



주거용 건물의 에너지 실사용량의 불확실성을 내포한 설명변수 인자에 대한 빅데이터 분석 기반의 정량화 방법 - 서울지역의 공동주택을 중심으로

*The Method of Quantitative Analysis Based on Big Data Analysis for Explanatory Variables Containing Uncertainty of Energy Consumption in Residential Buildings
- Focused on Apartment in Seoul Korea*

최준우* · 안승호** · 박병희*** · 고정림**** · 신지웅*****

Choi, Jun-Woo* · Ahn, Seung-Ho** · Park, Byung-Hee*** · Ko, Jung-Lim**** · Shin, Jee-Woong*****

* Division of New Business Dev. & Abroad Business, EAN Technology, Seoul, Korea (cjw@eantec.co.kr)

** Corresponding author, Division of New Business Dev. & Abroad Business, EAN Technology, Seoul, Korea (sha@eantec.co.kr)

*** Coauthor, Division of New Business Dev. & Abroad Business, EAN Technology, Seoul, Korea (pbh1111@eantec.co.kr)

**** Coauthor, Division of New Business Dev. & Abroad Business, EAN Technology, Seoul, Korea (jko@eantec.co.kr)

***** Coauthor, EAN Technology, Seoul, Korea (sjw@eantec.co.kr)

ABSTRACT

Purpose: The energy consumption of apartment units is affected by the lifestyle of the residents rather than system technology. In this study the numerical analysis of assumed energy consumption correlation factors with arbitrary value due to uncertainty. It is intended to be used as a simulation correction value which can be utilized as a predicted value of actual energy usage. The correction value of the simulation is set in the developed form of the existing process that derives the actual usage amount. The simulation results used in the existing evaluation system are used to maintain the useful value as the current system evaluation scale and predict the actual capacity. **Method:** The method of the study is to statistically analyze the data frames of all complexes capable of collecting the annual energy usage and to reconstruct the population by adding the variables that are expected to be correlated. Repeat the data frame configuration with variables that are assumed to be highly correlated with energy use levels. Determine whether there is correlation or not. The intensity of the external characteristics of the building equipment related to the energy consumption is presented as the quantitative value. **Result:** The correlation between electricity consumption and trading price since 2010 is analyzed as (Correlation coefficient 0.82). These results are higher than (Correlation coefficient 0.79), which is the correlation between residential area and trading price. This paper signifies the starting point of the methodology that broadens the field of view of verification of simulation feasibility limited to the prediction technique focused on the simulation tool and the element technology scope. The diversified phenomenon reproduction method develops the existing energy simulation method. It can be completed with a simulation methodology that can infer actual energy consumption.

KEYWORD

주거건물
에너지 사용량
통계
빅데이터
데이터기반 분석

Residential building
energy consumption
Statistic
Big data
Data evidence analysis

ACCEPTANCE INFO

Received May 18, 2017
Final revision received June 14, 2017
Accepted June 19, 2017

© 2017 KIEAE Journal

1. 서론

1.1. 연구의 배경 및 목적

친환경에 대한 중요성 인식이 높아짐에 따라 건축물의 에너지 사용량은 친환경성 평가에 활용된다. 에너지 사용량 예측을 위한 방법으로 대표적인 건축물 에너지 시뮬레이션기법은 각 기준과 방법론을 달리하며 평가 주체별로 독립적으로 운영된다. 건축물 에너지 시뮬레이션 결과 값인 에너지 사용량은 등간척도(Interval scale)¹⁾ 수준에서 의미를 찾을 수 있다. 등간척도 위에

의 결과는 수많은 확률론적 변수들을 약속된 조건으로 가정하고, 시간의 개념이 소거되어 현 시점에서의 건축요소기술 및 설비 시스템에 대한 성능평가 범위로 한정된다. 이러한 범위 한정 은 편향성(Bias)을 제거하고 유효성(Efficiency)을 확보하며 일치성(Consistency)을 높이는 효과가 있다. 따라서 법제도 상에서 기준 도달 유무를 평가하는 유용한 평가 척도로 작동한다.

