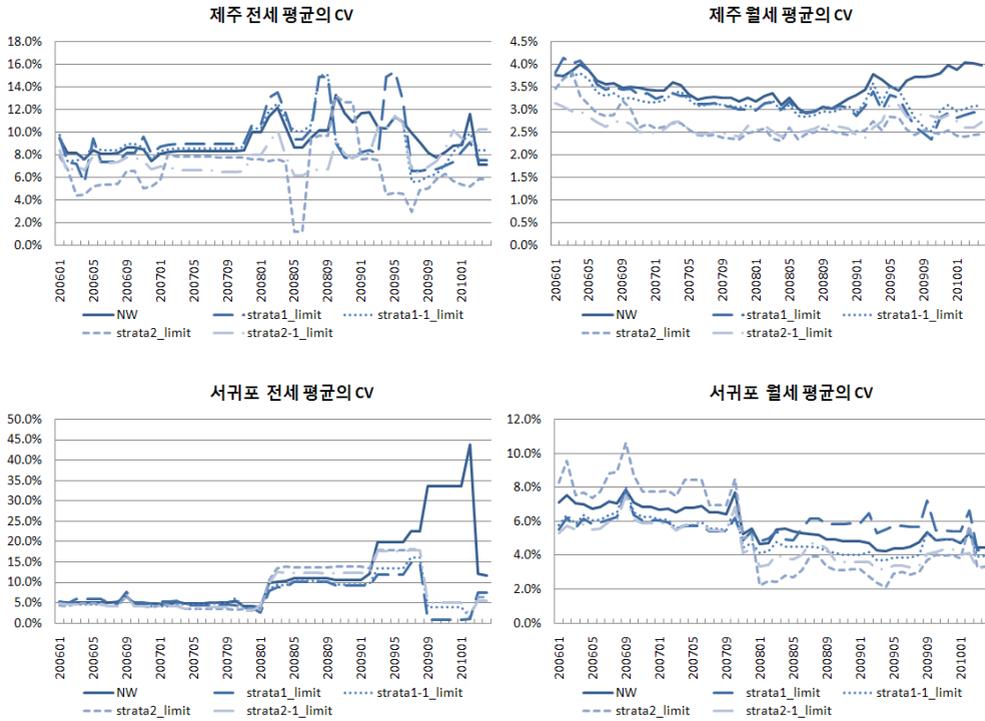
[부도 3-1] 도시별 전월세 평균 추정의 CV(계속)



[부도 3-1] 도시별 전월세 평균 추정 CV(계속)



[부도 3-1] 도시별 전월세 평균 추정치의 CV(계속)



[부도 3-1] 도시별 전·월세 평균 추정 CV(계속)