

Functional Forecasting of Seasonality

Geung-Hee Lee^{a,1}

^aDepartment of Information Statistics, Korea National Open University

(Received May 18, 2015; Revised June 15, 2015; Accepted June 24, 2015)

Abstract

It is important to improve the forecasting accuracy of one-year-ahead seasonal factors in order to produce seasonally adjusted series of the following year. In this paper, seasonal factors of 8 monthly Korean economic time series are examined and forecast based on the functional principal component regression. One-year-ahead forecasts of seasonal factors from the functional principal component regression are compared with other forecasting methods based on mean absolute error (MAE) and mean absolute percentage error (MAPE). Forecasting seasonal factors via the functional principal component regression performs better than other comparable methods.

Keywords: functional time series, functional principal component analysis, X-11, seasonal factor, current seasonal adjustment

1. 머리말

경제시계열에는 매년 반복되어서 나타나는 계절변동이 포함되어 있다. 계절변동은 경제시계열에 존재하는 기저적 움직임을 파악하는데 제약요인으로 작용한다. 따라서 계절변동이 포함된 경제시계열을 분석할 때 시계열에서 계절변동을 제거한 계절조정계열을 중심으로 살펴볼 필요가 있다 (Bloem 등, 2001; Lee, 1998). 통계청과 한국은행 등 통계작성기관에서는 우리나라 주요 분기 및 월 통계에 대해 계절조정통계를 원통계와 같이 공표하고 있다.

계절조정계열 작성방식은 크게 동시적 계절조정(concurrent seasonal adjustment)와 현시적 계절조정(current seasonal adjustment, factor projected seasonal adjustment)으로 구분된다. 동시적 계절조정은 매월 또는 매분기별 경제통계 작성 시마다 계절조정 작업을 실시하여 계절조정계열을 작성하는 방식이다 (McKenzie, 1984). 현시적 계절조정은 연간 경제통계 작성시 시계열을 예측한 후 계절조정방법을 적용하여 1년 후 계절변동을 예측하고 향후 1년간 이를 원계열에서 제거하여 계절조정계열을 작성하는 방식이다. 통계청과 한국은행 등 통계작성기관에서는 현시적 계절조정 방식으로 계절조정계열을 작성하고 있다. 이는 동시적 계절조정의 경우 최근의 계절조정계열이 매월 개정해야 되기 때문에 발생하는 개정 횟수를 줄이고자 원계열의 수정이 이루어지는 연간통계 작성시 1회 개정하는 현시적 계절조정을 채택하고 있다. 현시적 계절조정을 할 경우 계절변동을 효과적으로 예측하는 것이 계절조정계열의 품질 향상을 위해 무엇보다 중요하다. X-11에서는 단순 방법에 의해 1년앞 계절변동을 예측하

The work was supported by the Korea National Open University in 2014.

¹Department of Information Statistics, Korea National Open University, 86, Daehak-ro, Jongno-Gu, Seoul 110-791, Korea. E-mail: geunghee@knou.ac.kr

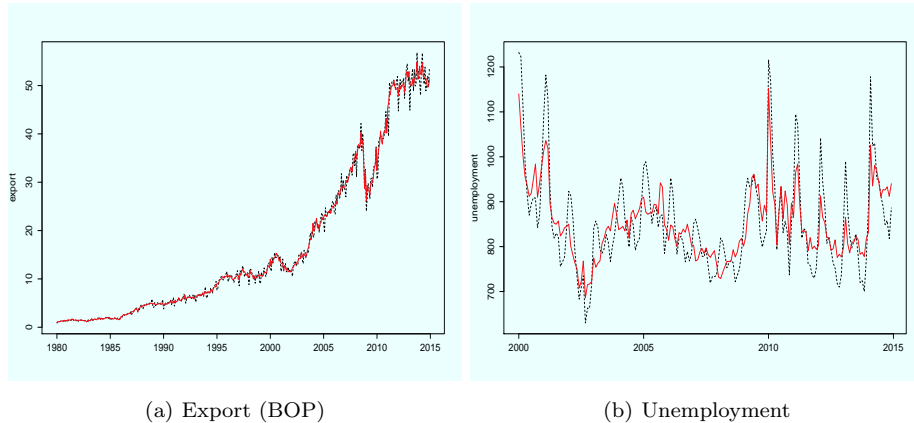


Figure 2.1. Monthly export and monthly unemployment (black dotted line: original series, red bold line: seasonally adjusted series).

