

공장용지 수요 추정 모형 개발 및 수요예측

정형철¹ · 한근식² · 김성용³

¹수원대학교 통계정보학과, ²한신대학교 컴퓨터공학부, ³고려대학교 세종캠퍼스 경제통계연구소

(2011년 2월 접수, 2011년 4월 채택)

요약

본 연구에서는 공장용지 면적을 예측하기 위한 통계적 추정을 다루었다. 공장용지에 대해서는 1981년부터 2003년까지 자료가 존재하며, 2004년 이후에는 공장용지보다 좁은 개념인 산업단지 면적에 대한 조사 자료만 존재한다. 한국산업단지공단에서는 2009년 10월 표본조사를 실시하여 당해의 공장용지 면적을 추정하였으며, 동 조사 시 향후 5개년의 공장용지면적에 대한 수요를 조사한 바 있다. 본 연구에서는 과거 절단된 자료를 여러 통계모형을 사용하여 적절히 대체할 수 있는 수요예측모형을 도출하고, 표본조사에 의한 추정치와 통계적 모형에 의한 대체값들을 융합하는 평활기법으로 향후 공장용지 수요를 예측하는 방법을 다루었다.

주요용어: 공장용지 면적, 표본조사, 계량 모형, 스플라인 회귀평활.

1. 서론

1980년대 전통제조업의 생산성과 기술력 향상 및 1990년대 정보통신 부문의 비약적 발전으로 국내 제조업은 꾸준한 성장을 거듭해 왔다. 제조업의 성장은 입지수요에 반영되어 1981년 전국 제조업 부지면적은 234km²에서 2003년 520.8km²로 증가하는 등 해당 기간 동안 공장용지 면적은 약 2배 이상 증가하였다. 그런데, 여기서 2003년까지라고 언급한 이유는 우리나라 전체 공장용지면적 통계가 2003년까지만 존재하기 때문이다 (한근식과 정형철, 2010).

1981년부터 2003년까지 “광공업 통계조사”에서 전국 제조업체 부지면적이 조사되어 효율적인 공장부지의 공급이 이루어졌으나, 2004년에 “광공업 통계조사”가 광업 및 제조업 통계조사로 변경되면서, 전국의 공장용지에 대한 조사가 산업단지 면적 조사로 축소되었다. 산업단지 면적이란 국가나 지방자치단체가 지정한 산업단지 내의 면적을 의미하며 일반적으로 이를 계획입지면적이라 한다. 그리고 산업단지 이외에 존재하는 제조업의 공장면적을 개별입지면적이라 한다. 공장용지란 계획입지면적과 개별입지면적이 합쳐진 보다 큰 개념의 면적을 의미한다.

본 연구에서는 2003년 이후 단절된 전국 제조업의 공장용지 면적을 추계하는 방법을 다루었다. 2007년 건설교통부에서는 1981년, 1986년, 그리고 1990년부터 2003년까지 조사된 통계청의 공장용지 면적을 이용하여 2004년부터 2017년까지의 공장용지 면적을 예측한 연구를 실시한 바 있다. 예측모형으로 홀트윈터스의 지수평활법과 ARIMA 모형을 사용하였으며, 각 모형별로 4가지 시나리오를 가정하여 예측

본 연구는 한국산업단지공단(2009)의 “산업입지 수요추정 모형 개발 및 입지수요 추정” 연구결과 중 일부를 수정 보완한 것임.

¹교신저자: (445-743) 경기도 화성시 봉담읍 와우리, 수원대학교 통계정보학과, 부교수.

E-mail: jhc@suwon.ac.kr

하였다 (건설교통부, 2007; 국토해양부, 2009). 하지만, 2004년부터 2007년까지의 실질 공장용지 면적을 알 수 없기에 앞의 연구를 사용하여 2007년 이후의 중장기 공장용지를 추정할 모형은 그 타당성과 신뢰성이 낮다고 평가되었다. 또한 장기예측에 대한 모형의 예측타당성을 비교 평가할 수 없는 한계점이 존재하였다. 이러한, 문제로 인해 한국산업단지공단에서는 2009년 10월에 공장용지 수요조사를 실시하여 2009년 조사시점의 공장용지를 과거의 전수조사 방법이 아닌 표본조사로 측정하였으며, 2014년까지의 미래 공장용지에 대한 구매 의향 및 실질 구매 예상 면적을 조사하였다. 이 조사 결과는 한국산업단지공단 (2009), 한근식과 정형철 (2010)에 제시되어있다.

