

정보통신기기 수출 예측을 위한 하향식(Top-down) 모형에 관한 연구 A top-down forecasting model for analyzing the export market of information and telecommunication products

지형구*, 주영진**, 조병선**, 김찬규*, 이영호*, 김영휘*
*고려대학교 산업공학과, **한국전자통신연구원

Abstract

이 연구는 정보통신기기 수출량에 관해 하향식(Top-down) 방법에 기초한 예측 모형을 제시한다. 하향식 방법은 전체 수출량과 전체를 구성하는 개별 항목간에 계층적 관계를 바탕으로 순차적으로 예측을 수행하는 방법이다. 전체와 개별 항목간에 관계는 데이터의 시계열 특성과 데이터에 영향을 주는 요인들에 의해서 만들어진다. 이러한 관계를 바탕으로 하는 하향식 예측은 전체 수출량을 먼저 예측한 후 이 예측치를 바탕으로 하여 개별 항목에 대한 예측을 수행한다. 하지만 하향식 방법은 가장 아래 계층의 예측치를 산출하기 위해 필요한 것이며 최종 예측치는 가장 마지막 계층에서부터 예측 데이터를 합산해서 얻을 수 있다. 결국 하향식 예측 방법은 전체와 개별 항목 사이에 상관관계가 높고 계층화되어 있는 구조에 적합하다. 이 예측 대상이 되는 정보통신기기 수출량에 대한 적용 사례를 살펴보자. 계층 구조를 보면 정보통신기기 전체 수출량과 전체를 구성하는 개별 항목으로 정보통신기기 분류별(유선기기, 무선기기, 방송기기, 정보기기, 기타부품기기)과 국가별(미국, 일본, 중국 등 7개국)로 나뉘어진다. 다시 이 아래 계층으로는 국가와 정보통신기기의 행렬 구조(예: 미국-유선, 일본-부품 등)에 의해 35개로 나뉘어진다. 각 단계별 예측 방법을 보면 전체 수출량은 시계열 특성과 거시적 변수를 반영한 시계열 모형, 그 아래 계층인 국가별과 분류별 모형에는 전체 수출량 시계열 특성과 국가별과 분류별에 영향을 주는 관련 변수를 반영한 회귀모형, 그리고 행렬 구조에 대한 예측은 상위 계층의 시계열 특성과 행렬 구조 데이터의 계절성이 반영된 다중 회귀모형을 이용하였다.

I. 서론

이 연구는 정보통신기기 수출 예측을 위한 하향식(Top-down) 예측 모형이다. 정보통신기기는 통신기기, 정보기기, 방송기기, 부품기기로 대분류된다. 다시 통신기기는 유선기기와 무선기기, 부품기기는 반도체와 기타부품(나머지부품) 등의 중분류 항목으로 나뉘어진다. 다시 유선기기는 유선전화기, 교환기, 전송기기 등으로 소분류 되어지고 다른 중분류 제품들도 유선기기와 같이 소분류된다. 따라서 정보통신기기는 전체 수출량과 전체를 구성하는 개별 항목간에 계층적인 관계를 가지고 있다. 또한 각 분류된 항목들은 국가별로 나뉘어지기 때문에 하나하나 모두를 예측하기는 대단히 어렵다. 이와 같이 계층 구조를 가진 데이터를 예측하는 방법은 상향식과 하향식 접근방법이 있다[5]. 하향식은 상위계층의 예측값을 하위계층의 독립변수로 이용하는 것을 바탕으로 전체에서 개별 항목으로 순차적으로 예측을 수행하는 방법이며 상향식은 각각의 개별 항목을 예측하여 합산하는 예측방법이다. 하향식 접근방법의 장점은 정보통신기기와 같이 예측 대상이 많은 계층 구조에서 예측 작업의 수를 줄이므로 시간과 비용을 줄일 수 있다는 것이다. 그러나 정확성 문제가 제기 될 수 있다. 왜냐하면 전체를 이루는 개별항목의 시계열 경향이 전체 경향과 다를 때 예측의 정확성이 떨어지기 때문이다. 그러나 정보통신기기 계층 구조는 전체적으로 높은 상관관계를 가지고 있으므로 하향식 접근방법을 적용한다. 전체 수출량과 국가별·제품별 수출량과의 상관관계는 표 1에 나타나있다. 표 1에서

보면 전체 수출량과 국가별·제품별 수출량과의 상관관계는 전체적으로 높다.