고, X-12-ARIMA와 TRAMO-SEATS는 경제시계열의 예측결과를 이용하여 계절변동을 예측하고 있다. 계절변동은 1년 단위로 비슷한 함수적 형태를 지니면서 서서히 변하고 있으므로 계절변동은 일종의 함수적 시계열(functional time series)로 정의될 수 있다. 함수적 시계열은 함수적 주성분분석 등의 방법으로 함수 자체를 예측할 수 있다 (Hyndman과 Shang, 2009). 함수적 시계열을 이용한 연구로는 Hyndman과 Ullah (2007), Hyndman과 Shang (2009), Shang (2011, 2012, 2013)을 참조하면 된다. 본 연구에서는 함수적 시계열 모형을 적용하여 향후 1년간 계절변동을 예측하는 방안을 마련하고 기존의 방법과 비교하여 유용성을 파악하고자 한다. 본 연구에서는 함수적 형태가 복잡한 월별 시계열의 계절변동을 중심으로 살펴보았다.

본 연구의 구성은 다음과 같다. 제2장에서는 우리나라 시계열의 계절변동의 특성을 살펴본다. 제3장에서는 계절변동을 함수적 시계열모형을 바탕으로 모형화한다. 제4장에서는 함수적 시계열모형에 의한 향후 1년의 계절변동을 예측하고 다른 예측모형과 예측오차 측면에서 비교하여 그 유용성을 평가한다. 마지막으로 제5장에서 연구결과를 정리하고 향후 연구방향을 정리한다. 함수적 시계열분석을 위해서는 R의 ftsa 패키지를 이용하였다 (Hyndman과 Shang, 2015).

2. 우리나라 시계열의 계절변동의 특징

우리나라 경제시계열은 대부분 계절변동을 포함하고 있다. 계절변동은 X-13ARIMA -SEATS와 같은 계절조정방법에 의해 식별되고 있고 이를 원계열에서 제거하여 계절조정계열을 산출하고 있다 (Lee와 Lee, 2013; U.S. Census Bureau, 2012). Figure 2.1은 월별 수출(국제수지 기준)과 월별 실업자의 원계열과 계절조정계열이다. 이를 보면 월별 수출과 월별 실업자는 계절변동을 포함하고 있으며 계절조정계열은 평활하게 나타났다. Figure 2.2는 월별 수출(국제수지 기준)과 월별 실업자의 원계열과 계절조정계열의 스펙트럼인데 점선으로 표시된 계절 주파수에서 원계열의 경우 큰 값을 보이나 계절변동이 제거된 계절조정계열은 그렇지 않은 것으로 나타났다.

계절변동은 원계열을 계절조정계열로 나누거나 빼서 구하는데 여기에는 순수 계절변동 외에도 영업일수 변동, 명절변동 등이 포함되어 있다. 계절변동은 시간에 따라 표현하거나 x 축을 월로 두고 연도별로 계절변동을 표현할 수 있다. 전자는 시계열적 표현이며 후자는 1년간 계절변동을 월의 함수로 보는 함수적 시계열 표현이다. Figure 2.3은 2000년 이후 월별 수출과 월별 실업자의 계절변동을 시간에 따라 표현한 그래프이고, Figure 2.4는 2000년 이후 월별 수출과 월별 실업자의 계절변동을 x 축을 월로 하여

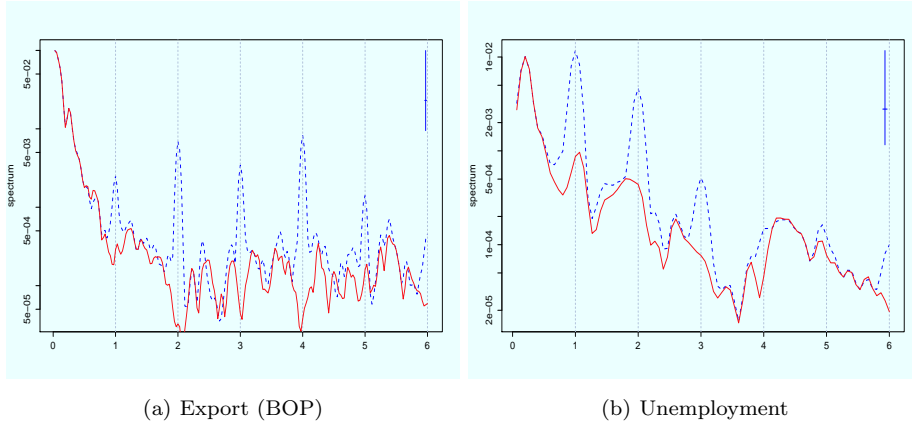


Figure 2.2. Spectrum of export and unemployment(blue dotted line: original series, red bold line: seasonally adjusted series).

