

시계열 해석

[영화 관객 예측 모델을 활용한 극장·VOD 동시상영 시점 예측]

2018110498 김평진

2018110478 정성훈

1. 분석배경 및 목적

국내 영화산업은 1 차시장인 영화관과 최근 급 부상중인 VOD, OTT 와 같은 2 차 시장의 상영매체로 구분된다. 전통적인 영화산업의 수익구조는 영화 – VOD – 순으로, 다른 매체로 넘어가기 전까지 각 매체 별 상영기간을 '홀드백' 이라고 한다. 일반적으로 90 일의 홀드백을 통해 상영사의 상영권리를 바탕으로 극장의 수익을 보장해주던 옛날과 달리, 2021 년 미국의 최대 극장 체인 AMC 에서는 변화된 인터넷 시장의 속도에 맞춰 홀드백 기준을 90 일에서 17 일로 단축하는 계열을 체결했다. 영화가 인터넷 매체와 동시 개봉하는 사례가 많으며, 초반 흥행 실패 시 빠른 VOD 전환으로 수익을 채우는 것이 최근 영화산업의 동향이다. 이러한 VOD 시장의 성장으로 기존 손익분기점의 기준 또한 낮아졌다. 기존 손익분기점은 평균 제작비 100 억당 관객수 300 만명이 기준이었으나, 최근 영화의 경우 VOD 수익을 반영하여 250 만명으로 감소하였다. 실제로 영화 '#살아있다'는 기존 손익분기점이 220 만명이었으나, VOD 의 높은 수익으로 인해 190 만으로 손익분기점을 돌파하였다. 이처럼 영화의 수익구조가 바뀐 지금 영화의 적절한 홀드백 시점은 배급사의 수익성을 극대화 할 수 있다. 이에 본 조는 영화추세에 대한 해석과 이를 통한 개봉 후 적절한 홀드백 시점을 예측하여 제시하고자 한다.'

2. 분석방법

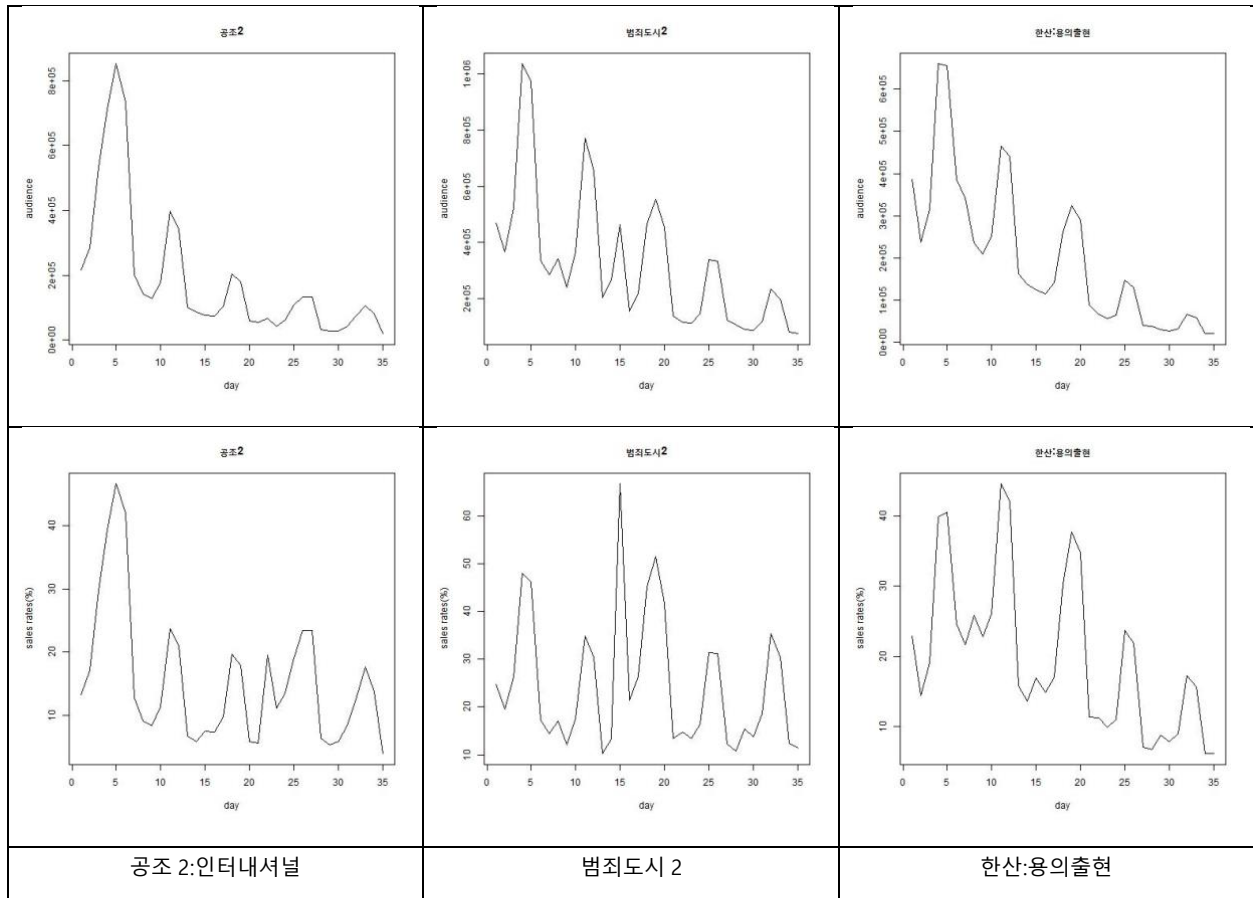
데이터는 KOBIS 영화관입장권통합전산망 홈페이지의 점유율 통계자료를 사용하였다. 빠른 홀드백은 극장의 수익을 감소시키며 너무 늦으면 영화의 인기가 시들어 VOD 수익이 감소할 수 있다. 이를 산정하기 위해서 배급사와 상영사의 손익분기점 계약에 대한 정보가 필요하나, 계약내용은 비공개이며 영화마다 천차만별 이기에 본 분석에서는 이를 고려할 수 없다. 따라서 수익성과 인기를 가늠할 수 있는 지표로 관객수와 좌석판매율을 이용하였다. 분석 기간을 길게 설정할 경우 신뢰도는 높으나 이미 관객수가 적은 단계에서 예측을 하는 것이므로 분석의 유용성이 떨어지며 너무 적은 기간 또한 분석의 신뢰성을 떨어트린다.

