

중국 내 자동차 산업 동향과 월별 판매량 시계열 분석⁺

(Analysis of Automobile Industry Trends and Demand Forecasting of Monthly Automobile Sales in China)

왕 첸 양¹⁾, 이 세 원^{2)*}

(Chenyang Wang and Se Won Lee)

요약 본 연구에서는 급변하고 있는 세계 경제 환경 하에서 중국 자동차 산업의 발전 현황과 자동차 산업과 관련한 중국 정부의 정책을 살펴보고, 중국 내 소비자들의 자동차 구입에 대해 소비자 동향 조사를 실시하였다. 중국 정부의 강력한 국가 배출가스 규제정책과 내연기관 자동차 제조·판매 기준의 강화에도 불구하고 소비자들은 다양한 이유로 앞으로 자동차를 구매 시 내연기관차를 선택하겠다는 응답비율이 59.6%에 달하는 등 정부 정책과 소비자 인식 사이에는 적지 않은 차이가 존재하고 있음을 확인하였다. 또한, 최근의 중국 내 자동차 판매량의 감소 추세를 발견하여 2010년 1월부터 2020년 12월까지 월별 판매량을 학습용 데이터로, 2021년 1월부터 2022년 11월 동안의 판매량을 평가용으로 구분하여 향후 중국의 자동차 수요를 예측하는 시계열 모형들을 제안, 평가하였다. 그리고 각 시계열모형을 적용하였을 때의 2023년도의 월별 예측 판매량을 보였다.

핵심주제어: 중국 자동차 산업, 수요예측, 자동차 월별 판매량, 시계열분석

Abstract In this study, we introduced the development status and the government policy of the Chinese automobile industry under the rapidly changing global economic environment. We conducted a consumer trend survey on automobile purchases by consumers in China. Despite the Chinese government's strong national emission control policy and stricter standards for manufacturing and selling internal combustion engine vehicles, 59.6% of respondents saying they would choose an internal combustion engine vehicle when purchasing a vehicle in the future for various reasons. It was confirmed that there is a significant gap between government policies and consumer perceptions. In addition, we have discovered the recent declining trend of automobile sales in China, and used the monthly sales volume from January 2010 to December 2020 as training set, and the sales volume from January 2021 to November 2022 as a test set. We proposed and evaluated a time-series model for predicting future automobile demand in China. Then, we showed the monthly sales forecast for 2023 when each model was applied.

Keywords: The Chinese automobile industry, demand forecasting, monthly automobile sales volume, time series analysis

* Corresponding Author : swlee@pknu.ac.kr

+ 이 논문은 부경대학교 자율창의학술연구비(2022년)에 의하여 연구되었음

Manuscript received January 18, 2023 / revised

February 01, 2023 / accepted February 08, 2023

1) 부경대학교 일반대학원 경영학과

2) 부경대학교 경영학부, 교신저자

1. 서 론

중국은 2009년부터 2022년까지 연속 14년 동안 1위를 지켜온 세계 최대의 자동차 시장으로 중국 내에서 자동차산업은 국가 경제에 가장 큰 영향을 미치는 산업 중 하나이다. 2019년을 기점으로 최근 몇 년간 중국 정부의 전통적인 내연기관 자동차 배출기준에 대한 엄격한 통제와 2020년 초부터 전 세계적으로 확산된 COVID-19의 영향으로 중국 내 자동차 생산과 판매는 일시적으로 주춤하였으나 여전히 중국은 세계 시장에서 큰 비중을 차지하고 있다.

중국의 자동차 판매량 수요예측과 관련한 연구들은 중국 자동차 산업이 급성장하기 시작한 2010년부터 2014년 동안의 연구들이 주를 이루고 있는데(Ahn et al., 2014; Chen, D. 2011; Zeng and Zeng, 2013; Wang et al., 2012), 성숙기에 접어들고 있다고 평가받는 최근의 연구는 거의 찾아보기 어렵다(Gao et al., 2018; Wang, 2022).

Ahn et al. (2014)은 정부의 환경규제정책과 같은 구매 제한이 없다는 가정을 기반으로 하여 중국 자동차 연간 생산량과 보유량을 ARIMA 모형과 VEC(vector error correction) 모형으로 예측하였다. 이들은 2014년 총 생산량이 전년 대비 18.8% 증가할 것으로, 2020년에 자동차 보유량을 6.1억대로 예측하며 중국의 자동차 시장의 성장 잠재력이 크다고 평가하였다. 실제 2020년도 중국 자동차 보유량은 약 2.8억 대였다. Chen, D. (2011)는 2001년 1월부터 2011년 6월까지 중국 자동차 시장의 월간 판매량 데이터를 ARMA 모형으로 예측하였다. 본 연구의 분석기간과 18개월 겹치는 부분이 있으나 월별 판매량 데이터를 실제보다 100배 큰 데이터로 예측하여 고평가된 예측 값을 제시하고 있어 문제가 있다. Zeng et al. (2013)은 처음으로 Bass 확산 모형을 사용하여 중국의 전기 차 연간 보유량을 예측하였다. 이들을 향후 10년간 중국의 전기 차 보유량이 급성장할 것이라는 연구 결과를 내었다. Wang et al. (2012)은 유럽, 브라질, 일본, 한국, 미국으로 벤치마킹 국가의 수를 줄이고 이들 국가의 1950년대 이전의 데이터를 찾는 데 많은 노력을 기울여 현재 중국과 유사한 상황을 포착하였다. 이를 토대로 S곡선(S-Curve)을 사용하여 중국의 1인당 자동차 보유량을 전망하였다. 이들은

2000-2022년 기간 중국의 연간 자동차 증가율이 13-17% 수준을 유지하며 2022년에는 총 자동차 보유량이 4억 1,900만 대에 이를 것으로 전망하였다. 실제 2022년도 중국 자동차 보유량은 4억 1,700만 대로 Wang et al.(2012)의 연구결과가 매우 정확했음을 알 수 있다. Gao et al. (2018)은 계량경제학 모형을 사용하여 중국 자동차 판매, 중국 국내 대표 브랜드 자동차(체리) 판매 및 경제 변수 간의 동적 관계를 분석하였다. VAR(Vector auto-regression) 및 VECM (Vector Error Correction Model) 알고리즘을 사용하여 경제 변수와 중국 자동차 판매를 분석하고 성능평가를 통해 VECM의 성능이 VAR 및 ARMA보다 우수함을 보였다.