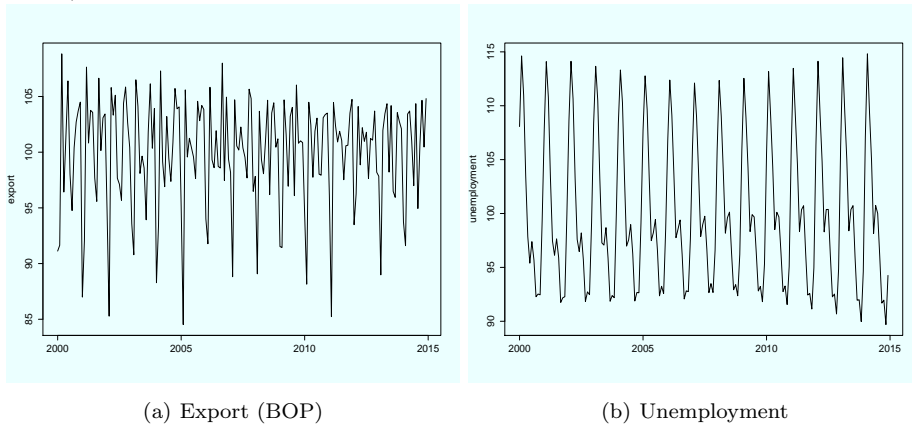


Figure 2.3. Seasonal factors of export and unemployment viewed as time series.

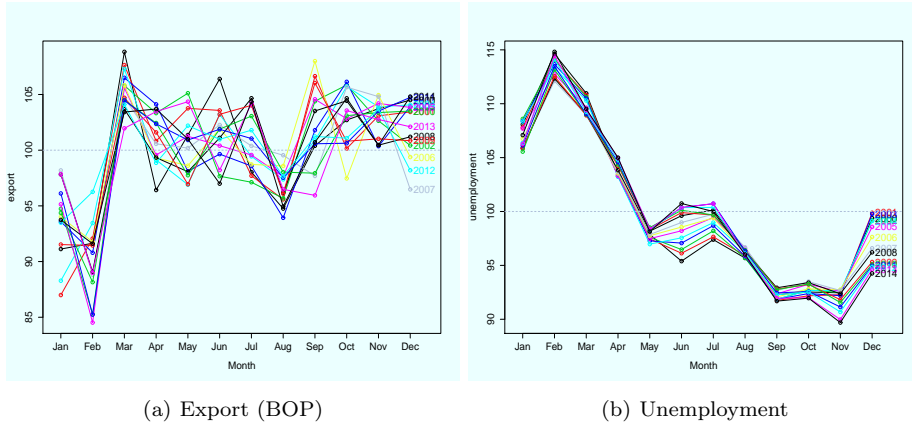


Figure 2.4. Seasonal factors of export and unemployment viewed as functional time series.

함수적으로 표현한 그래프이다. Figure 2.3과 Figure 2.4를 보면 계절변동이 시간에 따라 변하고 있음을 알 수 있다. 특히 Figure 2.4의 함수적 시계열 그래프를 보면 계절변동의 움직임을 보다 잘 파악할

수 있다. 수출의 계절변동의 경우 시간에 따른 패턴이 보이지 않으나 실업자는 시간에 따라 어느 정도 일정한 패턴으로 움직이는 것으로 나타났다. 이를 보면 계절변동의 함수적 형태 그 자체를 예측하는 것이 유용함을 알 수 있다.

