翻譯資料
90-01-001

경제기획원
조사통계국
도서관장서

計量經濟學入門

1990. 1.

調査統計局 統計分析課

경제기획원조사통계국
접
1990. 3. 28
수

머 리 말

우리국에서는 1981년 3월부터 국내 景氣變動의 측정과 예측을 위하여 景氣綜合指數를 매월 작성해 오고 있으며, 최근에는 경기에 대한 예측력을 제고하고자 기업경영자들이 판단하는 경기전망을 기초로 企業實査指數(BSI)를 편제하게 되었습니다.

그러나 여러가지 경제의 내·외적 요인에 의하여 발생하는 경기변동을 보다 합리적으로 측정하기 위해서는 더욱 다양한 測定方法이 요구되고 있으며, 특히 경제정책 효과의 時差分析과 장·단기 예측 및 속도측정이 가능한 計量經濟模型의 개발 필요성이 날로 증대되어 우리국에서는 1989년부터 계량경제모형 개발에 착수하여 현재 構造方程式 도출작업을 진행중에 있습니다.

이 자료는 동 모형개발과 관련하여 이에 참여하고 있는 직원들의 전문성을 향상시키고 개발작업을 원활히 수행하게 하기 위하여 계량경제학에 관한 基礎理論書의 하나인 " Introduction to Econometrics- Principles and Applications "를 완역한 것입니다.

아무쪼록 본 자료가 담당직원들과 계량경제모형 설계에 관심있는 분들의 이론습득을 용이하게 하고 나아가 개발되는 계량모형을 獨自的으로 운용할 수 있는 능력을 배양하는 데 많은 도움을 줄 수 있기를 기대합니다.

1990. 1.

統 計 分 析 課 長

차 례

제 2 판 서 문

새로운 제 2 판의 목적은 세가지이다. 첫째로, 이 책에서 개발한 계량경제학 기법에 대한 응용과 實例의 수와 범위를 넓히는 것이다. 둘째로, 非線型模型(non-linear model)의 추정을 포괄하는 분석의 확장이다. 셋째로, 제 1 판에 있었던 약간의 이론적 쟁점에 관한 처리를 올바르게 하고 명확히 하는 것이다. 결과적으로 제 2 판에는 相關 및 回歸分析 (correlation and regression analysis)의 몇가지 새로운 예를 편입시켰으며, 비선형체계의 추정에 관해서 새로이 제 8 장을 추가하였다.

제 1 판에 익숙한 독자들은 새로운 실례가 다양하다는 점을 발견할 것이다. 그 예들은 어떤 경우에는 학생들에게 추정한 母數(parameter)에 대한 실제적인 계산을 보여주고, 다른 사례에서는 중요한 모수를 추정하기 위해, 그리고 미시경제학 및 거시경제학의 몇가지 중심적인 假說을 검증하기 위해 회귀분석을 어떻게 이용하였는지를 보여주는 경제학 문헌에 직접 근거하고 있다. 새로운 예들은 單純相關分析 (simple correlation analysis), 실물 재화와 화폐잔고에 관한 수요곡선들의 추정, 총계(aggregation)로 인하여 발생하는 이분산성(heteroscedasticity)에 관한 사례연구를 포함하고 있다. 이렇게 추가된 사례들이 학생들로 하여금 計量經濟 模型의 정식화와 실제 자료의 사용, 그 결과의 해석을 더 잘 이해하는 데에 도움이 되기를 바란다.

실제로 대개의 계량경제모형은 線型이 아니다. 그러나 대부분의 학부와 대학원 교재는 非線型模型이 너무 어려워서 소개하지 못하고 명백한 가정을 토대로 線型模型만을 다루고 있다. 우리는 비선형모형이 어렵지 않음을 확신하고 제 8 장에서 학부수준정도로 비선형모형에 관한 많은 논의를 소개하였다. 교재내에 이전부터 있었던 題材에 의거해서 8 장에서는 비선형모형

에서의 識別(identification), 推定(estimation), 假說檢定 (hypothesis testing)의 문제를 탐구하고, 이러한 기법을 경제문제에 응용하는 것을 소개한다.

제 2판에서의 기본적 목적은 제 1판과 다르지 않다. 우리는 기초적인 수학과 통계학 기술에만 의거하여서 광범위한 계량경제학 기법을 소개하려고 노력하였다. 책의 서술방식에 관해서는 바로 뒤에 있는 제 1판 서문을 참고하기 바란다.

새롭게 추가된 제 8 장의 초고에 관하여 건설적인 논평을 하여준 골드펠트 (Stephen Goldfeld), 오쟈카 (Ronald Oaxaca), 퀀트 (Richard Quandt) 에게 감사를 표하고 싶다. 또한, 저서「조사연구자를 위한 통계학적 방법」(Statistical Methods for Research Workers) 으로부터 <표 2> 를 재인용하는 것을 허락하여준 올리버 (Oliver) 와 보이드 (Boyd), 에딘버그 (Edinburgh) 와 저작권자인 故 피셔 (A. Fisher, F.R.S.)경의 도움을 받았다.

해리 H. 켈레지언

윌리스 E. 오츠

제 1 판 서 문

가장 최근의 경제학 발달은 계량경제학 기법의 괄목할만한 발전과 그 기법의 경제학적 분석에의 사용을 포괄하여 왔다. 계량경제학은 과거 선택된 소수 사람만이 보유했지만, 이제는 거의 모든 경제학도들의 교육의 기본적 구성 내용이 되었다.

이러한 계량경제학적 분석의 용도가 광범위하게 늘어났음에도 불구하고, 그중에서 합리적인 결과의 영역을 다루는 대부분의 교재들은 여전히 많은 학생들이 보유하고 있는 數學실력을 능가하여 실제로 어느정도의 수학수준을 전제로 하고 있다. 이 책에서의 목적은 학생들에게는 오로지 적당한 수학수준만을 요구하면서도 광범위한 題材를 개발하는 데에 있다. 더 분명히 하자면, 이 책은 微分이나 行列代數를 사용하지 않는다. 대체적인 수학수준은 고등학교 2 학년 수준이면 족하리라고 본다.*

비록 설명이 기초적인 수학 기법에 의거하고는 있지만, 이 책에서 다루고 있는 題材의 범위는 계량경제학의 전형적인 대학원 과정에 상응한다.

예를 들면, 다루고 있는 주제들은 A.S. 골드버거 (A.S. Goldberger) 의 「계량경제학 이론」(Econometric Theory, Wiley, 1964) 과 J. 존스톤 (J. Johnston) 의 「계량경제학 방법론」(Econometric Methods, 2d ed., McGraw-Hill, 1972) 에서 소개하는 것들과 대체로 같다.

기본적인 결과들은 "代變數" 접근 (" instrumental-variable " approach) 을 사용하여 도출하였다. 이 기법은 통상적으로 보다 널리 쓰이고 있는 最小自乘法 (least-squares method) 에 비하여 두가지 중요한 長點을 갖고 있다. 첫째로, 그 접근방식의 제시에서 微分을 필요로 하지 않는다.

---

* 加算法에 관한 약간의 중요한 전제에 익숙하지 않은 학생들을 위하여 1장의 부록에서 加算法에 관해서 필요한 약간의 결과를 전개하였다.

둘째로 각각의 가정이 추정과정에서 행하는 역할을 학생들이 역력하게 볼 수 있도록 하여 준다. 예를 들면, 正規方程式(normal equations)과 回歸模型(regression model)의 기본 가정간의 대응을 전개하였다. 이것은 매우 유용한 것인데, 그 이유는 뒤에 어떤 가정을 위배할 경우에 생기는 결과를 학생들이 직접 볼 수 있고, 추정과정을 수정하는 데에 채택되는 기법에 관해서 더 잘 이해할 수 있게 하여 준다. 또한 이 책의 모든 내용에 걸쳐서 대변수 접근법을 사용하였다. 이는 自己相關(autocorrelation), 多重共線性(multicollinearity), 異分散性(heteroscedasticity) 체계문제(systems problems) 등을 통합 처리할 수 있게 한다. 왜냐하면, 이 모두는 본질적으로는 같은 방식으로 처리되기 때문이다.

이 교재의 강조점은 "엄격한 직관"(rigorous intuition)이라고 부를 수 있을 것이다. 결과는 단순히 주어지는 것이 아니다. 결과는 가능한한 느슨한 결말을 남기지 않으려는 시도속에서의 직관적인 방법으로 도출된다. 비록 표준적 사례와 결과는 제시되지만, 다음의 사항을 강조하고자 한다. 첫째로, 결과를 얻는 절차이고, 두번째가 그 결과의 실제 추정문제에의 응용이다. 각각의 새로운 기법을 소개한 뒤에 그 용도를 경제학 문헌에서의 수치예와 실제 연구들을 가지고서 설명하기로 한다. 결과적으로 이 교재를 통하여 학습하는 학생은 사물이 어떻게 행해지고 그 이유는 무엇인가라는 두가지 사항 모두에 관하여 상당한 감을 잡게 될 것이다.

각 장의 말미에 있는 문제들 역시 잘 짜여지도록 하였으며, 모든 문제의 해답을 작성하여 교재의 끝부분에 제시하였다. 성실한 학생이라면 이 문제들을 풀었으면 한다. 그 이유는 교재의 결과를 실증적으로 실행하고 적합한 개념들을 조작하고 이해하는 데에 이 연습문제들이 관련이 있기 때문이다. 더구나 모형을 말로써 묘사한 다수의 문제를 제시하고 있고, 학생

들은 이것을 회귀모형으로 정식화하도록 요구되고 있다. 이 문제들은 학생들에게 경제학 모형의 상세한 설명에 포함된 몇가지 어려운 내용을 더 잘 이해시켜 줄 것이다.

이 책은 학부나 석사과정에서의 계량경제학을 1학기에 마치기 위한 교재로 사용되도록 저술되었다. 여기서 석사과정이란 보다 선진적인 계량경제학의 공식적인 양식을 학습하기 위한 數理的 또는 統計學 훈련을 하지 않은 학생을 대상으로한 대학원 계량경제학 과정을 의미한다. 이 교재를 학습하는데 필요한 사전지식은 대체로 1학기짜리 기초 통계학의 전반부 2/3 정도의 내용수준이다. 예를 들면 W.C. 군터 (W.C. Gunther)의 「통계학적 추론」(Concepts of Statistical nference)의 처음 5장에서 다루는 내용인 것이다. 교재 (이에 더하여 통계학의 기본 개념에 대한 간결한 재정리)를 학습하는 데에 필요한 추가적인 내용은 제1장의 부록에 제시되어 있다.

이 교재의 원고가 작성되게된 배경을 간략히 기술하는 것이 이 책의 잠재적인 용도에 대해서 더 잘 알게 하여 줄 것이다. 이 책의 출발점은 프린스턴 (Princeton)대학교의 학부 계량경제학과정을 위한 켈레지언 (Kelejian)의 강의노트를 모은 것이었다. 이 강의노트 자체를 학생들에게 등사판의 양식으로 배부하였고, 프린스톤과 여타 학생들로 부터의 반응이 이 책의 저술을 고무시켜 주었다. 이 책 맨 처음의 원고와 마찬가지로 그 강의노트는 뉴욕 (New York)대학교의 계량경제학 비전공자를 위한 대학원과정에서 사용되었으며, 또한 프린스턴우드로우 윌슨스쿨 (Woodrow Wilson School)에서 공중업무 석사프로그램 (Masters Program in Public Affairs)의 定量技法 (quantitive techniques)과정에서도 이용되었다. 이상의 과정들은 이 책이 적당하다고 느껴진다.

이 책을 위한 특별한 공통연구가 이 책이 갖는 목적을 반영하고 있다.

해리 켈레지언은 자신의 주된 관심분야로서 계량경제학을 전공하였다. 그의 연구와 강의 노력은 이 분야에서 정평이 나있다. 월레스 오츠(Wallace Oates)의 기본 관심사는 財政問題에 있다. 그의 계량경제학과의 친숙은 주로 관심을 실제 경제문제의 定量的 분석에 두고 있는 전문직업가로서 비롯된 것이다. 이러한 관심사의 혼합으로써 계량경제학적으로 건전하고, 계량경제학의 지식이 없는 학생들에게 쉽게 접근할 수 있는 책을 생산하고자 하였다.

원고작성에 도움과 귀중한 조언을 하여준 찰스 비취(Charles Beach), 래리 허쉬(Larry Hersh), 윌리엄 로렌스(William Lawrence), 로버트 프로트닉(Robert Plotnick), 리처드 퀀트(Richard Quandt), V. 선더라라잔(V. Sundararajan), 이라 존(Ira Sohn), 마이어 소콜러(Meir Sokoler)에게 감사한다. 그러나 이들 중의 누구도 이 최종 생산물이 가질 어떠한 결점에 대해서도 책임이 없다. 또한, 「조사연구자를 위한 통계학적 방법」(Statistical Methods for Research Workers)으로 부터 <표 2 >를 재인용하는 것을 허락하여준 올리버(Oliver)와 보이드(Boyd), 에딘버그(Edinburgh)와 저작권자인 故 피셔(A. Fisher, F.R.S.)경에게 빚을 지고 있다. 마지막으로 바쁜 경황속에도 노련하게 타자를 쳐준 베티 카미니스키(Betty Kaminiski)부인에게 큰 신세를 졌음을 밝혀둔다.

해리 H. 켈레지언

월리스 E. 오츠

제 1 장 서 론

어떠한 과학에서도 핵심적인 활동은 사실에 대비한 이론의 체계적인 검증이다. 경제학도 예외가 아니다. 더구나, 경제학에서의 최근 수십 년간 가장 놀라운 발전은 經濟問題의 分析을 위한 統計學的 技法의 개발과 사용에 더 많은 역점이 주어진 것이었다. 경제변수간의 이론적 관계는 유형상으로 수학의 형식으로 표현된다. 그러나 이러한 관계에 실증적인 내용을 부여하기 위하여 경제학자들은 이러한 관계에 대한 假說 (hypotheses) 을 검증하고 실제 크기를 추정하며, 또한 이러한 추정치를 경제적 사안에 대한 量的 예측을 하는 데에 이용하는 통계학적 분석을 점점 더 많이 쓰게 되었다. 이러한 분석방식이 우리가 계량경제학 (econometrics) 이라는 용어로써 의미하는 것이다.

윌리엄 베버리지 (William Beveridge) 경은 1937 년 자신의 런던 경제대학 (London School of Economics) 학장 이임사에서 다음과 같은 이유로 경제학을 하는 동료들을 훈계하였다. 즉, "경제학 (political economy) 이 백년을 내려오면서, 사실들이 이론을 통제하는 것으로서 다루어져온 것이 아니라 설명으로서 취급되었다. ..... 사회에 대한 사실이 이용가능하지 않는 한 사회에 관한 이론은 존재할 수가 없게 된 것이다." 그러나, 베버리지의 연설 이후로 量的 분석방식 (quantitative methods of anaylsis) 의 발전과 경제이론들을 검증할 수 있는 자료 (data) 의 축적에서 엄청난 진전이 있었다. 어느 경제학 잡지를 보더라도 논문의 저자들이 자신들의 주장을 계량경제학적 분석을 가지고서 설명하는 수 많은 논문들을 접하게 된 것이다.