그러나 전과정 평가(LCA)에서 요구하는 에너지사용량은 현재 시점에서의 건축 및 설비요소의 시스템평가를 위한 등간척도가 아닌 실사용량에 대한 예측이 요구된다. 즉, 등급설정 및 수준 평가를 위한 절대치의 비율척도(Ratio scale)²⁾ 수준을 필요로 한

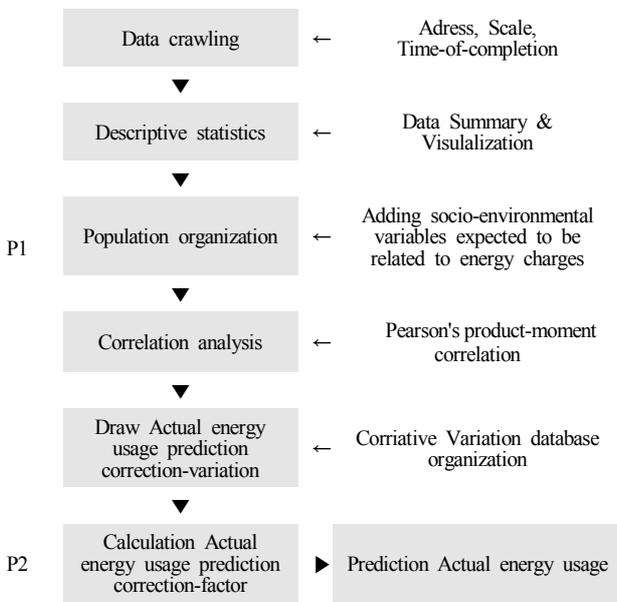
(온도, 물가지수, 옷사이즈 등)

2) 영점을 가지는 수치적 척도, 절대적 크기를 가진(체중, 나이, 교통사고 건수, 신

pISSN 2288-968X, eISSN 2288-9698
http://dx.doi.org/10.12813/kieae.2017.17.3.075

1) 자료가 가지는 특성의 양에 따라 순위를 매길 수 있는 척도, 상대적 크기를 가진

Table 1. Flow of research



2. 공동주택 데이터 수집 및 분석 개요

2.1. 데이터 수집 개요

서울시내 공동주택의 1차 수집 벡터는 구, 동, 단지명, 준공년도, 규모(주거용면적), 세대수, 단지 전기사용요금, 전기사용량(kWh/yr)이다. 규모를 세대수로 나누어 단지별 단위세대 구성의 대표성을 가지도록 평균적 단위세대규모로 치환하였다. 치환된 단위세대면적을 기준 값과 단지 전기사용요금으로 세대별 요금, 단위면적당 요금을 추가로 산출 하여 변수로 설정 하였다. 세대별 사용량과 단위면적당 에너지사용량 또한 세대요금 산출 방식과 같은 방법을 통해 추가변수로 설정 하였다.

2.2. 데이터 분석 환경

데이터 분석은 오픈 소스 R 프로그래밍 언어를 사용하였다. R 은 S 언어에서 유래되었으며, S 언어는 그래픽과 통계모형 적합 능력이 결합된 전문 통계 프로그래밍 언어였다. S의 상업적 통계 분석 프로그램으로 S-PLUS 가 개발되었으나, 너무 비싼 비용으로 인해 라이선스를 구입하기에 어려움이 있었다. 이러한 문제를 해결하기 위해 뉴질랜드 오클랜드 대학의 통계학자인 로스 이 하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)이 새로운 S 버전을 개발하였다. 이것을 개발자 이름의 알파벳을 따라 R 로 명명하였으며 2000년 2월 29일 버전 1.0.0이 최초로 배포되었다.

본 연구에서 전기에너지 비용을 반응변수(Response variable) 로 설정하고 나머지 설명변수(Explanatory variable) 간의 상관 분석을 위해 R에서 기본 제공되는 통계패키지와 시각화 라이브러리를 활용 하였다.