기존의 연구들은 대부분이 중국의 자동차 시장이 급속도로 성장하던 시기에 이루어진 것이어서 이후로도 지속적인 성장을 예측하고 있으나 실제 판매량은 2019년을 기점으로 2년 연속 감소하는 추세를 보이고 있어 이러한 특성을 고려한 새로운 모형을 확인할 필요가 있다. 따라서 본 연구에서는 추세에 영향을 미치는 요인들에 어떤 것들이 있는지 소비자 인식조사와 중국 내 자동차 산업 관련정책 등을 통해 살펴보고, 시계열 분석을 통하여 예측 모형들을 제안, 평가하고자 한다.

이러한 점에서 본 연구는 최근 발생한 중국 내 자동차 판매량의 감소추세를 반영하여 세 가지 단기예측모형을 제시하는 시의적절한 이론적 연구이다. 또한 실제 중국 내 소비자들의 자동차 구매에 대한 인식조사와 중국 내 자동차 산업 관련 정책 등을 통해 최근 판매량에 영향을 미치는 요인들을 발굴하기 위한 기초연구의 성격을 갖는다.

2. 중국 자동차 정책과 소비자 동향

2.1 중국 정부의 자동차산업 관련 정책

중국 자동차산업은 2001년 12월, WTO 회원국이 된 이후 가파른 성장 추세를 보인다. 2002-2010년 연평균 증가율은 15%를 넘었고 2008년 금융위기 여파로 판매량이 6.7%로 대폭 하락했으나 다시 고속 성장세를 이어나갔다.

2001년, 중국 정부는 국가 하이테크 연구개발 '863 계획'에 신에너지 자동차를 중심으로 기술개

발 계획에 착수한다고 발표하였다. 그리고 3종3형(3縱3橫)국면을 수립하였다. 여기서 3종은 순수 전기차, 플러그인 하이브리드차, 연료전지 자동차를, 3형은 배터리와 관리시스템, 모터드라이브와 전력·전자, 커넥티티스와 스마트화 개념을 뜻한다.

이어서 중국 정부는 국가 중장기 기술발전 계획(2006-2020년), 신에너지 자동차 산업 발전 계획(2021-2035년) 등의 정책을 제시하였고 공식적으로 신에너지 자동차 산업에 뛰어들며 신에너지 자동차 개발 자금을 대량 투자하였다. ‘863 계획’ 시행 이래 약 20억 위안을 투입하여, 2009년 중국 국무원 신에너지 자동차 전략을 제시하였으며 100억 위안을 추가 투자하였다.

이러한 노력에도 불구하고 2019년 중국은 전 세계 온실가스 배출량의 27%를 차지해 단일국가로서 전체 선진국의 온실가스 배출량을 능가했다. 2020년에는 COVID-19 사태의 영향으로 대부분의 국가에서 탄소 배출량이 하락하는 추세를 보이지만(미국의 경우, 44억 5700만 톤으로 전년 동기 대비 11.6% 감소) 중국은 오히려 탄소 배출량이 98억 9900만 톤으로 전년 동기 대비 0.6% 증가하였으며, 전 세계 탄소 배출량에서 차지하는 비중도 30.7%로 3.7% 포인트 높아졌다.

이에 따라 중국은 강력하게 국가 배출가스 규제를 시행하며 내연기관 자동차 제조·판매 기준을 강화시켰다. 현재 중국 내에서 신에너지 자동차는 어느 곳이든 주행할 수 있지만, 내연기관 자동차는 도심부에 진입하기 위해서는 미리 허가를 받아야 하는 번거로움이 있다.

2.2 중국 내 자동차 소비자 동향 조사

모집단을 최대한 반영하기 위하여 현재 중국에 거주 중이며 국적이 중국인 응답자들로 중국 34개 성 단위의 행정구역에 설문을 배포하여 324개의 표본을 수집하였다. 조사는 중국의 설문 전용 애플리케이션 “問卷星(Sojump)”을 통해 온라인으로 2021년 11월 한 달 동안 실시하였다. 설문 결과에 따르면 응답자의 74.4%가 내연기관 자동차를, 11.1%가 신에너지 자동차를, 나머지, 14.5%는 자동차를 소유하고 있지 않은 것으로 나타났다. 앞으로 자동차를 구매할 계획이 있는 응답자

들 중 59.6%가 내연기관차를 구매하겠다고 하였으며 주된 이유는 1) 원거리를 자유롭게 주행가능하기 때문에, 2) 신에너지차 충전소가 많지 않아서, 3) 신에너지차는 충전시간이 길기 때문에, 4) 내연기관 자동차 제조기술이 보다 성숙되어 있기 때문에, 5) 통행제한 구역에 갈 일이 자주 없어서 순으로 나타났다.

반면 신에너지 차를 구입하겠다고 응답한 경우는 플러그인 하이브리드차가 20.4%, 순수 전기차가 15.7%, 수소자동차가 1.2%로, 아직까지는 플러그인 하이브리드차를 선호하는 것으로 나타났다. 신에너지차를 구매하는 이유는 빈도순으로 1) 내연기관차는 통행제한이 많아서, 2) 최신 기술을 사용하고 싶어서, 3) 국가보조금을 받을 수 있어서, 4) 배출가스 정책 때문에, 5) 최근 유가가 많이 올라서 등으로 나타났다.

또한 새로 자동차를 구입할 경우, 20만 위안(약 3700만원) 이상의 자동차를 구매하겠다는 응답자의 비율이 내연기관차 57.1%, 신에너지 자동차 34.6%로 다른 가격대에 비해 가장 높았다.

중국 정부에서는 신에너지 자동차를 판매하고자 하는 강력한 정책을 펼치고 있으나 소비자들은 여전히 충전 인프라 부족, 짧은 주행거리와 같은 해결해야 할 문제들을 제시하고 있음을 확인할 수 있다. 본 장의 논의는 Wang(2022)을 그대로 따른다.

3. 중국 내 자동차 판매량 예측모형

3.1 분석 데이터 요약

자동차 시장의 규모와 미래 추세를 예측하기 위해 일반적으로 자주 사용하는 모형으로는 시계열 모형(time series model), 소비자 선택모형(customer selection model), 확산모형(diffusion model) 등이 있다. 본 논문에서는 2010년 1월부터 2022년 11월까지의 중국내 자동차 월별판매량 데이터를 사용하여 시계열모형을 개발한다. 분석에 사용할 원 자료(original data)의 출처는 중국 자동차 협회와 공신부로 Table 1과 같다(CAA; China Association of Automobile Manufacturers).