이것은 현재 이루어지고 있는 경제학의 연구작업 (자기 스스로의 실증작업에서와 마찬가지로) 을 이해하고 평가하는 능력을 얻기 위해서는 계량경

제학에 친숙하여야 하는 것이 필수적임을 의미한다. 예를 들면, 총산출량과 총고용 수준에 영향을 미치는 데에 있어서 통화정책과 재정정책의 상대적인 효과에 대한 소위 통화론자 (Monetarist) 와 신케인즈주의자 (Neo-Keynesians) 간의 흥미롭고 중요한 논쟁은 본질적으로 사실, 즉 경제구조와 그 경제구조의 이들 두 유형의 정책에 대한 반응의 문제이다. 그것으로서, 문제는 대체로 실증적 증거에 힘을 빌려서 해결되고, 이 논쟁의 참가자들은 크게 계량경제학적 분석기법에 의존하였다. 이는 곧, 만일 이 논쟁을 쫓아 이미 제출된 증거를 비판적으로 검사하기를 원한다면, 계량경제학의 지식 (그 지식의 경제문제에 대한 합법적인 응용으로부터 나온 결과의 해석은 물론 그 한계와 잘못된 사용을 인지하는 것을 포함한 지식) 을 반드시 가져야만 한다.

이상과 같이 계량경제학은 경제행위에 관한 양적 분석을 다루는 경제학의 한 분야이다. 이로써 계량경제학은 두 가지 중대한 기능을 하여왔다. 첫째로, 이론의 검증과 반박을 위한 기법을 제공하였다. 어느 경제학 이론이나 또는 경제학자의 전문용어 투성이인 모형 (model) 은 그가 특정 유형의 사상 (event) 들을 설명하기 위하여 사용할 수 있는 정의와 가정의 집합이다. 전형적으로 방정식들의 집합으로 표현되는 경제이론은 어떤 경제적 변수가 상호 작용하는 기능 (mechanism) 을 묘사한다. 예를 들면 소비자 선택이론은 소비자가 구매할 특정 재화의 양이 소비자의 선호, 소득, 문제가 되는 재화의 가격, 다른 재화 및 용역의 가격에 의존한다는 것을 시사한다. 그 이론은 재화의 가격이 오르면 구입되는 양은 전형적으로 (typically) 줄어들 것임을 기대하게 해준다.* 거시경제학에서는 총

---

* "전형적" 이라고 해야만 하는 이유가 있다. 즉, 正의 소득효과가 가격상승으로 인한 負의 대체효과를 능가하는 경우에는 결과적으로 소비자가 가격이 오른 재화의 구입을 실제로 늘릴 것임을 알수 있기 때문이다.

투자 수준이 이자율에 의존한다는 것을 의미하는 이론들이 있다. 더 자세히 보면, 이들 이론들은 이자율이 더 높아지면 실질자본형성(투자)에 관한 지출이 위축될 것임을 지적하고 있다.

이들 이론의 유용성을 평가하기 위해서는 경제적 事象을 예측하는 데에 이들 이론의 신뢰성을 측정해야만 한다-좀 전에 인용하였던 보기에서처럼, 경제이론들은 일반적으로 事象들의 의미있는 인과적 계기(causal sequence)를 뚜렷하게 지정함으로써 검증가능한 형식으로 되어 있다. 즉, 이렇다면 그렇다는 것이다(예컨데, 이자율이 상승하면, 투자지출은 줄어 든다). 이것은 빈번히 한 변수가 다른 변수의 함수라고 하고 그 관계의 일반적 특성을 명기하는 것에 의하여 수학적 용어로 표현될 것이다. 예를 들면, I를 투자지출수준, R을 이자율, 그리고 a와 b를 양(plus)의 값을 갖는 상수라고 할 때, $I = a - bR$이라고 할 수 있을 것이다. 이러한 형식으로 이론은 그 예측의 정확성을 실증적으로 검증할 수 있게 된다.

이러한 방법으로 경제이론을 검증하는 것은 거의 드문 단순한 문제라고 말하여 두어야 할 것이다. 앞에서 설명한 종류의 인과적인 진술은 다른 관련 요인들은 변함이 없다는 가정에 전형적으로 기초한 것이다. 예를 들면 이자율의 상승이 투자수준을 낮춘다는 명제는(여러 검증에서) 총수요가 일정하다는 가정에 기초한 것이다. 만일 이자율이 상승한 동시에 수요가 증가하면, (증대하는 수요에 맞추어)투자의 증가가 이자율의 상승과 동반됨을 발견할 수 있다. 이것이 반드시 이론의 반박을 의미하는 것은 아니다. 왜냐하면 분명히 증가된 이자율의 負의 효과가 증대된 총수요의 正의 영향을 상쇄하는 것보다 더 클 수 있기 때문이다. 이러한 측면에서 경제학자들이 직면하는 문제는 자신들의 엄청난 대부분의 자료가 통제된 실험실의 실험으로부터 나오는 것이 아니라 매일매일의 경험에서 유래하는 것이다. 그리고 현실 세계에서는 "다른 것이 변하지 않는다"는 경우가 거의

없다. 이러한 까닭으로 계량경제학자들은 문제가 되는 변수에 대한 다른 영향을 인위적으로 불변이게 하는 효과적인 통계학 기법을 고안하여야만 하였다. 이러한 방식으로 계량경제학자들은 한 변수의 다른 변수에 대한 효과를 측정할 수가 있다. 이 책에서 분명하게 되듯이, 이 문제는 量的 기법을 경제학 자체의 특성에 맞추어 주는 것을 돕는 것이다.

이상과 같은 이유로 계량경제학은 경제이론의 검증 또는 반박에 기본적인 중요성을 갖는 것이다. 그 두번째 기본적인 기능은 변수들간의 관계가 갖는 중요도를 양적으로 추정하여 제공하는 것이다. 어느 특정 이론은 가격의 상승이 수요량의 감소를 가져오거나, 조세수준의 인하가 총지출과 산출량을 촉진한다고 시사할 수가 있다. 이러한 관계들이 갖는 일반적인 성격은 매우 중요하기는 하지만, 실제 의사결정에는 자주 만족스럽지가 못하다. 어느 사업가는 가격을 10 % 올렸을 경우 판매 단위가 얼마나 많이 줄어들 것인가를, 그 결과로 이러한 선택이 자신의 이윤에 미칠 충격을 알고자 한다. 마찬가지로 경제자문가는 일정한 조세삭감이 얼마나 많은 총지출의 증가를 가져올 것인가를 추정해야만 한다. 만일 조세삭감이 너무 작다면, 과잉실업과 생산설비가 충분히 활용되지 않는 상태가 지속될 것이며, 반면에 조세가 너무 많이 준다면, 인플레이션이 발생할 것이다. 이러한 이유로, 경제학에서의 양적 기법은 경제이론이 전형적으로 시사하는 보다 일반적인 명제에 관한 평가를 제공하는 것에 더하여 크기를 추정하여 나갈 수 있어야만 한다.

특수하고, 더구나 상당히 중요한 사례를 살펴보는 것이 계량경제학적 문제의 일반적인 본성을 소개하는 데에 도움이 된다. 위에서 시사하였듯이 우리가 특정한 크기의 개인소득세의 삭감이 제안되었을 때, 그 삭감으로 인한 소비자 지출의 증가를 추정하는 임무를 맡은 경제자문가라고 가정하자. 우리가 맡은 과제를 풀기 위한 출발점으로서 이론적인 틀을 케인즈주의자

의 소비함수를 참고하기로 선택하였다고 하자. 이 소비함수는 소비지출이 가처분소득 수준에 의존한다고 본다. 최초의 추정이므로 단순하게 하기 위하여 관계가 선형이라고 가정하자. 여기서 C는 소비지출, Y_d는 가처분소득, 그리고 a와 b는 모수(즉, 상수)이다. 모수 b의 값이 매우 중요

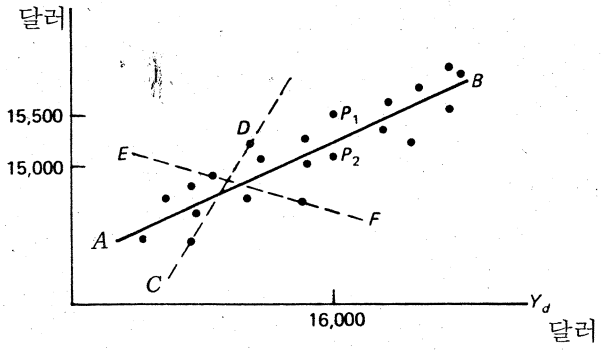
$$C = a + bY_d \tag{1.1}$$

하다는 것은 분명하다. 조세삭감은 가처분소득을 증가시킬 것이고, 이는 다시 소비지출을 자극할 것이다. 그리고 b는 '한계소비성향'(marginal propensity to consume : MPC)으로 잘 알려진 것으로서, 그 값은 가처분소득이 추가적으로 늘어났을 때 개인들이 소비에 지출하려는 가처분소득추가분의 분수를 가리킨다. 지출수준에 대한 조세삭감의 영향을 평가하려면 b의 추정치가 필요함이 명백하다.

거시경제이론은 대강의 지침을 제공한다. 예를 들면, 그 이론은 한계소비성향의 값 b가 반드시 0과 1사이의 어딘가에는 있음을 시사한다. 가처분소득의 추가분은 아마도 소비지출의 증가를 가져올 것이지만, 그 일부는 또한 저축됨으로써 소비는 가처분소득의 증가분보다는 어느정도 작을 것이다. 그러나, 소비지출에 미치는 조세삭감의 영향은 만일 b가 0.9(즉, 소비자가 추가적인 소득의 90퍼센트를 소비에 지출하고 단지 10퍼센트만을 저축할 경우)라면 b가 0.5일 때보다 크기 때문에, 이것을 훨씬 더 잘 추정하는 것이 필요로 한다.

따라서, 우리의 최초 과제를 b의 추정된 값을 측정하는 것으로 한다. 설득력있는 진행방법은 가처분소득수준이 다른 개인들의 소비와 저축행위를 검사하는 것이다. 그러한 정보로부터 소비가 어떻게 가처분소득수준에 따라 변하는가를 추정할 수가 있다. 일군의 가구에 관한 가구예산 조사를 통하여 소비지출과 소득에 대한 정보를 얻을 수 있다고 가정하자. 정보를

조립하고 도표를 만들어서 그리면 <그림 1.1>이 된다. <그림 1.1>과 같은 그림은 다음의 장들에 반복해서 사용할 것인데, 이를 산포도 (scattered diagram)라고 한다. 각 점들은 관찰된 두변수 값의 위치를 표시한다. <그림 1.1>에서 예를 들면, P_1은 가처분소득 16,000달러를 갖고, 15,500달러를 소비지출하는 예산 조사상의 한 가구를 가리키는 점이다.



〈그림 1.1〉 가설적인 예산연구로부터의 정보

방정식 (1.1)로 묘사된 소비함수를 가지고서 <그림 1.1>로 주어진 정보를 고려하여 보자. 첫째로, 소비함수를 수리적으로 명시한 것이 엄격하였음을 주지할 수 있다. 방정식 (1.1)은 각각의 가처분소득에 대응하는 특정한 소비지출수준이 있을 것임을 말하여 준다. 그러나 인간의 행위는 그렇게 정확하지가 않다. 사실상 <그림 1.1>의 결과는 같은 소득수준의 가구가 거의 다 어느정도 소비에 지출하는 액수가 다를 것임을 시사한다. 예를 들면 P_1 점과 P_2 점은 둘다 가처분소득이 16,000달러인 가구를 가리키지만 P_1의 가구는 15,500달러를 지출하는 반면에 P_2의 가구는 단지 15,000달러만을 지출하고 1,000달러를 저축한다.

불일치하는 정보로 보이는 덩어리로부터 어떻게 a와 b의 추정치를 측

정할 수 있는가? 초보적인 수학으로부터 두점이 한 직선을 결정한다는 것은 알고 있다. 그러므로, 방정식 (1.1) 에서의 a와 b값을 결정하는 데에 필수적인 것으로 보이는 것은 두 개의 관찰치이다. 얼핏보면, 문제는 너무 많은 정보를 가짐으로써 비롯된 것처럼 보인다. 적합한 자료를 무시하는 것은 비합리적인 것처럼 (실제로 그렇다) 보인다.

이들 a와 b를 구하였던 특정한 두 점에 의존하여 이들 모수에 대한 다른 값을 분명히 얻을 것이다. <그림 1.1>에서 예를 들면 선 CD, EF 또는 다른 어떠한 선들, 우리가 선을 결정하려고 뽑은 특정한 두 점에 의거하여 얻을 수 있었다.

그러나, <그림 1.1>의 흩어져 있는 점들을 면밀히 검사하면 C와 Y_d 사이에 어떤 종류의 관계가 있음을 시사할 것으로 보인다. 즉, 가처분소득이 증가함에 따라 평균적으로 소비지출의 수준도 증가하는 것으로 나타나는 것이다. C와 Y_d의 관계는 정확한 것과는 거리가 멀지만, 그럼에도 불구하고 두 변수 사이에는 어떤 "전형적인" 관계가 나타난다. 단순히 검사에 의해서 또는 보다 공식적인 기법으로써 AB와 같은 "전형적인" 예산의 행위를 나타내는 선을 흩어져 있는 점들에 맞출 수가 있다. 이 선은 a와 b의 추정치를 제공하여 주는데, 이들 각각은 수직선의 절편과 기울기를 가리킨다.

이상이 계량경제학이 관심을 갖는 종류의 문제이다. 그리고 사실상 이후 장들의 촛점은 두 변수 (더 뒤에는 여러 변수) 간의 이러한 "전형적인" 관계를 추정하기 위해서 체계적이고 사리에 맞는 기법을 전개하는 데에 있다. 이러한 과정에서 중요한 의미를 갖는 해답이 필요하여지는 관계에 관한 수많은 의문이 있다는 것을 발견할 것이다. 앞의 가상적인 사례에서 독자들 자신이 제안된 조세감축에 관한 경제자문가로서의 역할을 하고 있다고 생각할 것을 요청하였는데, 총지출에 대한 조세삭감의 효과의 신뢰성있

는 예측에 앞서 b의 추정에 대하여 해결하여야 할 몇몇 문제가 더 남아 있음이 명백하다. 서론을 완결하기 위해서는 이들 문제의 목록과 간결한 논의를 제공하는 것이 도움을 줄 것이다. 왜냐하면 이들 문제의 해결이 이 책이 다루고 있는 모든 것이기 때문이다.

1. 가설검정 (hypothesis testing)

두 변수 사이의 인과관계를 의미하는 이론을 가지고 있다고 가정하자. 어떻게 이 두 변수 모두와 또 다른 적합한 변수들의 자료를 가지고, 이들 변수간에 실제로 존재하는 관계에 어느정도의 신뢰도를 설정할 수 있겠는가? 예를 들면, <그림 1.1>의 산포도에서 보이는 C와 Y_d의 명백한 "전형적인"관계가 그럴싸한 것에 지나지 않아서 이러한 특정한 표본에서 우연히 나타난 믿을 수 없는 결과인지를 알 수가 있는가?

2. 모수의 추정치 (estimates of parameters)

두 변수간의 어떤 관계가 있다면, 어떻게 이용가능한 자료를 가장 잘 이용하여 그 크기를 정확하게 얻어 낼 수 있겠는가? <그림 1.1>에 깔려 있는 정보에 기초한다면, a와 b의 추정치를 얻어 내는 가장 효과적인 방법은 무엇인가? 이에 덧붙여 평균으로 부터 경제행위가 얼마만큼이나 이탈하는가를 알아내어, 결과적으로 a와 b의 추정치가 얼마나 유용한가에 관한 평가를 얻을 수 있기 바란다.