따라서 여러 시점을 설정해 예측 값을 비교한 결과 개봉 후 5주차인 35일의 데이터가 적합하다고 판단하였으며, 이를 통해 관객수 데이터와 좌석판매율 데이터를 분해법, 차분을 이용해 2개 방식으로 모델링한다. 70일 이후로는 스크린수가 한자릿수로 줄어들어 분석의 신뢰성이 부족할 있다고 판단하여 36일부터 약 70일까지의 값을 예측했으며, 이후 산출된 예측 값과 실제 값의 RMSE 와 AIC 를 비교하여 최종모델을 선정하였다. 관객수 데이터에서 실제 값이 예측 값의 구간추정치인 예측구간을 벗어나기 시작하는 시점부터는 기존의 추세만큼 관객들이 영화관을 찾지 않아 VOD 가 영화관의 수익에 큰 영향을 주지 않는 최소 홀드백 시점으로 지정하였으며, 좌석 판매율 데이터는 팔린 좌석/전체 좌석으로 영화의 인기를 나타내는 척도로서 마찬가지로 예측 값의 구간 추정치를 벗어나는 시점부터는 기존만큼 사람들이 찾지 않아 영화의 인기가 떨어지는 것으로 해당 시점에 VOD 를 개봉해야 되는 최대 홀드백 시점으로 지정하였다.

최종적으로 최소에서 최대 홀드백 사이를 적정 홀드백 기간으로 산정한 뒤 실제 VOD 개봉 시점과 비교하여 분석결과의 타당성을 확인할 것이다. 분석 대상은 영화 '범죄도시 2', '공조 2:인터내셔널', '한산:용의출현' 으로 2022 년 국내영화 중 누적관객수 700 만 이상인 홀드백기간이 유의한 영화로 선정했다.

3. 분석과정

1) 데이터 탐색 및 해석



위 데이터는 35일간 일별 관객수, 좌석판매율 그래프이다. 두 그래프 모두 주말에 값이 높고 주중에 값이 낮아지는 주기가 7인 계절성분을 확인할 수 있다. 또한 시간이 지날수록 관객수는 급격하게 감소하는 것과 달리 좌석판매율은 크게 감소하지 않는 추세를 확인할 수 있으며 이를 통해 관객수와 인기를 파악할 수 있다. 또한 폭이 감소하는 이분산성 역시 확인되어 종속변수 변환을 고려하였다.

3개 영화 모두 특정 주차 수요일에 값이 증가하는 것을 확인할 수 있으며 이는 매달 마지막 주 수요일에 문화가 있는 날의 영화할인으로 인해 관람객이 증가하는 현상으로 파악된다. 또한 추석 등 공휴일에 관객수가 높아지는 경향이 확인되므로 문화가 있는 날, 공휴일에 대한 지시함수를 모형에 반영하였다. 공조 2의 경우 전체 관객수는 줄었지만 인지도는 유지가 되고 있음을 보이고 한산의 경우 개봉 초기보다 인지도가 감소했음을 알

수 있다. 범죄도시 2의 경우 다른 두 영화보다 관객수가 완만하게 감소하며 인지도가 비교적 유지됨을 확인할 수 있다.

2) 데이터 전처리

공조 2:인터내셔널	범죄도시 2	한산:용의출현
0.019	0.005	0.092
0.128	-0.02	0.142

<Boxcox lambda>

3개 영화에 대해 일별 관객수, 좌석 판매율에 대해 이분산성해소를 위해 변수변환을 진행하였다. 변환 강도를 파악하기 위해 Box-Cox lambda 값을 확인했으며 0일 때 log 변환, 0.5일 때 제곱근 변환, 1일 때 원변수 그대로, 그외에 Box-Cox 변환하도록 기준을 정하였다. 6개 데이터에 대해 lambda 값을 확인한 결과 0에 근사한 값으로 확인되어 log 변환과 큰 차이가 없으므로 log 변환을 진행하였다.