Table 1 Monthly Sales of Automobiles in China during Jan. 2010–Nov. 2022 (Unit: 10,000 units)

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Jan	166.42	189.43	138.98	203.45	215.64	231.96	250.06	252.00	280.90	236.70	194.10	250.3	253.1
Feb	121.15	126.7	156.71	135.46	159.64	159.33	158.10	193.90	171.80	148.20	31.00*	145.5	173.7
Mar	173.52	182.85	183.86	203.51	216.91	224.06	244.00	254.30	265.60	252.00	143.00	252.6	223.4
Apr	155.52	155.2	162.44	184.17	200.42	199.45	212.20	208.40	231.90	198.00	207.00	225.2	118.1
May	143.84	138.28	160.72	176.15	191.12	190.38	209.20	209.60	228.80	191.30	219.40	212.8	186.2
Jun	141.21	143.59	157.75	175.41	184.58	180.31	207.10	217.20	227.40	205.60	230.00	201.5	250.2
Jul	124.4	127.53	137.94	151.63	161.81	150.3	185.20	197.10	188.90	180.80	211.20	186.4	242.0
Aug	132.23	138.11	149.52	164.89	171.56	166.45	207.10	218.60	210.30	195.80	218.60	179.9	238.3
Sep	155.67	164.61	161.74	193.58	198.36	202.48	256.40	270.90	239.40	227.10	256.50	206.7	261.0
Oct	153.86	152.48	160.60	193.26	198.72	222.16	265.00	270.40	238.00	228.40	257.30	233.3	250.5
Nov	169.71	165.6	179.1	204.39	209.09	250.88	293.90	295.80	254.80	245.70	277.00	252.2	232.8
Dec	166.67	168.96	180.99	213.42	241.01	278.55	305.73	306.00	266.15	265.80	283.10	278.6	-

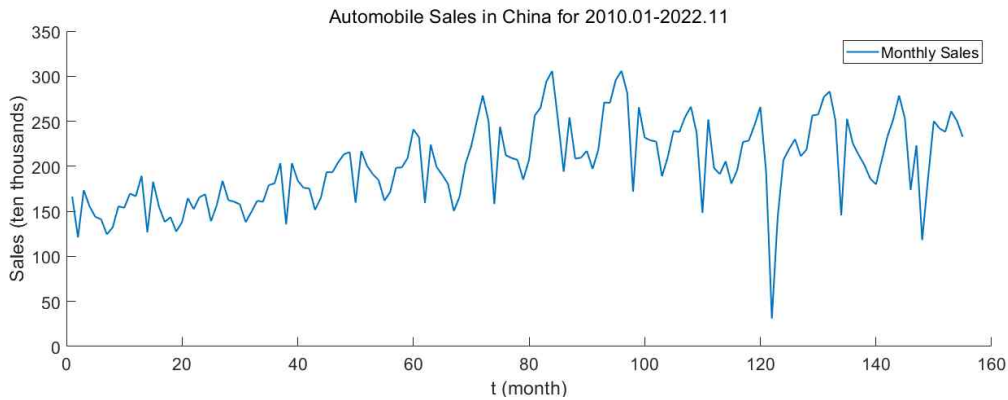


Fig. 1 Time Series Data of Monthly Automobile Sales in China

Table 1에서 파란색, 빨간색 음영은 각각 일 년 중 월별 최대판매량과 최소판매량을 나타낸다. 또한 *는 분석기간 동안의 최솟값으로서 COVID-19로부터 야기된 이상치로 보인다.

자동차 판매량은 우편물의 양, 완구, 맥주, 고속도로 통행량, 호텔 예약 등과 더불어 전통적으로 계절변동을 갖는 데이터로 잘 알려져 있다 (Stevenson and Chuong, 2014). Fig. 1은 본 연구에서 다루는 중국 내 월별 자동차 판매량 역시 계절성(seasonality)을 갖는 시계열 데이터일 뿐만 아니라 추세(trend)까지 포함하고 있음을 보여준다.

시계열 분석에 앞서 간략히 데이터의 특성을 살펴보자. 중국의 경우 월별 자동차 판매량이 가장 적은 달은 주로 2월, 7월로 나타나며, 2월, 7월을 기점으로 판매량이 줄어들었다가 다시 증가하는 추세를 보인다. 특히 2020년도 2월의 경우는 COVID-19의 확산으로 중국 내 자동차 판매

량이 지난 13년 동안 최저치를 기록했다. 반대로 분석기간 동안 일 년 중 판매량이 가장 많은 달은 대체로 12월, 1월인 것으로 나타났다.

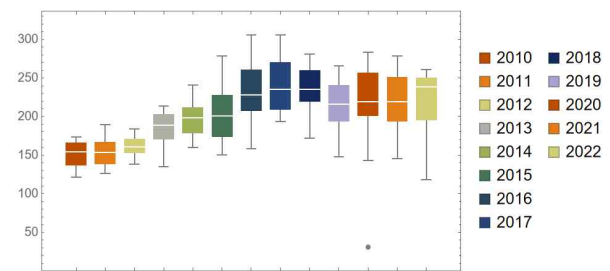


Fig. 2 Box-Whisker Plot by Year of Data(Unit: 10,000)

Fig. 2는 분석데이터의 연도별 특징을 확인하기 쉽도록 그린 Box-Whisker plot이다. 각 직사각형(box)의 아랫변과 윗변은 해당년도 월별 자

동차 판매량의 1사분위수(Q1)와 3사분위수(Q3)를, 박스 안쪽의 흰색 선은 중위판매량(2사분위수)을 나타낸다. 박스의 길이는 해당년도 데이터의 사분범위(interquartile range; IQR = Q3-Q1)로서 해당년도 데이터의 50%가 이 범위 안에 있음을 나타내며 나머지는 박스 외부의 수염(whisker)부분에 각각 25%씩 위치하는데 이 수염에 포함되지 않는 데이터를 이상치로 간주한다. 박스 및 수염의 길이와 중위수의 위치로부터 데이터 분포의 형태를 파악할 수 있고 여러 가지 데이터집합의 분포를 동시에 확인할 수 있다. 예를 들어 2012년도의 데이터는 중위수를 중심으로 좌우대칭의 형태를 갖는 분포를 가지며 데이터의 변동성이 가장 적다. 또한 2022년 월별 판매량은 50%의 데이터가 중위수 위쪽부분에 몰려 있어 분포 상으로는 왼쪽으로 꼬리가 긴 형태이다. 2011년도를 제외한 2010-2014년 동안에는 판매량의 사분범위가 비슷하며, 연중 판매량은 2015년 이후 변동성이 증가하여 2016년에 최대가 되었다가 이후에는 변동성이 줄어들고 COVID-19가 발생한 후인 2020년부터 다시 증가하고 있음을 확인할 수 있다. 또한 판매량의 중위수(median)는 2018년까지 서서히 증가하다가 2019년에 감소한 후 2020년 소폭 증가하여 유지되다가 2022년에는 다시 증가하고 있음을 확인할 수 있다. 앞에서 언급한 2020년도 2월의 판매량은 이상치로 확인되어 2010년부터 2019년까지의 2월 평균매출로 대체하여 분석하였다.