3. 계절변동의 함수적 시계열모형

이 절에서는 계절변동을 함수적 시계열모형을 이용하여 예측하는 방법을 정리하였다. 계절변동을 연도와 월/분기를 이용하여 함수적 시계열로 표현될 수 있다. Hyndman과 Ullah (2007)의 표현방식에 따르면 연도 t 와 월/분기 s 에서 측정된 계절변동 $y_t(s)$ 는 기저의 평활한 계절변동 함수와 오차로 구성되어 있다고 가정된다. 분기 시계열의 경우 $s = 1, 2, 3, 4$ 이고, 월별 시계열의 경우 $s = 1, 2, \dots, 12$ 이다. 관측된 계절 변동을 $\{s, y_t(s)\}$ 와 같이 측정된다면 측정된 계절변동 $y_t(s)$ 은 식 (3.1)과 같이 표현된다.

$$y_t(s) = f_t(s) + \sigma_{t,s}\epsilon_{t,s}, \quad (3.1)$$

여기서 $f_t(s)$ 는 기저의 계절변동 함수이며, $\epsilon_{t,s}$ 는 서로 독립적인 표준정규분포를 따른다. $\sigma_{t,s}$ 는 계절에 따른 표준편차로 오차항의 이분산성을 표현한다. $f_t(s)$ 는 평균함수와 직교 함수 주성분(functional principal components)과 상관관계가 없는 주성분 점수의 곱으로 식 (3.2)와 같이 표현된다.

$$f_t(s) = \mu(s) + \sum_{k=1}^{\infty} \beta_{t,k} \phi_k(s) + e_t(s), \quad (3.2)$$

여기서 $\mu(s)$ 는 평균 모함수이고 $\beta_{t,k}$ 는 k 번째 주성분 점수이고, $\phi_k(s)$ 는 k 번째 정규직교(orthonormal) 주성분 함수이다. $e_t(s)$ 는 평균이 0이며 독립적인 오차항이다. 기저의 계절변동 함수 $f_t(s)$ 는 식 (3.3)과 같이 추정된다.

$$\hat{f}_t(s) = \hat{\mu}(s) + \sum_{k=1}^K \hat{\beta}_{t,k} \hat{\phi}_k(s) + \hat{e}_t(s), \quad (3.3)$$

여기서 $\hat{\mu}(s)$ 는 추정된 평균함수이다. $\hat{\phi}_k(s)$ 는 추정된 k 번째 직교정규(orthonormal) 고유함수인 주성분 함수이며 식 (3.4)의 함수적 시계열의 공분산을 바탕으로 추정된다 (함수적 데이터의 주성분 함수는 일반적인 다변량 주성분분석과 마찬가지로 분산을 최대로 하고 서로 직교하도록 구해진다). $\hat{\beta}_{t,k}$ 는 추정된 t 년의 k 번째 주성분 함수의 점수이다. $\hat{\beta}_{t,k}$ 와 $\hat{\beta}_{t,l}$ ($k \neq l$)의 상관관계는 0이며 K 는 최적 차수이다. 함수적 데이터의 주성분 함수는 일반적인 다변량 주성분분석과 마찬가지로 서로 직교적이면서 분산이 최대가 되도록 하여 구해진다. 자세한 내용은 Ramsay와 Silverman (2005)을 참조하면 된다.

$$\hat{\Gamma}(s) = \frac{1}{n} \sum_{t=1}^n [f_t(s) - \hat{\mu}(s)][f_t(s) - \hat{\mu}(s)]. \quad (3.4)$$

시계열이 1 ~ n 년까지 존재할 때 h 년 앞 함수 시계열모형의 예측값은 식 (3.5)와 같이 표현된다.

$$\hat{y}_{t+h|n}(s) = \hat{\mu}(s) + \sum_{k=1}^K \hat{\beta}_{n+h|t,k} \hat{\phi}_k(s), \quad (3.5)$$

여기서 $\hat{\beta}_{n+h|t,k}$ 는 $\hat{\beta}_{1,k}, \dots, \hat{\beta}_{n,k}$ 를 바탕으로 시계열모형으로 예측된다.

Figure 3.1은 월별 수출의 2차 함수주성분과 관련된 그래프이다. Figure 3.1의 상단을 보면 평균 함수(main effect)는 1, 2월이 작고, 12월은 큰 모습을 보이고 있다. 첫 번째 주성분함수는 1, 2, 10월이