한근식과 정형철 (2010)에서는 2009년 조사시점 시 전체 공장용지는 618.30km²이며, 2010년 53.11km², 2011년 25.28km², 2012년 22.06km², 2013년 10.35km², 2014년 38.54km²의 추가 공장용지 구매가 있으리라 예측하였다. 그 결과 2010년부터 2014년까지 공장용지는 671.41km², 696.69km², 718.75km², 729.11km², 767.65km²로 확대된다고 발표되었다. 그런데, 표본조사 결과에서, 2010년에 공장용지 수요가 높으며 이후 점차적으로 추가 수요가 감소하다가 다시 2014년에 공장용지 수요가 증가하는 것으로 나타나고 있음을 발견할 수 있다. 하지만, 이는 2011년 이후 공장용지 추가 수요가 감소하는 것이 아니라, 미래 수요에 대한 불확실성 때문에 사업체 경영진이 현재 조사 시점에서 미래 공장용지 구입 여부를 판단하지 못하여, 가장 늦은 시점인 2014년에 공장용지가 필요하다고 응답하였음을 짐작하게 하는 결과라 생각된다. 즉, 표본조사에는 어느 정도의 비표집오차가 존재하고 있다고 생각된다. 이를 개선하기 위해 표본조사 결과와 과거자료와의 연관성을 살펴볼 필요가 있다.

본 연구에서는 공장용지 수요예측을 위해 다양한 통계적 모형을 설정하고, 모형에 의한 예측 결과와 표본조사 결과를 비교하여 예측 모형 및 수요조사의 예측타당성을 파악하는데 연구 목적을 두었다. 더 나아가 모형을 통한 예측치와 표본조사를 통한 예측치를 융합하여 타당성과 신뢰성이 보다높은 예측모형을 제시하고자 한다. 또한, 공장용지 면적의 수요 예측을 위해 (1) 1990년에서 2003년까지 공장용지 면적을 이용한 지수평활모형, (2) 시계열 회귀모형, (3) 1994년부터 2008년까지 산업단지 면적을 이용한 시계열 회귀모형을 설정하고, 이를 무추세, 선형 추세, 로그추세, 그리고 이차추세 시나리오 등으로 설정하여 각 시나리오별로 공장용지를 예측하였다. 그리고 각 모형에 의한 예측값을 표본조사 결과와 비교하여 적절한 시나리오를 선택한 후, 스플라인 회귀평활을 사용하여 공장용지 면적을 최종적으로 예측하기로 하였다.

본 연구의 2장에서는 산업단지 면적과 공장용지 면적, 표본조사 결과를 살펴보고, 3장에서는 다양한 상황에 따라 통계적 모형을 사용한 공장용지 면적의 예측을 살펴보았다. 4장에서는 스플라인 회귀평활을 통한 최종 예측치를 제시하였다.

2. 공장부지 면적 및 산업단지 분양 면적

본 연구에서 다루고자하는 공장용지 면적은 연별 자료로 1981년, 1986년, 1990년-2003년의 실 관측자료와 2009년부터 2014년까지 수요조사에 의한 예측값이 존재한다. 그런데, 공장용지에 대한 모형 수립을 위해 이용하고자 하는 시계열 모형에서는 동일한 시간 간격을 가정하고 있기 때문에 1990년 이전의 자료를 활용할 수 없다. 표 2.1에서 알 수 있듯이 1990년부터 2003년까지 공장부지 면적은 증가하는 추세에 있다. 그러나 2004년 이후에는 광공업 통계조사의 중단으로 공장용지 조사 자료가 없으며, 이로 인해 최근의 추세를 파악할 수 없다. 그러나 공장용지의 70% 이상을 차지하는 산업단지 분양 면적의 경우, 한국산업단지공단에서 1994년부터 2008년까지 연별 자료를 제공하고 있으며, 2000년 4분기부터는 분기 자료로 산업단지 분양면적을 제공하고 있다(참고: <http://www.kicox.or.kr>). 그러므로 산업단지 분양 면적 및 공장용지의 관계를 파악하고, 이를 통해 공장용지의 추이를 파악할 수 있으리라 예상된다.

표 2.1. 산업단지 분양 면적, 공장용지 면적 및 표본조사 결과 (산업단지 분양면적은 매년 12월 기준이며 국토해양부 (2009)의 자료임)

연도	산업단지 분양면적	공장용지 면적	자료출처
1990	-	349.30	
1991	-	371.74	
1992	-	395.65	
1993	-	460.74	
1994	328.07	427.07	
1995	341.38	447.96	
1996	350.80	480.57	
1997	366.61	503.84	통계청 조사
1998	379.80	473.32	
1999	392.22	495.72	
2000	402.90	485.28	
2001	414.75	512.04	
2002	423.47	535.51	
2003	433.26	529.20	
2004	441.94	-	
2005	459.94	-	
2006	462.94	-	공장용지에 대한 자료없음
2007	465.32	-	
2008	461.28	-	
2009	480.03	618.30	
2010	-	671.41	
2011	-	696.69	수요조사
2012	-	718.75	(표본조사 결과)
2013	-	729.11	
2014	-	767.65	

산업단지 분양 면적과 공장용지 면적의 상관계수는 0.89로 매우 높으며, 표 2.1에서 알 수 있듯이 2006년 이후 산업단지 분양 면적의 증가 추세가 감소하기에 최근의 공장부지 면적 또한 감소할 것으로 예상된다. 이제 이들의 관계를 반영하여 적절한 계량 모형을 설정하고, 이를 산업입지 수요조사 결과와 비교하기로 한다.