3.2 예측모형

본 연구에서 사용하는 분석 데이터는 추세와 계절성을 동시에 갖는 시계열 데이터이므로 이 둘을 함께 고려한 예측모형들을 유도하여 각 방법을 비교 분석한다. 비교 분석하고자 하는 모형은 다음과 같으며, 분석에는 R과 MATLAB을 사용하였다. 2010년 1월부터 2020년 12월까지 데이터를 training set으로, 2021년 1월부터 2022년 11월까지 데이터를 test set으로 사용하여 실제 판매량과 비교하였다. Training set과 test set을

이와 같이 분류한 이유는 COVID-19가 발생하여 확산된 2020년도의 데이터를 반영시켜 추세를 구하는 것이 타당하다고 보았기 때문이며, 2021년도 자료까지 학습데이터로 사용하게 되면 평가용 데이터가 11개로 너무 적어지기 때문이다. 실제로 기간을 달리 하여 분석해 본 결과 2020년도 데이터를 training set에 포함시키지 않은 경우는 예측오차가 크게 나타나는 것을 확인하여 구분하였다. 예측 방법은 다음과 같다:

- (1) 모형 1: 계절지수를 활용한 계절변동모형
- (2) 모형 2: Holt-Winters 모형
- (3) 모형 3: ARIMA 모형

3.2.1 계절지수를 활용한 계절변동모형

먼저 계절지수를 이용하여 모형을 적합(fitting)하고자 한다. 분석절차는 다음과 같다:

- 단계 1) 중심이동평균(centered moving average) 기법을 사용하여 계절지수를 구한다.
- 단계 2) 원 데이터를 해당 계절지수로 나누어 계절성을 제거한다.
- 단계 3) 계절성이 제거된 데이터(deseasonalized data)로부터 추세를 파악한다.
- 단계 4) 단계 3에서 구한 추세 방정식에 해당 계절지수를 곱하여 최종 예측 값을 구한다.

중심이동평균을 사용하여 계절지수를 구하는 방법은 다음과 같다. j 번째 실제 데이터를 a_j 라 하자. 먼저 n 개의 계절을 갖는 시계열 데이터의 이동평균값들을 식 (1)과 같이 구한 후, 인접한 평균값들의 산술평균수열 $\{c_i | i = 1, 2, \dots\}$ 를 구하여 계절지수 $S_{i+6} = a_{i+6}/c_i$ 를 계산한다. 이와 같이 a_{i+6} 으로 계산하는 이유는 중심이동평균을 사용할 때 계절의 개수가 짝수인 경우 이동평균값을 $\frac{n}{2} + i$ 번째 예측 값처럼 사용하기 때문이다.

$$\bar{X}_i = \frac{\sum_{j=1}^{n+i-1} a_j}{n}, \quad i = 1, 2, \dots \quad (1)$$

$$c_i = \frac{\bar{X}_i + \bar{X}_{i+1}}{2}, \quad i = 1, 2, \dots \quad (2)$$

결과적으로 우리에게 필요한 계절지수는 12개의 값이므로 각 계절에 해당하는 S_i 들의 평균을 정규화(normalization)하여 Table 2와 같은 결과를 얻는다. S_i 의 값은 i 월의 실제 판매량과 중심 이동평균값으로 구한 예측 판매량의 비율로서 1보다 큰 경우 평균추세보다 실제 판매량이 많음을 의미하며, 1보다 작은 경우는 그 반대로 해석이 가능하다. 중심이동평균을 사용할 때의 장점과 보다 자세한 내용은 Cachon and Terwiesch (2019), Stevenson and Chuong (2014), Swink et al. (2016)을 참고하기 바란다.

Table 2 Monthly Seasonal Index

Month	Seasonal Index (S.I.)	Symbol
January	1.102028	S_1
February	0.791022	S_2
March	1.089452	S_3
April	0.978867	S_4
May	0.948836	S_5
June	0.950096	S_6
July	0.817558	S_7
August	0.890153	S_8
September	1.047047	S_9
October	1.049616	S_{10}
November	1.140027	S_{11}
December	1.195297	S_{12}

본 연구의 출발점은 중국의 자동차 산업이 시작된 이래 지금까지 보지 못했던 중국 내 자동차 판매량의 감소추세를 예측모형에 반영하는 것이므로 단계 2와 3을 거쳐 추세방정식을 구할 때 1차 함수부터 4차 함수까지 고려하며 모형을 찾았고, 다항방정식 중 가장 성능이 좋은 것은 2차 추세방정식으로 나타났다. Table 2의 계절지수와 단계 2,3을 거쳐 구한 추세함수는 각각 다음과 같다.

1차 함수: $y_t = 150.9 + 0.7267t$

2차 함수: $y_t = 130.3 + 1.65t - 0.00694t^2$

3차 함수: $y_t = 148.4 + 0.04681t + 0.02307t^2 - 0.00015t^3$

4차 함수: $y_t = 164.8 - 2.342t + 0.1032t^2 - 0.001085t^3 + 0.000003514t^4$

Fig. 3은 계절성을 제거한 데이터와 이를 기반으로 구한 2차 함수의 결과를 95% 신뢰구간과 함께 표시한 그래프이다. Fig. 4는 2차 추세 방정식에 계절성을 반영하여 기존 데이터와 예측모형의 결과를 함께 나타낸 것으로 2019년 이후의 감소추세가 잘 반영되고 있음을 확인할 수 있다. 이 모형은 중국의 배기가스 규제정책 강화로 인한 일부 소비자들의 불만족 증가, 충전시설 인프라 확보 실패, 순수 전기차 대비 내연기관 자동차, 플러그인 하이브리드카에 대한 선호도 우세, 차량용 반도체 부족 등의 요인으로 향후 중국 내 자동차 판매량이 감소할 수 있다는 시나리오 하에 적용하여 모니터링 할 수 있다.