3. 예측을 위한 추정치의 사용

무슨 조건 또는 제한하에서 이들 추정치를 예측하는 데에 사용할 수 있는가? 다시 한번 <그림 1.1>를 참조하면, 소비지출에 미치는 조세삭감의 효과를 평가하기 위한 가구 예산 조사로부터 추정한 b값을 사용하기 위

해서는 무슨 가정을 해야만 하는가? 또는 이와 관련된 문제로서, 이 정보에 기초해서 어떤 특정화된 신뢰도 (degree of confidence)를 가지고 Y_d의 수준이 주어졌을 때 C가 어떻게 될 것인지를 예측할 수 있는가?

4. 함수형식

관계를 나타내는 적절한 함수형식은 무엇인가? 우리는 단순하게 하기 위하여 C와 Y_d를 단순 선형관계로 가정했지만, 이것은 분명 참된 관계가 아니다. 아마도 MPC, 즉 식 (1.1)의 b는 소득이 증가함에 따라 감소할 것이다. 참된 관계는 $C = a + Y^{\frac{1}{2}}$일지도 모른다. 추정하려는 변수의 특수한 함수관계를 선별하기 위해서는 어떻게 이론적인 결과와 이용가능한 자료를 사용할 수 있는가?

5. 자료의 불완전성

이용가능한 자료의 불완전성 (예컨데 측정상의 오차)이 결과에 무슨 영향을 미치는가? 이 불완전성이 추정치를 무효로 하는가?

6. 피드백 (feedback) 관계

변수 X가 Y에 미치는 효과를 측정하려 하는데, X가 Y에 영향을 주는 것 뿐만 아니라 Y 또한 X에 영향을 준다고 가정하자. 이 경우에 모수의 추정치가 X의 Y에 대한 영향을 반영하는 것인지 Y의 X에 대한 영향을 반영하는 것인지를 식별하기가 어렵다. 이러한 종류의 피드백 관계는 경제학에서 자주 발생한다. 가격이 재화의 수요량을 결정하지만 수요 역시 가격에 영향을 미친다. 경제의 총지출수준은 총산출량과 소득에 강력한 효과를 가지나 다시 총산출량과 소득은 지출수준에 영향을 미친 것 등등이다. 이것이 이른바 체계문제 (systems problem)라는 것이다. 경제문제

는 전형적으로 한 경제체제의 기능을 특징짓는 상호의존성을 반영하기 때문에, 자주 체계 유형의 문제를 가진다. 그러나 뒤에서 보겠지만 이 문제는 계량경제학자에게 심각한 문제를 일으킨다. 계량경제학자는 이러한 상호의존성을 양적으로 파헤치는 시도를 해야만 한다.

이상이 대강의 계량경제학 문제이며, 뒤에서는 이들 문제를 다루는 기법을 전개할 것이다.

부록A. 加算法 사용에 관한 명제

서문에서 지적하였듯이 이 책은 고등 수학 또는 통계이론을 사용하지 않고 있다. 그러나, 일부 독자는 익숙하지 않거나 기억이 어렴풋한 더하기 부호를 가지고 계산하는 것과 관련된 약간의 명제가 있다.

특수한 명제들은 꽤 광범위 하게 사용할 것이기 때문에, 여기서 미리 다루어서 뒤에서 나왔을 때 여러분이 익숙하다면, 분석을 촉진할 수 있을 것이다.

이 책의 처음부터 끝까지 加算을 지칭하는 시그마(sigma : $\sum$)자를 사용할 것이다. 예컨데 Q_1를 1년에 생산된 특정 상품의 양이라고 하거나, 일반화시켜서 Q_t를 t년에 생산된 양이라고 하자. 그러면 1, 2, 3년에 걸친 총생산량은 $(Q_1+Q_2+Q_3)$로 표현될 수 있다. 이를 다음과 같이 $\sum_{t=1}^{3}$으로 표시하면, 표현을 간략하게 할 수 있다.

$$\sum_{t=1}^{3} Q_t = Q_1 + Q_2 + Q_3 \qquad (1A.1)$$

보다 일반적으로 $\sum_{t=1}^{n} Q_t$라는 표현이 변수 Q의 n번째까지의 합을 정의한다. 이러한 표기를 단순하게 확장하면 다음과 같이 중간의 합을 표현할

수 있다.

$$\sum_{t=3}^{7} Q_t = Q_3 + Q_4 + Q_5 + Q_6 + Q_7 \tag{1A.2}$$

계속 하기전에 아래가 맞는지를 스스로 확인하여야 한다.

$$\sum_{t=1}^{16} Q_t - \sum_{t=3}^{17} Q_t = Q_1 + Q_2 - Q_{17} \tag{1A.3}$$

이제 이 책에서 반복적으로 사용하게 될 약간의 전제를 전개하도록 하겠다.

명제Ⅰ. 만일 c가 상수 (예컨데 $c=5$) 이면,

$$\sum_{t=1}^{n} cX_t = c\sum_{t=1}^{n} X_t$$

이것을 확인하기 위해서는 정의상 $\sum_{t=1}^{n} cX_t$ 는 n 번째까지의 X값을 각각 c를 곱해서 더한 것임을 주지하면 된다. 그러므로,

$$\begin{aligned}\sum_{t=1}^{n} cX_t &= cX_1 + cX_2 + \cdots + cX_n = c(X_1 + X_2 + \cdots + X_n) \\ &= c\sum_{t=1}^{n} X_t\end{aligned} \tag{1A.4}$$

명제Ⅱ. X와 Y가 두 변수라면,

$$\sum_{t=1}^{n} (X_t + Y_t) = \sum_{t=1}^{n} X_t + \sum_{t=1}^{n} Y_t$$

명제Ⅱ는 X와 Y의 합은 X의 합과 Y의 합을 더한 것에 지나지 않는다는 것을 말한다. 이는 다음과 같은 이유로 참이다.

$$\sum_{t=1}^{n}(X_t + Y_t) = (X_1 + Y_1) + (X_2 + Y_2) + \cdots + (X_n + Y_n)$$
$$= (X_1 + X_2 + \cdots + X_n) + (Y_1 + Y_2 + \cdots + Y_n) \qquad (1A.5)$$
$$= \sum_{t=1}^{n} X_t + \sum_{t=1}^{n} Y_t$$

명제Ⅰ과 Ⅱ를 동시에 일반화하면,

$$\sum_{t=1}^{n}(aX_t + bY_t + cZ_t) = a\sum_{t=1}^{n} X_t + b\sum_{t=1}^{n} Y_t + c\sum_{t=1}^{n} Z_t$$

이다. 여기서 a, b와 c는 상수이고 X, Y와 Z는 변수이다.

명제Ⅲ. $\overline{X}$가 변수 X의 n번째까지의 단순평균이고 그래서 $\overline{X} = (\sum_{t=1}^{n} X_t)$ / n이면,

$$\sum_{t=1}^{n}(X_t - \overline{X}) = 0$$

이를 증명하기 위해서는 맨 먼저 다음을 주지하기 바란다.

$$\sum_{t=1}^{n}(X_t - \overline{X}) = \sum_{t=1}^{n} X_t - \sum_{t=1}^{n} \overline{X} \qquad (1A.6)$$

우측의 첫번째 항에 n을 곱하고 다시 n을 나누어 주면,

$$\frac{n\sum_{t=1}^{n} X_t}{n} = n\overline{X} \qquad (1A.7)$$

다음, 두번째 항은,

$$\sum_{t=1}^{n} \overline{X} = \overline{X} + \overline{X} + \cdots + \overline{X} = n\overline{X} \qquad (1A.8)$$

(1A.6)식에 (1A.7)과 (1A.8)를 대입하면 다음의 결과를 얻는다.

$$\sum_{t=1}^{n}(X_t - \bar{X}) = \sum_{t=1}^{n} X_t - \sum_{t=1}^{n} \bar{X} = n\bar{X} - n\bar{X} = 0$$

이러한 논의로부터, 만일 K가 어떤 상수일 경우에는 일반적으로 더 명확하게 다음과 같이 된다.

$$\sum_{t=1}^{n} K = nK \tag{1A.9}$$

명제Ⅳ. $\bar{X}$와 $\bar{Y}$가 X, Y 두 변수의 n번째까지의 단순 평균이면,

$$\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^{n}(X_t - \bar{X})Y_t$$

이것을 확인하려면 먼저 다음에 주의하면 된다.

$$\begin{aligned}\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y}) &= \sum_{t=1}^{n}[(X_t - \bar{X})Y_t - (X_t - \bar{X})\bar{Y}] \\ &= \sum_{t=1}^{n}(X_t - \bar{X})Y_t - \sum_{t=1}^{n}(X_t - \bar{X})\bar{Y}\end{aligned} \tag{1A.10}$$

이제 우측의 두번째 항이 0임을 보일 수 있다.

$$\sum_{t=1}^{n}(X_t - \bar{X})\bar{Y} = \bar{Y}\sum_{t=1}^{n}(X_t - \bar{X}) = \bar{Y}\cdot 0 = 0 \tag{1A.11}$$

식(1A.11)은 명제Ⅰ과 Ⅲ에서 연유한 것이다($\bar{Y}$가 상수임에 유의). 이것은 명제Ⅳ를 증명한다.

이상을 한단계 더 나아가면,

$$\sum_{t=1}^{n}(X_t - \bar{X}_t)Y_t = \sum_{t=1}^{n} X_tY_t - \sum_{t=1}^{n}\bar{X}Y_t \tag{1A.12}$$

우측의 두번째 항에 n을 곱하고 동시에 n을 나누어 주면,

$$\sum_{t=1}^{n}(X_t - \bar{X})Y_t = \sum_{t=1}^{n} X_t Y_t - n\bar{X}\bar{Y} \qquad (1A.13)$$

독자들은 명제Ⅳ에 관한 두 계 (Corollary)를 증명하기 바란다.

$$\sum_{t=1}^{n}(X_t - \bar{X})^2 = \sum_{t=1}^{n}(X_t - \bar{X})X_t$$

$$\sum_{t=1}^{n}(X_t - \bar{X})^2 = \sum_{t=1}^{n} X_t^2 - n\bar{X}^2$$

〔 힌트 : $\sum_{t=1}^{n}(X_t - \overline{X})^2$ 를 $\sum_{t=1}^{n}(X_t - \overline{X})(X_t - \overline{X})$ 로 표시하시오. 〕

부록B. 통계학 개념의 복습

편의를 돕기 위하여 이 책에서 반복해서 사용하는 기초적인 통계학 개념을 간결하게 다시 살펴 보도록 하겠다. 이 부록은 통계학 기초강의를 대신하려고 만들어진 것이 아님을 강조하여 둔다. 이 부록의 목적은 통계학 개념의 일부를 뽑아 간략하게 재검토하는 데에 있다.

가. 確率變數 (random variable)

확률변수는 결과가 우연하게 나오는 실험결과에 의하여 그 값이 결정되는 변수를 일컫는다. 다른 말로 하면, 확률변수의 가능한 값은 그것이 출현할 특정한 확률과 연관된다. 예를 들면, 확률변수의 값은 동전 던지기에 의존하는 것으로도 볼 수 있다. 던지기 (실험)의 결과는 "앞면"이거나 "뒷면"이다. 그러면, 확률변수는 앞면이 나타날 경우는 1이고 뒷면이 나타나면 0인 값이 되는 변수 Y로 정의될 수 있다. Y의 가능한 값이 0과 1이라는 말은 자주 Y는 $y=0, 1$ 값을 취하는 확률변수라고 일컬

어진다. 여기서 Y와 y가 혼동되면 안된다. Y는 그 값이 실험 결과에 종속되는 확률변수이다. y는 단지 Y가 취할 지도 모를 특정한 값(수)을 나타낸다.

또 다른 확률변수로서 W를 들어 보자. 여기서 W는 주어진 집단으로부터 임의적으로 선발된 사람의 몸무게이다. 이 경우에 W의 가능한 값은 w일 것이며, 사람들이 성인으로 구성되어 있다면 아마도 50 파운드$\leq w \leq$1000 파운드이다. 비록 Y와 W가 확률변수이지만, 둘 사이에는 중요한 차이점이 있다. W는 연속적인 값의 범위내에서 어떠한 값도 취할 수가 있다. 반면에 Y는 연속적인 구간에 걸쳐서는 그 값이 정의되지 않는다. Y와 같이 값의 연속체(continum)를 취하지 않는 변수를 이산형(discrete) 확률변수라고 한다.

이 부록에서 다루는 제재는 이산형 확률변수 식인 것이다. 그 이유는 연속형 확률변수의 분석은 미분의 사용을 요구하기 때문이다. 그러나 우리의 목적을 추구하기 위하여 뒤에서는 오로지 이산형 변수 식에서만 필요한 개념을 전개할 수가 있다.*

나. 확률(또는 밀도)함수

확률변수와 연관된 것은 확률변수가 각각의 가능한 값을 취하게 될 확률을 부여하는 확률함수(때로는 확률밀도함수라고 불린다)이다. 확률함수는 대체로 방정식이나 표의 형식으로 표현된다. 예를 들면, 위의 동전 던지기에서 변수 Y의 가능한 값은 $y=0$, 1이었으며, 이에 관련된 확률은(동전은 균형이 잡혀있고 "공정"하게 던져진다면) 1/2과 1/2이다. 그러

---

* 수학적 배경 지식을 갖고 있는 독자는 가산 기호($\sum$)를 적분 기호($\int$)로 모두 대체하면, 이하에서 다루고 있는 소재를 연속형 확률변수의 내용으로 변환시킬 수 있다.

므로, Y의 확률함수는 $f(y)=\frac{1}{2}$, $y=0$, 1로 쓸수 있을 것이다. 만일 확률변수 Y가 확률함수 $f(y)$를 가진다면, $f(1)=\frac{1}{2}$이라는 말은 "$Y=1$인 확률은 $\frac{1}{2}$이다."

또 다른 예가 도움이 될 것이다. Z를 주사위를 굴렸을 때 보이는 수라고 하자. 그러면 Z값의 범위는 $z=1, 2, 3, 4, 5, 6$이고, 그 확률함수는 $g(z)=\frac{1}{6}$, $z=1, \cdots\cdots, 6$이다. 이러한 정보는 대안적으로 다음과 같은 표의 형식으로 주어질 수 있다.

z	1	2	3	4	5	6
$g(z)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(1B.1)

또한

$$g(1) + g(2) + \cdots + g(6) = \tfrac{1}{6} + \tfrac{1}{6} + \cdots + \tfrac{1}{6} = 1 \qquad \text{(1B.2)}$$

Z는 정수 1, 2, ……, 6중의 하나임이 틀림없기 때문에, 관련된 확률의 합은 1과 같다. 간단히 줄여 말한다면, (1B.2)는 z가 정수 1, 2, ……, 6중의 하나라는 확률이 1과 같음을 말하는 것이다.

이상이 일반적인 결과이다. 어느 확률변수의 모든 가능한 값에 상응하는 확률의 합은 1과 같다. 확률은 0과 같거나 크다고 정의하기 때문에, 또 다른 일반적인 결과의 확률함수도 그렇게 정의되어야만 한다는 것이다. 예를들면, 위의 주사위의 예에서 비록 말은 하지 않았지만 $Z=\sqrt{363}$인 확률은 0이기 때문에 $z=\sqrt{363}$에 상응하는 확률함수의 값은 0일 것이다.