하향식 방법을 이용한 정보통신기기 예측 모형 모델링에서 고려해야 할 중요한 3 가지 문제가 있다. 첫째는 정보통신기기 예측에 하향식 접근 방법을 적용 할 때 상위계층의 예측값을 하위계층의 독립변수로 이용할 때 독립변수와 종속변수 사이의 관계 모형을 어떻게 모델링 할 것 인가이다. 둘째로 정보통신기기는 계층적 구조 뿐만 아니라 시계열의 데이터의 특성을 가진다. 이 시계열 데이터 특성은 경향(Trend)과 계절성(Seasonality)이다. 이와 같이 시계열 데이터의 경향과 계절성을 모형에 어떻게 반영할 것인가 이다. 그리고 셋째로 정보통신기기 수출량의 시계열 데이터는 또한 여러 가지 요인들에 의해서 영향을 받는다. 예를 들면 정보통신기기 전체 수출량은 환율이나 유가의 영향을 받을 수 있다. 따라서 예측력을 높이기 위해서는 이와 같은 요인들을 각 계층의 예측 모형에 반영해야 한다.

최근까지 정보통신기기 수출량에 대한 예측 모형에 관한 연구는 전혀 없었다. 단지 통계적인 수치에 의존하는 정도였다. 이 정보통신기기 예측 모형수립은 선행적 연구이다. 그리고 예측모형으로부터 나온 예측값은 앞으로의 정보통신기기 수출산업의 흐름을 파악할 수 있을 뿐만 아니라 시장계획의 방향을 제시해 줄 수 있으며 정책결정에 기초 자료로 이용될 수 있다.

아래 II에서는 정보통신기기 하향식 예측모형을 설명한다. III에서는 모형의 적용 결과이고 IV에서는 결론 및 추후 연구 과제이다.

[표 1] 총 수출량과 국가별·제품별과의 상관관계

유선	무선	정보	나머지 부품
0.6678	0.8244	0.9368	0.9761
미국	일본	중국	유럽
0.9660	0.8421	0.9541	0.9802

II. 정보통신기기 하향식 예측 모형

이 연구에서 예측대상을 정리하면 표 2와 같은 행렬 구조가 된다. ①은 첫째 계층으로서 정보통신기기 전체 수출량을 의미한다. 그리고 ②,③은 둘째 계층으로서 행렬구조의 행과 열의 합을 나타내는 국가별(미국, 일본, 중국, 유럽, 남미, 기타아시아, 나머지국가), 제품별(유선기기, 무선기기, 정보기기, 방송기기, 기타 부분품) 수출량을 의미하며 마지막으로 ④는 셋째 계층으로서 행렬구조의 각 셀에 해당하는 것으로 각 국가의 제품을 의미한다. 계층사이에 상관관계가 높다면 다음과 같이 하향식 예측 방법을 적용할 수 있다. 이 연구에서는 정보통신기기 전체 수출량을 예측하고 그 예측값을 독립변수로 하여 국가별, 제품별 예측을 수행한다. 마지막으로 국가별, 제품별 예측 데이터를 독립변수로 하여 국가 제품별 예측을 한다. 그런 후 국가 제품별 예측데이터를 합산하여 최종 국가별, 제품별 예측데이터를 구하고 다시 이를 합산하여 최종 전체 예측데이터를 얻어낸다. 이것이 정보통신기기 하향식 예측 순서이다.

[표 2] 세부적인 예측 대상

	미국	일본	유럽	...	기타	제품별 합
유선기기	V_{11}	V_{12}	V_{13}		V_{1n}	$Y_{1.}$
무선기기	V_{21}	V_{22}	V_{23}		V_{2n}	$Y_{2.}$
정보기기	V_{31}	V_{32} ④	V_{33}		V_{3n}	$Y_{3.}$ ③
방송기기	V_{41}	V_{42}	V_{43}		V_{4n}	$Y_{4.}$
기타부품	V_{51}	V_{52}	V_{53}		V_{5n}	$Y_{5.}$
국가별 합	$Y_{.1}$	$Y_{.2}$ ②	$Y_{.3}$		$Y_{.n}$	$Y_{..}$ ①