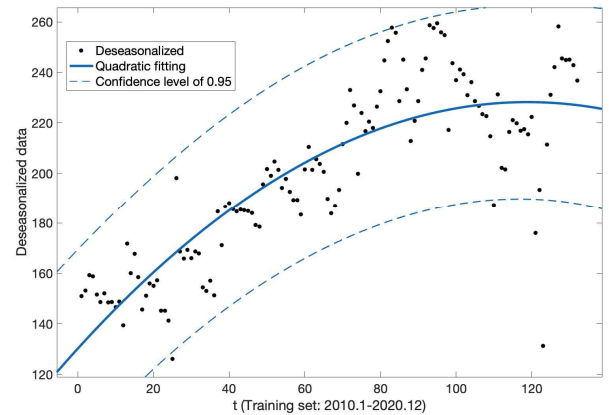


Fig. 3 Quadratic Fitting of Deseasonalized Training Data: $y_t = 130.3 + 1.65t - 0.00694t^2$

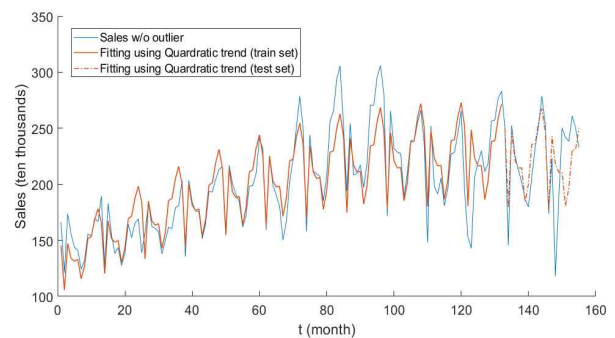


Fig. 4 Fitting Results by Adding Seasonality to the Quadratic Trend Equation

3.2.2 Holt-Winter의 지수평활모형

고급 지수평활기법(advanced exponential smoothing) 중 하나로 Holt-Winter의 지수평활기법이 있는데 이 방법은 이중지수평활법의 확장형으로서 k 시점 후의 예측은 또한 $t+k$ 시점에서의 계절성을 고려한다. 추세에 대하여는 가법(additive) 및 승법(multiplicative) 계절성을 고려한 모형들이 있으며 승법 계절성의 경우 다른 계절의 값들이 특정 비율만큼 다르다고 가정하는 반면, 가법 계절성의 경우에는 고정된 양만큼 차이가 난다고 가정한다 (Shmueli and Lichtendahl, 2016).

Holt-Winter의 지수평활기법은 R forecast package의 ets 함수를 이용하여 분석하였다. 여기서, ets 함수명의 각 문자는 시계열 데이터의 오차(error), 추세(trend), 계절성(seasonality)을 뜻한다. 시계열에 ets 함수를 적용하면 학습용과 평가용 기간의 예측값과 예측오차(residual)가 생성된다.

Fig. 5는 2010년부터 2020년까지 중국내 자동차 월별 판매량 데이터를 Holt-Winter의 지수평활모형에 적합(fitting)시킨 후, 이로부터 얻은 2021년 1월부터 2022년 11월까지의 예측 판매량(하늘색 점선)을 실제 월별 판매량(파란색 실선)과 함께 비교하여 그린 그래프이다.

R 분석결과 본 연구에서 다루는 시계열 데이터는 Holt-Winter의 지수평활모형 중 승법 오차(M), 승법 계절성(M)이 있는 것으로 확인되었으며 추세 성분(N)은 크게 영향을 주지 않는 것으로 나타났다.

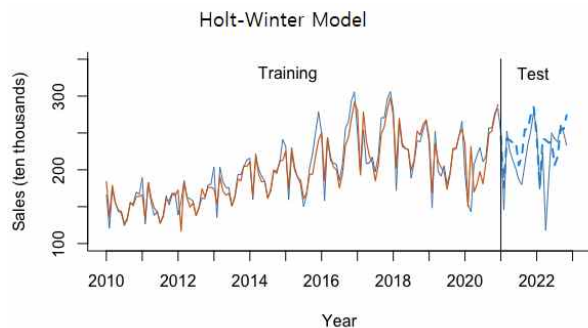


Fig. 5 Fitting by Holt-Winter's Exponential Smoothing

```

> hwin
ETS(M,N,M)

Call:
ets(y = train.ts, model = "ZZZ")

Smoothing parameters:
alpha = 0.3537
gamma = 0.2552

Initial states:
l = 165.5634
s = 1.0775 1.0893 1.0148 1.0421 0.8965 0.8393
    0.9462 0.9514 1.0185 1.1636 0.8483 1.1126

sigma: 0.0877

      AIC      AICc      BIC
1407.315 1411.453 1450.557
    
```

Fig. 6 Summary of Holt-Winter's model

Fig. 6은 모형의 결과를 요약한 것으로 두 가지 평활모수의 추정치($\alpha = 0.3537$, $\gamma = 0.2552$)와 이의 초기상태를 보여주고 있다. α 값으로부터 시계열 수준이 지역적인 요소를 갖고, γ 로부터 최근의 관측값을 더 우선적으로 고려한다는 것을 알 수 있다.

위의 결과들은 Holt-Winter ETS(M,N,M) 모형의 각 모수 값들로서 수학적 모형은 다음과 같다 (Hyndman and Athanasopoulos, 2018)

$$y_t = l_{t-1}s_{t-m}(1 + \epsilon_t) \tag{3}$$

$$l_t = l_{t-1}(1 + \alpha\epsilon_t) \tag{4}$$

$$s_t = s_{t-m}(1 + \gamma\epsilon_t) \tag{5}$$

여기서, y_t 는 시점 t 의 예측 값, l_t 는 시점 t 에서의 시계열 수준, s_t 는 시점 t 에서의 계절요소, m 은 계절의 개수를 나타내며, α , γ 는 평활상수를 의미한다. Fig. 6의 결과를 사용하여 모형을 정리하면 최종적으로 다음을 얻는다.

$$y_t = l_{t-1}s_{t-12}(1 + \epsilon_t) \tag{6}$$

$$l_t = l_{t-1}(1 + 0.3537\epsilon_t) \tag{7}$$

$$s_t = s_{t-12}(1 + 0.2552\epsilon_t) \tag{8}$$

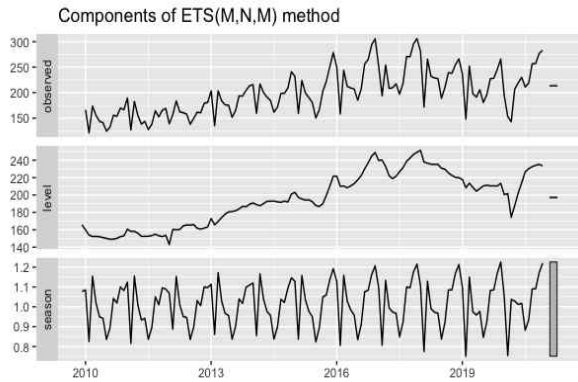


Fig. 7 Components of ETS Method

Fig. 7은 관측데이터와, 데이터의 수준, 계절적 요소를 따로 구분하여 보여주고 있다. 수준을 살펴보면 2018년까지 상승하다가 2019년 이후 감소하여 2021년을 기점으로 다시 상승하는 현상을 볼 수 있으며 계절성 또한 명확하게 드러난다.