다. 독립과 종속

몇몇 확률변수간의 관계를 다룰 경우에 문제가 종종 발생한다. 예를 들며는 임의적으로 주사위를 두번 던졌다고 가정해보자. 첫번째와 두번째 던

졌을 때에 출현한 값을 각각 Z_1과 Z_2라고 하자. 이 경우에는 Z_2의 값이 Z_1의 값의 영향을 받았다거나 그 역이라고 예상하지 않는다. 예를 들어, 주사위가 온전하다면 두번째 던졌을 때 3을 얻을, $Z_2=3$의 확률은 첫번째 던진 결과와 무관하게 $\frac{1}{6}$일 것이다. Z_1과 Z_2와 같은 방식으로 확률이 무관한 두 변수는 서로간에 독립되었다고 말한다. 따라서, Z_1과 Z_2는 독립적인 확률변수라고 한다. 만일 두 변수가 독립적이지 않다면, 종속적이라고 말한다. 예를 들면, 평범한 한 벌의 카드에서 카드 한 장을 집어 냈다고 하자. 그림카드(picture card)가 뽑혔으면 $P=1$이라고 하고, 그렇지 않은 경우에는 $P=0$이라고 한다. 추가적으로 킹(king)이 뽑혔을 경우 $K=1$이라 하고 그외에는 $K=0$이라고 한다. K와 P는 종속적인 확률변수이다. 예를 들어, $P=1$이면 $K=1$일 확률은 $\frac{4}{12}$일 것이며, $P=0$이면 0이다. P에 관한 정보가 주어져 있지 않다면, $K=1$일 확률은 $\frac{4}{52}$이다. 간략히 말해서, 만일 둘중의 어느 한 변수에 관한 정보가 다른 변수에 관련된 확률을 변경시킨다면, 두 변수는 종속적이다.

이상의 정의를 일반화하는 것은 수월하다. 만일 X_1이 어떠한 값을 취할 확률이 X_2, ……, X_n이 취하는 특정한 값에 전혀 영향을 받지 않는다면, 확률변수 X_1은 확률변수 X_2, ……, X_n과 독립적이다. 만일 X_1의 확률에 어떠한 영향이 있다면, X_1과 변수 X_2, ……, X_n의 적어도 하나는 종속적이다.

라. 期待(expectations)

확률변수 X의 수리적인 기대(종종 기대값으로 불려진다)는 그 가능한 값이 $x=x_1, x_2, \cdots\cdots, x_n$이고 그 확률함수가 $f(x)$라고 하면, $E(X)$로 표기되며 다음과 같이 정의된다.

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) \tag{1B.3}$$

(1B.3) 에서 X의 기대값이 X의 가능한 값의 가중평균으로 정의되는데, 그 가중치는 연관된 확률이다. (1B.3)의 기호 E는 기대값 연산자(expected value operator)로 불린다. 예를 들어 앞의 주사위 보기에서 변수 Z의 기대값은

$$E(Z) = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + \cdots + 6(\tfrac{1}{6}) = \tfrac{21}{6} = 3\tfrac{1}{2} \tag{1B.4}$$

확률변수의 기대값은 종종 확률변수의 평균(mean)이라고 불리며, 고려되는 확률변수를 가리키는 하첨자를 함께 붙여서 μ자로 표기한다. 예를 들면, $E(Z)=\mu_z$, $E(X)=\mu_x$이다. 어느 정도 직관적으로 변수의 평균은 변수의 중심적인 경향 또는 위치를 측정하는 것이다. 만일 실험이 수없이 많이 시행되면, 그 평균은 모든 실험에 걸쳐서 변수의 평균이 될 것으로 기대된다.

확률변수 X의 분산(variance)은, X의 가능한 값이 $x = x_1, x_2, \cdots\cdots, x_n$이고 X의 확률함수가 $f(x)$일 때, 통상적으로 σ_x^2로 쓰여지고 다음과 같이 정의된다.

$$\begin{aligned} \sigma_X^2 &= E(X - \mu_X)^2 \\ &= (x_1 - \mu_X)^2 f(x_1) + (x_2 - \mu_X)^2 f(x_2) + \cdots + (x_n - \mu_X)^2 f(x_n) \end{aligned} \tag{1B.5}$$

여기서 $E(X)=\mu_x$이다. (1B.5)로부터 분산은 평균에서 변수가 떨어진 부분을 자승한 것의 기대값임을 알 수가 있다. 어떤 의미에서는 분산이 확률변수의 평균에 대하여 그 확률변수의 퍼짐(dispersion)을 측정하는 것이다. 즉, 평균적으로 확률변수의 값이 그것의 평균으로부터 얼마나 멀리

떨어져 있겠는가를 가리키는 것이다. 관련된 계산의 보기로서, 앞의 Z가 갖는 분산은

$$\begin{aligned}\sigma_Z^2 &= E(Z-3.5)^2 \\ &= (1-3.5)^2\left(\tfrac{1}{6}\right) + (2-3.5)^2\left(\tfrac{1}{6}\right) + \cdots + (6-3.5)^2\left(\tfrac{1}{6}\right) \\ &= \frac{17.50}{6} = 2\tfrac{11}{12}\end{aligned} \tag{1B.6}$$

마. 기대에 관한 약간의 명제

이 절에서는 이 책에서 자주 쓰이는 기대의 속성을 간결하게 요약하려 한다. 먼저, 상수(c)의 기대값에 관한 자명한 것부터 시작하도록 하자.

$$E(c) = c \tag{1B.7}$$

만일 c가 5라는 값을 갖는 상수라고 하면, (1B.7)은 단지 c의 기대값이 5라고 말하는 것이다. c는 5이외의 것이 될 수 없기 때문에, 5라는 값은 1의 확률을 가져서 $E(5)=5\cdot f(5)=5(1)=5$이다.

새로운 확률변수를 Y라고 하고, Y가 또 다른 확률변수에 상수를 곱한 것과 같다고 하는 예를 고려하기로 하자. 예를 들어 주사위를 던지는 보기로써, $Y=15Z$이라고 한다. 만일 주사위 숫자가 4로 나왔다면 $Y=15\cdot 4=60$이다. Y의 기대값은

$$\begin{aligned}E(Y) = E(15Z) &= 15(1)\left(\tfrac{1}{6}\right) + 15(2)\left(\tfrac{1}{6}\right) + \cdots + 15(6)\left(\tfrac{1}{6}\right) \\ &= 15(3.5) = 52.5\end{aligned}$$

따라서 $Y=15Z$의 기대값은 단지 Z의 기대값의 15배임을 발견하게 된다. 이것이 일반적인 결과이다. b가 상수이고 X가 확률변수이면,

$$E(bX) = bE(X) = b\mu_X \tag{1B.8}$$

이상의 두 명제를 확장하여, X_1, X_2, ……; X_n 이 각각의 평균을 μ_1, μ_2, ……; μ_n으로 갖는 n개의 확률변수라고 하자. 변수 Y를 a_0, a_1, ……; a_n 이 상수라고 했을 때 다음과 같이 정의한다.

$$Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n \tag{1B.9}$$

이것은 변수 Y가 X의 선형결합 (linear combination)으로 정의된 것이다. 이제 증명없이 다음과 같이 정리하기로 하자.

$$\begin{aligned} E(Y) &= E(a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &= E(a_0) + E(a_1X_1) + E(a_2X_2) + \cdots + E(a_nX_n) \\ &= a_0 + a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\ &= a_0 + a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \end{aligned} \tag{1B.10}$$

따라서, Y가 어느 확률변수의 집합으로 이루어진 선형결합이면, Y의 기대값은 함수형식인 기대값들의 합에 불과하다.

Z가 주사위를 던졌을 때 나오는 값이고, Z^2이 또 다른 확률변수 Q와 같다고 정의하기로 하자. 그러면 Q의 값은 주사위에서 보이는 수의 자승과 같다. Q의 기대값은,

$$E(Q) = E(Z^2) = 1(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + \cdots + 36(\tfrac{1}{6}) = 15\tfrac{1}{6} \tag{1B.11}$$

$E(Z) = 3.5$ 이었던 것을 기억하면,

$$[E(Z)]^2 = (3.5)^2 = 12.25 \neq E(Z^2) = 15\tfrac{1}{6}$$

따라서 $[E(Z)]^2 \neq E(Z^2)$ 이다. 말로 하면, Z의 기대값의 자승은 Z^2의 기

대값과 같지 않다.

더 일반적인 결과의 설명은 아래와 같다. 만일 $Y = g(X)$이고 $g(X)$가 확률변수 X의 비선형 (nonlinear) 함수이면, 일반적으로

$$E(Y) = E[g(X)] \neq g[E(X)] \tag{1B.12}$$

예를 들면, 좀전에 보았듯이 $E(X^2) \neq [E(X)]^2$이고 또 다른 예는 $E(e^X) \neq e^{E(X)}$이다. 기대에 관한 한가지 진전된 결과가 필요하게 될 것이다. Y를 이제 확률변수 집합의 원소를 모두 곱한 것으로 하자.

$$Y = (X_1 X_2 \cdots X_n) \tag{1B.13}$$

이 경우에 변수 $X_1, X_2, \cdots\cdots, X_n$이 상호 독립적이 아니면, 일반적으로

$$E(Y) = E(X_1 X_2 \cdots X_n) \neq E(X_1)E(X_2)\cdots E(X_n) \tag{1B.14}$$

이미 이러한 명제의 한 예를 본 적이 있다. 즉,

$$E(Z^2) = E(Z \cdot Z) \neq E(Z)E(Z) = [E(Z)]^2$$

그러나 X가 독립적이면, X의 원소를 모두 곱한 것은 그것들의 기대값을 모두 곱한 것과 같다. 모든 X_i가 독립적일 때,

$$E(Y) = E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2)\cdots E(X_n) \tag{1B.15}$$

바. 확률표본(random sample)

온전하지 못한 동전을 가지고 있다고 가정하자. 그 동전에 관한 한, 앞면이 나올 확률을 P라 했을 때 일반적으로 P는 반드시 $\frac{1}{2}$이 되지는 않는다. 뒷면이 나올 확률은 $(1-P)$일 것이다. P와 같이 어떤 종류의

계산식이나 확률모형에 등장하는 상수를 모수(parameter)라고 한다.

모수 P의 값을 모르나 대략 그 추정치를 얻을려고 한다고 가정하자. 이를 위해서 동전을 많이, 한 백번 정도 던져서 P의 추정치인 $\hat{P}$을 취한다고 한다. 이때 $\hat{P}$은 총 던진 횟수에 대비하여 앞면이 나온 횟수의 비율, 즉 (앞면의 횟수) / 100 이라고 한다. 이를 공식화하기 위하여, 처음에 뒷면이 나오면 그 값이 0 이고 앞면이 나오면 값이 1 이 되는 확률변수를 X_1 이라고 한다. 마찬가지로 $X_2, \cdots\cdots; X_{100}$을 각각 2 번째에서 100 번째 던졌을 때 그 값이 0 과 1 로 대응되는 확률변수들이라고 한다. 즉, 만일 i 번째 던졌을 때 뒷면이 나오면 $X_i = 0$ 이고, 앞면이 나오면 $X_i = 1$ 이다.

던져서 앞면이 나올 확률이 어떤 경우에도 달리 던졌을 때의 결과에 영향을 받지 않는다고 가정하면, 변수 $X_1, X_2, \cdots\cdots, X_{100}$ 은 독립적이다. 더구나 변수들은 모두 동일한 확률함수를 갖는다. 즉 X_i의 확률함수는 다음과 같을 것이다.

x_i	0	1
$f(x_i)$	$1 - P$	P

(1B.16)

$X_1, X_2, \cdots\cdots, X_{100}$ 과 같이 독립적이고 동일한 확률함수를 가진 확률변수들을 확률표본을 구성한다고 말한다. 표본이 추출되는 모집단(population)은 확률변수와 공통적인 확률함수를 가진 것으로 설명된다. 앞의 경우에서 표본이 추출된 모집단 (우리가 알려고 하는 모집단)은 다음과 같을 것이다.

x	0	1
$f(x)$	$1 - P$	P

(1B.17)

사. 추정량 (estimators)

앞의 보기에서 추정치 (estimate) P, 즉 $\hat{P}$을 구하는 방식은 확률변수 $X_1, X_2, \cdots\cdots; X_{100}$으로써 묘사될 수 있다.

$$\hat{P} = \sum_{i=1}^{100} \frac{X_i}{100} \qquad \text{(1B.18)}$$

예를 들면, 100번 던졌을 때 앞면이 80번 나오면 X_i 중의 80은 1이고 20은 0이어서, $\hat{P}$는 $\frac{80}{100} = 0.8$일 것이다.

기술적인 문헌에서는 앞의 예에 대하여 모수 P의 추정치가 0.8이라고 말할 것이다. 다른 한편으로 우리가 그 추정치, 즉 (1B.18)의 $\hat{P}$을 얻으려고 사용한 계산식은 추정량이라고 불린다. 즉, 추정치는 추정량 (또는 계산식)을 기초로 계산된 특정한 수이다. 예를 들어, 위의 $\hat{P}$은 100개의 확률변수인 $X_1, X_2, \cdots\cdots; X_{100}$의 값을 다 더해서 100으로 나눈 것을 말한다. 이것이 추정량이다. 그러나 0.8과 같은 특정한 값을 얻기 위해서는 실제로 동전을 100번 던져야만 100개의 확률변수의 관찰된 값을 얻을 수 있다. 그러므로 직관적으로는 추정치를 추정량의 사후적 또는 실현된 값으로 생각할 수 있을 것이다.

일반적으로는 $\hat{P}$과 같은 추정량은 확률변수의 함수가 된다. 〔 (1B.18)을 보라〕. 그러므로 추정량은 반드시 그 자체가 일종의 확률변수이다. 예를 들어, $\hat{P}$은 동전을 100번 던진 결과에 의존하여 0, 0.01, 0.02, ……, 0.99, 1.00중의 어느 값이라도 가질 수 있다. 따라서 100번의 던지기를 하나의 커다란 확률실험으로 본다면 $\hat{P}$은 한 확률변수가 된다.

아. 불편추정량 (unbiased estimators)

만일 어떤 추정량의 기대, 또는 평균이 모집단의 모수와 같다면, 그 추정량을 불편추정량이라고 한다. 즉, $\hat{b}$이 b의 추정량이고 다음의 (1B.19)

와 같다면, $\hat{b}$은 불편추정량이다.

$$E(\hat{b}) = b \tag{1B.19}$$

한편 $E(\hat{b}) \neq b$이면, $\hat{b}$은 b의 편의추정량 (biased estimator)이다.

예로써, 다시 P의 추정량, 즉 (1B.18)에서 정의된 $\hat{P}$을 보자. 기대에 관한 (1B.10)을 사용하면,

$$\begin{aligned} E(\hat{P}) &= E(\tfrac{1}{100}X_1 + \tfrac{1}{100}X_2 + \cdots + \tfrac{1}{100}X_{100}) \\ &= \tfrac{1}{100}[E(X_1) + E(X_2) + \cdots + E(X_{100})] \end{aligned} \tag{1B.20}$$

각 X_i의 확률함수가 (1B.16)이므로,

$$E(X_i) = 0(1 - P) + 1(P) = P \tag{1B.21}$$

(1B.21)을 (1B.20)에 대입하면,

$$E(\hat{P}) = \tfrac{1}{100}(100P) = P \tag{1B.22}$$

따라서 $\hat{P}$이 P의 불편추정량임을 보인 것이다.

편의추정량의 예로써 P^*, $P^* = e^P$의 값을 추정하는 문제를 고려하여 보자. 얼핏 보면, 이것은 우리가 지금 막 다룬 문제를 거의 바꾸어 놓지 않은 것처럼 보인다. 그러나 P^*는 비선형 유형으로 관련된 것이기 때문에, 사소하게 바꾸어 놓은 것은 아니다.