표 2에서 첫번째 예측 대상인 $Y_{..}$ 는 정보통신기기 전체 수출량을 의미한다. 그리고 두 번째 예측 대상인 제품별 합은 $Y_{i.}, i=1,2,3,\dots, M$ 와 같고 국가별 합은 $Y_{.j}, j=1,2,3,\dots, N$ 으로 표현이 된다. 세 번째는 i 제품이 j 국가에 대한 수출량(국가 제품별 수출량)으로 각 셀에 해당하는 것으로 $Y_{ij}, i=1,2,3,\dots, M, j=1,\dots, N$ 와 같이 표현된다.

각 계층의 하향식 접근 방법의 모형은 다음과 같다. 첫째 전체 정보통신기기 수출량 예측 모형이다. 예측 모형은 자기상관회귀(ARIMA)모형이다. 따라서 정보통신기기 수출량 예측은 자기회귀모형을 이용하여 증감을 예측한다. 증감은 전체 시계열 수출량 데이터의 증가 또는 감소경향(Trend)을 반영하기 위해서이다. 예측 모형은 식 (1)과 같다.

$$\Delta Y_{..}(t) = \phi_1 \Delta Y_{..}(t-1) + \phi_2 \Delta Y_{..}(t-2) + \dots + \phi_p \Delta Y_{..}(t-p) + e(t) \quad \text{식 (1)}$$

식 (1)에서 전체 정보통신기기 수출량에 영향을 미치는 요인들을 반영해야 한다. 변수 반영은 고려해야 할 변수를 개별적으로 미치는 영향을 살피는 것이 아니라 정성적인 방법을 통해 산출한다. 다시 말해서 반영해야 할 변수를 종합적으로 판단하고 고려해서 수출량에 미치는 영향이 증가인지 또는 감소인지 아니면 현재와 동일할 것인지를 판단해서 그 결과를 식(1)에 반영해 준다. 수출량에 미치는 영향 정도를 판단하기 위해서 먼저 전문가에게 앞으로 변해갈 방향에 관한 의견을 듣는다. 각각의 전문가 의견을 반영하여 변수별로 가중치를 반영한 후 최종적으로 전체 수출량에 미치는 영향을 판단한다. 그 결과가 수출량에 (+)요인으로 작용한다면 예측기간동안 수출량 예측값에 $\sigma \times$ (변수 증가율)를 반영하고 (-)요인으로 작용한다면 $\sigma \times$ (변수 감소율)를 반영하여 최종 예측값을 구할 수 있다.

$$Y_{..}(t) = \phi_1(Y_{..}(t-1) - Y_{..}(t-2)) + \phi_2(Y_{..}(t-2) - Y_{..}(t-3)) + \phi_3(Y_{..}(t-3) - Y_{..}(t-4)) + Y_{..}(t-1) + e(t) + \dots$$

$$+ \sigma \times \Delta(\text{변수})$$

σ : 오차의 표준편차

둘째 국가별, 제품별 예측이다. 예측은 표 2에서 $Y_{.j}, j=1,2,\dots, N$ 와 $Y_{i.}, i=1,2,\dots, M$ 을 예측하는 것이다. 이 두 변수는 다음과 같이 정의할 수 있다.

$$Y_{.j} = \sum_{i=1}^M Y_{ij}, \quad Y_{i.} = \sum_{j=1}^N Y_{ij}$$

국가별 예측과 제품별 예측은 동일하다. 그러나 반영되는 요인들은 서로 다르다. 그러므로 여기에서는 제품별 수출량 예측모형만 고려한다. 제품별 수출량 예측모형은 전체 수출량을 독립변수로 하는 회귀 모형이다. 따라서 기본식을 보면 다음과 같다.