3.2.3 SARIMA (Seasonal ARIMA) 모형

ARIMA(자기회귀누적이동평균) 모형은 지수평활법과 함께 시계열 예측에 널리 사용하는 접근 방법 중 하나이다. 자기회귀(AutoRegressive)와 이동평균(Moving-Average) 모형을 기반으로 시계열을 체계적으로 식별, 추정, 진단 및 예측하는 경험 기반적인 방법이며 상호보완적으로 다루는 접근법이라 할 수 있다. Box와 Jenkins의 노력으로 예측 연구와 실습에서 가장 인기 있는 방법의 하나가 되었으며, 정적 시계열 데이터에 대한 판매 예측에 더 적합하다고 알려져 있다 (Swink et al., 2016; Suh et al., 2004; Sana et al., 2017).

비정상 시계열은 ARIMA 모형으로 분석하는 과정에서 분해(decomposition) 및 차분(differencing)하여 정상(stationary) 시계열로 변환한 뒤 분석을 진행할 수 있다.

본 절에서는 R 프로그램의 decompose 함수와 로그차분을 이용하여 시계열 데이터를 분해하고 Augmented Dickey-Fuller test (ADF 검정)를 통해 변환한 데이터가 정상 시계열임을 확인하였다 (Fig. 8-9).

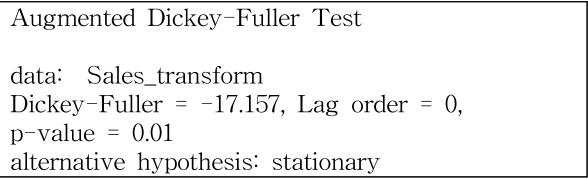


Fig. 8 Result of Augmented Dickey-Fuller Test

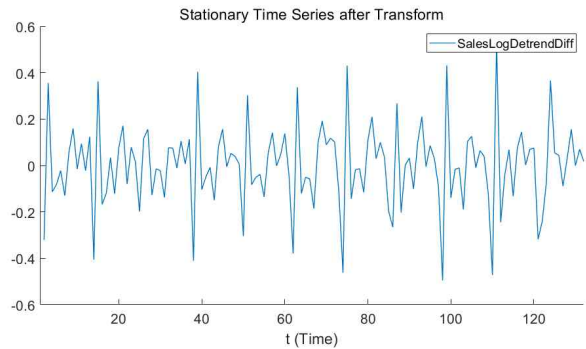
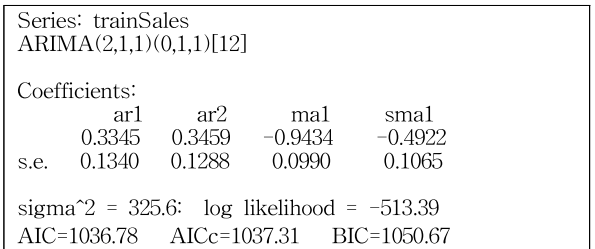


Fig. 9 Stationary Time Series

최적의 ARIMA 모형을 얻기 위해 절차에 따라 모수를 확인하는 작업을 거친 후 forecast 패키지의 auto.arima 함수를 사용하여 검토한 결과 본 연구의 test data는 계절의 수가 12개인 다음의 ARIMA(2,1,1)(0,1,1)[12] 모형으로 적합 시킬 수 있음을 확인하였다. 모형의 수식은 식 (9)와 같으며 여기서 B 는 후방이동(backshift) 연산자로서 $By_t = y_{t-1}$ 이다.



$$(1 - 0.3345B - 0.3459B^2)(1 - B)(1 - B^{12})y_t = (1 + 0.9434B)(1 + 0.4922B^{12})\epsilon_t \quad (9)$$

Fig. 10은 생성한 모형의 차수와 모수 추정이 제대로 되었는지 확인하기 위해 tsdiag 함수를 통해 얻은 결과를 나타낸다. Box-test 함수는 자기상관이 모두 0이라는 귀무가설을 검정할 때 사

용하는데, Fig. 11의 p-value로부터 잔차(residual)들의 상관관계가 0이라는 귀무가설을 기각할 수 없다. Fig. 10-11의 결과로부터 SARIMA 예측모델은 시계열 데이터를 잘 적합시킨 것으로 볼 수 있다 (Kwahk, 2019).

Fig. 12는 Training set 데이터를 Seasonal ARIMA(2,1,1)(0,1,1)[12]모형으로 적합(파란색 실선)시켜 이를 바탕으로 2021.1월부터 2022.11월까지 월별 예상판매량(갈색 점선)을 구하여 그린 그래프이다.

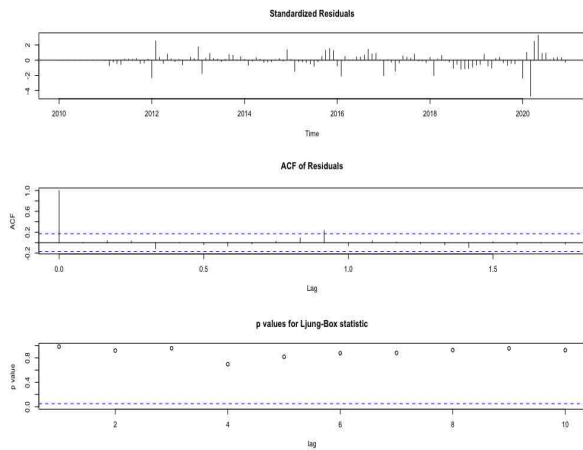


Fig. 10 Residual Diagnostic

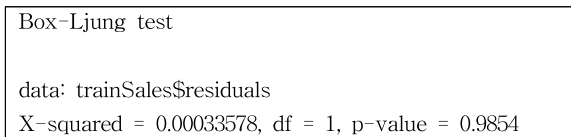


Fig. 11 Result of Box-Ljung Test

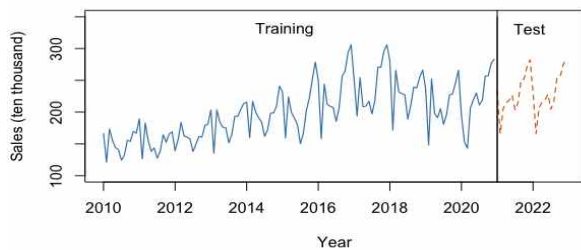


Fig. 12 Seasonal ARIMA and Forecasting Plot

4. 예측모형 성능평가

예측 성능의 주요 척도로 사용되는 예측오차는 주어진 기간의 실제 수요 값과 예측 값의 차이로 정의된다.

