

스펙트럼 분석과 계절성 선형 모델을 이용한 Intra-Day 콜센터 통화량 예측

신태수¹ · 김명석²

한국씨티은행¹, 서강대학교 경영학과²

(2010년 10월 접수, 2011년 1월 채택)

요약

본 논문에서는 스펙트럼 분석과 계절성 선형 모델을 이용하여 intra-day 콜센터 통화량 예측에 필요한 계절성 변수를 찾아내는 방법을 제시한다. 제시한 방법을 북미 지역의 어느 은행의 5분 단위 콜센터 통화량에 실증 적용하여 기존의 통계적 방법으로는 입증할 수 없었던 월 단위 계절성 변수가 유의함을 보인다. 새로이 찾아진 변수가 intra-day 콜센터 통화량 예측능력을 향상시키는지 확인하기 위해서 새로운 변수를 포함하는 계절성 선형 모델과 이 변수를 포함하지 않은 계절성 선형 모델의 일일 통화량 예측능력을 비교 평가한다. 평가결과 새로운 변수를 포함한 모델이 우수하다는 결과를 얻었다.

주요어: 계절성 변수, 계절성 선형 모델, 스펙트럼 분석, 콜센터 통화량 예측, intra-day 콜센터 통화량.

1. 서론

콜센터의 통화량 예측 및 모델링에 관한 연구는 많은 학자들에 의해서 실행되어져 왔다. 초기에는 Bianchi 등 (1993)와 Andrews와 Cunningham (1995)에 의해서 autoregressive integrated moving average(ARIMA) 모델이 일별 통화량 예측에 사용되었다. 그 후, 많은 콜센터 통화량에서 계절성 주기가 발견되면서 Box-Jenkins의 계절성 ARIMA 모델들이 활발히 활용되어지고 있다. 한편, 시간 단위 또는 분 단위와 같이 하루 단위 보다 작은 단위(intra-day) 통화량 자료를 획득할 있게 되면서 부터는 일별 통화량 보다는 intra-day 통화량의 모델이 학계에서 활발하게 분석되어지고 있다. Tych 등 (2002)는 dynamic harmonic regression 모델을 intra-day 통화량 예측에 응용하였다. Taylor (2008)는 exponential smoothing 모델을 1시간 단위 콜센터 통화량 예측에 사용할 것을 제안하면서 계절성 ARIMA 모델, 주기성 AR 모델, dynamic harmonic regression 모델들과 비교 평가하였다. Weinberg 등 (2007)는 베이지안 방법론을 5분 단위 콜센터 통화량 모델에 활용하였다. Shen과 Huang (2008)는 singular value decomposition을 15분 단위 콜센터 통화량 예측에 적용하였다.

유리한 계절성 변수를 찾아내는 것은 예측의 정확성을 높인다는 측면에서 매우 중요한 의미를 갖는다. 기존의 많은 문헌들은 그림을 이용한 비교를 통하여 계절성 변수를 선정하였다 (Gans 등, 2003; Weinberg 등, 2007; Shen과 Huang, 2008). 이와 같은 방법은 변수선택에 대한 통계적 증거를 제시할 수 없었다. 최근에는 엄격한 통계적 방법을 동원하여 유의한 계절성 변수를 찾아낸 경우가 있었다. Taylor

이 논문은 첫 번째 저자의 서강대학교 석사학위 논문을 바탕으로 작성되었음. 이 연구는 2008년도 서강대학교 연구비 지원에 의한 연구임(200810048.01).

²교신저자: (121-742) 서울 마포구 신수동, 서강대학교 경영학과, 부교수. E-mail: myungkim@sogang.ac.kr

(2008)는 Schwarz Bayesian Criterion(SBC)을 ARIMA 모델에 적용하였고, Tych 등 (2002)와 Taylor (2008)는 스펙트럼 분석을 변수 선택에 사용하였다. 그러나, 이러한 방법들의 한계는 이러한 방법을 동원하여 찾아낸 유의한 변수가 기존의 많은 문헌에서 그림 비교를 통해서 찾아낸 변수인 요일 효과 및 시간대별 효과에 그친다는 것이다. 그 이상의 변수에 대한 유의성을 밝히지 못하고 있다. Gans 등 (2003)는 그림 비교를 통해서 월 단위 효과를 제시하였으나, 이 효과에 대한 통계적 유의성은 아직까지 제시되어지지 못하고 있다.