3) 분해법

분해법은 시계열 변동성분을 추세, 계절, 순환, 불규칙성분으로 나누어 예측하는 전통적인 분석방법이다. 일반적으로 추세-계절-불규칙성분 순으로 제거하며 이는 추세성분과 계절성분을 제거함으로써 정상성을 만족시킨 후 ARMA 모형을 적합시키기 위함이다.

먼저 필요한 추세성분을 파악하기 위해 t 항, t^2 항 포함모형을 AIC 차이와 adjust_R^2 값을 통해 비교했으며 검정결과는 다음과 같다.

영화	변수	Anova(t2-t)	Anova(t-1)	AIC(t2-t)	AIC(t-1)
공조 2: 인터내셔널	일별 좌석수	0.22	<0.0001	0.35	-30.04
	좌석판매율	0.3	0.02	0.82	-4.14
범죄도시2	일별 좌석수	0.57	<0.0001	1.63	-24.73
	좌석판매율	0.54	0.19	1.57	0.13
한산: 용의출현	일별 좌석수	0.09	<0.0001	-1.27	-48.31
	좌석판매율	0.09	<0.0001	-1.21	-17.6

<추세성분 변수선택 ANOVA 분석>

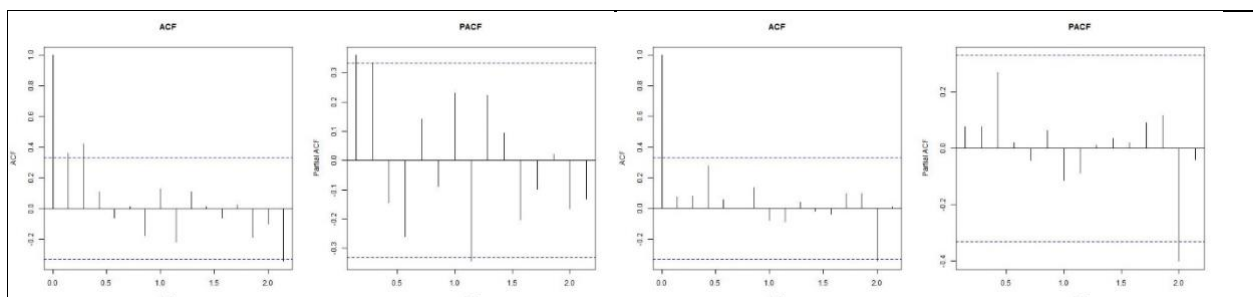
각 변수의 유의성 검정값과 AIC 차이를 통해 추세성분을 확인하였다. 일별좌석수의 경우 t 일차항 변수가 공통적으로 유의하며 좌석판매율의 경우 공조 2:인터내셔널과 범죄도시 2가

추세성분이 유의하지 않는 것으로 확인할 수 있다. 따라서 두 경우는 추세성분을 적합하지 않고 이분산 변환한 값 그대로 계절성분을 적합한다. 다른 경우들은 추세성분을 적합한 나머지 잔차에 계절성분을 적합한다.

추세성분을 제외한 잔차에서 계절성분과 앞서 파악한 문화가 있는 날, 공휴일에 대한 지시함수를 같이 반영하였다. 또한 공조 2:인터내셔널과 범죄도시 2의 좌석판매율은 원인을 알 수 없는 값이 증가하는 구간에 대해 지시함수를 넣어 정상화한다. 계절성분 주기항은 검정결과와 adj_R^2 , AIC 값을 비교해 선택했고 분해법에 사용할 최종 변수는 다음과 같다.

영화	데이터	추세성분	계절성분
공조 2: 인터내셔널	일별 좌석수	t	2 주기항 + 문화가 있는 날 + 공휴일
	좌석판매율	-	2 주기항 + 개봉 1 주차 + ind1 + 문화가 있는 날 + 공휴일
범죄도시 2	일별 좌석수	t	2 주기항 + ind1 + 공휴일
	좌석판매율	-	3 주기항 + 공휴일
한산: 용의출현	일별 좌석수	t	2 주기항 + 문화가 있는 날 + 공휴일
	좌석판매율	t	2 주기항 + 문화가 있는 날 + 공휴일

추세성분과 계절성분의 독립관계를 가정할 근거가 없으므로 한번에 회귀모형을 적합했고 잔차에 대한 백색잡음과정 검정을 실시했다. 백색잡음과정은 등평균 검정, 등분산 검정, 자기상관계수 검정인 포트먼트 검정, ACF, PACF 그래프, auto.arima 인자값을 통해 확인하였다. 검정결과는 다음과 같이 확인된다.

						
영화	데이터	등평균	등분산	포트먼트(lag=7)	평균=0	Auto.arima
한산	관객수	0.264	0.147	기각	0.5	Arima(1,0,0)
	판매율	0.436	0.145	기각	0.5	Arima(1,0,0)
공조 2:인터내셔널	관객수	0.891	0.31	0.215	0.5	Arima(0,0,0)
	판매율	0.61	0.578	0.6817	0.5	Arima(0,1,1)(0,0,1)[7]
범죄도시 2	관객수	0.618	0.901	0.402	0.5	Arima(0,0,0)
	판매율	0.673	0.523	0.123	0.5	Arima(0,0,0)