$$Y_{i.}(t) = \beta_0 + \beta_1 Y_{..}(t) + e(t) \quad \text{식 (2)}$$

식 (2)에서 $e(t)$ 는 평균 0, 분산 σ_e^2 이다. 단 순회귀분석에서 가정은 바로 $e(t)$ 의 합이 0으로 백색잡음(White noise)을 따르는 것이다. 그러나 0을 중심으로 랜덤하게 나타나지 않고 경향을 나타내는 경우가 있다. 다시 말해서 오차항이 자기 상관관계를 가진다는 것을 의미한다. 듀빈-와슨(Duubin-Watson)통계량을 이용하여 오차항이 자기 상관관계를 있는지 판단 가능하다[2]. 또 다른 방법은 분산으로 알 수 있다. 만약 오차항에 자기 상관관계가 존재한다면 회귀방정식에 코크라인 앤 오커트(Cochrane & Orcutt) 변환방법을 사용하여 반영해야 한다[6]. 다음은 식(2)에서 오차항이 자기 상관관계가 있어서 코크라인 앤 오커트 변환을 적용한 수식이다.

$$Y_{i.}(t+1) - \phi Y_{i.}(t) = \beta_0'(1 - \phi) + \beta_1'(Y_{..}(t+1) - \phi Y_{..}(t)) + \delta(t) \quad \text{식 (3)}$$

식(3)에서 $\delta(t)$ 는 평균 0 이고 분산이 σ_e^2 이다. 국가별 또는 제품별 수출량 예측을 위해서는 관련 변수를 반영해야 한다. 반영해야 할 변수는 전체 수출량에 영향을 미치는 요인을 제외하고 국가별과 제품별에만 의미를 갖는 변수만을 반영해야 한다. 국가별 수출량에 영향을 미치는 변수를 보면 국가별 GDP, 환율, 관세, 국가 시장 수준 등이라 할 수 있다. 또한 제품별 수출량 예측에 영향을 주는 변수를 살펴보면 신상품, 신기술 개발, 상대가격, 대체재 존재 유무, 경쟁 업체 등이라 할 수 있다. 변수의 반영은 전체 수출량 예측 모형에서 변수 반영하는 방법과 동일하다.

셋째는 단위 셀 예측(각 국가 제품별 수출량 예측)이다. 각 셀을 예측하기 위해서 기본방법은 다중회귀분석을 이용한다. 다중회귀를 위해 반영하는 변수로는 같은 기간에 대한 국가별 수출량과 제품별 수출량을 반영한다. 셀이라는 것은 표 1에서 보듯이 각 제품별에 대한 국가별 수출량을 의미하는 것으로 이것은 당연히 국가별 수출 경향과 제품별 수출 경향에 많은 영향을 받는다. 하지만 전체 수출량에서 차지하는 국가나 제품의 차지하는 비율은 다르다. 따라서 비중이 높은 부분에는 더 많은 가중치를 주고 비중이 낮은 부분에는 비중을 낮게 주는 것이 필요하다. 이렇게 가중치를 주는 방법으로 국가와 제품이 전체 수출량에서 차지하는 비율을 반영해 준다. 반영하기 위하여 다음과 같은 변수를 정의한다. 각 국가와 제품에 해당하는 인덱스를 i 라 하고 국가에 해당하는 인덱스를 j 라 할 때, t 시점에서 i 제품에 대한 j 국가의 수출량을 $Y_{ij}(t)$ 라 하자.

● pw_{it} - t 시점에 총수출량 중 i 제품기기가 차지하

는 비율
$$pw_{it} = \frac{Y_{i..}(t)}{Y_{..}(t)}$$

● nw_{jt} - t 시점에 총수출량 중 j 국가가 차지하는 비율

$$nw_{jt} = \frac{Y_{.j}(t)}{Y_{..}(t)}$$

년도 분기		1989	...	1994	1995	1996	1997
1			...				
2			...				
3			...				
4			...				

[그림 1] 데이터의 경향과 계절성 반영 방법

다음으로 셀의 다중회귀모형의 독립변수로 사용되는 국가나 제품 데이터는 경향과 계절성을 갖는다. 따라서 경향과 계절성이 반영이 되어야 한다. 경향은 시계열 데이터가 매년 또는 매 분기마다 나타나는 특성을 의미하고 계절성은 같은 분기에서 나타나는 데이터 특성을 의미한다. 이를 반영하는 방법으로 그림 1에 나타나 있듯 i) 경향과 ii) 계절성을 반영한다. 다시 말하면 i)은 $t-1$ 분기에서 t

분기까지의 증감량을 나타낸다. ii)는 $t-4$ 분기와 t 분기 사이에 얼마만큼 증가를 했는지 비를 나타낸다. 국가나 제품이 반영되는 형태가 동일하므로 제품에 대해서 보면 아래와 같이 나타난다.