본 논문에서는 스펙트럼 분석을 계절성 선형 모델에 연동시켜서 유의한 계절성 변수를 찾는 방법을 제시하고자 한다. 제시된 방법을 북미 지역의 한 은행의 콜센터에 걸려온 5분 단위 통화량 예측에 적용하여 기존의 요일 및 시간대별 효과와 더불어 새로운 계절성 변수인 월 단위 효과를 찾아낼 수 있음을 보여 준다. 제시된 방법을 통하여 얻은 계절성 변수는 계절성 선형 모델에 적용되어 요일 및 시간대별 효과만을 포함하고 있는 기존의 계절성 선형 모델과 비교 평가된다. 분석 결과 새로이 발견된 계절성 변수를 포함한 계절성 선형 모델은 기존에 알려진 계절성 변수만을 포함하고 있는 계절성 선형 모델보다 일일 통화량 예측에서 보다 정확한 예측 결과를 나타내었다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 콜센터 통화량 자료의 일반적으로 알려져 있는 특성을 설명한다. 계절성 선형 모델과 스펙트럼 분석을 이용한 변수 선택방법은 제 3장에서 소개되어지며, 예측에 대한 평가 방법도 이 장에서 설명되어진다. 북미 지역의 한 은행의 콜센터 통화량을 이용한 실증 자료 분석은 제 4장에서 이뤄지며, 새로이 발견된 계절성 변수를 이용한 계절성 선형 모델의 예측 정확도가 평가되어진다. 결론 및 향후 연구 방향이 마지막 장에서 언급된다.

2. 콜센터 통화량 자료의 특징

본 연구에서는 북미에 위치한 어느 대형 상업 은행의 콜센터 자료를 이용하였다. 자료는 2003년 3월 3일에서 2003년 10월 24일까지 총 165일에 해당하는 자료이다. 이 자료는 Weinberg 등 (2007)에 의해서 분석된 적이 있다. 콜센터는 24시간 운영되어지나, 대부분의 전화는 오전 7시에서 오후 9시 사이에 걸려오므로, 본 연구에는 Weinberg 등 (2007)와 마찬가지로 오전 7시에서 오후 9시 5분까지의 통화량을 이용한다. 5분 간격으로 통화량을 집계하였으며, 하루는 169개의 구간으로 이루어져 있다.

콜센터 통화량의 매우 잘 알려져 있는 특징인 요일별 효과와 시간대별 효과를 우리의 자료에서도 그림 2.1을 통해서 쉽게 관찰할 수 있다. 그림 2.1의 (a)는 전체 자료의 요일별 평균 통화량의 분포를 box plot을 통해서 구현한 것이다. 요일별 평균 통화량을 살펴보면, 월요일이 평균적으로 가장 통화량이 많은 것을 볼 수 있다. 이는 주말동안 은행이 영업을 하지 않음으로 인해서, 주말동안 해결하지 못한 은행 업무에 대한 고객들의 통화가 집중되어 발생하는 것으로 추정된다. 또한 수요일, 목요일을 통해서 감소된 통화량이 금요일 다시 증가하는 것을 볼 수 있는데, 이는 휴일을 앞에 두고, 휴일 전에 은행 업무를 해결하기 위한 통화에 의한 것으로 추정된다.

그림 2.1의 (b)는 5분 단위 시간대별 평균 통화량을 요일별로 나타낸 것으로 요일 및 시간에 따른 통화량 분포를 확인할 수 있다. 요일에 상관없이 오전 7시부터 통화량은 꾸준히 증가하여, 오전 10시대가 하루 중 가장 많은 통화량을 보인다. 10시대에 정점에 도달한 통화량은 이후 5시까지 완만하게 감소하다가, 5시 이후에는 통화량이 급격히 감소한다. 이를 통해서 고객들은 대부분 오전 10시에서 오후 5시 사이에 은행 업무를 해결하고, 이 과정에서 콜센터를 이용하고 있다고 생각할 수 있다. 요일별로 관찰하였을 경우, 월요일 오전 7시에서 8시 사이의 통화량은 다른 요일에 비해서 적게 나타남을 알 수 있다. 반면에 금요일 오후 7시 이후에는 다른 요일에 비해서 통화량이 급격히 감소함을 관찰할 수 있다. 요일과 시간대가 서로 상호 작용(interaction)을 하는 것처럼 보인다.