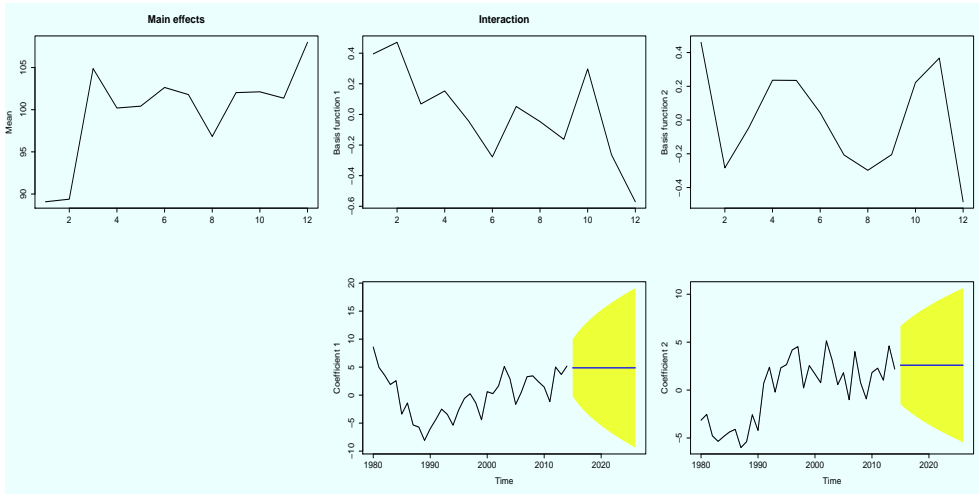


Figure 3.1. Basis function and associated coefficients for seasonal factors of monthly export.

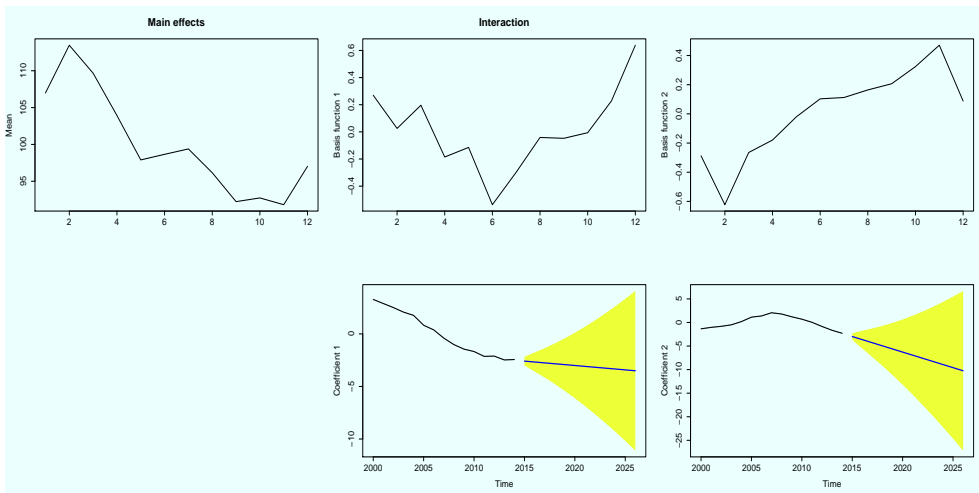


Figure 3.2. Basis function and associated coefficients for seasonal factors of monthly unemployment.

크고, 12월이 작으며, 두 번째 주성분 함수는 1, 4, 5, 10, 11월이 큰 모습을 보이고 있다. Figure 3.1의 하단은 주성분함수의 점수 계열의 그래프인데 예측값과 신뢰구간이 나타나 있다 (주성분함수의 점수 계열은 ARIMA모형으로 예측되었으며 신뢰구간의 신뢰수준은 95%이다). 이를 보면 점수의 예측에 특별한 패턴이 없게 나타났다. Figure 3.2는 월별 실업자의 2차 함수주성분과 관련된 그래프이다. Figure 3.2의 상단을 보면 평균 함수(main effect)는 1, 2, 3월이 크고, 9, 10, 11월이 작은 모습을 보이고 있다. 첫 번째 주성분함수는 6월이 작고, 12월은 큰 모습을 띠고, 두 번째 주성분 함수는 2월은 작고, 11월은 큰 모습을 보이고 있다. Figure 3.2의 하단을 보면 주성분함수 점수 계열의 예측값이 하락하는 체계적 변화가 있음을 볼 수 있다 (계절변동 함수의 패턴이 서서히 움직이는 경우 주성분 함수 점수의 체계적 변동이 나타나며, 이를 시계열모형으로 예측하는 경우 단위근이 나타나므로 계절변동 함수를 차분하고 차분계열로 예측하는 경우 보다 유용하다).