<한산 관객수 및 공조 2:인터내셔널 좌석판매율 ACF PACF 그래프>

3-1) 영화 데이터별 성분제거 잔차검정

범죄도시 2 데이터와 공조 2 일별 관객수 데이터는 분해법만으로 백색잡음과정을

만족하였으며 한산데이터는 자기상관을 따르지않아 ARMA 적합이 필요해보인다. 공조 2

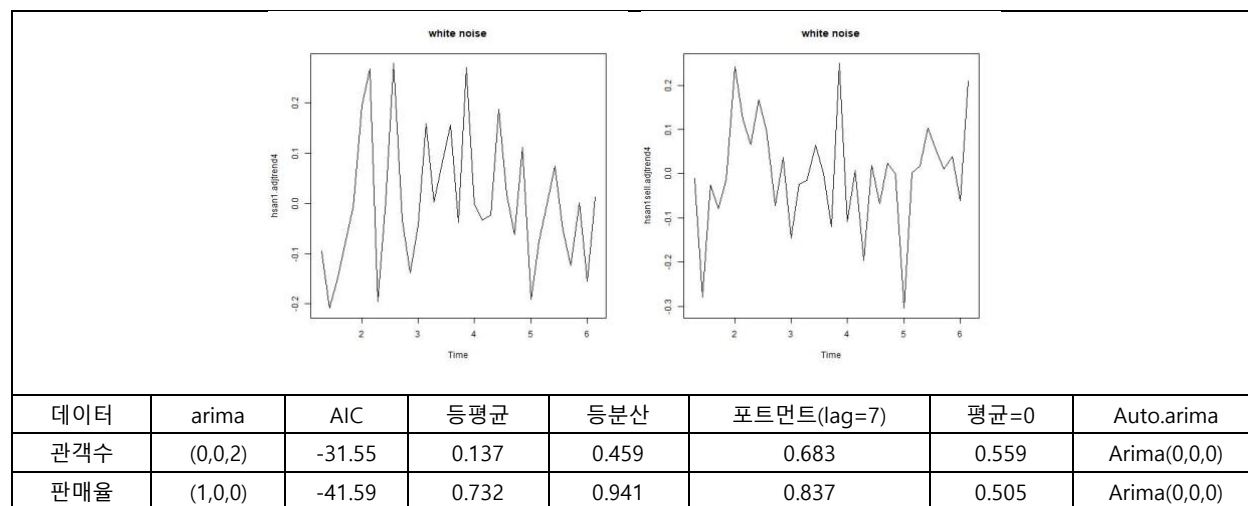
좌석판매율의 경우 auto.arima 결과는 계절차분이 필요하지만 acf, pacf 및 다른 검정값을 볼 때 정상성 및 자기상관을 만족한다고 보아 백색잡음과정으로 판단하였다.

한산:용의출현을 제외한 4 개 데이터의 잔차를 백색잡음과정으로 판단해 최종모델로

선정했고 등분산, 등평균검정만 만족한 경우 ARMA 모델을 적합하였다. Auto.arima 를 통해

차분이 필요없음을 확인했으며 $0 \leq p, q \leq 2$ 하에서 AIC 가 가장 낮은 인자를 사용하여

ARMA 모델을 적합한다.



<한산:용의출현 ARMA 모형 잔차 검정>

관객수 데이터의 경우 auto.arima 에서 적합값 arima(1,0,0)모형보다 arima(0,0,2)가 AIC 값이

더 낮아 해당모형을 선정했다. 좌석판매율의 경우 arima(1,0,3)모형이 더 AIC 값이 작으나

35 개 소 표본이기 때문에 많은 모수는 신뢰성이 부족하므로 AIC 가 높고 모수의 개수가

적은 arima(1,0,0)모형을 사용한다. 두 모형의 잔차 모두 백색잡음과정의 조건을 만족하므로

최종모델로 선정했다.