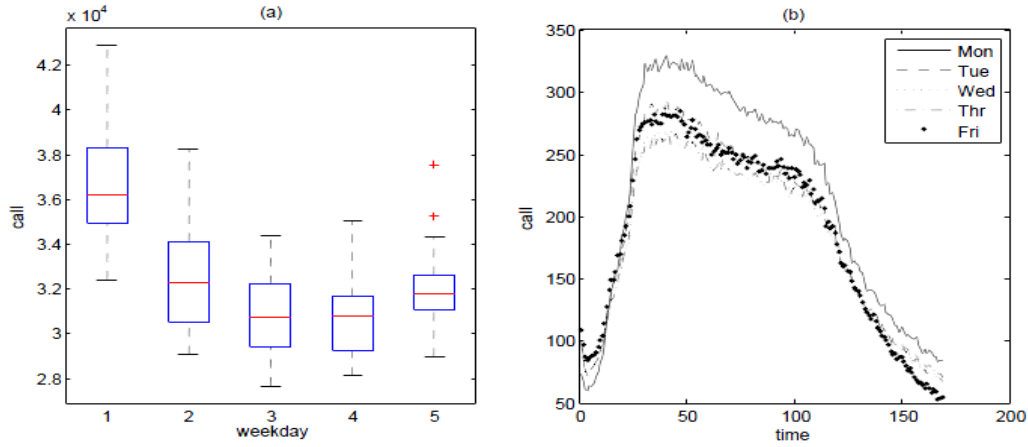


그림 2.1. (a) 요일별 평균 통화량의 분포를 box plot으로 나타냄. 횡축의 1 부터 5는 각기 월요일부터 금요일을 명시함. (b) 월요일부터 금요일까지의 5분 단위 시간대별 평균 통화량을 나타냄.

요컨대 그림 2.1(a)와 (b)는 각기 요일별 효과와 시간대별 효과를 나타내며, 특히, (b)에서는 두 효과의 상호작용 효과도 관찰할 수 있었다. 그러나, 이러한 특징들은 그림을 통해서 해석되어지는 것으로 이러한 계절성 변수들의 유의성에 대한 통계적 근거를 제시한 것은 아니다. 서론에서 언급한 바와 같이 본 연구에서는 통계적 방법인 스펙트럼 분석과 계절성 모델을 이용해 유의한 계절 변수를 찾아내고자 한다.

3. 계절성 선형 모델과 스펙트럼 분석

3.1. 계절성 선형 모델과 스펙트럼 분석을 이용한 변수 선택

N_{jk} 는 j 일의 시간 구간 $[t_{k-1}, t_k]$ 사이에 걸려온 통화량을 나타낸다. 여기서, $j = 1, 2, \dots, J$ 이고 $k = 1, 2, \dots, K$ 이다. Brown 등 (2005)에 의하면 통화량은 time inhomogeneous 포아송 분포를 따르며, 원래의 통화량 N_{jk} 을 $y_{jk} = \sqrt{N_{jk} + 0.25}$ 으로 변환할 경우, 분산의 안정화(variance stabilization)를 확보할 수 있다. 아울러 y_{jk} 는 통화량이 많아질 경우 정규분포로 근사한다고 주장하였다.

Weinberg 등 (2007)는 통화량을 Brown 등 (2005)의 방법을 따라서 변환을 시키고, 계절성을 더미 변수로 반영한 계절성 선형 모델(seasonal linear model)을 사용하였다. Weinberg 등 (2007)에서 사용된 시간 단위 효과와 요일 단위 효과, 그리고 이들의 상호작용 효과를 반영해주는 계절성 선형 모델은 다음과 같다.

$$\text{Model A: } y_{jk} = \mu + \alpha_{d_j} + \beta_k + \epsilon_{jk}$$

$$\text{Model B: } y_{jk} = \mu + \alpha_{d_j} + \beta_k + \gamma_{d_j k} + \epsilon_{jk}$$