한편, 산업입지 수요조사는 2009년 10월 기준 전국 5인 이상 제조업체 105,062개 사업체 중 9,734개 사업체를 조사한 것이다. 표본할당은 먹등할당을 사용하였으며, 자체가중 방법으로 2009년에서 2014년까지의 공장용지가 추정되었다 (한근식과 정형철, 2010).

3. 계량 모형을 통한 공장용지 예측

본 절에서는 계량 모형을 통해 공장용지 면적을 예측하고, 이를 표본조사 결과와 비교하기로 한다. 계량 모형은 지수평활, 회귀모형, 산업단지 면적을 이용한 회귀모형을 선정하였으며, 세 가지 시나리오를 사용해 2009년부터 2014년까지의 예측치를 구하였다. 각 모형 및 시나리오에 대한 정의는 표 3.1과 같다.

3.1. 지수평활

승법윈터스 방법을 이용하여 2014년까지의 공장부지 면적을 예측하였다. 분석 방법은 표 3.1과 같이 선

표 3.1. 공장부지 면적 예측 모형 및 시나리오

공장부지 면적 예측 모형	시나리오
지수평활모형	선형추세
	이차추세
	추세없음
회귀모형 I	선형추세
	로그추세
	이차추세
산업단지 분양 면적을 이용한 회귀모형 II	선형추세
	로그추세
	이차추세

표 3.2. 회귀모형 적합 결과

시나리오	모수	추정치	표준오차	t-값	유의확률	DW	결정계수
선형추세	β_0	393700	17764	22.16	<.0001	1.77	0.80
	β_1	10037	1790	5.61	0.0005		
로그추세	β_0	290422	33138	8.76	<.0001	1.88	0.82
	β_1	90231	14893	6.06	0.0003		
이차추세	β_0	438778	11526	38.07	<.0001	1.60	0.76
	β_1	510.37	102.14	5.00	0.0011		

형추세, 이차추세, 추세없음 세 경우이며, 해당 시나리오별 예측 결과는 그림 3.1 Smoothing에 주어졌다. 지수 평활의 경우, 선형추세 시나리오는 2009년 예측치가 589.43km²로 나타난다. 이차추세 시나리오의 경우 2009년 예측치가 364.14km²로 2003년의 529.20km²보다 공장부지 면적이 줄어드는 것으로 나타난다. 추세없음 시나리오의 경우 2009년 예측치가 502.80km²로 2003년보다 줄어드는 것으로 나타난다. 선형추세의 MAPE(Mean Absolute Percent Error)는 2.56%, 이차추세의 MAPE는 3.52%, 추세없음의 MAPE는 5.48%이다. 이차추세 및 추세없음의 MAPE가 상대적으로 높으며, 예측치 또한 2003년보다 감소하는 것으로 나타나 적절한 모형이 아닌 것으로 보인다.

3.2. 회귀모형 I

회귀모형을 이용하여 2014년까지의 공장부지 면적을 예측하였다. 여기서 종속변수는 공장용지 면적이며, 설명변수는 시간이다. 각 시나리오별 회귀모형은 다음과 같다.

- 선형추세 시나리오: $y_t = \beta_0 + \beta_1 \text{time} + \epsilon_t$
- 로그추세 시나리오: $y_t = \beta_0 + \beta_1 \log(\text{time}) + \epsilon_t$
- 이차추세 시나리오: $y_t = \beta_0 + \beta_1 \text{time}^2 + \epsilon_t$

모형적합 결과, 세 가지의 모형 모두 β_0 및 β_1 이 유의하며, 결정계수(R^2) 역시 75% 이상으로 높게 나타난다. DW(Durbin-Watson) 통계량이 2에 가까우므로, 1차 자기상관 역시 없는 것으로 판단된다. MAPE의 경우 선형추세 2.46%, 로그추세 2.24%, 이차추세 2.73%로 지수평활보다 높은 적합도를 보였다. 예측한 결과는 그림 3.1의 Regression I과 같다.

회귀모형을 통한 예측결과, 선형추세 시나리오는 2009년 예측치가 594.44로 2003년보다 증가하였으며, 95% 신뢰구간이 지수평활의 선형추세 시나리오를 포함하는 것으로 나타난다. 로그추세 시나리오의 경