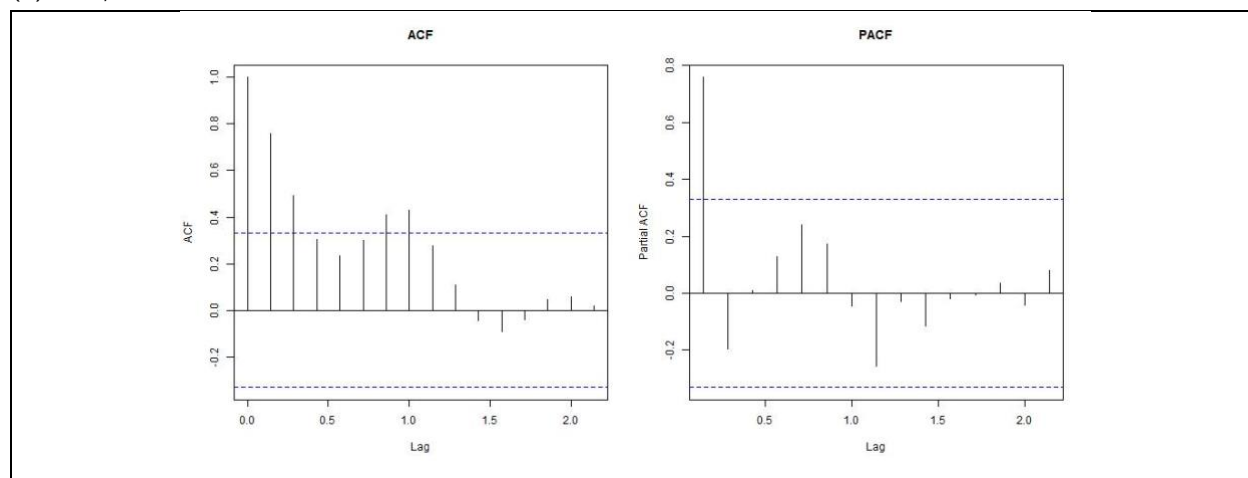
4) 계절차분

확률적 추세를 갖는 비정상시계열의 경우 추세성분, 계절성분으로 분해하기엔 모형이 복잡해진다는 단점이 있어 차분을 통해 정상화한다. 영화 데이터의 경우 계절성을 보이기 때문에 계절차분을 통해 계절성을 제거할 수 있다. 따라서 ind1, 공휴일, 문화가 있는 날의 효과를 제거한 잔차에 대해 계절차분을 진행한다. SARIMA 모형의 인자는 3 가지 방법을 적용한다.

(1) auto arima 로 모형판단

공조2:인터내셔널의 경우 auto.arma적용시 ARIMA(0,1,1)(0,1,0)[7]로 계절차분 후 1차차분과 ma(1)모형이 필요함을 알 수 있다.

(2) ACF, PACF 그래프



<공조2:인터내셔널 ACF, PACF 그래프>

ACF, PACF 그래프로는 어떤 SARIMA 모형이 적합한지 알 수 없다.

(3) AIC

영화	데이터	arma	AIC	등평균	등분산	포트먼트(lag=7)	평균=0
한산	관객수	(0,1,0)(0,1,0)[7]	-1.8	0.935	0.091	0.4	0.14
	판매율	(1,0,1)(1,1,0)[7]	-15.6	0.184	0.88	0.87	0.15
공조 2	관객수	(0,1,2)(0,1,1)[7]	-18.26	0.21	0.47	0.17	0.75
	판매율	(0,0,0)(0,1,1)[7]	-4.95	0.38	0.8	0.27	0.88
범죄도시 2	관객수	(0,1,1)(0,1,0)[7]	44.77	0.45	0.42	0.33	0.7
	판매율	(0,0,1)(0,1,0)[7]	33.7	0.76	0.9	0.86	0.31

<계절차분 백색잡음과정 검정>

계절차분은 계절성 인자와 비계절성 인자를 각각 구해야한다. 따라서 auto.arima 로 제공된 계절성 인자에 비계절성 인자를 3x3 행렬로 AIC 값을 나열해 가장 작은 값을 사용했다. 각 데이터 모두 백색잡음과정의 조건을 만족하므로 최종모델로 선정하였다.

5) 분해법 vs 계절차분

분해법 최종모델과 계절차분 최종모델을 비교할 때 적합 데이터의 시점이 차이이기 때문에 예측값의 RMSE 값으로 비교한다. 적합데이터인 개봉 후 35 일까지의 데이터를 trainset 으로, 예측대상인 개봉 후 36~70 일까지 데이터를 test set 으로 두고 RMSE 값을 확인한 결과는 다음과 같다.

영화	데이터	분해법 RMSE	계절차분 RMSE
공조 2:인터내셔널	Train	26995.88	57260.64
	Test	3160.6	11398.66
범죄도시 2	Train	34404.26	33428.12
	Test	21023.74	32874.48
한산:용의출현	Train	25578.36	56483.24
	Test	3598.98	11342.1

<일별 관객수 모델비교>

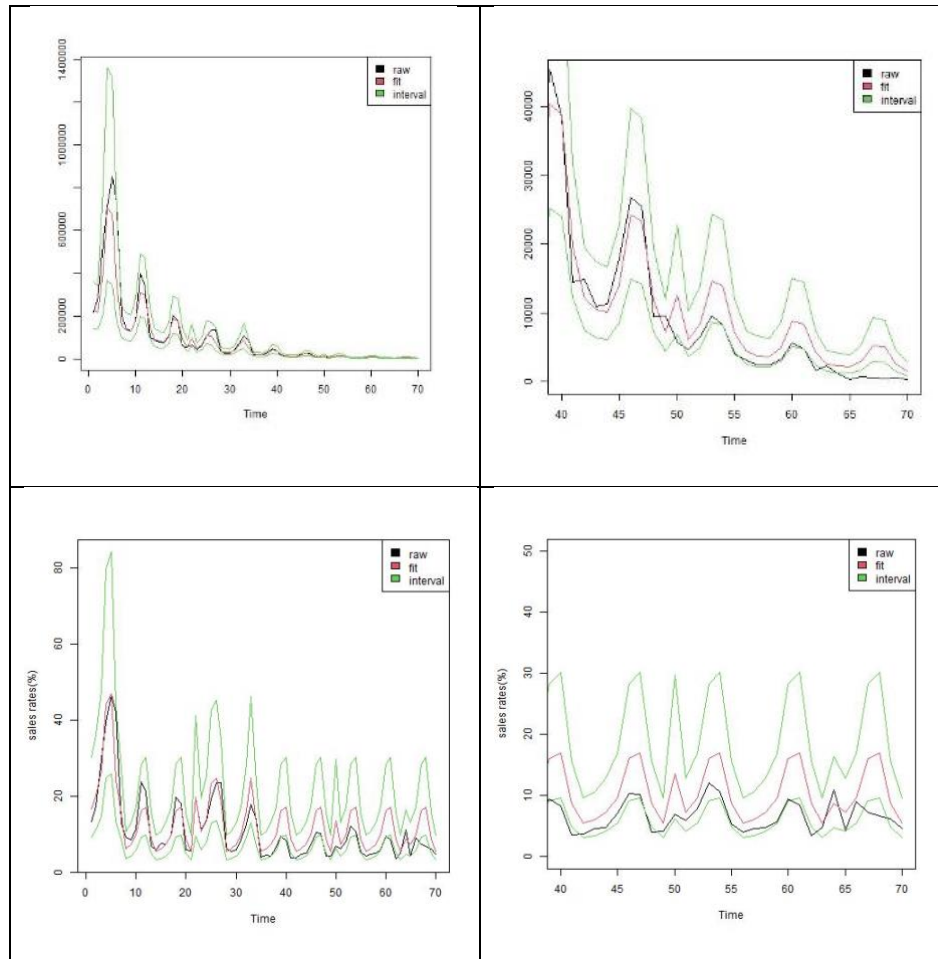
영화	데이터	분해법 RMSE	계절차분 RMSE
공조 2:인터내셔널	Train	5.07	8.88
	Test	4.61	5.22
범죄도시 2	Train	4.61	12.58
	Test	4.86	16.81
한산:용의출현	Train	2.81	3.19
	Test	13.67	13.91