여기서, $\epsilon_{jk} \sim N(0, \sigma^2)$ 이며 d_j 와 k 는 각기 j 일의 요일을 나타내 주는 변수와 시각을 나타내주는 변수를 의미한다. α_{d_j} 와 β_k 는 d_j 요일이 y 에 미치는 효과와 k 시각이 y 에 미치는 효과를 나타낸다. $\gamma_{d_j k}$ 는 d_j 요일의 k 시각이 y 에 미치는 효과로 요일과 시각의 상호작용 효과를 의미한다. Model A는 요일 효과와 시간 효과만 갖는 반면, Model B는 이들의 상호작용 효과도 포함하고 있다. 참고로, 콜센터 통화량에 휴일이 있는 경우, 일반적으로 휴일 다음날은 특별히 높은 통화량을 보이는 일종의 휴일 효과가 관찰되어진다. 이는 월요일이 다른 요일에 비해서 특별히 많은 통화량을 보이는 것과 같은 이유일 것이라 생

각된다. 본 연구에서 사용된 콜센터 통화량 자료에서도 이와 같은 효과는 명백하게 나타난다. 따라서, 이런 효과를 모델에 반영해주기 위해서 휴일 다음날의 요일 변수는 월요일로 취급하였다.

시계열 자료를 분석하는 방법은 크게 두 가지가 있다. 하나는 시간영역(time domain)에서 시간이 진행함에 따라 이를 분석하는 시계열 분석 방법이며, 다른 하나는 시계열이 어떤 주기를 갖고 변동되어 가는 것을 분석하는 것으로 진동수영역(frequency domain)에서의 스펙트럼 분석이 있다. 스펙트럼 분석은 시계열을 시간의 차원이 아닌 변동주기의 차원에서 접근해 나간다. 스펙트럼 분석 또는 주기 분석이란 시계열 자료를 스펙트럼(spectrum)이라는 필터에 여과시켜 주기(period) 또는 진동수(frequency)별로 변동 내용을 분리한 후 그 형태를 분석하는 과정을 말한다.

어떤 시계열 자료 X_t 가 $E(X_t) = 0$ 을 만족하면서 약 안정성(weakly stationary)의 성질을 가지면 시간영역에서 자기 공분산 $\rho_h = \text{COV}(X_t, X_{t+h})$ 을 얻을 수 있다. 여기서, $h = 0, \pm 1, \pm 2, \dots$ 이다. 한편, 진동수 영역에서는 이에 대응되는 개념으로 spectral density

$$f(\phi) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho_h \exp(-ih\phi)$$

을 얻을 수 있다. Spectral density는 아래의 periodogram $\hat{f}(\phi)$ 을 이용하여 추정할 수 있다.

$$\hat{f}(\phi) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \hat{\rho}_h \exp(-ih\phi). \quad (3.1)$$

실질적으로 식 (3.1)은 아래의 공식을 이용해서 구할 수 있다.

$$\hat{f}(\phi) = \frac{1}{2n\pi} \left| \sum_{t=0}^{n-1} x_t \exp(-ih\phi) \right|^2. \quad (3.2)$$

식 (3.2)는 Fourier 변환을 이용해서 쉽게 얻을 수 있다. 식 (3.2)를 이용하여 횡축을 진동수 영역으로 하고 종축을 periodogram으로 하는 그림을 그려서 유의한 계절성 주기를 찾을 수 있다. 진동수 영역 가운데 최대의 periodogram 또는 상대적으로 유의하게 커다란 periodogram을 가져다주는 값을 찾고 이에 대응되는 주기를 찾으면, 이것이 유의한 계절성 주기에 해당된다.

본 연구에서는 스펙트럼 분석과 계절성 선형 모델을 연동하여 유의한 계절성 주기를 찾아낸다. 유의한 계절성 주기를 찾는 과정은 다음과 같이 요약 정리된다.

- Step 1. 시계열 자료에 추세선이 있을 경우, 이를 제거한 뒤 분석을 실행한다.
- Step 2. 백색 잡음 검정(white noise test)을 통해서 유의한 주기 여부를 확인한다.
- Step 3. Intra-day 시계열 자료 및 이를 하루 단위로 합산하여 주기 분석을 실행하고, 유의한 계절성 주기를 찾는다.
- Step 4. 유의한 계절성 변수를 선형 모델에 적용하고, 계절성 변수 간에 상호작용 효과여부를 검정한다.
- Step 5. 유의한 계절성 변수들을 이용하여 계절성 선형 모델을 추정한 후 얻어진 잔차항에 대하여 백색 잡음 검정 및 주기 분석을 실행하여 새로운 유의한 변수를 찾는다.