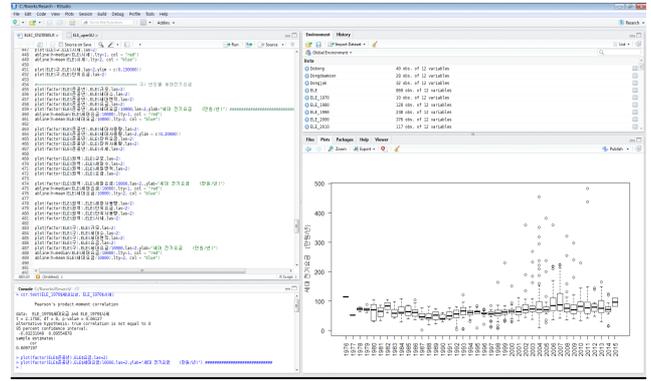


Fig. 3. R programming integrate develop language editor RStudio

3. 데이터 분석 결과

3.1. 기술통계 결과

Fig. 4. 은 준공 년도별로 구분되어 세대별 연간 전기사용 요금을 나타낸다. 2000년대 초반부터 현재까지 공동주택의 년도별 boxplot의 상위25% 1분위 범위와 4분위 범위에서 벗어난 이상치(outlier)값이 서울시내 공동주택 전기에너지 요금의 평균값(Mean) 및 중위값(Median)을 크게 넘어선다. 지속가능성(Sustainable)이 화두가 되기 시작하며 정부차원에서 건축물에서 에너지절약을 위한 건축기준이 강화되기 시작한 시점과 맞물려 있다. 이러한 결과로 비추어 볼 때, 주거건물에서 실제 에너지 소비량은 건물의 노후화, 법규정책에 따른 건물에너지 성능요향상과는 선형관계성이 없다. 1977년부터 2000년 이전 공동주택 가정의 전기에너지 요금은 서울시 전체 평균요금에 미치지 못하는 분포를 보인다. 그러나 1976년 준공된 단지 중위값이 모든 년도의 중위값을 상회하는 값으로 나타났다. 이는 강남구 압구정동 H 아파트 단지의 가정 전기요금이다.

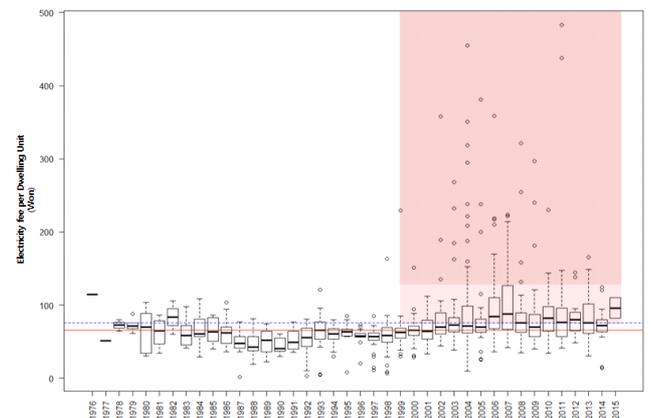


Fig. 4. Annual electric energy fee in household

서울 권역별 사용요금은 Fig. 5. 에서 같이 강남권역과 도심권역에서 전기사용요금의 평균치를 상회하였다.

여 데이터를 분석하였다. 따라서 Fig8~10 에 나타난 Boxplot의 세대전기요금은 단위세대규모의 오차가 제거된 값임으로 동별 세대전기요금수준을 판단하기에 적합하다.

3.2. 에너지비용 상관 예상 변수 추가

에너지 사용요금을 높게 지불하는 지역은 언론 매체등을 통해 일반적 상식선에서 매매 및 임대 시세가 높다고 인식되어지는 지역이다. 따라서 에너지 사용요금과 매매시세는 선형 상관을 보이며, 선형은 ‘에너지사용요금이 높은 지역은 매매시세가 높은 양(+)'의 상관을 가진다.’ 라는 가설을 설정하였다. 가설을 검증하기 위해 모집단의 매매시세를 웹 크롤링 하여 설명변수로 추가하였다. 구별 매매시세에 대한 Boxplot 결과(Fig. 10.) 강남구 서초구 용산구의 매매시세는 전기요금에서 분석된 패턴과 동일하였다. 이러한 패턴결과에 따라 ‘매매시세가 높은 주거시설은 에너지 소비또한 높다’ 라는 연구가설을 설정 할수 있는 타당성을 확보 하였다.