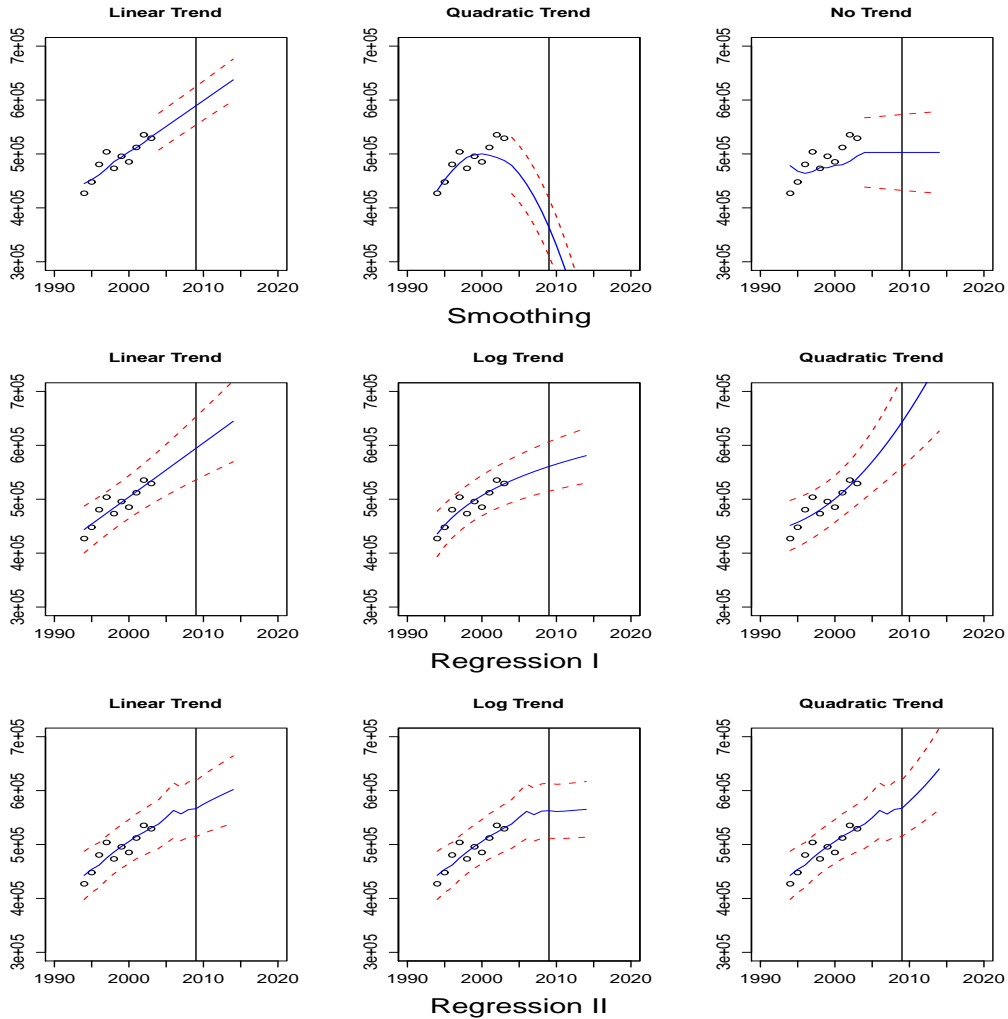


그림 3.1. 지수평활(Smoothing), 회귀모형 I(Regression I), 회귀모형 II(Regression II) 이용한 예측 결과

우 2009년 예측치는 560.73로 2003년보다 증가하나, 선형추세 시나리오보다는 증가추세가 낮으며, 이차추세 시나리오의 경우 2009년 예측치는 642.93로 선형추세 시나리오의 예측치보다 높게 나타났다.

3.3. 산업단지 분양 면적을 이용한 회귀모형 II

산업단지 분양 면적을 이용한 회귀모형은 3.2절의 회귀모형식과 동일하나 설명변수로 time, log(time), time²를 사용한 대신 산업단지 분양 면적의 예측치를 사용한 점에 차이가 있다. 따라서 미래의 공장용지 면적 예측을 위해서는 산업단지 분양 면적을 예측하여야 하며, 이를 위해 시계열 회귀 모형을 사용하였다. 각각의 시나리오 별 설명변수는 time, log(time), time²이며, 종속변수가 산업단지 분양 면적인 회귀모형이다. 자료는 2000년부터 2008년까지의 분기별 자료를 이용하였으며, DW 통계량이 선형추세

표 3.3. 산업단지 분양 면적 예측모형 적합 결과

시나리오	모수	추정치	표준오차	t-값	유의확률	DW	결정계수
선형추세	η_0	416479	6249	66.65	<.0001	1.80	0.95
	η_1	1919	309.93	6.19	<.0001		
	ϕ_1	-0.81	0.11	-7.38	<0.05		
로그추세	η_0	407509	6785	60.06	<.0001	1.43	0.94
	η_1	16817	2518	6.68	<.0001		
	ϕ_1	-0.71	0.13	-5.40	<0.05		
이차추세	η_0	428545	6883	62.26	<.0001	1.60	0.94
	η_1	48.21	11.66	4.13	0.0003		
	ϕ_1	-0.86	0.10	-8.92	<0.05		

표 3.4. 공장용지 면적 예측모형 적합결과

모수	추정치	표준오차	t-값	유의확률	DW	결정계수
β_0	168102	59114	2.84	0.0217	1.72	0.79
β_1	0.84	0.15	5.45	0.0006		

인 경우 1.8, 로그추세인 경우 1.43, 그리고 이차추세인 경우 1.6으로 1차 자기상관이 있는 것으로 판단됨에 따라 오차항이 AR(1)인 모형을 설정하였다. t 시점에서의 공장용지 면적을 y_t 라 하고, 산업단지 분양 면적을 x_t 라 하면, 공장용지 면적을 예측하기 위한 모형은 다음과 같다.