3.2. 모델 평가

스펙트럼 분석을 통하여 새로이 선택된 계절성 주기가 콜센터 통화량 예측의 정확도를 증가 시키는데 기여하는지 여부를 살펴보기 위하여 새로운 변수를 포함하여 예측한 계절성 선형 모델과 기존의 변수만

을 포함한 계절성 선형 모델의 예측 정확도를 비교 평가 한다. 예측의 정확도 평가 기준은 Weinber 등 (2007)이 사용하였던 아래의 방법을 이용한다.

$$\text{rmse}_j = \left(\frac{1}{k} \sum_{k=1}^K (N_{jk} - \hat{N}_{jk})^2 \right)^{\frac{1}{2}}, \quad \text{ape}_j = \frac{100}{K} \sum_{k=1}^K \frac{|N_{jk} - \hat{N}_{jk}|}{N_{jk}},$$

여기서 \hat{N}_{jk} 은 j 일의 k 시간에 예측되어진 콜센터 통화량으로 $\hat{N}_{jk} = \hat{y}_{jk}^2 - 0.25$ 으로 계산되어진다. 위의 두 공식은 각기 root mean square errors(RMSE)와 average percentage error(APE)으로 j 일 단위로 계산되어진다. 본 연구에서는 하루당 총 169개의 자료가 관측되어지므로 $K = 169$ 가 이용된다. 예측은 5분 단위로 통화량 자체를 예측하되, 예측의 평가는 일단위로 이루어진다. 참고로, 콜센터 통화량 예측의 주요 목적은 예측 통화량에 따라서 얼마나 많은 인원을 콜센터에 배치시키느냐가 중요하고, 인원 배치는 일차적으로 하루 단위로 정해진다. 본 연구에서는 익일 통화량 예측 기간은 2003년 7월 25일부터 10월 24일까지 총 64일이 고려되어진다. 결국, 64개의 RMSE와 APE가 추정되어지며 이들의 평균값, 중앙값, 최소값, 최대값 등의 통계량을 이용하여 모델 예측능력의 정확성이 평가되어진다.

4. 콜센터 자료 실증 분석

북미에 위치한 어느 대형 상업 은행의 콜센터에 5분 단위로 집계한 통화량에 대하여 앞 장에서 제시한 방법론을 적용해 보았다. 유의한 계절성 주기를 찾기 위해서는 스펙트럼 분석이 필요한데, 이에 앞서서 추세선의 존재 여부와 백색 잡음 검정을 실시하였다. 시간을 독립변수로 하고 통화량을 종속변수로 하는 회귀분석을 이용하여 5분 단위 통화량과 하루 단위 통화량에 대한 추세선의 존재 여부를 확인하였으나 유의한 추세선을 발견할 수 없었다. 5분 단위 통화량과 하루 단위 통화량에 대하여 추세선 교정 과정 없이 백색 잡음 검정을 실시한 결과, 두 개의 시계열 모두에서 Bartlett의 Kolmogorov-Smirnov 검정 통계량 (Bartlett, 1966; Durbin, 1967; Fuller, 1976) 및 Fisher의 Kappa 검정 통계량 (Fuller, 1976)에서 유의성을 발견할 수 있었다. 이는 의미 있는 주기의 존재 가능성으로 해석될 수 있다.

구체적인 계절성 주기를 찾기 위해서 5분 간격 통화량과 하루 단위의 통화량에 대해서 각각 주기분석을 실시하였다. 분석 결과는 그림 4.1를 해석함으로써 얻을 수 있다. 우선, 그림 4.1의 (a)를 보면 5분 단위 통화량에 대하여 주기 분석을 한 결과, 최대 periodogram에 대응하는 period는 169에 해당되었다. 이는 스펙트럼 분석을 실시한 자료가 169개마다 일정한 계절 변동을 나타냄을 의미하며, 분석에 이용된 자료의 하루 구간 수가 169개이므로 통화량은 하루 단위를 주기로 갖고 있다는 것을 의미한다. 하루 단위 통화량을 이용하여 주기분석을 실시한 결과, 그림 4.1의 (b)를 보면, 최대 periodogram에 대응하는 주기는 4.8일에 해당된다. 콜센터 자료는 주중 자료만을 가지고 있으므로 한 주는 5일로 계산되어지며, 주어진 자료는 3월에서 6월에 해당하는 자료로 이 사이에, 6일의 특별한 휴일이 존재한다. 콜센터의 통화량이 일주일을 계절 변동으로 갖는다면, 일주일의 근무 일수 5일에서 휴일로 인한 오차로 인해 4.8일 정도의 값을 갖게 될 것이다. 결국 콜센터의 통화량은 하루단위 계절성과, 한주단위 계절성을 갖고 움직이고 있음을 확인하였다. 이는 시간 효과와 요일 단위 효과로 해석되어질 수 있다.