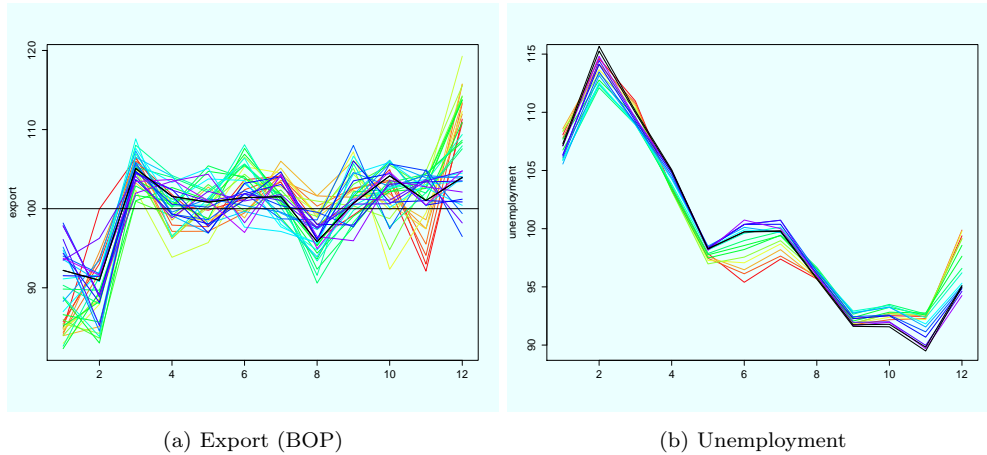


Figure 3.3. Functional forecasts of seasonal factors of export and unemployment.

Figure 3.3은 월별 수출과 월별 실업자의 계절변동을 2년간 함수적으로 예측한 결과이다. 검정색으로 표시한 결과가 예측이고 적색에서 보라색으로 갈수록 최신 계절함수이다. 이를 보면 월별 수출의 계절 변동 예측값은 평균적으로 움직이지만 월별 실업자의 계절변동은 1분기는 커지고 4분기는 작아지는 패턴을 반영하여 예측되었다.

4. 계절변동의 함수적 예측의 성과 비교

현시적 계절조정에서 이용되는 계절변동의 예측에 있어서 함수적 예측의 유용성을 파악하기 위해서 기존의 방법들과 예측력을 비교하였다. 이 연구에서는 8개의 월별 경제시계열을 이용했다. 월별 시계열로는 수출(국제수지 기준, export), 수입(국제수지 기준, import), M1, M2, 광공업생산지수, 생산자제품재고지수, 취업자, 실업자의 계절변동을 이용하였다 (수출과 수입은 국제수지 통계의 월별 통계이고, M1, M2는 통화통계이다. 광공업생산지수(production index), 생산자제품재고지수(inventory index)는 산업동향 통계이며, 취업자(employment)와 실업자(unemployment)는 경제활동인구조사의 통계이다. 광공업생산지수, 생산자제품재고지수, 취업자와 실업자는 통계청에서, 수출, 수입, M1와 M2는 한국은행에서 매월 원계열과 계절조정계열을 작성하여 발표하고 있다). 계절변동은 발표된 원계열을 계절조정계열로 나누어 구했다. 예측력을 비교하기 위한 예측기간은 2010년부터 2014년까지 5년간이며, 예측은 1년 단위로 향후 1년간 실시한 후 예측력을 비교하였다.

비교대상 예측방법으로는 식 (4.1)과 같은 임의보행모형(RW), ARIMA모형, 식 (4.2)의 X-11방법과 아울러 식 (4.3)에서 차수 K 를 1, 2, 3으로 달리한 함수적 시계열모형(FPCA)을 고려하였다. ARIMA모형에 의한 예측모형은 BIC를 최소로 하는 계절ARIMA모형을 이용하였다 (R 프로그램의 forecast 패키지의 auto.arima문을 이용하였다). 함수적 예측에서 취업자와 실업자와 같이 계절변동의 함수가 일정한 패턴으로 변하는 경우 식 (4.4)와 같이 계절변동의 차분계열을 예측한 후 이 값을 전년의 계절변동을 더하여 예측값을 구했다 (함수적 시계열에 일정한 패턴이 있는 경우 예측력 향상을 위해 함수적 시계열을 그대로 예측하지 않고 차분한 함수적 시계열을 바탕으로 예측한 후 이를 전환해 예측값을 구했다).

$$\text{임의보행모형} : \hat{y}_{t+1}(s) = y_t(s), \quad (4.1)$$

$$\text{X-11법} : \hat{y}_{t+1}(s) = y_t(s) + \frac{y_t(s) - y_{t-1}(s)}{2}, \quad (4.2)$$