모형의 성능을 비교하기 위한 목적으로 사용하는 척도로는 평균 예측오차(mean forecast error, MFE), 평균 백분율 오차(mean percentage error, MPE), 평균 절대 편차(mean absolute deviation, MAD), 평균 절대 백분율 오차(mean absolute percentage error, MAPE), 평균 제곱 오차(mean squared error, MSE), MSE의 제곱근(root mean squared error, RMSE) 등이 있다.

평균 절대 편차(MAD)는 계산이 가장 간편하나 모든 오차를 동일하게 취급한다. 평균 제곱 오차(MSE)는 오차를 제곱하여 처리함으로써 문제점을 많이 일으키는 큰 오차에 대해 보다 큰 가중치를 준다. RMSE는 MSE가 오차의 제곱을 다루므로 실제 평균오차보다 그 값이 더 커지는 현상을 조정할 때 사용한다.

본 연구에서는 모형의 성능기준으로 평균 절대 백분율 오차(MAPE)를 사용하였다. 그 이유는 다른 기준에 비하여 오차를 전체적인 관점에서 상대적으로 파악할 때 유용하고 백분율로 계산되어 이해하기 편하다는 장점을 가지고 있기 때문이다. 보다 자세한 각 척도의 특징과 한계점, 장단점에 대해서는 Swink et al. (2016), Shmueli and Lichtendahl (2016)을 참고하기 바란다.

본 연구에서는 예측모형의 정확도(accuracy)를 확인하는데 주로 사용하는 MAPE로 성능을 평가하였다. MAPE는 식 (10)과 같이 계산한다.

$$MAPE = \frac{\sum_{t=1}^k \left| \frac{A_t - F_t}{A_t} \right|}{k} \times 100(\%) \quad (10)$$

여기서, A_t 는 t 기의 실제 값, F_t 는 t 기의 예측 값, k 는 데이터의 개수이다.

세 가지 방법으로 구한 모형들의 성능척도를 비교하면 Table 3과 같다. 전체 데이터를 가장 잘 설명하는 모형은 SARIMA(계절성을 갖는 ARIMA)

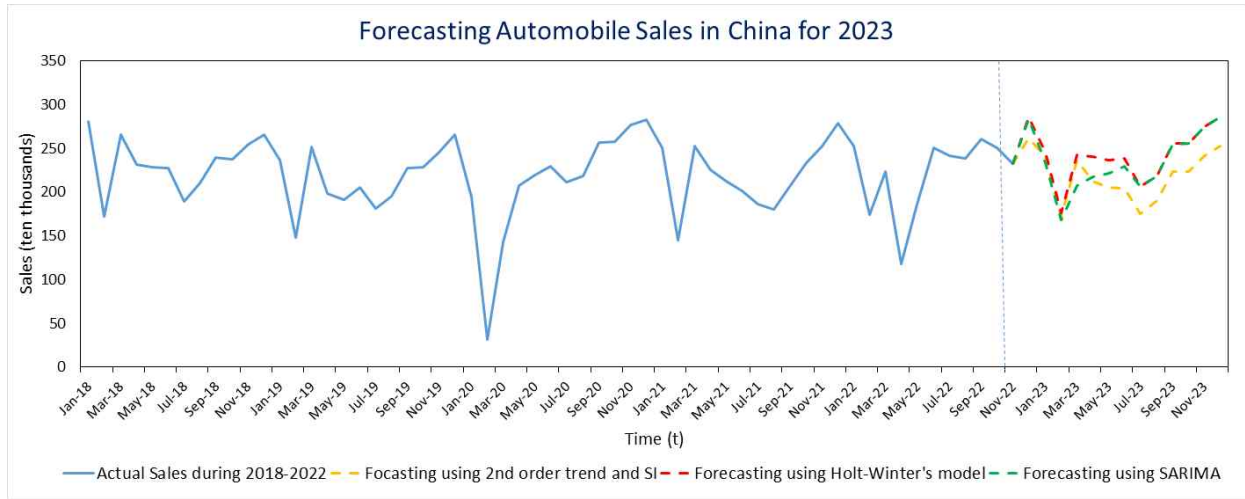


Fig. 13 Forecasting: Monthly Automobile Sales in China for 2023

Table 3 Evaluation of Forecasting Models

Model	MAPE (%)		
	Training Set	Test Set	Total
Using 1 st order trend equation and S.I.	7.692	19.719	9.477
Using 2 nd order trend equation and S.I.	7.276	11.460*	7.897
Using 3 rd order trend equation and S.I.	6.441	20.121	8.471
Using 4 th order trend equation and S.I.	6.293	14.057	7.445
Holt-Winter's exponential smoothing	5.465*	14.438	6.796
SARIMA Model	5.598	13.148	6.718*

(* is the minimum MAPE in each data set.)

가, 2010년 1월부터 2020.12월까지 Test set에 예측모형을 적용했을 때의 MAPE는 ‘계절지수와 2차 추세방정식을 이용한 예측모형’이 가장 좋은 것으로 나타났다. 이는 2019년을 기점으로 하락세를 갖는 중국 내 자동차 판매량이 아직까지는 그대로 적용할 수 있을 정도로 유지되고 있다고 조심스럽게 해석할 수 있다.

Fig. 13은 각 데이터 세트에서 성능이 가장 좋았던 3가지 모형(2차 추세선과 계절지수를 활용한 모형, Holt-Winter의 지수평활모형, SARIMA 모형)으로 구한 2023년도의 중국 내 월별 자동차 예측판매량을 2018년-2022년도의 실제 판매량과 함께 나타낸 그림이다. 2차함수를 추세방정식의

로 하여 계절지수를 적용한 모형(노란색 점선)이 다른 두 개의 모형보다 적은 예측판매량을 보이는데 이는 함수의 특성상 시간이 지날수록 더 작은 값을 결과로 보이므로 단기예측 목적으로 활용하는 것이 바람직해 보인다.