<일별 좌석판매율 모델비교>

RMSE 값을 비교했을 때 test set 을 잘 예측하는 것은 분해법이 계절차분보다 뛰어난을 알 수 있다. 35 개의 소표본 분석에 계절차분과 차분을 사용해 결과적으로 27 개의 표본을 사용하게 되었고 이 점이 예측성능의 차이를 만든 것으로 볼 수 있다.

따라서 모든 영화의 최종 모델은 분해법을 이용한 모델로 선정한다.

6) 예측 해석



<공조 2:인터내셔널 일별 관객수 및 좌석판매율 예측그래프>

앞서 분석 방법에서 언급한 최소 및 최대 홀드백 기간을 확인하기 위해 예측 점 추정치와 구간추정치, 원 자료 값을 시각화 하였다. 이를 통해 공조와 범죄도시의 홀드백 적기는 다음과 같으며 (공조 2 : 개봉 후 50 일 ~ 59 일) , (범죄도시 2 : 개봉 후 46 일 ~ 64 일) 해당 구간 값들의 정확한 차이를 계산해보면 다음과 같다.(부록 3)

하지만 (부록 4) 의 한산 좌석판매율을 확인해보면 관측 값이 예측구간의 하한 값이 아닌 상한 값을 넘는 것을 알 수 있다. 해당 현상에 대해 조사한 결과, 한산:용의출현은 투자사인 '쿠팡플레이' 와의 독점 계약으로 인해 영화 상영 기간 중 OTT 서비스에 동시 상영되었음을 알 수 있었다. 이로 인해 영화관에서의 인기가 급속히 줄어 다른 영화와는 다르게 선형

감소추세가 적합 되어 생긴 현상으로 해석 할 수 있다. 즉 해당 영화에서는 홀드백 적기를 판단하기엔 부적합한 것을 알 수 있다.

4. 결론

앞서 예측한 홀드백을 다시 정리해보면 아래와 같다

영화	홀드백 기간	실제 VOD 개봉일
공조 2	10 월 26 일(50) ~ 11 월 4 일(59)	10 월 29 일
범죄도시 2	7 월 2 일(46) ~ 7 월 20 일(64)	7 월 20 일
한산:용의 출현	9 월 5 일(41) ~	8 월 29 일

한산을 제외하고 실제 VOD 개봉일과 홀드백 기간을 비교한 결과 큰 차이가 없음을 확인할 수 있다. 즉 본 분석에서 설정한 홀드백 기간은 상영사의 수익과 배급사의 2 차 매체 수익을 모두 고려한 적절한 기간임을 실제 VOD 개봉일과 유사함을 통해 확인 할 수 있다. 또한 표본수가 크진 않지만 통상적으로 흥행에 성공한 영화에 한해서는 대략 개봉 후 2 달 전후가 VOD 출시 적기임을 알 수 있다. 따라서 관객수와 좌석판매율 두 지표를 활용한다면 영화 개봉 후 35 일이 지난 뒤에는 언제부터 극장, VOD 동시상영이 적합한지에 대해 제안 할 수 있다. 물론 최근 영화산업은 홀드백을 줄이는 추세이며 흥행에 실패한 영화는 VOD로의 전환이 빠를수록 좋으므로 해당 분석은 어느 정도 흥행한 영화에 대해서만 적용할 수 있으며, 다른 국가의 개봉과 다른 수익구조를 고려해야 되는 해외 영화는 국내의 홀드백을 정할 수 없다는 것이 것이 분석의 한계점 이며 이는 향후 과제로 남긴다. 하지만 영화관 내에서의 현실적인 수익성과 인기를 고려하여 상영사, 배급사 모두를 고려했다는 점에서 본 분석이 유의미하다고 말할 수 있다.