Table 4.1. Comparison of forecasting performance based on MAE

	RM	ARIMA	X-11	FPCA(1)	FPCA(2)	FPCA(3)
Export	2.299	2.123	2.879	1.836	1.836	1.751
Import	2.353	2.169	2.810	1.855	1.810	1.810
Production Index	1.748	1.239	2.548	1.171	1.171	1.110
Inventory Index	0.177	0.173	0.212	0.126	0.126	0.126
M1	0.174	0.389	0.250	0.154	0.154	0.154
M2	0.131	0.100	0.175	0.086	0.085	0.085
Employment	0.020	0.011	0.013	0.011	0.009	0.009
Unemployment	0.265	0.961	0.236	0.164	0.164	0.164

Table 4.2. Comparison of forecasting performance based on MAPE

	RM	ARIMA	X-11	FPCA(1)	FPCA(2)	FPCA(3)
Export	2.345	2.342	2.943	1.855	1.875	1.785
Import	2.369	2.243	2.283	1.865	1.821	1.821
Production Index	1.783	1.178	2.605	1.242	1.242	1.181
Inventory Index	1.768	1.749	2.116	1.260	1.260	1.260
M1	1.738	1.762	2.490	1.535	1.499	1.489
M2	1.306	1.003	1.750	0.940	0.927	0.927
Employment	0.0200	0.0106	0.0130	0.0316	0.0168	0.0136
Unemployment	0.2652	0.2342	0.2367	0.4438	0.2337	0.2338

$$\hat{y}_{t+h|n}(s) = \hat{\mu}(s) + \sum_{k=1}^K \hat{\beta}_{n+h|t,k} \hat{\phi}_k(s), \tag{4.3}$$

$$\Delta \hat{y}_{t+1|n}(s) = \hat{\mu}(s) + \sum_{k=1}^K \hat{\beta}_{n+1|t,k} \hat{\phi}_k(s) \rightarrow \hat{y}_{t+1|n}(s) = y_t(s) + \Delta \hat{y}_{t+1|n}(s). \tag{4.4}$$

예측력을 식 (4.5)의 평균절대오차(MAE)와 식 (4.6)의 평균절대퍼센트오차(MAPE)를 이용하여 비교하였다. 여기서 T_1, T_2 은 각각 2010, 2014이며 월별데이터이므로 $S = 12$ 이다. 또한 $\hat{y}_{n+t|n+t-1}(s)$ 는 $n + t - 1$ 년에서 1년 앞 s 월의 예측값이다.

$$AME = \frac{1}{S(T_2 - T_1 + 1)} \sum_{t=T_1}^{T_2} \sum_{s=1}^S |y_{n+t}(s) - \hat{y}_{n+t|n+t-1}(s)|, \tag{4.5}$$

$$MAPE(\%) = \frac{1}{S(T_2 - T_1 + 1)} \sum_{t=T_1}^{T_2} \sum_{s=1}^S \left| \frac{y_{n+t}(s) - \hat{y}_{n+t|n+t-1}(s)}{y_{n+t}(s)} \right|. \tag{4.6}$$

Table 4.1은 월별 데이터 계절변동의 예측오차 즉 절대평균예측오차(MAE)를 구한 것인데 이를 보면 식 (3.5)의 함수적 시계열모형에 의한 예측이 다른 방법에 의한 예측보다 우수하게 나타났다. 또한 식 (3.5)의 함수적 시계열모형의 차수 K 에 따라 함수적 예측오차가 크게 차이나지 않는 것으로 나타났다. Table 4.2는 월별 데이터 계절변동의 평균절대퍼센트오차(MAPE)를 구한 것인데 이를 보면 취업자와 실업자를 제외한 시계열의 경우 함수적 시계열모형에 의한 예측이 다른 방법에 의한 예측보다 우수하게 나타났다. 한편 취업자와 실업자의 경우 전체적으로 예측오차가 전반적으로 작았고 함수적 시계열모형에 의한 예측이 기존 예측과 비슷하게 나타났다.