- 선형추세 시나리오: $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$
 $x_t = \eta_0 + \eta_1 \text{time} + \nu_t, \quad \nu_t = \phi_1 \nu_{t-1} + a_t$
- 로그추세 시나리오: $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$
 $x_t = \eta_0 + \eta_1 \log(\text{time}) + \nu_t, \quad \nu_t = \phi_1 \nu_{t-1} + a_t$
- 이차추세 시나리오: $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$
 $x_t = \eta_0 + \eta_1 \text{time}^2 + \nu_t, \quad \nu_t = \phi_1 \nu_{t-1} + a_t$

여기서 a_t 는 오차항으로, 서로 독립이며 $N(0, \sigma^2)$ 을 따른다. x_t 에 대한 모형 적합의 결과는 표 3.3에 주어져있는데, 세 모형 모두 η_0, η_1, ϕ_1 이 유의하며, R^2 역시 95% 이상으로 높게 나타났다. 또한, DW 통계량이 2에 가까우므로, 1차 자기상관 역시 없는 것으로 판단되었다. 위 모형을 통해 2009년부터 2014년까지의 산업용지 면적을 예측한 후, 이를 연별로 정리한 결과를 이용하여 공장부지 면적을 예측하였다.

산업단지 분양 면적은 2009년까지는 실제값을 이용하였으며, 이후의 값은 예측값을 사용하였다. 공장용지 예측 시, 선형추세 시나리오에서는 설명변수로 산업단지 분양 면적 예측값의 선형추세 예측치를, 로그추세 시나리오에서는 산업단지 분양 면적 예측값의 로그추세 예측치를, 이차추세 시나리오에서는 산업단지 분양 면적 예측값의 이차추세 예측치를 사용하였다. 모형 적합 결과는 표 3.4와 같다.

산업단지 분양면적을 이용한 공장용지 추정 결과는 표 3.4와 같이 모두 동일하게 나타난다. 이는 공장용지 면적이 2003년까지만 존재하기에, 모형 적합 시 시나리오와 상관없이 모두 동일한 자료를 사용하기 때문이다. 모형 적합결과 β_0, β_1 모두 유의하며, R^2 는 78.79%로 높게 나타났다. DW 통계량이 2에 가까우므로, 1차 자기상관 역시 없는 것으로 판단된다. MAPE는 2.5%로 지수평활보다 높은 적합도를 보인다. 회귀모형 II의 예측 결과는 그림 3.1 Regression II와 같다. 선형추세 시나리오는 2009년

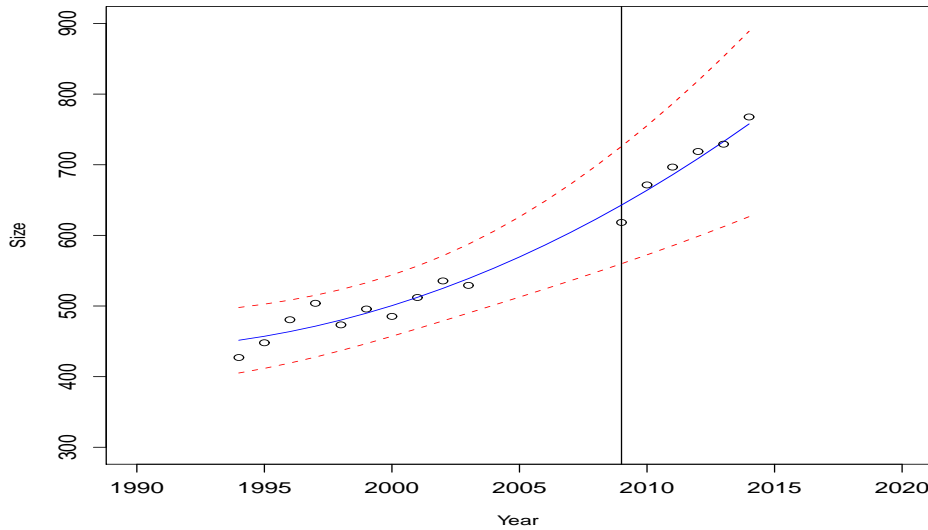


그림 3.2. 회귀모형 I의 이차추세 및 표본조사 예측치

예측치가 566.45km^2 로 회귀모형 I의 선형추세 및 이차추세 예측치보다 낮게 나타난다. 로그추세 시나리오 경우 2009년 예측치는 562.72km^2 로 선형추세 예측치보다 낮으며, 이차추세 시나리오의 경우 567.27km^2 로 선형추세 보다 약간 높게 나타났다.