이보다 복잡도가 높은 Holt-Winter의 지수평활모형(빨간색 점선)과 SARIMA 모형(초록색 점선)은 2023년도 4월을 제외하고는 거의 같은 예측 값을 보임을 알 수 있다. 흥미로운 사실은 Holt-Winter의 모형이 SARIMA보다 이상치에 덜 민감한 강건성(robustness)을 보인다는 것이다. Table 1의 자료를 보면, 2022년도 4월 중국 내 COVID-19 확진자 수 증가로 단행한 재봉쇄 정책의 영향이 자동차 판매량에 일부 반영되어 분석기간 4월 평균판매량 19만 5천대의 60%에 불과함을 확인할 수 있다. 이러한 특수 상황을 SARIMA 모형은 23년 4월 판매량 예측에 반영하여 Holt-Winter의 모형보다 더 작은 판매량으로 예측하므로 향후 예측모형 선택에 참고할 수 있다.

5. 결론 및 추후연구

본 연구에서는 세계 최대의 자동차 시장인 중국의 국내 자동차 시장에서 2019년 이후 감소추세를 보이는 월별 자동차 판매량에 직간접적으로

영향을 미칠 수 있는 요인을 자동차 산업관련 정책과 함께 소비자 동향 조사를 통해 살펴보았다.

내연기관자동차를 중국 정부의 강력한 국가 배출가스 규제정책과 내연기관 자동차 제조·판매 기준의 강화에도 불구하고 정부 정책과 소비자들의 구매 선호도 사이에는 차이가 존재하고 있음을 확인하였다.

또한, 2010년 1월부터 2022년 11월까지의 데이터를 바탕으로 세 가지 방법의 시계열 분석을 통해 예측모형을 개발, 평가하여 2023년도 월별 판매량에 대하여 단기 예측값을 구하였다. 추세선과 계절지수를 사용한 모형개발은 Holt-Winter의 지수평활모형이나 Seasonal ARIMA 모형보다 변화하는 추세에 대한 적응력이 낮아 단기 예측이 적합하고, Holt-Winter의 모형이 SARIMA 보다는 이상치에 강건한 예측결과를 나타냄을 확인하였다.

기존의 중국 내 자동차 판매량 관련 예측모형 연구들은 중국의 자동차 시장이 급속도로 성장하는 시점에 이루어져 2019년 이후에 나타난 감소 추세를 반영시킨 모형의 연구는 시의적절하며, 중국 자동차 시장이 차지고 있는 비중을 볼 때 그 의미를 갖는다.

단, 본 연구에서는 중국 내에서의 자동차 월별 판매량 데이터만을 기반으로 하여 예측모형을 개발하였기에 추후 유가, 경제 지수, COVID-19의 재확산으로 인한 추가봉쇄 시점 등 중국 내 자동차 판매량에 영향을 끼치는 추가변수들을 함께 고려한 모형의 고도화 작업이 필요하다. 또한 본 연구에서는 전체 자동차 판매량을 대상으로 하였으므로 신에너지차와 내연기관차 각각에 대한 예측모형의 개발을 통해, 이를 기반으로 전체 판매량 예측을 하는 통합모형의 분석을 추후 연구과제로 고려할 수 있다.

References

- Ahn, B.-K., Choi, Y.-H., Lee, Z.-H. (2014). A Study on Forecasting of Chinese Auto Industry using Series Model. *The Northeast Asia Economic Association of Korea*, 26(3), 37-62.
- Cachon, G., Terwiesch, C., *Matching Supply with Demand*, McGraw-Hill Education, 2019.
- Chen, D. (2011). Chinese Automobile Demand Prediction Based on ARIMA Model. *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2197-2201.
- China Association of Automobile Manufactures, http://www.caam.org.cn/chn/4/cate_30/list_1.html
- Gao, J., Xie, Y., Cui, X., Yu, H., Gu, F. (2018). Chinese automobile sales forecasting using economic indicators and typical domestic brand automobile sales data: A method based on econometric model. *Advances in Mechanical Engineering*, 10(2), 1-11.
- Hyndman, R. J. & Athanasopoulos, G., *Forecasting: Principles and Practice*, OTexts, 2018.
- Kaya, A., Kaya, G., Çebi, F. (2019). Forecasting Automobile Sales in Turkey with Artificial Neural Networks. *International Journal of Business Analytics*, 6(4), 50-60.
- Kwahk, K.-Y., *Statistical Data Analysis with R*, Chungnam, 2019.
- Shmueli, G. & Lichtendahl, K. C. Jr., *Practical Time Series Forecasting with R: A Hands-On-Guide*, Axelrod Schnall Publishers, 2016.
- Stevenson, W. J. & Chuong, S. C., *Operations Management*, McGraw-Hill Education, 2014.
- Swink, M., Melnyk, S., Cooper, M. B., Hartley, J. L., *Managing Operations: Across the Supply chain*, McGraw-Hill Education, 2016.
- Suh, S. C., Saffer, S. I., Li, D., Gao, J. M. (2004). A new insight into prediction modeling systems. *J. of Integrated Design and Process Science*, 2, 85-104.
- Sana, P. S., Hassan, M. K., Yang, Z. (2017). Annual Automobile Sales Prediction Using ARIMA Model. *International Journal of Hybrid Information Technology*, 10(6), 13-22.

Wang, C. (2022). *Analysis of development trends and demand forecasting of Chinese automobile industry*, MBA Thesis, Graduate School of Pukyong National University, Busan, Korea.

Wang, Y., Jacob, T., Daniel, S. (2012). Will China's Vehicle Population Grow Even Faster than Forecasted. *ACCESS Magazine*, 41, 29-33.

Zeng, M., Zeng, F., Zhu, Q., Xue, S. (2013). Forecast Of Electric Vehicles In China Based On Bass Model. *Electric Power*, 46(1), 36-39.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.



왕첸양 (Chenyang Wang)

- 부경대학교 일반대학원 경영학과 박사 과정
- 관심분야: 수요예측, 공급사슬 관리, 스마트공장



이세원 (Se Won Lee)

- 종신회원
- 성균관대학교 산업공학과 학사
- 성균관대학교 산업공학과 석사
- 성균관대학교 산업공학과 박사
- 부경대학교 경영대학 경영학부 부교수
- 관심분야: 대기행렬이론, 확률모형, 시스템 최적화