i) 경향 : $\Delta Y_{i.}(t) = Y_{i.}(t) - Y_{i.}(t-1)$

$$\Delta Y_{.j}(t) = Y_{.j}(t) - Y_{.j}(t-1)$$

ii) 계절성 : $ap_{i.}(t) = \frac{Y_{i.}(t)}{Y_{i.}(t-4)}$

$$ap_{.j}(t) = \frac{Y_{.j}(t)}{Y_{.j}(t-4)}$$

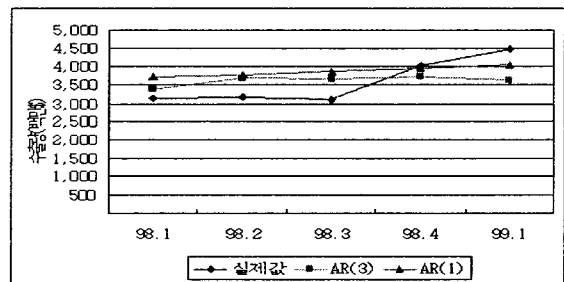
반영되는 독립변수는 증감형태이다. 따라서 셀의 예측값 또한 증감 형태의 결과가 나온다. 그러므로 모형은 다음과 같다.

$$\Delta Y_{ij}(t) = \beta_0 + \beta_1 \times pw_{it} \times ap_{i.}(t) \times \Delta Y_{i.}(t) + \beta_2 \times pw_{jt} \times ap_{.j}(t) \times \Delta Y_{.j}(t) + e(t)$$

이 모형에서도 각 셀에 영향을 미치는 변수를 반영해 주어야 한다. 반영방법은 국가별, 제품별 수출량을 예측모형에서와 동일한 방법으로 반영하면 된다.

III. 모형의 적용 결과

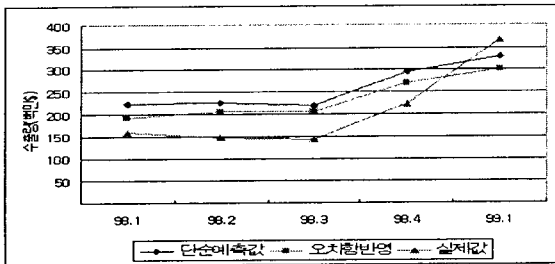
주어진 데이터는 94년 1/4 분기부터 99년 1/4 분기로 총 21개의 분기별 시계열 데이터이다. 시계열 데이터 중에서 94년 1/4 분기부터 97년 4/4 분기를 훈련기간으로 하여 98년 1/4 분기부터 99년 1/4 분기까지 5 분기를 실제데이터와 비교하여 모형을 검증하였다. 첫째로 정보통신기기 전체 수출량 모형의 경우 AR(3)가 타당한지를 AR(1)모형과 비교하였다.



[그림 2] AR(3)와 AR(1)의 예측값과 실제값의 비교

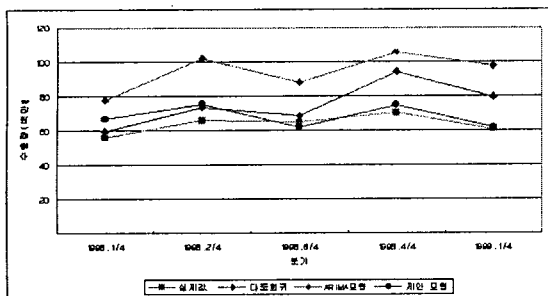
분산을 보면 AR(1)은 9.04×10^{10} 이고 AR(3)는 5.29×10^{10} 이다. 또 하나 AIC 값을 보면 AR(1)은 423.263 이고 AR(3) 모형은 418.774로 AR(3) 모형보다 AR(1)모형보다 더 적합하다. 그리고 그림 2에서 AR(3)모형과 AR(1)모형 및 실제값을 비교한 것이다. AR(3)모형의 예측값이 AR(1)모형의 예측값 보다 더 실제값에 가깝다. 따라서 AR(3)모형이 더 적합하다. 둘째로 국가별, 제품별 모형의 경우는 일본