3.4. 모형 선택

3.1절에서 3.3절의 모형을 비교하기로 한다. 오차를 기준에서는 회귀모형 I이 좋게 나타나고 있다. 반면, 미래 추정치 경우 회귀모형 II의 추정치가 비교적 낮으며, 다음으로 지수평활모형 그리고 회귀모형 I의 추정치가 높게 나타나고 있음을 볼 수 있다. 단순히 추정치 크기의 순으로 살펴보면, 회귀모형 I이 산업단지 분양 면적을 이용한 회귀모형 II에 비해 과대추정 되는 것처럼 보이므로 산업단지 분양면적을 사용한 회귀모형 II가 의미 있을 수 있으나, 2004년부터 2008년까지 공장부지 면적에 대한 참값을 알 수 없으므로 어느 모형이 적절한지 판단할 수는 없다. 그러므로 표본조사를 통한 수요예측 결과를 바탕으로 적절한 추세모형을 선택하기로 한다.

모형 선택에서는 건설교통부 (2007)에서 미래예측으로 사용한 홀트윈터스 이모수평활 방법과 ARIMA 모형을 같이 사용하여 표본조사 결과와 차이를 비교하였다. 그런데, ARIMA 모형수립은 건설교통부 (2007) 결과와 달리 ARIMA(1, 1, 1)으로 설정하였으며, 각 지역별 추정 결과를 합한 값을 사용하였다. 모형 판단의 기준은 표본조사 결과를 참값으로 가정하고, 표본조사 예측치와 가장 유사한 모형을 찾는 방법을 사용하였다. 이에 의하면 회귀모형 I의 이차추세 모형 예측치가 표본조사 결과와 가장 유사함을 볼 수 있었다. 따라서 본 연구에서는 2003년부터 2008년까지의 단절된 공장용지 면적의 값을 이차추세 시나리오의 회귀모형 예측치로 대체(imputation)하고, 2009년부터 2014년까지의 값은 표본조사의 예측치, 1990년부터 2003년까지는 통계청의 공장용지 면적을 사용하여 공장용지에 대한 새로운 추정을 시도하기로 한다.

그림 3.2는 회귀모형 I의 이차추세 및 표본조사 결과를 보여준다. 실선이 이차추세 예측치이며, 2009년 이후의 값은 표본조사 결과이다.

표 3.5. 표본조사 결과와 각 모형의 미래 예측치와의 오차

연도	표본조사	홀트윈터스	ARIMA	지수평활		
				선형	이차	무추세
2009	618.30	716.55	649.68	589.43	364.14	502.80
2010	671.41	738.46	667.95	599.03	330.32	502.80
2011	696.69	760.37	687.26	608.63	292.87	502.80
2012	718.75	782.29	705.51	618.23	251.79	502.80
2013	729.11	804.20	724.70	627.83	207.08	502.80
2014	767.65	826.11	743.45	637.43	158.74	502.80
잔차합		-426.06	23.36	521.33	2596.97	1185.11

연도	표본조사	회귀모형 I			회귀모형 II		
		선형	로그	이차	선형	로그	이차
2009	618.30	594.44	560.73	642.93	566.45	562.74	567.27
2010	671.41	604.47	565.13	663.85	574.82	561.20	580.23
2011	696.69	614.51	569.33	685.80	582.08	561.85	593.72
2012	718.75	624.55	573.34	708.76	588.86	562.96	608.09
2013	729.11	634.58	577.18	732.75	595.43	564.11	623.52
2014	767.65	644.62	580.87	757.76	601.93	565.20	640.13
잔차합		484.74	775.33	10.06	692.34	823.85	588.95