주기 분석을 통하여 발견한 시간 단위 효과와 요일 단위 효과를 계절성 선형 모델에 더미 변수를 이용한 독립변수로 적용할 경우, Model A가 된다. 두 개의 효과의 상호작용 효과도 고려하여 이를 계절성 선형 모델에 포함시킬 경우, Model B가 된다. 상호작용 효과의 유의성은 분산분석을 통하여 검정할 수 있다. 분산분석 결과는 표 4.1에 요약 정리되어 있다. 분산분석 결과, 상호작용 효과가 유의함을 알 수 있다.

두 개의 모델 Model A와 Model B를 이용하여 콜센터 통화량을 추정된 뒤, 실제치와 추정치의 차이인

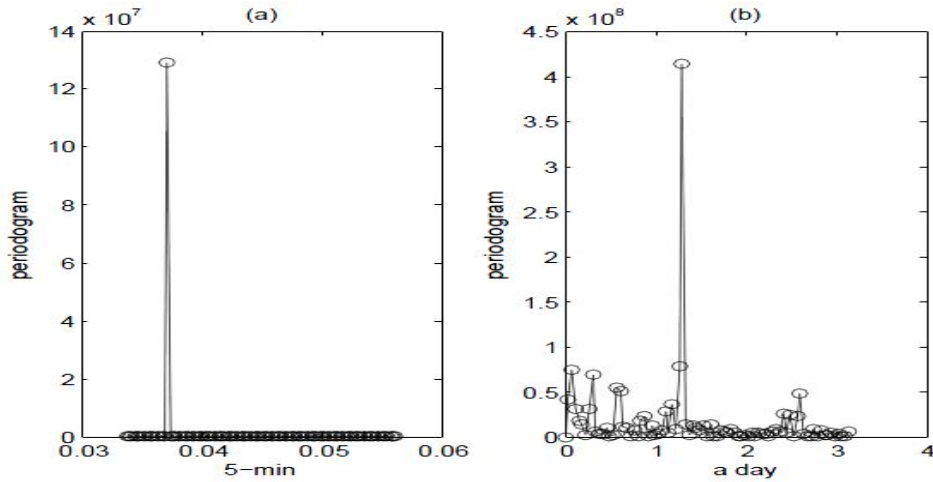


그림 4.1. (a) 5분 단위 통화량의 periodogram 그림으로 최대 periodogram 값에 해당되는 주기는 169로 하루에 해당됨. (b) 하루 단위 통화량의 periodogram 그림으로 최대 periodogram 값에 해당되는 주기는 4.8로 약 일주일에 해당됨.

표 4.1. 통화량 자료의 요일 효과, 시간 효과 및 상호작용 효과에 대한 분산분석 결과

source	DF	SS	MS	F stat	P-value
day	4	4057058.1	1014264.5	2303.88	<.0001
time	168	152111670.9	905426.6	2056.66	<.0001
day×time	672	2136286.5	3179.0	7.22	<.0001

잔차항들에 대하여 백색 잡음 검정을 실시하였다. 백색 잡음 검정 결과, 두 개의 잔차항 모두에서 유의한 주기의 가능성이 발견되었다. 실제값과 예측값의 차이가 일정한 주기를 보인다는 사실은, 모델이 자료가 갖고 있는 실제 계절 변동요인을 모두 포함시키지 못하였음을 의미한다. 이로 인해 실제값이 갖고 있는 계절성을 예측값은 포함시키지 못하고 있는 것이고, 이것이 잔차를 통해서 나타나게 되는 것이라 할 수 있다. 구체적인 주기를 찾기 위해서 주기 분석을 실행한 결과, 두 개의 잔차항 모두에서 최대의 periodogram 값에 해당하는 주기값은 약 3605이다. 분석에 사용된 자료는 하루를 169개의 구간으로 나누고 있으므로 3605는 $21.3(=3605/169)$ 일에 해당하는 값이다. 우리의 분석 자료는 일주일이 5일로 구성되어 있으므로, 21일은 약 한 달에 해당한다. 따라서 Model A와 Model B의 실제 값과 예측 값의 차이인 잔차는 한 달을 주기로 일정한 계절적 변동을 하고 있음을 의미한다. 이는 월 단위의 새로운 계절성 변수가 존재함을 의미한다.