수출량에 대하여 회귀모형과 오차항의 자기상관관계를 반영한 회귀모형 및 ARIMA 모형을 비교하였다. AR(2) 모형의 표준오차는 34,990.22 이다. 반면 단순회귀 모형의 표준오차는 26,630.46 이고 상관계수는 0.9138 이고 조정 결정계수는 0.8233 으로 예측력이 상당히 높다. 그러나 오차항을 반영한 회귀모형의 표준오차는 26,625.87 이므로 일본 수출량의 오차항이 자기상관관계를 가진 경우이다. 그림 3 은 회귀모형과 오차항을 반영한 회귀모형의 예측값과 실제값을 나타낸 것이다. 여기에서 오차항을 반영한 회귀모형이 보다 더 실제값에 근접한 것을 알 수 있다.



[그림 3] 일본 수출량 예측결과 비교

마지막으로 셀 모형(국가 제품별 모형)은 기 타아시아-유선기기에 대해 제안모형과 다중회귀모형 및 ARIMA 모형을 서로 비교 하였다. 세가지 모형에 대한 표준오차를 살펴보면 제안 모형은 13,770, 다중회귀모형은 21,993, AR(2)모형은 17,060 이다. 또 제안 모형과 다중회귀모형에서 모형의 조정계수를 살펴보면 제안 모형의 경우는 0.7889 이고, 다중회귀모형의 조정계수는 0.3801 이다. 그림 4 을 보면 다중회귀모형과 ARIMA 모형에 비해 제안 모형이 실제값에 대단히 근사해 가는 것을 볼 수 있다. 따라서 비교 대안보다 제안모형이 우수하다고 할 수 있다.



[그림 4] 기타아시아-유선기기 모형별 예측값

그리고 모형별로 설정한 방법을 이용하여 1994년부터 1999년 1/4까지 데이터를 이용하여 2000년까지 예측한 결과는 다음 표와 같다. 표 3는 전체 수출량 최종 예측 데이터이다. 전체적으로 약 간씩 증가하고 있다.

[표 3] 전체 수출량 예측 데이터 (단위:천불)

분기	전체 매출액(단위 천불)
1999. 2/4	5,055,845
1999. 3/4	4,686,753
1999. 4/4	4,699,728

2000. 1/4	4,803,602
2000. 2/4	4,867,432
2000. 3/4	4,763,483
2000. 4/4	4,695,565

IV. 결론

이 연구에서는 계층구조를 가진 정보통신기기 수출량을 하향식 접근방법으로 모델링 및 예측하였다. 모델링은 계층의 데이터의 시계열 특성과 데이터에 영향을 주는 요인들을 반영하였다. 이 모델링의 결과는 정보통신기기 전체 수출량, 국가별 수출량, 제품별 수출량 및 국가 제품별 예측 데이터로써 수출이나 산업 활성화 등에서 정책결정의 기초자료로써 이용될 수 있다. 이 연구는 선행적 연구로서 의미를 가지며 앞으로는 변수 반영 측면에서 보다 정량적인 반영과 상관관계가 낮은 대상에 대한 연구가 수행되어야 할 과제이다.

[참고 문헌]

- [1] Douglas C. Montgomery, Lynwood A. Johnson and John S. Gardiner, Forecasting and Time Series Analysis, Second Edition, McGraw-Hill, 1990.
- [2] Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman, Forecasting- Methods and Applications, Third Edition, John Wiley & Sons, 1998.
- [3] Young J. Joo and Duk Bin Jun, "Forecasting A Daily Time Series with Varying Seasonalities: An Application to Daily Visitors to Farmland in Korea," Computers Industrial Engineering, Vol. 30, No. 3, pp 365-373, 1996.
- [4] 박성현, 회귀분석, 민영사, 1995.
- [5] 주영진, 전재호, 임명환, 박명철, "총량적 접근방법에 의한 정보통신기기시장 예측," 대한산업공학회, 97 춘계공동학술대회, pp 762-766, 1997.4.
- [6] 최기현, 이종협, SAS/ETS를 이용한 시계열 분석과 그 응용, 고려대학교 통계연구소 통계분석 강의총서 15, 자유아카데미, 1994.
- [7] 하인봉, "정보통신산업이 국가경제 및 산업의 경쟁력에 미치는 영향 분석," 정보통신학술 연구과제, 1996. 3.