$P^* = e^P$의 명백한 추정량은 $\hat{P}^* = e^{\hat{P}}$일 것이며, 여기서 $\hat{P}$은 (1B.18)로 주어진다. 그러나 앞에서의 논의를 회상하면,

$$E(\hat{P}^*) = E(e^{\hat{P}}) \neq e^{E(\hat{P})} = e^P = P^* \tag{1B.23}$$

그러므로 $E(\hat{P}^*) \neq P^*$ 이어서, $\hat{P}^*$ 는 P^* 의 편의추정량이다. 요약하면, $\hat{P}$ 이 P 의 불편추정량이라는 것은 $\hat{P}$ 을 가지고서 직접 P 의 비선형함수의 불편추정량을 구할 수 있음을 의미하는 것은 아니다. 이 책에서 보게 되겠지만, 비선형변환 (nonlinear translation) 문제는 계량경제학의 중요한 문제의 하나이다.

자. 일치성 (consistency)

實例가 되는 (1B.18)의 추정량 $\hat{P}$ 은 확률표본의 크기가 100 인 것에 기초하고 있다. 만일 100 번 대신에 동전을 n 번 던졌다면, 우리는 $\hat{P}$ 을 다음과 같이 정의할 수 있을 것이다.

$$\hat{P}_n = \sum_{i=1}^{n} \frac{X_i}{n}. \tag{1B.24}$$

(1B.10)을 사용한다면, $E(\hat{P}_n) = P$ (즉, $\hat{P}_n$ 이 불편임)를 보이는 것은 어렵지 않다. 더 나아가서 2 장의 부록에서 전개할 식을 사용한다면, $\hat{P}_n$ 의 분산이 아래와 같음을 보일 수 있다.

$$\sigma^2_{\hat{P}_n} = \frac{1}{n^2}(\sigma_1{}^2 + \sigma_2{}^2 + \cdots + \sigma_n{}^2), \tag{1B.25}$$

여기서 $\sigma_i{}^2$ 은 X_i 의 분산이다. (1B.16) 에서 정의된 X_i 의 확률함수로부터 다음을 계산할 수가 있다.

$$\begin{aligned} \sigma_i{}^2 &= E[X_i - E(X_i)]^2 = E(X_i - P)^2 \\ &= [(0 - P)^2(1 - P) + (1 - P)^2 P] = P(1 - P). \end{aligned} \tag{1B.26}$$

(1B.26)을 (1B.25) 에 대입하면,

$$\sigma^2_{\hat{P}_n} = \frac{1}{n}[P(1 - P)]. \tag{1B.27}$$

(1B.27)로부터 표본크기가 무한대로($n \to \infty$)로 접근하면 $\hat{P}_n$의 분산이 0에 접근함을 알 수 있다.* 이러한 결과는 직관적으로 $\hat{P}_n$의 평균이 P라는 결과와 함께, $n \to \infty$함에 따라 $\hat{P}_n$의 값으로 유망한 것이 P임을 시사한다. 이를 기호화하면,

$$\lim_{n \to \infty} \text{Prob}(|\hat{P}_n - P| > \varepsilon) = 0 \qquad \text{(1B.28)}$$

여기서 ε는 사전에 정한 수이지만, 그 크기가 작은 수이다.

추정량이 (1B.28)과 같은 조건을 만족할 경우, 그 추정량을 대응하는 모수의 일치추정량이라 한다. 따라서, $\hat{P}_n$은 P의 일치추정량이다. 일반화하면, 다음의 (1B.29)가 성립하면 $\hat{b}$은 b의 일치추정량이다.

$$\lim_{n \to \infty} \text{Prob}(|\hat{b} - b| > \varepsilon) = 0 \qquad \text{(1B.29)}$$

(1B.29)에 비추어 보아 일치성의 조건은 자주 $P \lim \hat{b} = b$로 쓰인다. 이는 다시 만일 표본의 크기가 무한하면 $\hat{b}$이 b 이외의 것이 될 확률은 0이 될 것임을 말해준다. 마지막으로 어떤 추정량, $\hat{c}$이 일치추정량이 아니면,

$$\lim_{n \to \infty} \text{Prob}(|\hat{c} - c| > \varepsilon) \not\to 0 \qquad \text{(1B.30)}$$

이러한 추정량은 불일치추정량이라 한다.

---

* 이를 달리 표현하면, 표본이 무한의 크기를 가질 경우 P_n의 분산은 0이 될 것이라고 할 수 있다.

문 제

1. $X_1 = 0$, $X_2 = 5$, $X_3 = 6$, $X_4 = 1$ 일 때 $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ 임을 보여라.

2. 다음을 보여라.

$$\sum_{i=1}^{n}(aX_i + bY_i + cZ_i) = a\sum_{i=1}^{n}X_i + b\sum_{i=1}^{n}Y_i + c\sum_{i=1}^{n}Z_i$$

3. $\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$ 를 $\sum_{i=1}^{n}X_i(Y_i - \overline{Y})$ 로 표현할 수 있음을 보여라.

제 2 장 이변수회귀모형

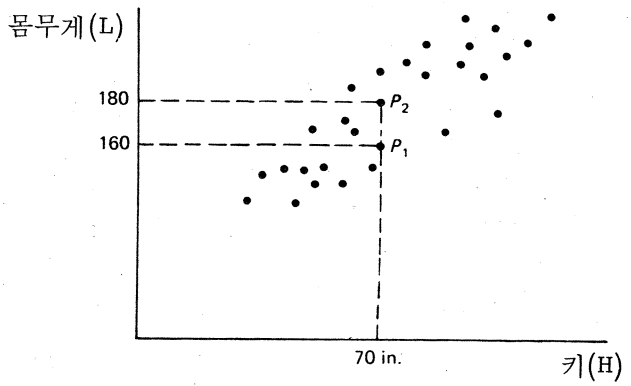
계량경제학의 중심이 되는 문제중의 하나는 경제변수들간의 양적 관계를 효과적으로 추정하는 기법들을 개발하는 것이다. 1장의 보기에서 보았듯이, 우리가 필요한 것은 소비함수의 모수 a, b의 신뢰성있는 추정치를 얻는 방법이고, 그래서 여러 것중에서도 가처분소득수준에 따라 어떻게 소비가 변할 것인가를 예측할 수가 있다. 이 2장에서 두 변수간의 관계를 추정하기 위한 기본적인 원리를 설명하기로 한다. 이 책에서 가장 중요한 부분이 이 2장임을 강조하여 두고 싶다. 여기서와 3장의 첫 절에서는 추정과 가설검정의 기본적인 개념의 구조를 소개하고 있다. 이어지는 장에서 다루고 있는 것은(예를 들면, 여러 변수들을 포함하고 있는 관계에 대한 추정) 기본적으로 이변수의 경우에서의 분석을 바로 직접 확대한 것으로써 구성된다.

1. 二變數間의 통계학적 관계 측정 : 공분산과 상관

우선, 두 변수간의 통계학적 관계를 묘사하는 데에만 관심이 있는 것으로 가정하자. 두 변수간의 어떤 종류의 인과관계를 포함하는 가설은 없다고 한다. 이 점에서 구하고자 하는 것은 오로지 두 변수가 어떠한 종류의 체계적인 연관 유형(pattern)을 보여주는 지를 파악하는 것이다.

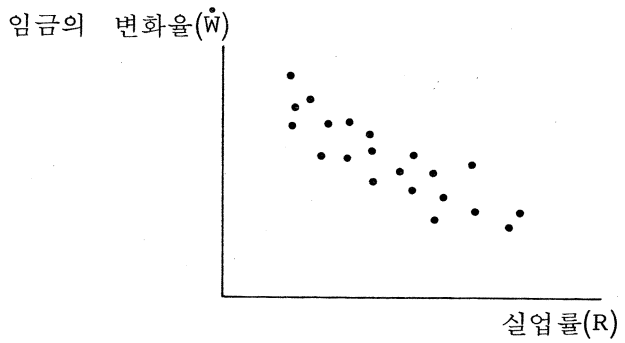
예로서, 임의로 선발된 30명의 몸무게(L파운드)와 키(H인치)를 기록하였다고 가정하자. 1장에서와 같이 관찰치를 그려놓은 <그림 2.1>의 산포도를 사용하기로 하고, 전과 마찬가지로 두 변수간에는 엄밀한 관계가 없다는 것을 밝혀 둔다. 일반적으로 키가 동일한 두 사람의 몸무게는 정확히 일치하지 않는다. 예를 들면, <그림 2.1>에서 볼 수 있듯이, 키가 70인치로 동일한 두 사람을 나타내는 점 P_1, P_2는 P_1이 160파운드인

반면에 P_2는 180파운드의 몸무게를 보이고 있다. 그럼에도 불구하고 L과 H 사이에는 동일한 유형의 관계가 드러난다. 키가 큰 사람일수록 대체로 키가 작은 사람보다 몸무게가 더 나가는 것이다. 따라서 평균적으로 L과 H는 正의 방향으로의 관련을 보이고 있다. H의 값이 클수록 전형적으로 L의 더 큰 값과 연관된다.



〈그림 2.1〉

대조적으로 <그림 2.2>의 산포도에서 고려되는 두 변수, 즉 임금의 퍼센트 변화율($\dot{w}$)과 실업률(R)과는 逆의 방향으로 관련되어 있다. 키를 가리키



〈그림 2.2〉

는 점은 왼쪽에서 오른쪽으로 가게 됨에 따라 감소하고, 이는 임금이 더 신속하게 증가하는 것은 전형적으로 더 낮은 실업률과 연관됨을 지적하는 것이다. 이러한 발견은 크게 놀랄만한 일이 아니다. 경기가 호황이고 사용 가능한 실업 노동자가 없으면, 고용주들은 자신의 생산품에 대한 고수준의 수요에 맞추어 산출량을 늘리기 위한 시도에서 상대적으로 신속하게 임금을 높일 것을 예상할 수 있다. 역으로 총수요가 적고, 그 결과로 실업이 보다 높은 수준에 있다면, 임금의 상승 압력은 훨씬 덜한 경향이 나타나게 된다. 덧붙여 말하자면, 이렇게 흩어져 있는 점들에 맞는 곡선이 필립스곡선 (Phillips Curve)으로 알려진 것은 A.W. 필립스가 최초로 영국에서 $\dot{w}$과 R의 관계에 주목한 이후의 일이다.*

가. 共分散 (covariance)

이제 두 변수에 관하여 묻고자 하는 것은 이 두 변수가 正의 관계를 갖는지 負의 관계를 갖는지에 관한 문제이다. 한 변수가 전형적인 값보다 더 큰 값을 가질때 다른 변수의 전형적인 값보다 더 큰 변수와 연관되는가(正의 관계)? 또는 정상적으로는 첫번째 변수의 더 큰 값을 두번째 변수의 더 작은 값이 따르게 되는가 (負의 관계)? 이러한 플러스 또는 마이너스 관계를 포착하는 모수가 공분산이다. 평균이 $E(X)=\mu_X$ 와 $E(Y)=\mu_Y$ 인 두 변수 X, Y에서 공분산 ($\sigma_{X,Y}$)는 공식적으로 다음과 같이 정의된다.

$$\sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] \tag{2.1}$$

---

* A.W.Phillips, "The Relation Between Unemployment and the Rate of Change of Money Wage Rate in the United Kingdom,1861-1957", Economica 25 (Nov.1958) pp.283-299를 보라.

즉, 공분산은 $(X-\mu_x)\times(Y-\mu_Y)$의 곱(Product)이 갖는 기대값이다. 만일 공분산이 正인 $\sigma_{x,Y}>0$이면, X가 X의 평균보다 더 큰 값인$(X-\mu_x)>0$은 통상적으로 Y가 Y의 평균보다 더 큰 값인$(Y-\mu_Y)>0$과 연관되며, 그 반대도 성립한다. 직관적으로 보면, 이 점은 단순하다. 만일 $\sigma_{x,Y}$가 양이면, $(X-\mu_x)$와 $(Y-\mu_Y)$ 두 항은 전형적으로 둘 다 양이거나 둘 다 음이어야만 한다. 그러므로 X가 자신의 평균인 μ_x보다 크다면, Y는 전형적으로 자신의 평균 μ_Y보다 클 것이다. 따라서 X와 Y는 正의 관계를 갖고 있다. 대조적으로 만일 $\sigma_{x,Y}<0$이면, X의 평균값보다 더 큰 X값은 통상적으로 Y의 평균값보다 더 작은 Y값이 따르게 되는데, 이는 두 변수간의 負의 관계를 가리키는 것이다.

물론 중간의 보기로서 $\sigma_{x,Y}=0$인 경우가 있을 것이다. 이 경우에 X의 평균값보다 더 큰 X값은 Y의 평균값보다 더 작은 Y값이 따르는 것과 마찬가지로 평균보다 더 큰 Y값이 따르게 된다. 이러한 경우가 발생할 수 있는 경우는 두 가지이다. 첫번째는 두 변수가 독립적일 경우이다. 예컨데, X와 Y가 독립적이면,

$$\begin{aligned}\sigma_{X,Y} &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E(X-\mu_X)E(Y-\mu_Y)=0\end{aligned} \tag{2.2}$$

왜냐하면

$$E(X-\mu_X)=E(X)-E(\mu_X)=\mu_X-\mu_X=0\ *$$

두번째는 두 변수가 서로간에 특수한 비선형 방식으로 관련이 맺어진 경

---

* 만일 두 변수가 독립적이면 그것들의 곱의 기대값은 두 변수의 기대값의 곱과 같다는 것을 독자들이 기억해야만 한다(1장의 부록 B). 또한 X의 평균, 즉 μ_x는 상수이어서, $E(\mu_x)=\mu_x$이다.

우이다. 이러한 예를 아래에서 보게 될 것이다. 그러나, 이 점에서 두 변수간의 공분산이 0이라는 것이 곧 그 두 변수가 독립적임을 의미하는 것이 아니라는 점에 유의해야 한다. 즉, 비선형의 경우도 있기 때문이다. 대신에 0의 공분산은 오로지 두 변수가 선형으로 관련된 것이 아니라는 것만을 의미한다.