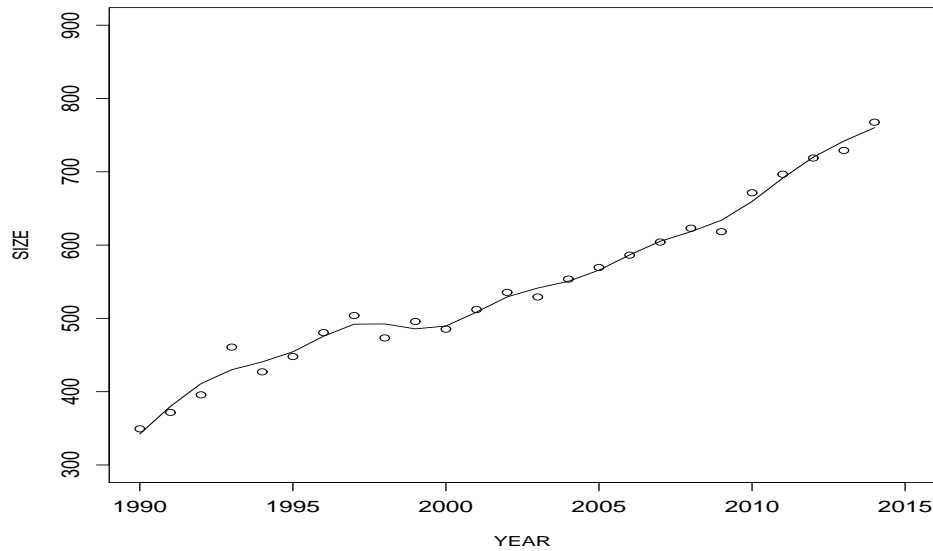


그림 3.3. 스플라인 회귀평활결과, 자유도 = 10

4. 스플라인 평활에 의한 공장용지 예측

본 절에서는 이차추세를 가정한 회귀모형에 의한 대체값과 표본조사 결과에 의한 예측치를 융합하여 미래 공장용지를 예측하기로 한다. 사실, 수요예측에서 선택한 모형에 의한 분산과 표본조사의 오차를 가중치로 놓고, 두 예측값의 가중평균으로 미래 공장용지에 대한 예측을 수행할 수 있으나, 본 연구에서는 자료의 추세를 감안하여 스플라인 회귀평활로 공장용지를 추정하기로 하였다. 즉, 스플라인 평활에

표 4.1. 스플라인 회귀 평활을 사용한 공장용지 예측값

연도	사용자료	이차추세 회귀모형 I			스플라인 회귀 평활 예측값
		예측값	95% 신뢰하한	95% 신뢰상한	
1994	427.07	451.54	405.17	497.90	440.65
1995	447.96	457.15	411.69	502.61	454.44
1996	480.57	463.79	419.22	508.35	475.59
1997	503.83	471.44	427.67	515.21	492.07
1998	473.32	480.12	436.91	523.32	492.43
1999	495.72	489.82	446.80	532.83	485.67
2000	485.28	500.53	457.20	543.87	489.61
2001	512.04	512.27	467.95	556.60	508.59
2002	535.51	525.03	478.93	571.13	529.32
2003	529.20	538.81	490.05	587.57	541.41
2004	553.61	553.61	501.29	605.93	550.54
2005	569.43	569.43	512.63	626.23	565.77
2006	586.27	586.27	524.12	648.43	586.60
2007	604.14	604.14	535.79	672.48	605.16
2008	623.02	623.02	547.70	698.34	618.32
2009	618.30	642.93	559.90	725.95	633.88
2010	671.41	663.85	572.43	755.27	659.53
2011	696.69	685.80	585.32	786.27	691.27
2012	718.75	708.76	598.62	818.91	719.99
2013	729.11	732.75	612.34	853.17	741.90
2014	767.65	757.76	626.50	889.02	760.17

사용된 자료는 1990년에서 2003년까지 통계청의 공장용지 자료, 2004년에서 2008년까지 회귀모형 I의 이차추세 모형으로 대체된 자료, 2009년에서 2014년까지 표본조사에 의한 수요예측 자료이다. 이처럼 서로 다른 3조각의 자료를 융합하였다. 스플라인 평활에서는 사용하는 자료의 크기(자유도)에 영향을 받는데, 자유도 3에서 자유도 12까지 변화를 주면서 예측값의 변화를 관찰하였다. 자유도가 낮을수록 공장용지 예측값은 이차모형보다 더 급격한 수준으로 공장용지가 증가할 수 있다는 결과를 유도하였다. 즉, 자유도가 낮을수록 공장용지 수요가 시간의 제공에 비례하여 급격히 증가하는 경향으로, 자유도가 높을수록 미래예측치는 표본조사 예측 결과를 더 반영하는 방향으로 움직임을 볼 수 있었다. 그러므로 자료적합적인 방법으로 자유도 10을 최종 예측모형으로 선택하였다. 그림 3.3은 자유도 10에 대한 공장용지 평활 결과이다. 또한, 표 4.1은 이차추세 예측치와 최종 스플라인 평활값을 보여준다.

스플라인 평활결과 2009년 공장용지 수요는 표본조사 예측치 618.30km²보다 다소 높은 633.88km²로 추정되며, 2010년은 표본조사 예측치 671.41km²보다 다소 낮은 659.53km²로 추정되고 있음을 볼 수 있다. 이는 표본조사 시 조사시점의 문제로 2009년 수요값의 일부가 2010년 결과에 포함되어, 2010년 공장용지가 과대추정되는 문제를 스플라인 평활로 예측치를 다소 완화시키도록 한 결과임을 의미한다. 또한, 표본조사 결과에는 2014년의 미래 공장용지 수요가 비교적 높게 나타나는 비표본오차를 포함하고 있는데 반해, 자유도 10의 스플라인 평활을 사용함으로써, 2014년에는 표본조사 결과보다 다소 낮게 공장용지가 예측되며 2014년의 추세가 반영된 2013년에는 표본조사 결과보다 공장용지 수요가 더 높게 예측되도록 하는 효과가 나타나도록 하였다. 즉, 스플라인 회귀평활이 미래투자 시점의 불투명성 때문에 2014년에 수요가 증가하는 추세를 감소시키고, 전년도의 값을 다소 올려주는 추세평활을 하고 있음을 볼 수 있다.