5. 부록

참고문헌

(1) : 영화관을 외면하는 영화들, 변화하는 '홀드백' 제도 : <https://blog.ibk.co.kr/2556>

(2): '#살아있다', 손익분기점 돌파 "VOD 예상 매출로 220 만→190 만으로 낮아져" :

https://www.ytn.co.kr/_ln/0117_202008041415373130

(3) 대박영화도 60 일 이내 VOD 개봉...'진짜 안방극장 시대 활짝' :

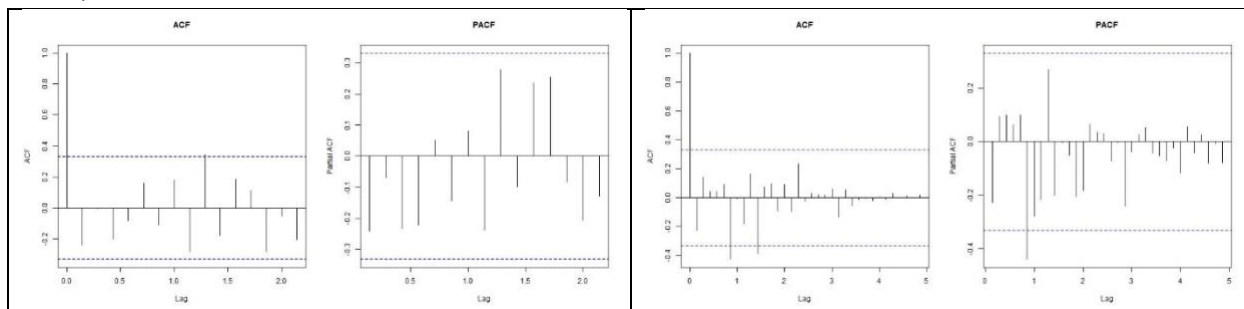
<https://zdnet.co.kr/view/?no=20160923145445>

(4) VOD 늦으면 '시들'...흥행 이어가려면 두 달 뒤가 '최적':

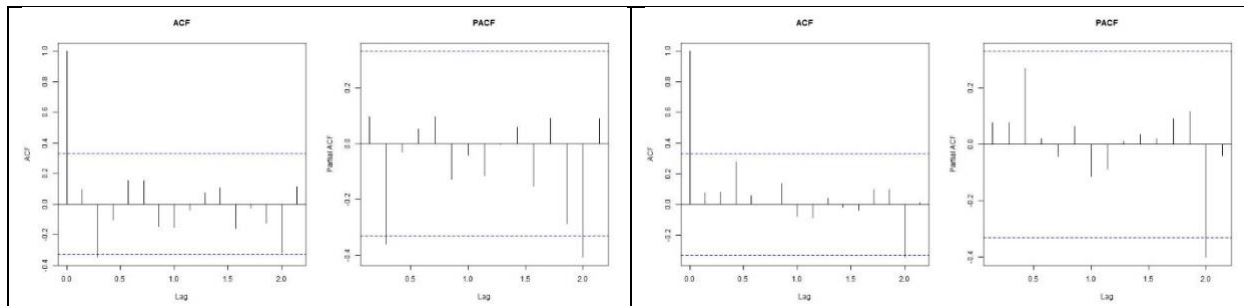
<https://m.khan.co.kr/culture/movie/article/201601262121335#c2b>

(5) 구훈영, 이희정, 이근철. (2021). 극장과 VOD 흥행에서 홀드백을 포함한 영화 흥행 요인의 영향. 경영과학, 38(2), 47-61.

부록1) 분해법 성분제거 후 잔차 ACF PACF

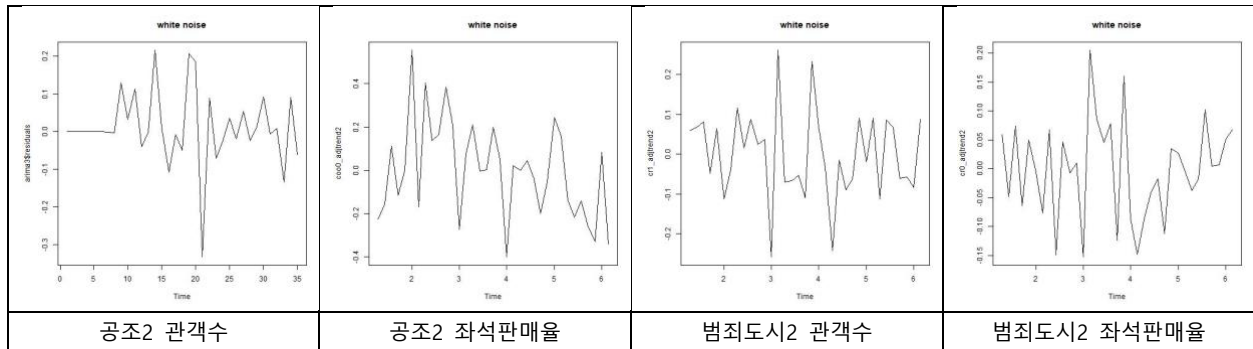


<범죄도시 2 일별 관객수, 좌석판매율 분해 후 잔차 ACF, PACF 그래프>



<공조 2:인터내셔널 일별 관객수, 좌석판매율 분해 후 잔차 ACF, PACF 그래프>

부록2) 분해법 백색잡음과정



부록3) 홀드백 구간 예측값 차이

영화/데이터	50	51	52	53	54	55	56	57	58	59	60
공조 2/관객	-1263	1037	1497	889	-113	-180	543	304	306	373	-19