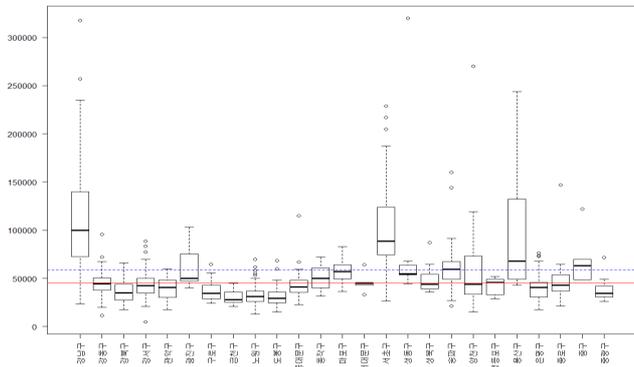


Fig. 10. Home market price by Gu

매매 시세와 세대별 전기요금이 평균을 상회하는 강남, 서초, 용산구에대한 매매시세와 전기사용요금에 대한 관계를 다음 Fig. 11. 과 같이 시각화 하였다. 시각화 결과에서 보이는 바와 같이 모집단의 분포는 매매 시세와 세대 전기요금간의 양의 상관을 가지는 선형 분포를 나타낸다.

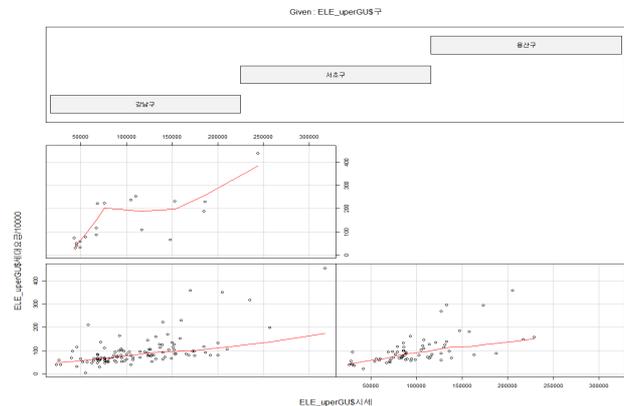


Fig. 11. Correlation plot between market price and Annual electric energy fee

3.3. 상관분석 결과

연구조사의 목적과 자료의 종류에 따라 상관정도를 평가하는 기준은 상이하나 사회과학 분야에서 사용되는 일반적 기준은 다음 Table 2. 와 같다.

Table 2. Criteria for evaluation of correlation coefficient

Value	Level of Agreement
0 ~ 0.2	None
0.2 ~ 0.4	Weak
0.4 ~ 0.6	Moderate
0.6 ~ 0.8	Strong
Above. 0.8	Almost perfect

선형의 양의 상관 분포를 나타내는 결과에 따라 상관계수 분석을 수행하였다. 수행과정은 분석 스케일에 따라 전체 상관도가 오염될 가능성을 고려하여 년도 구간별 검정통계량을 산출하여 상관의 유의성을 검증하였다. 유의성 검증을 위한 가설 설정은 다음과 같다. 연구가설: 매매시세와 세대전기사용량 간에는 선형관계가 있다. 귀무가설: 매매시세와 세대전기사용량 간에는 선형관계가 없다. 년도 구간 스케일별 산출된 피어슨의 상관 계수(cor)와 자유도(DF)에 따라 검정통계량 t값을 임계치와 비교하면, 검정통계량 t값이 우측임계치보다 스케일 구분별 케이스에서 모두 크기 때문에 귀무가설을 기각하고 연구가설을 채택한다. 따라서 매매시세와 세대전기사용량의 상관계수 값으로 산출된 값은 유의수준(α) 0.05에서 검정한 결과 통계적으로 유의하다. 그러나 1970~1979 년도 스케일의 p-value 값은 0.05 이하임으로 통계적 유의성에 대해 재검증이 필요하다.

Table 3. Significance Verification by Yearly range scale for Correlation analysis

Scale	p-value	DF	t-test	cor
1970~1979	0.06127	8	2.175	0.609
1980~1989	2.20E-16	126	13.338	0.765
1990~1999	1.67E-13	236	7.828	0.454
2000~2009	2.20E-16	374	22.754	0.761
2010~2015	2.20E-16	115	15.418	0.820

1980년 이전 준공된 단지는 Fig. 12.의 상관분석 그래프에서 보이는바와 같이 매매 시세와 세대별 전기요금간의 상관도는 0.61의 강한 상관도를 보인다. 그러나 선형성을 판단하였을 때 일부 모집단의 케이스에서 이상치를 나타냈다. 이는 재건축 예정 단지의 경우 매매가는 높으나, 임대가는 낮은 경향을 보인다. 따라서 실거주자가 임대거주일 비율이 높은 단지의 경우 매매가와 에너지 사용패턴이 선형적이지 않은 결과가 나올 가능성이 높다. 본 연구에서는 매매시세와 에너지 사용량의 관계를 분석한

다. 일반적으로 매매가와 임대가는 선형관계를 보임으로 매매시세와 임대시세를 동일한 성향의 변수로 가정한다. 그러나 재건축 예정단지의 경우 매매가와 임대가가 선형관계를 보이지 않으므로 1970년대 모집단을 분석에서 제외시킨다.