나. 공분산의 추정량 (covariance estimator)

실제로는 일반적으로 $\sigma_{X,Y}$의 값을 알지 못할 것이다. 전형적으로, 임의의 처리에 따라 X와 Y의 관찰된 값으로 이루어진 확률표본을 가질 뿐이다. 예를 들어 앞에서 특정한 수, n명의 사람을 임의로 선발한 적이 있었다. 그러한 경우에 우리는 L과 H에 관하여 n 크기의 표본을 가졌다고 말한다. 이제 X와 Y에 대하여 n 크기의 표본을 가졌다고 가정하자. 우리가 필요로 하는 것은 이 관찰된 표본으로부터 $\sigma_{X,Y}$의 값을 추정하는 방식이다. $\sigma_{X,Y}$는 변수들이 자신의 평균과의 편차를 곱 (product) 한 것의 기대값으로 정의되기 때문에 $E[(X-\mu_X)(Y-\mu_Y)]$, $\sigma_{X,Y}$를 추정하는 뚜렷한 방식은 X, Y변수의 표본평균으로부터의 X와 Y의 편차를 곱한 것이 갖는 표본 평균의 계산이 될 것이다. 더 공식적으로 말하자면, 우리의 X와 Y 표본의 관찰된 값 n개를 $X_1, X_2, \cdots\cdots, X_n$과 $Y_1, Y_2, \cdots\cdots, Y_n$*이라고 하자. 그러면 X와 Y 사이의 공분산은,

$$\hat{\sigma}_{X,Y} = \frac{\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y})}{n-1} \qquad (2.3)$$

여기서

---

* 예를 들면, X_5는 X의 다섯번째 관찰된 값이다. 앞에서의 설명한 방식에 따르면, H가 한 사람의 키이고 L이 몸무게일 때, H_5는 다섯번째 관찰된 사람의 키이고 L_5는 동일인의 몸무게이다.

$$\bar{X} = \frac{\sum_{t=1}^{n} X_t}{n} \quad \text{이고} \quad \bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} \quad \text{임.}$$

아래에서 (2.3)식이 왜 n이 아니라 $n-1$로 나누어졌는가를 설명하기로 한다. 그러나, 이 점에서 $(X_t - \bar{X})(Y_t - \bar{Y})$가 표본 평균으로부터의 변수들의 편차에 관한 t번째 관찰된 곱이기 때문에, $\hat{\sigma}_{X,Y}$는 단지 그러한 곱의 평균이라는 것을 밝혀 둔다.*

이 점에서 잠시 멈추어서 이상의 표기법을 설명하는 것이 도움이 될 것이다. 이 책에서 어떤 변수나 모수 위에 있는 ^기호는 "~의 추정량"을 가리킨다. 따라서 $\hat{\sigma}_{X,Y}$는 $\sigma_{X,Y}$의 추정량이다. (2.3)의 우변은 $\hat{\sigma}_{X,Y}$가 표본 전체, $t = 1, \cdots\cdots, n$에 걸친 곱 $(X_t - \bar{X})(Y_t - \bar{Y})$의 총합에 기초하고 있음을 지적한다. 표기를 간단히 하기 위해서 여기서부터는 번거롭게 $\sum_{t=1}^{n}$이라고 쓰지 않고, 만일 별도의 지적이 없다면 가산 과정이 표본의 모든 n개의 관찰치까지 전개된다고 이해하고 간단히 $\sum$만을 사용한다.

다. $\hat{\sigma}_{X,Y}$의 불편성

$\hat{\sigma}_{X,Y}$가 X와 Y 사이의 공분산의 추정량이라는 것을 알았다. 더구나, $E[\hat{\sigma}_{X,Y}] = \sigma_{X,Y}$임을 보일 수 있다. 즉, $\hat{\sigma}_{X,Y}$는 $\sigma_{X,Y}$의 불편추정량이다. 여기서 얻는 생각은 $\hat{\sigma}_{X,Y}$가 확률표본 X와 Y에서 발생되기 때문에, 그 자체가 값이 표본에 따라 변화하는 확률변수라는 점이다. 예를 들어 임의로 선발된 30명의 사람의 몸무게와 키, L과 H가 선택된 특정한 사람들에 종속한다면, L과 H 사이의 공분산의 추정량이 갖는 값도 또한 선

---

* 만일 우리의 표본에서 5번째 관찰된 사람이 표본 전체의 키와 몸무게 평균보다 3인치 크고 15파운드 덜 나간다고 하면, $(H_5 - \bar{H}) \times (L_5 - \bar{L}) = 45$이다. L과 H의 공분산에 대한 우리의 추정량 $\hat{\sigma}_{L,H}$는 단지 표본 전체에 걸친 이와같은 곱의 평균이다.

발된 사람에 의존한다. 따라서 그러한 추정량의 값은 표본에 따라 변화할 것이다. 보다 공식적으로 말하면, 추정량은 추정량의 표본분포(sampling distribution)라고 불리는 확률함수를 가진다. 앞서 말한 결과, 즉 $E(\hat{\sigma}_{X,Y}) = \sigma_{X,Y}$는 $\hat{\sigma}_{X,Y}$의 표본분포의 평균이 모수 $\sigma_{X,Y}$의 값임을 의미한다.

비록 공식적인 증명은 이 책의 수준을 넘어서지만, 한 예로써 $\hat{\sigma}_{X,Y}$가 $\sigma_{X,Y}$의 불편추정량이라고 말하는 것이 무엇을 의미하는가를 보다 직관적인 수준에서 명확히 하는 데 도움을 줄 것이다.* 앞의 서술을 확장해서 30명으로 이루어진 M개의 표본들을 추출하고 각 표본에 속하는 사람들의 몸무게 L, 키 H를 측정하였다고 가정하자. 그러면 각 표본으로 $\hat{\sigma}_{L,H}$를 구하는 값을 계산할 수 있었는데, 그 계산에서 L과 H 사이의 공분산에 관한 M개의 구분되는 추정치를 얻게 되었을 것이다. 다양한 $\hat{\sigma}_{L,H}$는 일반적으로 상이하다는 것을 유의해야 한다. 특히, 우리의 추정치들 중의 일부는 $\sigma_{L,H}$보다 크고 일부는 작을 것임을 예견할 수 있을 것이다. 이제 $\hat{\sigma}_{L,H}$의 기대값이 $\sigma_{L,H}$라는 기억을 되살려 보자. 이것은 만일 우리가 이들 M개의 추정치의 평균을 취하면, 평균값이 모수 $\sigma_{L,H}$가 될 것임을 예상할 것임을 의미한다. 어느 정도로 더 공식화하면, $(\bar{\hat{\sigma}}_{L,H})$가 그 평균이라고 하자.

$$\bar{\hat{\sigma}}_{L,H} = \sum_{i=1}^{M} \frac{\hat{\sigma}_{L,H_i}}{M} \tag{2.4}$$

여기서 $\hat{\sigma}_{L,H_i}$는 i번째 표본에 기초한 $\sigma_{L,H}$의 추정량이다. 그러면

---

* 뒤의 보기들은 불편성 개념에 관한 "직관적인 정의"가 아니다. 대신에, 그 보기들은 일반적인 조건 아래에서 불편성이 의미하는 일정한 결과를 직관적으로 제시하는 것이다.

$E(\bar{\hat{\sigma}}_{L,H}) = \sigma_{L,H}$이다. 더 나아가, 일반적인 조건하에서, 만일 M이 무한한 크기이면, $\bar{\hat{\sigma}}_{L,H}$와 $\sigma_{L,H}$가 조금이라도 차이가 날 확률은 0이 될 것임을 보일 수 있다.

이러한 결과의 해석은 수월하다. 실제로는 전형적으로 단지 하나의 표본만을 가진다. 이러한 표본에 기초하여서 (2.3)으로 주어진 일반적인 식을 가지고 공분산을 추정한다. 만일 이 표본이 임의로 추출되었다면, 무한한 수의 표본중의 어느 하나(예컨데, 위의 M개 표본중의 어느 것)가 될 수 있었다. 그러나, 앞에서 본 평균화된 결과 때문에, 우리가 추정한 것이 대응하는 공분산 모수의 값을 능가하거나 작다고 믿을 이유가 없다. 대조적으로, 만일 $E(\hat{\sigma}_{X,Y}) = \sigma_{X,Y} + 5$라면 우리가 계산한 값이 $\sigma_{X,Y}$를 넘을 것이라고 예견할 수 있었을 것이다. 만약 이러한 경우라면, 이 편의는 $\sigma_{X,Y}$의 추정량으로서 $(\hat{\sigma}_{X,Y} - 5)$를 취하여 조절할 수 있을 것이다.

아마 독자들은 아직까지도 (2.3)식의 분모가 표본 크기의 전부인 n이 아니라 $(n-1)$이라는 사실에 당혹할 것이다. 정상적인 경우 평균을 계산할 때는 값들의 합계를 계산에 포함된 항의 수로 나눈다. 그러나 이런 경우에는 비록 분자의 합계에서 n개 항이 계산되더라도 이 n항은 동일한 합계를 갖는 $(n-1)$항으로 축소될 수 있다. 어떤 의미에서는 단지 $(n-1)$항의 정보 "조각"(bits)만이 있는 것이다. 그 이유는 표본 평균인 $\overline{X}$와 $\overline{Y}$가 X와 Y의 관찰된 값과 함께 분자에 있기 때문이다. 직관적으로 (2.3)에서 처음의 $(n-1)$항까지를 형성하기 위해서는, $X_1, X_2, \cdots\cdots; X_{n-1}, Y_1, Y_2, \cdots\cdots, Y_{n-1}$과 $\overline{X}$, $\overline{Y}$를 알아야만 한다. 여기서 $\overline{X}$과 $\overline{Y}$는,

$$\overline{X} = \frac{\sum(X_1 + \cdots + X_n)}{n}, \qquad \overline{Y} = \frac{\sum(Y_1 + \cdots + Y_n)}{n}$$

이러한 정보를 가지고서 X_n과 Y_n의 값이 얼마가 되어야 하는지를 정확하게 결정할 수 있다. 이 상황 아래에서 (2.3)의 마지막 항, 즉 $(X_n - \overline{X}) \times (Y_n - \overline{Y})$는 아무런 새로운 정보를 갖고 있지 못하다. 요컨데, 우리는 처음 $(n-1)$항까지에 담긴 정보로부터의 합계에서 마지막 항의 값을 이미 알고 있거나 계산할 수 있는 것이다. 이 조건은 자주 (2.3)의 분자가 $(n-1)$의 자유도(degree of freedom) 만을 가진다고 표현된다. 이는 단지 $(n-1)$개의 독립적인 정보 조각을 갖고 있음을 의미한다. 분자가 $(n-1)$의 자유도만을 가지고 있기 때문에, 다음과 같이 보일 수가 있다.

$$E[\sum (X_t - \overline{X})(Y_t - \overline{Y})] = (n-1)\sigma_{X,Y} \tag{2.5}$$

결과적으로, $(n-1)$로 나누는 것이 $\hat{\sigma}_{x,Y}$를 $\sigma_{x,Y}$의 불편추정량으로 만드는 것이다.

라. $\hat{\sigma}_{x,Y}$의 일치성

$\hat{\sigma}_{x,Y}$가 대표본에서 갖는 속성들을 살펴보는 것이 유용하고, 또한 뒤에서의 분석에서 귀중하다는 것이 입증될 것이다. 그 속성은 $\hat{\sigma}_{x,Y}$가 기초한 표본의 크기가 무한하게 커질 경우에 $\hat{\sigma}_{x,Y}$의 행태를 의미한다.

예를 들어, X의 한 확률표본의 표본 평균 $\overline{X}$를 고려하여 보자. 우리는 기초 통계학에서 $\overline{X}$의 평균이 μ_x이고 $\overline{X}$의 분산이 σ^2_x/n임을 배웠다. 여기서 μ_x와 σ^2_x은 확률변수 X의 평균과 분산이며, n은 표본의 크기이다. 표본크기 n이 커짐에 따라, $\overline{X}$의 분산, 즉 σ^2_x/n가 점점 작아지고 n이 한계를 넘어 증가하면 0이 됨을 알 수 있다. 요점은 표본이 커짐에 따라 표본 평균 $\overline{X}$가 모집단 평균 μ_x에 대한 특정한 구간(interval)내에 존재할 확률이 점차 높아진다는 점이다. 표본의 크기가 무

한대가 되는 극한에 이르면, $\overline{X}$의 분산은 0이어서 $\overline{X}$가 μ_x와 다를 확률이 0이 된다. 이러한 이유로 $\overline{X}$는 μ_x의 一致推定量이다. 보다 일반적으로는(1장의 부록 B에서 논의하였듯이) 일치성은 표본이 무한대의 크기를 갖는 극단적 경우에서 추정량이 대응하는 모수와 조금이라도 다른 값을 가질 확률이 0인 것이다. ($\overline{X}$와 같은)추정량이 일치추정량이면, 그 추정량은 그 대응하는 모수(μ_x)에 "확률상 수렴한다(converge in probability)"고 말한다.

적어도 직관적으로는 (2.3)식의 경우에서 $\hat{\sigma}_{x,y}$가 $\sigma_{x,y}$의 한 일치추정량임을 쉽게 알 수 있다. 표본크기가 무한대로 됨에 따라서 $\overline{X}$와 $\overline{Y}$는 각각 μ_x와 μ_y에 확률상 수렴한다. 그러므로, $\hat{\sigma}_{x,y}$는 무한한 크기의 표본에 기초한 $(X-\mu_x)(Y-\mu_y)$의 표본 평균이 된다. $E(X-\mu_x)(Y-\mu_y)=\sigma_{x,y}$이므로, 일반적인 조건 아래에서 만일 표본이 무한대의 크기를 가진다면,

$$\hat{\sigma}_{X,Y} = \sigma_{X,Y}$$

이며 그 확률은 1과 같다.*

---

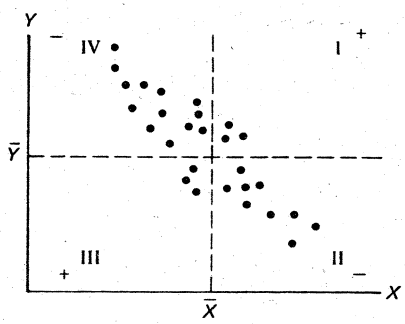
* 더 많은 것을 아는 독자에 대한 사항으로서, 1장의 부록 B에서 논의하였던 일치성의 속성에 더해서 다른 수렴 형식이 있다. 그것중의 하나가 "확률 1을 갖는 수렴"으로 불리는 것이다. 앞에서는 이러한 수렴의 형식을 언급하지 않았다. 대신에 표기를 단순하고 직관적으로 알 수 있도록, 아래와 같이 묘사하고,

$$\lim_{n\to\infty} \text{prob}\,(|\hat{\sigma}_{X,Y} - \sigma_{X,Y}| > \varepsilon) = 0$$

말로는 "표본이 무한대의 크기를 가지면 $\sigma_{x,y}$와 같게 될 것이며 그 확률은 1과 같다"고 하였다.

마. $\hat{\sigma}_{x,Y}$의 해석

<그림 2.3>의 흩어져 있는 점들에 비추어 이제 $\hat{\sigma}_{x,Y}$를 해석하도록 하자.



〈그림 2.3〉

<그림 2.3>에서 점선은 X와 Y의 관찰치가 갖는 평균이며, 이 점선으로 <그림 2.3>을 네 영역으로 분할하였다. 각 영역에서 관찰된 것의 특징은 아래와 같다.

Ⅰ영역 : $(X-\bar{X})>0$ 또한 $(Y-\bar{Y})>0$,

그러므로 $(X-\bar{X})(Y-\bar{Y})>0$

Ⅱ영역 : $(X-\bar{X})>0$ 또한 $(Y-\bar{Y})<0$,

그러므로 $(X-\bar{X})(Y-\bar{Y})<0$

Ⅲ영역 : $(X-\bar{X})<0$ 또한 $(Y-\bar{Y})<0$,

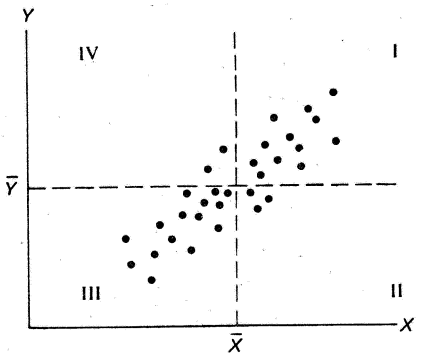
그러므로 $(X-\bar{X})(Y-\bar{Y})>0$

Ⅳ영역 : $(X-\bar{X})<0$ 또한 $(Y-\bar{Y})>0$,

그러므로 $(X-\bar{X})(Y-\bar{Y})<0$

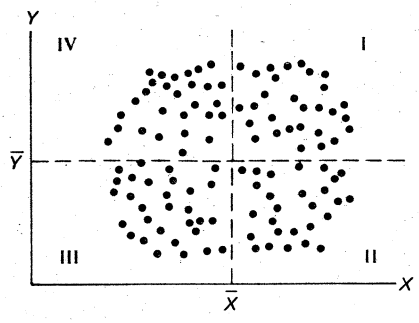
<그림 2.3>의 관찰에 따르면, 두 변수간의 負의 관계가 제시된다. 이 관계는 상대적으로 Ⅰ과 Ⅲ 영역에 소수의 관찰기록이 있으면서, Ⅱ와 Ⅳ

영역에 점들이 집중되는 양상을 취하고 있다. $(X-\overline{X})(Y-\overline{Y})$가 Ⅱ와 Ⅳ에서 陰이고, Ⅰ과 Ⅲ에서 陽이기 때문에, 이러한 경우에서의 $\hat{\sigma}_{X,Y}$는 음일 것으로 기대된다. 또는 다른 말로 하면, 각 평균으로부터 편차의 곱 $(X-\overline{X})(Y-\overline{Y})$이 갖는 평균이 음이 될 것이다. 다른 한편으로 <그림 2.4>에서 처럼 영역 Ⅰ과 Ⅲ에 점들이 집중하여 있다면, 앞과 유사한 논리로 $\hat{\sigma}_{X,Y}$가 양일 것으로 예견된다.