5. 맺음말

통계청과 한국은행 등 통계작성기관은 발표되는 원통계에서 예측된 계절변동을 제거하여 계절조정계열을 작성하고 있다. 따라서 1년 앞 계절변동을 정확히 예측한다면 품질 높은 계절조정계열을 작성할 수 있다. 이제까지 계절변동은 원통계를 ARIMA모형으로 예측한 후 X-11방법을 적용하여 예측되었다. 그런데 계절변동의 패턴이 안정적이지 않은 경우 기존의 예측방법으로 예측하는 데에는 제약이 있다. 계절변동이 함수적 특성을 가지므로 계절변동을 함수적 시계열로 전환하고 이를 함수적 시계열모형으로 예측하는 것을 시도해 볼 필요가 있다. 본 연구에서는 함수적 시계열모형을 이용하여 8개의 월별 경제 시계열의 향후 1년간 계절변동을 예측하고 이를 기존의 예측방법에 의한 예측과 비교했다. 비교 결과 함수적 시계열모형에 의한 예측이 기존의 예측에 비해 대체로 우수하게 나타났다.

본 연구의 함수적 예측은 계절조정은 물론 다른 경제시계열분석으로 확장될 수 있다. 계절조정을 할 때 계절변동에는 명절변동, 영업일수변동이 포함되어 있으므로 이를 분리하여 각각 함수적 시계열모형으로 예측하면 계절변동을 보다 정교히 추정할 수 있다. 또한 다양한 경제 시계열을 함수적으로 살펴보고 이를 함수적 시계열모형을 통해 분석, 예측할 수 있다.

References

- Bloem, A. M., Dippelsman, R. J. and Maehle, N. O. (2001). *Quarterly National Accounts Manual-Concepts, Data Sources, and Compilation*, International Monetary Fund.
- Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series (with discussion), *Journal of the Korean Statistical Society*, **38**, 199–221.
- Hyndman, R. J. and Shang, H. L. (2015). Package *ftsa*, <http://cran.r-project.org/web/packages/ftsa/>
- Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach, *Computational Statistics and Data Analysis*, **51**, 4942–4956.
- Lee, G.-H. (1998). X-12 ARIMA seasonal adjustment in Korean economic time series, *Economic Analysis*, **4**, 205–242.
- Lee, G.-H. and Lee, H. (2013). Seasonal adjustment in Korean economic time series with X-13ARIMA-SEATS, *Quarterly National Accounts Review*, **4**, 205–242.
- McKenzie, S. K. (1984). Concurrent seasonal adjustment with census X-11, *Journal of Business and Economic Statistics*, **2**, 235–249.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed, Springer, New York.
- Shang, H. L. (2011). A survey of functional principal component analysis, Working Paper 06/11, Monash University.
- Shang, H. L. (2012). Point and interval forecasts of age-specific fertility rates: A comparison of functional principal component methods, *Journal of Population Research*, **29**, 249–267.
- Shang, H. L. (2013). Functional time series approach for forecasting very short-term electricity demand, *Journal of Applied Statistics*, **40**, 152–168.
- U.S. Census Bureau (2012). *X-13ARIMA-SEATS Reference Manual*, Statistical Research Division, U.S. Census Bureau.

계절변동의 함수적 예측

이금희^{a,1}

^a한국방송통신대학교 정보통계학과

(2015년 5월 18일 접수, 2015년 6월 15일 수정, 2015년 6월 24일 채택)

요약

통계청과 한국은행 등 통계작성기관에서 이용되고 있는 계절조정 연간 경제통계 작성시 시계열을 예측한 후 계절조정방법을 적용하여 1년 후 계절변동을 예측하고 원통계 작성시 원통계에서 이를 제거하여 계절조정계열을 작성하고 있다. 이 경우 계절변동을 효과적으로 예측하는 것이 계절조정계열의 품질 향상을 위해 무엇보다 중요하다. 계절변동은 1년 단위로 비슷한 함수적 형태를 지니면서 변하므로 계절변동은 일종의 함수적 시계열이다. 함수적 시계열은 함수적 주성분분석을 바탕으로 한 함수적 시계열모형으로 예측할 수 있다. 본 연구에서는 함수적 시계열 모형을 이용하여 향후 1년간 계절변동을 예측하는 방안을 마련하고 X-11 방식 등 기존의 예측방법과 비교하여 유용성을 파악하였다.

주요용어: 함수적 시계열, 함수적 주성분분석, X-11, 계절변동, 현시적 계절조정

이 논문은 2014년 한국방송통신대학교 전기 학술연구비의 재정지원을 받아 작성된 것임.

¹(110-791) 서울특별시 종로구 대학로 86, 한국방송통신대학교 정보통계학과. E-mail: geunghee@knou.ac.kr