5. 결론

공장용지 면적 수요추정은 ‘산업집적활성화 및 공장설립에 관한 법률’에 의거 지식경제부가 국토해양부에 송부하고, 국토해양부는 ‘산업입지 및 개발에 관한 법률’에 따라 16개 시도부에 송부하여 각 시도가 해당 지역의 ‘산업입지 수급계획’에 해당 결과를 반영토록 하는 매우 중요한 사안이다. 그런데, 2003년까지 통계청에서 공장용지 면적을 산출하였으나, 2004년부터 전국 공장용지 면적 추계가 제외되었다(한국산업단지공단, 2009). 이에 공장용지를 추정하기 위한 과거의 계량적 시계열모형은 실제 참값을 모르는 상황에서 정확한 판단을 내리는데 한계가 있었다. 그러므로 2009년 산업부지 수요조사 목적으로 한국산업단지공단에서 실시한 표본조사는 그동안의 계량적 모형이 어느 정도 잘 적용되어 왔는가에 대한 매우 중요한 판단근거를 마련해 준다. 또한, 약 1만개 정도의 대규모 표본조사 결과를 활용함으로써 신뢰성 있는 통계를 제공하여주는 장점이 있었다. 본 연구에서는 표본조사 결과를 활용하여 2003년까지의 공장용지 면적 자료에 근거한 계량적 예측 모형에 대해 어느 것이 좋은 모형인가에 대한 판단 근거를 마련하고 추정모형에 대한 바로미터 역할을 수행하도록 하였다. 또한, 수립된 계량 모형과 표본조사 결과를 융합하여 미래 공장용지 수요를 예측하였다. 전체적 분석과정 다음과 같다.

첫째, 2003년까지의 자료를 기초로 여러 계량 모형을 수립하였다. 둘째, 표본조사에 의한 미래 공장용지 수요 예측치를 분석에 활용하였다. 셋째, 통계청 조사가 실시되지 않았던, 2004년부터 2008년까지의 공장용지를 계량모형으로 대체하였다. 넷째, 표본조사와 계량모형의 적절한 가중치로 미래 공장용지를 예측하였다. 이때는 자료분석적 방법으로 스플라인 회귀 평활방법을 사용하였다. 스플라인 회귀평활 결과 2010년부터 2014년까지 공장용지는 659.53km², 691.27km², 719.99km², 741.90km², 760.17km²로 추정되었다. 사실, 표본조사를 매년 실시한다면, 계량모형에 대한 꾸준한 수정을 가할 수 있으며 우리나라 공장용지 면적에 대한 희망수요를 산출할 수 있는 장점과 더불어 제조업 전수조사에 비해 훨씬 적은 비용으로 정확한 공장용지 면적을 예측하는데 많은 도움이 될 것이다. 하지만, 표본조사를 매년 할 수 없는 경우를 가정한다면, 우리나라 공장용지 추세에 대한 모형을 설정하고 이를 기초로 표본조사와 계량모형과의 융합을 시도하는 본 연구 역시 나름의 의미가 있다고 판단한다. 결론적으로 본 연구를 통해 지역별, 입지유형별, 업종별 산업입지 수급계획 수립에 실질적 도움이 되길 희망한다.

참고문헌

- 건설교통부 (2007). <산업입지 공급계획 수립지침 연구>, 국토연구원.
 국토해양부 (2009). <2009 전국산업단지 통계>, 국토연구원.
 한국산업단지공단 (2009). <산업입지 수요추정 모형 개발 및 입지수요 추정>, 한신대학교.
 한근식, 정형철 (2010). 공장용지 추정을 위한 표본설계 및 표본조사에 의한 수요예측, *Journal of the Korean Data Analysis Society*, **12**, 997-1008.
 한국산업단지공단, <http://www.kicox.or.kr>

Forecasting the Demand Areas of a Factory Site: Based on a Statistical Model and Sampling Survey

Hyeong Chul Jeong¹ · Geunshik Han² · Seongyong Kim³

¹Department of Applied Statistics, University of Suwon

²Department of Computer Engineering, Hanshin University

³Economics and Statistics Institute, Korea University Sejong Campus

(Received February 2011; accepted April 2011)

Abstract

In this paper, we have considered the problems of the estimation of the gross areas of a factory site relating to the areas of industrial complex lands based on a statistical forecasting model and the results of a sampling survey. In respect to the data of a gross areas of a factory site, we have only the sizes from 1981–2003. In 2009, the *Korea Industrial Complex Corp.* conducted a sampling survey to estimate its bulk size, and investigate the demands of its sizes for the next five years. In this study, we have adopted the sampling survey results, and have created a statistical growth model for the gross areas of a factory site to improve the prediction for the areas of a factory site. The three-different parts of data: the results of areas of a factory site by Korea National Statistical Office, imputation results by the statistical forecasting model, and sampling survey results have used as the basis for analysis. The combination of the three-different parts of data has created a new forecasting value of the areas of a factory site through the spline smoothing method.

Keywords: Areas of a factory site, forecasting model, sampling survey, spline smoothing.

¹Corresponding author: Associate Professor, Department of Applied Statistics, University of Suwon, Hwaseong, Gyeonggi 445-743, Korea. E-mail: jhc@suwon.ac.kr