〈그림 2.4〉

이제 X와 Y가 독립적이어서, 이들간에 아무런 연관도 없는 사례를 보도록 하자. Y의 큰 값은 X의 큰 값이 따르는 것과 마찬가지로 X의 작은 값도 수반된다. 이러한 경우에, X와 Y의 산포도는 상향하거나 하향하는 추세를 보이지 않을 것으로 예상된다. <그림 2.5>는 그러한 산포도를 그린 것이다. 그림에서 점들이 네 영역으로 상대적으로 균등하게 있음을 본다. 결과적으로 영역 Ⅰ과 Ⅲ에서의 점으로 산출된 양의 값의 $(X-\overline{X})(Y-\overline{Y})$는 Ⅱ와 Ⅳ의 점으로 발생한 음의 값에 의하여 상쇄되는 경향이 있다. 따라서 계산된 $\hat{\sigma}_{X,Y}$값은 0에 근접하는 경향이 있다.



〈그림 2.5〉

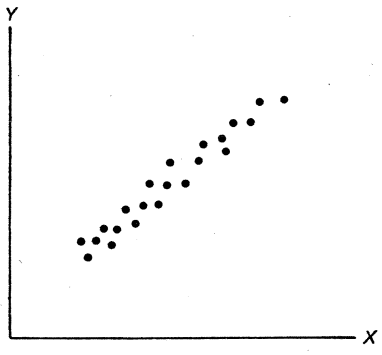
바. 상관계수 (correlation coefficient)

두 변수의 正 또는 負의 관련 여부를 아는 것에 더하여, 우리는 일반적으로 두 변수가 얼마나 강하게 관련을 맺고 있는지를 알기 원한다. 예를들어 <그림 2.6>과 <그림 2.7>은 X와 Y의 正의 관계에 관한 두 사례이다. 그러나 어떤 의미에서, 후자보다 전자의 正의 연관이 더 강하다. 더 자세히 말하자면, 일단 X의 값이 알려진 경우에서 <그림 2.6>의 Y 값이 변동하는 것이 <그림 2.7>에서 그려져 있는 Y값의 변동보다 작음을 알 수 있다.* 이러한 X와 Y의 특징적인 관계를 측정하는 수단을 갖는 것이 매우 바람직하다. 공분산의 측정은 불행히도 그 값이 변수가 측정되는 특정한 단위에 의존하기 때문에, 연관의 강한 정도(degree)를 지시하는 데에는 부적합하다. 예를 들어 키와 몸무게 사이의 공분산은 단위를 피트 (feet)와 파운드 (pound)로 사용했을 경우보다 인치 (inch-

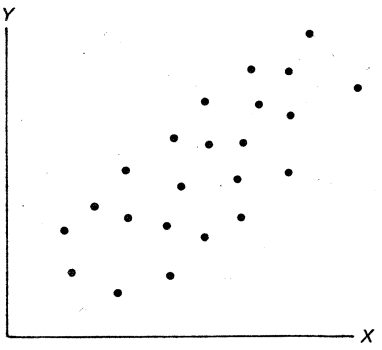
* 더 많은 지식을 가진 독자라면, 어떤 특정한 조건 아래에서 X가 주어졌을 때, Y의 조건부 분산(conditional variance)은 상관계수의 자승에 따라 역으로 변동함을 볼 수 있다.

es)와 온스(ounces)로 키와 몸무게를 재었을 경우 훨씬 더 크다.*

두 변수간의 연관이 갖는 강도를 재는 의미있는 지수(index)를 얻기 위해서는 특정한 측정단위와 무관한 값을 갖는 모수가 필요하다. 그



〈그림 2.6〉



〈그림 2.7〉

---

* 예를 들어, 측정단위가 X대신에 $Z = aX$이고 a가 상수라고 가정하자. 그러면,

$$E(Z - \mu_Z)(Y - \mu_Y) = E(aX - a\mu_X)(Y - \mu_Y) = aE(X - \mu_X)(Y - \mu_Y) = a\sigma_{X,Y}$$

따라서 $\sigma_{Z,Y} \neq \sigma_{X,Y}$

러한 모수는 X와 Y의 공분산을 그들 변수의 표준편차(standard deviation)로 나누어 줌으로써 얻을 수 있다. 더 정확히 말하자면, X와 Y 사이의 관계가 갖는 강도는 상관계수, $\rho_{x,Y}$에 의하여 지시된다.

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{2.6}$$

여기서 σ_x와 σ_Y는 X와 Y의 표준편차이다.

$$\sigma_X = +\sqrt{E(X-\mu_X)^2} \quad \text{또한} \quad \sigma_Y = +\sqrt{E(Y-\mu_Y)^2}$$

이제 상관계수의 속성을 살펴 보자.*

첫째로, 상관계수는 항상 공분산과 같은 부호를 갖는다는 점이다. (2.6) 식의 분모는 항상 양이기 때문에, $\rho_{x,Y}$의 부호는 분자의 부호와 동일할 것이며, 분자는 두 변수간의 공분산이다. 만일, X와 Y가 正의 관계를 가지면, $\rho_{x,Y} > 0$이다. 만일, X와 Y가 陰의 관계를 가지면, $\rho_{x,Y} < 0$이다. X와 Y가 독립적이라면, $\sigma_{x,Y} = 0$이므로 $\rho_{x,Y} = 0$이다. 상관계수는 변수들 간에 존재하는 유형의 관계를 지시하는 점에서 공분산의 모든 특징을 가지고 있음을 여기서 알았다.

그러나 상관계수는 공분산과 다르게 그 가능한 값의 범위에 한계가 있다. 자세히 말하자면, $\rho_{x,Y}$의 값은 반드시 +1과 −1 사이에 놓여 있다. 더구나 $\rho_{x,Y}$가 +1이나 −1에 가까워질수록 변수들 사이의 선형 연

---

* 앞의 각주에서처럼, $Z = aX$를 X대신에 사용함으로써 측정단위를 바꾸어도, 모수 $\rho_{x,Y}$는($\sigma_{x,Y}$와는 달리) 영향을 받지 않는다. 즉, $\rho_{z,Y} = \rho_{x,Y}$. 이의 증명은 독자들의 연습문제로 남겨 놓는다.

관(負이거나 正이거나)은 더욱 강해진다. $\rho_{X,Y}$가 0의 근처로 가면, 그 관계 즉, $\rho_{X,Y}=0$은 변수 사이의 어떠한 선형 연관도 없음을 보여주는 것이므로, 더 약해진다. 예를 들어 <그림 2.6>과 <그림 2.7>로 보면, <그림 2.6>의 상관계수가 <그림 2.7>의 경우보다 더 클 것임을 예견할 수 있다. 물론 두경우 모두 상관계수는 陽이다.

이제, X와 Y가 완전한 그리고 선형인 관계를 가지고 있다면, $\rho_{X,Y}$는 $+1$ 또는 -1의 값을 가지는 것을 보이기로 하자. 이것을 확인하기 위하여,* 다음의 실제 관계를 고려하여 보자.

$$Y = a + bX \tag{2.7}$$

이 경우에, 산포도상의 모든 점은 기울기 b와 절편 a를 갖는 직선 위에 정확하게 놓여 있다. X와 Y의 상관계수는,

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

σ_X로 $\sigma_{X,Y}$와 σ_Y를 모두 바꾸어 줌으로써 만일 $b>0$이면 $\rho_{X,Y}=+1$임을 보이기로 하자. 먼저 μ_Y를 끌어 내자.

$$E(Y) = E(a + bX) = a + bE(X) = a + b\mu_X = \mu_Y \tag{2.8}$$

따라서 공분산은

---

* 다음의 예는 X와 Y가 완전히 선형으로 관계를 맺고 있다면, $\rho_{X,Y}$가 $+1$ 또는 -1임을 보여 준다. 불행히도 $\rho_{X,Y}$가 어떠한 상황에서도 절대값 1을 넘지 못한다는 것의 증명은(비록 직관적으로 볼 때 합리적이지만) 이 책의 수준을 넘어 선다.

$$\begin{aligned}\sigma_{X,Y} &= E[(Y-\mu_Y)(X-\mu_X)] \\ &= E[(a+bX-a-b\mu_X)(X-\mu_X)] \\ &= E[b(X-\mu_X)^2] = b\sigma^2{}_x\end{aligned} \tag{2.9}$$

정의에 따르면, Y의 분산은

$$\sigma^2{}_x = E[(Y-\mu_Y)^2] \tag{2.10}$$

이 분산은 다음과 같이 나타낼 수가 있다.

$$\begin{aligned}\sigma^2{}_Y &= E[(Y-\mu_Y)^2] = E[(a+bX-a-b\mu_X)^2] \\ &= E[b^2(X-\mu_X)^2] = b^2\sigma^2{}_x\end{aligned} \tag{2.11}$$

Y의 표준편차는 陽의 제곱근, $\sigma_Y = b\ \sigma_X$ 이다. 이제 상관계수를 정할 수가 있다.

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_Y\sigma_X} = \frac{b\sigma^2{}_x}{(b\sigma_X)\sigma_X} = 1 \tag{2.12}$$

독자들은, $b<0$ 이면 $\rho_{X,Y} = -1$ 이라는 것을 증명하여 보기 바란다(힌트 : 만일 $b<0$ 이면, $\sigma_Y = -b\ \sigma_X > 0$).

요약하면, 상관계수는 두 변수간의 선형 관계가 갖는 부호와 강도를 동시에 지시하는 것이다. $\rho_{X,Y}$의 陽과 陰의 값은 각각 正과 負의 관계를 표시하고, $\rho_{X,Y}$가 $+1$ 또는 -1에 가까이 갈수록 선형관계가 더 강해지거나, 또는 통상적으로 일컬어 지듯이 두 변수는 한층 더 높은 상호 관련을 갖는다.

만약 두 변수가 독립적이면, 공분산은 물론 상관계수도 0이 될 것임을 보았다. 이제 그 역은 성립하지 않음을 보이기로 하자. 즉, 두 변수가 비선형 방식으로 관련을 맺고 있어도 상관계수가 0일 것이다. 상관계수는

두 변수 사이의 선형 관계를 측정하는 수단임을 다시 강조하여 둔다.

〈표 2.1 〉

X	$P(X)$
−1	$\frac{1}{3}$
0	$\frac{1}{3}$
1	$\frac{1}{3}$

예를 들어 X가 확률변수이고 $Y=X^2$ 이라 하자. 그러면 X값을 앎으로써 Y값을 정확히 예측할 수 있다는 점에서 X와 Y가 완전히 관련되어 있다는 것이 분명하다. 이제 <표 2.1 >과 같이 X의 확률함수가 주어졌다고 가정하자. 즉, X는 동일한 확률로 −1, 0과 1의 값을 취한다. 이제 X와 Y의 공분산을 정하도록 하자.

$$\sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)]$$

<표 2.1 >로 부터, $\mu_x = 0$ 임을 알 수 있기 때문에 $\sigma_{x,Y}$ 의 식은 다음과 같이 단순화된다.

$$\sigma_{X,Y} = E[X(Y - \mu_Y)] = E(XY) - E(X\mu_Y) \tag{2.13}$$

가정에서 $Y=X^2$ 이므로,

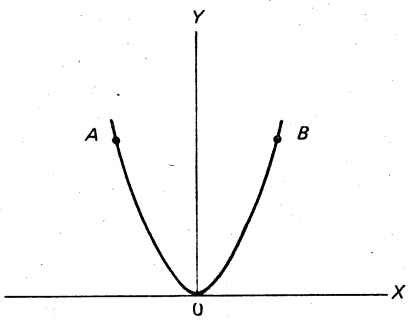
$$\sigma_{X,Y} = E(X^3) - \mu_Y E(X) = E(X^3) \tag{2.14}$$

왜냐하면 $E(X)= 0$ 이기 때문이다. $E(X^3)$을 알려면 X^3 이 〈표 2.1 〉의 X처럼 동일한 출현 확률을 갖는 동일한 값을 정확하게 취한다는 사실에 유의하면 된다. 그러므로,

$$E(X^3) = -1(\tfrac{1}{3}) + 0(\tfrac{1}{3}) + 1(\tfrac{1}{3}) = 0 \qquad (2.15)$$

$\sigma_{X,Y}$ 는 따라서 0 이다. 이로부터 $\rho_{X,Y}$ 또한 0 이다.

직관적으로 지금까지 거론된 내용을 알려면, 좀 더 일반적인 경우로서 $Y=X^2$ 로서 X가 0을 주위로 대칭적인 확률함수를 갖는 조건의 제약을 받으면서 모든 범위의 값을 취할 수 있다고 하자. 확률함수가 0을 중심으로 대칭적이라는 것은 단지 X값이 어떤 두 양수 사이, 예컨데 5와 10에 있을 확률이 정확하게 그 양수에 대응하는 음수, 예컨데 −5와 −10 사이에 있을 확률과 동일하다는 것을 의미한다. $Y=X^2$ 은 X축에 원점에서 접하는 <그림 2.8>에 그려진 포물선의 식이다. <그림 2.8>로 부터 Y와 X같은 변수가 일반적으로 0의 상관을 갖는 이유가 명백하게 된다. $Y=X^2$ 이기 때문에, X와 Y의 모든 관찰치는 반드시 포물선위에 놓여 있다. X의 확률함수가 0을 주위로 대칭적이기 때문에, A와 같은 사



〈그림 2.8〉

상(event)은 동일한 확률을 갖고 출현할 수 있는 B와 같이 대응하는 사상이 존재할 수가 있다. 결과적으로, X가 양의 값으로 제한될 경우

의 X와 Y 사이의 양의 상관은 X가 음의 값으로 제한될 경우의 음의 상관에 의하여 상쇄된다. 그러므로, X와 Y 사이의 전체에 걸친 상관은 0이 될 것이다.

이상의 경우들에서 0의 상관계수를 발생시키는 X와 Y의 실제 관계를 보았다. 앞에서는 X와 Y가 독립적일 때 $\rho_{X,Y}=0$임을 보였다. 여기서는 $\rho_{X,Y}=0$이 두 변수가 독립적이 되기 위한 필요조건이지 충분조건은 아님을 강조하여 둔다. 환언하면, 만약 $\rho_{X,Y}=0$이면, X와 Y가 독립적일 지는 몰라도 반드시 그런 것은 아니다. 그러나, X와 Y가 독립적이면, $\rho_{X,Y}=0$이다. 이것이 계량경제학의 분석에 의미하는 모든 것은(뒤에서 논의할 것이다), 만일 두 변수간의 낮은 상관을 발견하더라도, 그 변수 사이에 어떤 비선형 관계가 존재할 가능성이 있다는 것에 위안을 삼아야만 한다는 점이다.