영화/데이터	42	43	44	45	46	47	48	49	50	51	52
범죄도시 2/관객	9024	11184	8086	9931	-6225	-424	-2039	7849	-9681	-4219	-7545

영화/데이터	36	37	38	39	40	41	42	43	44	45	46
한산/관객	13727	1850	4047	7048	7387	-2266	750	-4314	-22	-4682	-16502

<일자 별 관객 수 실제 값과 예측구간 하한의 차이(실제 값 - 하한)>

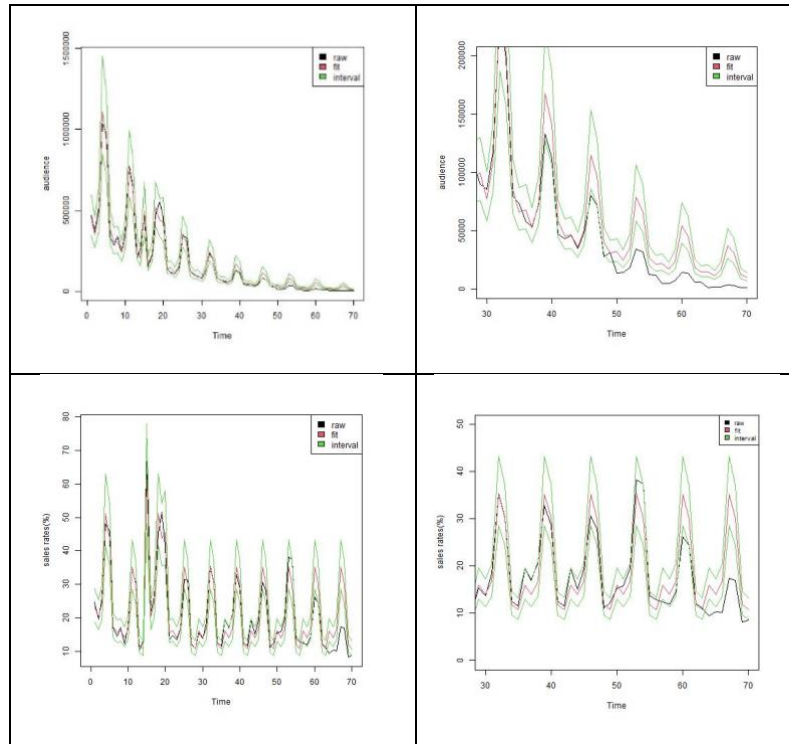
영화/데이터	59	60	61	62	63	64	65	66	67	68	69
공조 2/좌석	0.29	0.31	-1.12	-1.54	1.67	6.26	0.13	3.49	-1.88	-2.91	-1.25

영화/데이터	59	60	61	62	63	64	65	66	67	68	69
범죄도시 2/관객	0.83	-2.33	0.07	2.44	2.36	-3.58	-1.11	-3.36	-11.1	-7.47	-1.45

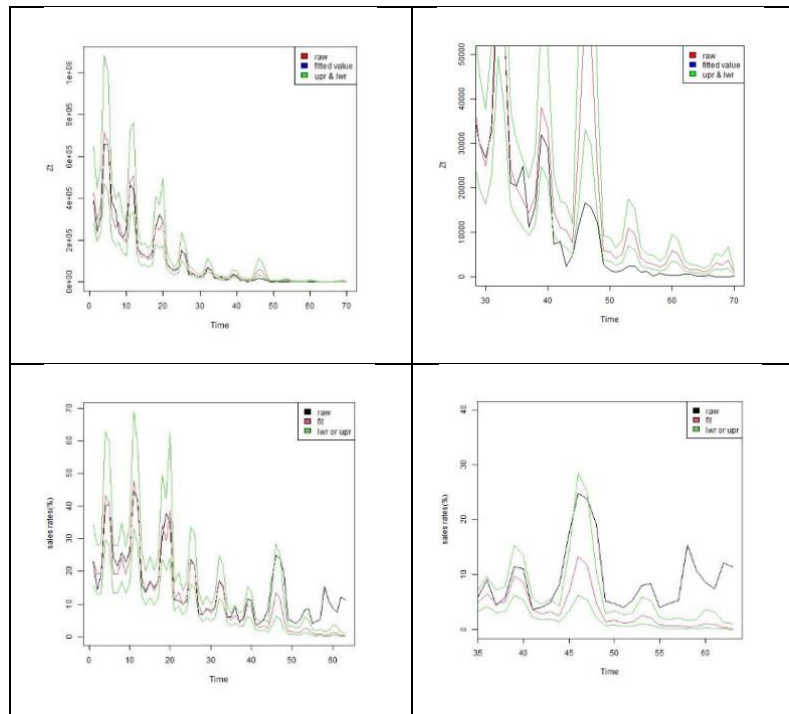
영화/데이터	47	48	49	50	51	52	53	54	55	56	57
한산/관객	-14101	-1351	-909	-1943	-1528	-1749	-4492	-3660	-1528	-696	-1619

<일자 별 좌석 판매율 실제 값과 예측구간 하한의 차이(실제 값 - 하한)>

부록4) 영화별 예측그래프



<범죠포시 2 일별 관객수 및 좌석판매율 예측그래프>



<한산:용의 출현 일별 관객수 및 좌석판매율 예측그래프>