제시된 방법을 통해서 얻은 계절성 주기 변수가 실제로 콜센터 통화량 예측에 기여하는지를 살펴보기 위해서 새로이 발견된 월 단위 계절성 주기를 포함한 계절성 선형 모델과 기존의 변수만을 포함한 계절성 선형 모델의 예측 능력을 비교 평가 하였다. 월 단위 변수를 포함하는 새로운 계절성 선형 모델은 아래와 같다.

$$\text{Model C: } y_{jk} = \mu + \alpha_{d_j} + \beta_k + \delta_{m_j} + \epsilon_{jk}$$

$$\text{Model D: } y_{jk} = \mu + \alpha_{d_j} + \beta_k + \gamma_{d_j k} + \delta_{m_j} + \epsilon_{jk}$$

여기서 δ_{m_j} 은 그 달의 m_j 번째 일이 y 에 미치는 효과로 월 단위 효과를 반영한다. Model C와 Model D는 각기 Model A와 Model B에 월 단위 계절성 변수를 추가한 것이다. 각 모델의 예측 능력 평가는

표 4.2. Root mean square error(RMES)와 average percent error(APE)를 이용한 2003년 7월 25일부터 10월 24일까지의 Model A, B, C, D의 익일 통화량 예측 평가 결과

RMSE	최대값	최소값	평균값	중앙값
Model A	51.400	12.783	21.271	19.522
Model B	45.544	12.052	20.440	18.023
Model C	46.888	12.927	20.380	18.785
Model D	41.553	13.232	19.422	17.827
APE	최대값	최소값	평균값	중앙값
Model A	36.291	5.812	10.173	9.352
Model B	30.245	5.678	9.228	8.061
Model C	37.570	6.120	9.759	8.589
Model D	31.273	6.388	8.748	7.788

아래와 같은 방법으로 실행한다.

- 예측일 바로 앞의 100일치의 자료를 in-sample로 이용하여 익일 통화량을 예측한다.
- 2003년 7월 25일에서 10월 24일까지 총 64일의 통화량을 5분 단위로 예측한다.
- 익일 통화량을 예측한 뒤, 날짜별로 RMSE와 APE를 계산하여 예측 능력을 평가한다.

각 모델의 예측 능력에 대한 평가 결과는 표 4.2에 요약 정리되어 있다. 한 달 주기를 나타내주는 변수를 추가한 Model C는 Model A에 비해서 RMSE와 APE의 평균값 및 중앙값이 더 낮은 것을 볼 수 있다. 즉 이는 잔차의 스펙트럼 분석을 통해서 찾아낸 한 달 주기를 모델에 반영함으로써, 예측치의 정확도가 높아졌음을 의미한다. 마찬가지로 Model D도 Model B에 비해서 RMSE와 APE의 평균값 및 중앙값이 더 낮게 나왔다. 따라서 Model D가 Model B에 비해서 예측의 정확도가 높다고 할 수 있다. 한편, Model C는 Model B에 비해서 예측력이 떨어지는 결과를 보였다. 이와 같은 결과는 월 단위 계절성이 중요하기는 하지만 시간과 요일에 의해서 만들어지는 상호작용의 효과가 통화량 예측에서 보다 중요하다고 해석할 수 있을 것이다. Model A, Model B, Model C, Model D의 전체 비교에 있어서는 시간효과, 요일효과, 시간효과와 요일효과의 상호작용효과 및 한 달 단위 효과를 포함한 Model D에서 예측 오차가 가장 작게 기록되었다.

마지막으로, 한 달 단위 효과 이외에 다른 계절성 효과는 없는지를 검정하기 위해서 Model C와 Model D의 잔차항에 대한 백색 잡음 검정과 스펙트럼 분석을 실시하였다. 백색 잡음 검정 결과를 보면, Model C, Model D의 잔차는 아직도 유의한 자기상관성을 나타내고 있음을 확인하였다. 이는 Model C, Model D에서 반영한 하루 단위 계절성과 일주일 단위 계절성, 그리고 한 달 단위 계절성 이외에도 다른 계절성이 있을 수 있음을 의미한다. 주기 분석 결과, 두 개의 모델에서 최대 periodogram에 해당되는 주기는 똑같이 10816이었다. 이는 Model C 및 Model D를 통해서 예측된 값의 개수와 일치한다. 이는 추가적인 계절성 주기가 의심되지만, 분석에 사용된 자료의 범위의 한계 때문에 구체적인 주기를 찾는 어려운 것으로 해석되어 질 수 있다.