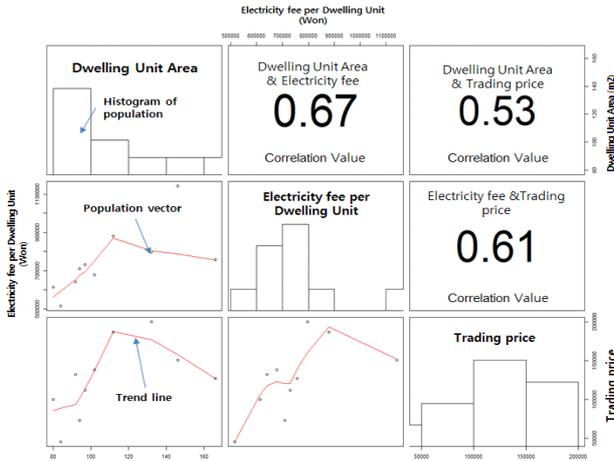


Fig. 12. 1970~1979 range scale of Correlation analysis between market price and Annual electric energy fee

1970년대 모집단이 제거된 상관분석 결과는 Fig. 13.과 같다. 매매시세와 세대전기요금의 상관계수는 0.69로 강한 상관관계로 되었다.

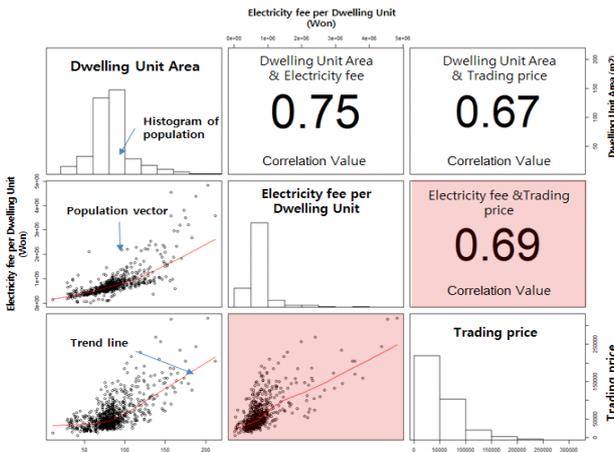


Fig. 13. 1980~2015 range scale of Correlation analysis between market price and Annual electric energy fee

Fig. 14.에서와 같이 2010년 이후 모집단의 상관계수는 0.82로 매매시세와 전기사용요금간의 거의 완벽한 상관관계를 나타낸다. 이 수치는 매매시세와 세대면적의 상관도가 0.71이며, 세대면적과 전기에너지 사용량 상관도가 0.79로 나타남과 비교하였을 때, 단위면적당 에너지사용량과 평당 매매단가 보다, 매매시세와 전기사용량이 오히려 깊은 상관도를 나타내는 것이다. 단위면적당 에너지 사용량을 원단위로 사용하여 면적에 따른 에너지 사용량을 유추하는 방법은 일반적으로 통용되는 에너지 사

용량 산출방법이다. 또한 거주 평수가 높을수록 주택 매매 시세는 높아지는 것이 일반적으로 수용되는 상식이다. 그러나 2010년 이후 준공된 공동주택에서는 세대면적과 매매시세의 상관도 그리고 세대면적과 에너지사용량의 상관도보다 매매시세와 전기사용량의 상관도가 가장 높게 분석된 결과로 나타났다. 물리적 범위의 상관요소보다 오히려 인문사회적 범위의 상관요소가 더 깊은 상관을 보이는 것으로 해석된다.