5. 결론 및 향후 연구방향

본 연구에서는 스펙트럼 분석과 계절성 선형 모델을 이용하여 계절성 변수를 찾아내는 방법을 제안하였고, 실증분석을 통하여 기존의 통계적 방법으로는 제시하지 못하였던 월 단위 효과를 찾아내었다. 새로이 발견된 계절성 변수는 계절성 선형 모델에 포함되어 이러한 변수를 포함하지 않는 기존의 계절성 선형 모델과 비교한 결과, 새로운 변수를 포함한 모델의 예측 정확성이 기존 모델보다 우수함을 보였다.

실증분석을 하는 과정에서 월 단위 계절성 보다 더 장기간 단위의 계절성이 존재할 수 있음을 백색 잡음 검정을 통해서 발견할 수 있었다. 그러나, 이러한 계절성은 본 연구에서 이용하였던 자료의 기간 범위로 는 명확히 주장할 수 없었다. 향후 보다 장기간의 범위를 가지고 있는 자료를 이용할 경우, 본 연구에서 제시한 방법으로 보다 장기간의 새로운 주기를 발견할 수 있을 것으로 기대된다.

참고문헌

- Andrews, B. H. and Cunningham, S. M. (1995). L.L. bean improves call center forecasting, *Interfaces*, **25**, 1–13.
- Bartlett, M. S. (1966). *An Introduction to Stochastic Process*, Cambridge University Press, London.
- Bianchi, L., Jarrett, J. and Choudary Hanumara, R. (1993). Forecasting incoming calls to telemarketing centers, *Journal of Business Forecasting Methods and Systems*, **12**, 3–12.
- Brown, L. D., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. H. (2005). Statistical analysis of a telephone call center: A queueing-science perspective, *Journal of the American Statistical Association*, **100**, 36–50.
- Durbin, J. (1967). Tests of serial independence based on the cumulated periodogram, *Bulletin of the International Statistics Institute*, **42**, 1039–1049.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*, John Wiley & Sons, New York.
- Gans, N., Koole, G. and Mandelbaum, A. (2003). Telephone call centers: Tutorial, review and research prospects, *Manufacturing and Service Operations Management*, **5**, 79–141.
- Shen, H. and Huang, J. Z. (2008). Interday forecasting and intra-day updating of call center arrivals, *Manufacturing and Service Operations Management*, **10**, 391–410.
- Taylor, J. W. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center, *Management Science*, **54**, 253–265.
- Tych, W., Pedregal, D. J., Young, P. C. and Davies, J. (2002). An unobserved component model for multi-rate forecasting of telephone call demand: The design of a forecasting support system, *International Journal of Forecasting*, **18**, 673–695.
- Weinberg, J., Brown, L. D. and Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous poisson process with applications to call center data, *Journal of the American Statistical Association*, **102**, 1185–1198.

Spectral Analysis Accompanied with Seasonal Linear Model as Applied to Intra-Day Call Prediction

Taek Soo Shin¹ · Myung Suk Kim²

¹Citi Bank; ²Sogang Business School, Sogang University

(Received October 2010; accepted January 2011)

Abstract

In this paper, a seasonal variable selection method using the spectral analysis accompanied with seasonal linear model is suggested. The suggested method is applied to the prediction of intra-day call arrivals at a large North American commercial bank call center and a significant intra-month seasonal variable is detected. This newly detected seasonal factor is included in the seasonal linear model and is compared with the seasonal linear models without this variable to see whether the new variable helps to improve the forecasting performance. The seasonal linear model with the new variable outperformed the models without it in one-day-ahead forecasting.

Keywords: Call volume forecast, intra-day call arrival, seasonal linear model, seasonal variable selection, spectral analysis.

This work was supported by the Sogang University Research Grant of 2008 (200810048.01). This paper is based on the first author's M.S. thesis at the Sogang University.

²Corresponding author: Associate Professor, Sogang Business School, Sogang University, Shinsu-dong, Mapo-gu, Seoul 121-742, Korea. E-mail: myungkim@sogang.ac.kr