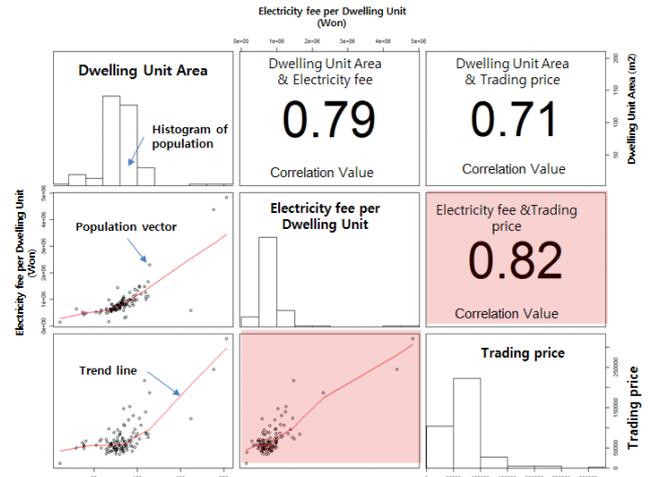


Fig. 14. 2010~2015 range scale of Correlation analysis between market price and Annual electric energy fee

4. 결론

기술통계 분석결과 최근 2000년 이후 신규 공동주택 단지에서 노후화된 1999년 이전 공동주택 단지보다 에너지 사용요금이 많이 납부되고 있었다. 2000년 이후 친환경인증을 받은 공동주택 단지가 누적 될수록, 서울 시내의 공동주택 에너지 실사용량은 폭증 하였다. G-SEED인증제도의 발전 및 정착과 더불어 실제 에너지 사용량이 폭증한 가장 큰 원인은 건축물 에너지 사용량을 실제의 '예측'이 아닌 요소기술범위에 한정된 '성능평가'가 중심이 된 원인으로 작동한 것이다.

본 논문은 실제 값을 예측하기 위해 '불확실성이 내제되어 있는 에너지시물레이션을 위한 인풋데이터는 기술요소범위 내에서 충분히 작동 되는가?'에 대한 의문에서 출발하였다. 이러한 출발점에서 인문 사회적 범주로 시각을 달리하여 건축물의 에너지 사용량을 바라보았다. 시물레이션 틀에 집중된 예측기법과 요소기술 범위로 한정된 시물레이션 타당성에대한 검증시야를 넓히는 방법론의 시작점이 되었다는 것에서 의미를 찾을 수 있다. 그러나 매매시세와 에너지 사용량과의 높은 상관도에 대해 증명 하였지만, 인과관계검증에 대한 연구를 수행 못한 것은 본 연구의 한계로 밝힌다. 인과관계검증에 대한 연구를 수행하지 못한 이유는 연구가설을 검증하는 과정에서 수집 불가능한 데이터가 존재하였기 때문이다. 예를 들어, '시물레이션의 주요 설정요소인 운영 스케줄은 거주자의 생활패턴에 큰 영향을 받는다.'는 이론을 바탕으로 '매매시세는 거주자의 생활패턴과 관련이 있

다.’ 라는 가설을 제시했을 경우, 본 논문에서 주요 연구 자료로 사용한 빅데이터만으로는 개개 가정의 생활패턴에 대해 조사하고 그 인과관계를 밝히기는 어려웠다. 따라서 해당 가설의 검증을 위해서는 본 논문과는 다른 방식의 후속연구(설문조사 등)를 통해 생활패턴과 매매시세와의 인과관계를 검증하고자 한다.

인문사회적 범위로의 시야확대를 통한 다양화된 시뮬레이션 현상재현 방식은 실제 에너지 소비량을 유추 할 수 있는 시뮬레이션 방법론으로 발전 될 수 있을 것이다.

Acknowledgements

본 연구는 국토교통부 국토교통기술촉진연구개발사업의 연구비 지원(16CTAP-C114806-01)에 의해 수행되었습니다.

Reference

- [1] 유충현, 홍성학, “R을 활용한 데이터 시각화”, 인사이트. 2016 // (Yu, Chung-Hyun, Hong, Sung-Hak, R-Visualization, Insight, 2016)
- [2] Michael J.Crawley, STATISTICS: An Introduction Using R, Second Edition, Wiley, 2015
- [3] Paul Teetor, R Cookbook, O’Reilly Media, 2011
- [4] Cohen, J, Statistical power analysis for the behavioural sciences Second Edition, Lawrence Erlbaum Associates, 1988