사. 상관계수의 추정량

공분산의 경우에서처럼, 일반적으로는 상관계수 $\rho_{X,Y}$의 값을 알지 못하므로, 추정문제가 다시 제기된다. $\rho_{X,Y}$의 명백한 추정량은*

$$\hat{\rho}_{X,Y}=\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X\hat{\sigma}_Y} \tag{2.16}$$

여기서 $\hat{\sigma}_X$와 $\hat{\sigma}_Y$는 X와 Y가 갖는 표준편차의 통상적인 추정량이다.

$$\hat{\sigma}_X=+\sqrt{\frac{\sum(X_t-\bar{X})^2}{n-1}} \quad \text{또한} \quad \hat{\sigma}_Y=+\sqrt{\frac{\sum(Y_t-\bar{Y})^2}{n-1}} \tag{2.17}$$

대표본에서 $\hat{\rho}_{X,Y}$의 속성을 고려하여 보자. 앞서서 n이 무한대로 접

---

* 어떤 책에서는 $\hat{\rho}$대신에 r를 기호로 사용하고 있다. 우리는 $\hat{\rho}$표기를, $\hat{\rho}$가 ρ의 추정량임을 강조하기 위하여 사용하도록 한다.

근함에 따라 $\overline{X}$와 $\overline{Y}$가 각각 μ_X와 μ_Y로 확률상 수렴한다는 결과를 상기하도록 한다. 결과적으로, $\sum(X_i-\overline{X})^2/(n-1)$은 무한의 표본에 기초하여 그 평균으로의 X의 평균적인 편차의 자승이 되기 때문에, X의 분산 σ^2_x에 확률적으로 수렴한다. 따라서, $\hat{\sigma}_x$는 확률상 σ_x로(그리고 유사하게 $\hat{\sigma}_Y$도) 수렴한다. 동시에 $\hat{\sigma}_{X,Y}$는 $\sigma_{X,Y}$에 수렴하기 때문에, 적어도 직관적으로 $\hat{\rho}_{X,Y}$는 $\rho_{X,Y}$의 한 일치추정량이다.

$\hat{\sigma}_{X,Y}$와 대조적으로 $\hat{\rho}_{X,Y}$는 일반적으로 불편추정량이 아니다. 즉,

$$E(\hat{\rho}_{X,Y}) \neq \rho_{X,Y} \tag{2.18}$$

그 이유는 $\hat{\rho}_{X,Y}$는 $\hat{\sigma}_{X,Y}$, $\hat{\sigma}^2_x$과 $\hat{\sigma}^2_Y$의 비선형 함수로써 구성되어 있다. 즉,

$$\hat{\rho}_{X,Y} = \frac{\hat{\sigma}_{X,Y}}{\sqrt{\hat{\sigma}_X{}^2}\sqrt{\hat{\sigma}_Y{}^2}} \tag{2.19}$$

이제 $E(\hat{\sigma}^2_x)=\sigma^2_x$와 $E(\hat{\sigma}^2_Y)=\sigma^2_Y$를 보일 수 있다. 그러나, 1장 부록 B에서의 비선형 함수에 대한 논의를 따르면, 다음과 같은 사실을 발견한다.*

$$\begin{aligned} E(\hat{\rho}_{X,Y}) &= E\left(\frac{\hat{\sigma}_{X,Y}}{\sqrt{\sigma^2_x}\sqrt{\sigma^2_Y}}\right) \\ &\neq \frac{E(\hat{\sigma}_{X,Y})}{\sqrt{E(\sigma^2_x)}\sqrt{E(\sigma^2_Y)}} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y} = \rho_{X,Y} \end{aligned} \tag{2.20}$$

---

* 본문의 증명은 공식적인 증명이 아니다. 단지 $\hat{\rho}_{X,Y}$가 편의를 가짐을 시사한다.

요약하면, $\rho_{X,Y}$ 는 편의추정량이지만 일치추정량이다. 이는 편의가 표본크기가 크다면 중요하지 않은 것으로 고려될 수 있다는 것을 의미한다. 왜냐하면, 일치성의 속성이 $\hat{\rho}_{X,Y}$ 가 $\rho_{X,Y}$ 에 근접할 높은 확률을 가진다는 점을 보증하기 때문이다. 뒤에서 보겠지만, 편의를 가지나 일치추정량인 문제는 계량경제학의 모형에서 상당히 잘 등장한다.

아. 自由度 (degree of freedom) 에 관한 노트

$\hat{\sigma}_{X,Y}$ 의 경우에서처럼, (2.17) 로서 정의된 분산의 추정량 $\hat{\sigma}^2_X$ 과 $\hat{\sigma}^2_Y$ 은 자신들의 "자유도"를 반영한 분모를 가지고 있음을 지적해야겠다. 앞에서처럼, $\hat{\sigma}^2_X$ 과 $\hat{\sigma}^2_Y$ 에 상응한 합계에는 n 항이 있지만, 오직 ($n-1$) 개의 독립적인 정보 조각만이 있다. X의 경우에 대하여 어느 정도 직관적으로 이 점을 보이면 다음과 같다.

$$\sum_{t=1}^{n}(X_t - \bar{X}) = 0$$

그래서 또한

$$(X_n - \bar{X}) = -(X_1 - \bar{X}) - \cdots - (X_{n-1} - \bar{X})$$

다른 말로 하면, 합계의 마지막 항은 완전히 ($n-1$) 항까지의 합에 의존하며, 따라서 아무런 새로운 정보도 가지지 않는다.

(관찰치의 수가 아니라) 자유도의 수로 나누는 것은, 부수적으로 한 변수의 불편추정량을 얻는 표준적인 과정이다. 다행히도 일반화를 시켜서 (이 책에서 고려되는 유형의) 분산 추정량의 자유도는 간단한 법칙에 의하여 획득할 수가 있다. 보다 엄밀히 말하자면, 그러한 추정량의 자유도는 일반적으로, n이 표본 크기이고 추정량의 분자를 평가하는 데에 추정해야만하는 모수의 수가 k 일때, ($n-k$) 가 된다. 예를 들어 분산 추정량, $\hat{\sigma}^2_X$

은 μ_x을 알지 못하기 때문에 그 분자에 $\overline{X}$를 가지고 있다. 이 경우에 $k=1$이다. 만약 μ_x가 알려져 있다면, X의 분산은 다음의 공식으로써 추정될 것임을 유의하도록 하자.

$$\sum_{t=1}^{n} \frac{(X_t - \mu_X)^2}{n}$$

자. 주의 사항

선형회귀모형에 들어가기 앞서서 상관계수의 해석에 관한 주의사항이 있다. 우리는 단지 두 변수 사이의 통계학적인 관련성만을 측정하는 것에 지나지 않는다는 점이다. 두 변수간의 인과관계에 대해서는 아무런 것도 말할 수 없다. 사실 두 변수 서로간에는 아무런 인과관계도 없으면서도 높은 상관을 보일 가능성이 아주 크다. 예를 들어, 미국에서의 교사들의 연간 봉급(달러)과 총 철강 산출량의 사이에 시간에 걸친 正의 상관을 의심의 여지없이 발견할 것이다. 아마도 이것은 한 변수의 다른 변수에 대한 어떤 종류의 직접적인 효과를 반영하지는 않고, 다만 대체로 다른 이유로 해서 이들 변수 둘다 시간에 걸쳐 양적으로 증가하는 사실의 결과에 지나지 않는다. 正 또는 負의 상관은(그것이 아무리 높은 상관을 갖더라도) 변수 사이의 인과적인 관련성이 존재한다는 것을 입증하지 않는다.

차. 實例

비록 두 변수 사이의 강한 상관이 존재하는 것은 어떠한 인과관계가 존재하는 것을 입증하지는 않지만, 그러한 상관은 그럼에도 불구하고 가설화된 관계를 실증적으로 지지할 만한 것을 제공한다. 예로써 많은 관찰자들이 중핵 도시가 갖는 병폐의 원천에 깔려 있는 것으로 주장하는 도시현

상을 들어 보자.* 그 주장은 우리의 대도시(metropolitan)의 교외화 (suburbanization)는 상대적으로 빈곤한 가구로 이루어진 잔여의 인구를 중핵 도시가 수용하도록 한다는 것이다. 이러한 현상은 재정상으로 비싼 댓가를 치르고 도시생활의 누적적인 부패과정을 발생시킨다는 것이다.

그러나, 이러한 주장을 지지하는 증거가 있는가? 이 점을 조명하기 위하여 우리는 이러한 과정의 관찰이 가능하고 측정이 가능한 결과가 무엇이 될 것으로 기대하는 것인지를 물을 것이고, 또한 실제로 그 결과가 이러한 기대와 일치한다면, 측정할 적합한 자료를 검사할 것이다. 이러한 경우에, 예를들어 우리는 교외화과정이 가장 많이 이루어짐으로써 어떤 점에서 교외와 비교하였을 때 상대적으로 큰 불이익에 처해 있는지를 알기위하여 미국의 상이한 도시들을 비교할 수가 있다. <표 2.2>가 이에 관한 일정한 자료를 제시하고 있는데, 이는 미국의 15개 대도시에서 얻은 것이다. 좀더 자세히 하면, P^c_i 를 i 번째 도시의 인구, P^s_i 를 i 번째 도시 주위 교외지역의 인구라고 하자. 그러면 15개 도시의 각각에 대하여 <표 2.2>는 도시 자체에 거주하는 전체 도시지역의 인구의 백분율, 즉 $100[P_i{}^c/(P_i{}^c+P_i{}^s)]$을 보이고 있다. $Y_i{}^c$와 $Y_i{}^s$는 각각 i 번째 도시의 평균 가구소득, i 번째 도시 교외의 평균 가구소득으로 한다. 그러면 15개 도시에 대하여 <표 2.2>는 도시-교외 소득 비율, $100(Y_i{}^c/Y_i{}^s)$도 보여 준다. 만일 "대탈출" (exodus)가설이 옳다면, 우리는 상대적으로 도시인구 부분이 적은 도시(상대적으로 $100[P_i{}^c/(P_i{}^c+P_i{}^s)]$의 값이 적은 도시)가 빈민이 지배적인 도시 인구를 수용하리라고 기대할 것

---

* 이 논점에 대하여 개진하고 있는 계량경제학적 연구는 다음을 보라. David Bradford and Harry Kelejian, "An Econometric Model of the Flight to the Suburbs", Journal of Political Economy, 81 (May-June 1973), pp.566-589.

〈표 2.2 〉

도 시	도시인구비율 (%)[a]	도시-교외 소득비율 (× 100)[b]
발 티 모 어	57	65
보 스 턴	24	69
시 카 고	50	73
클 레 브 랜 드	38	65
달 라 스	63	102
디 트 로 이 트	38	82
인디아나폴리스	91	107
로스앤젤레스	34	92
멤 피 스	94	104
뉴 욕	49	66
필 라 델 피 아	49	75
피 닉 스	67	103
샌 디 에 고	58	98
세 인 트 루 이 스	33	70
시 애 틀	45	81

a. (2)열은 전체 도시화된 지역의 인구 백분율로서 1970년 중핵 도시의 인구이다.

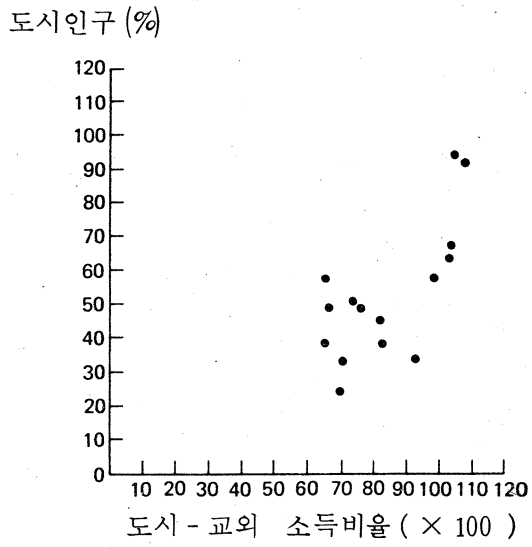
b. (3)열은 중핵 도시에서의 평균 가구소득을 중핵 도시 외부이지만 거대 도시지역에 속하는 곳의 평균 가계소득에 대비한 비율을 가리킨다 (Standford Metropolitan Statistical Area의 정의를 사용함).

출전: R.D.Norton, " City Life-Cycles and Municipal Expenditure Contrast", in Proceedings of the Seventieth Annual Conference on Taxation, National Tax Association-Tax Institute of America (Columbus, Ohio, 1978), p.328.

이다. 그러한 도시는 소득비율, $100(Y_i^c/Y_i^s)$가 상대적으로 낮을 것으로

예견된다. 요컨데, 변수 $100\,[\,P_i^c/(\,P_i^c+P_i^s\,)\,]$과 $100\,(\,Y_i^c/Y_i^s\,)$ 사이에는 正의 관련성이 기대될 것이다.

두 변수가 그러한 正의 관련성을 보이는지를 알 수 있는 자료를 검사하는 것이 확실한 절차이다. <그림 2.9>는 상당히 시사적인 산포도이다. 인



〈그림 2.9〉

구비율이 큰 값을 가질수록 평균적으로 상대적 소득 변수의 더 큰 값과 연관된 것으로 나타나 있다. 비록 그 연관의 유형이 "완벽한"것과는 거리가 있지만 말이다.

관련된 계산으로서의 설명은 이제 인구와 소득 비율 변수 사이의 상관을 추정하는 것이다. 표기를 간단히 하기 위해서, $X_i = 100[P_i^c/(P_i^c+P_i^s)]$ $Y_i = 100(Y_i^c/Y_i^s)$라고 하자. 그러면 (2.16)으로 부터 이들 변수간의 표본 상관은

$$\hat{\rho}_{X,Y} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})/n-1}{\sqrt{\sum (X_i - \bar{X})^2/n-1}\sqrt{\sum (Y_i - \bar{Y})^2/n-1}} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}} \quad (2.21)$$

1장 부록 A의 항등식〔자세히 말하면 (1A.10)과 (1A.13)〕을 약간 수정하여 (2.21)에 대입하는 데에 이용하면, 계산을 단순하게 할 수 있다.

$$\hat{\rho}_{X,Y} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum X_i^2 - n\bar{X}^2}\sqrt{\sum Y_i^2 - n\bar{Y}^2}} \quad (2.22)$$

실제 계산은 <표 2.3>과 같다. 두 변수간의 표본 상관은 0.71로 판명된다.

〈표 2.3 〉 표본 상관계수의 계산

	X_t	Y_t	$X_t Y_t$	X_t^2	Y_t^2
	57	65	3,705	3,249	4,225
	24	69	1,656	576	4,761
	50	73	3,650	2,500	5,329
	38	65	2,470	1,444	4,225
	63	102	6,426	3,969	10,404
	38	82	3,116	1,444	6,724
	91	107	9,737	8,281	11,449
	34	92	3,128	1,156	8,464
	94	104	9,776	8,836	10,816
	49	66	3,234	2,401	4,356
	49	75	3,675	2,401	5,625
	67	103	6,901	4,489	10,609
	58	98	5,684	3,364	9,604
	33	70	2,310	1,089	4,900
	45	81	3,645	2,025	6,561
총	790	1,252	69,113	47,224	108,052

$\bar{X} = 52.7$ $\bar{Y} = 83.5$

$$\hat{\rho}_{X.Y} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum X_i^2 - n\bar{X}^2}\sqrt{\sum Y_i^2 - n\bar{Y}^2}}$$

$$= \frac{69,113 - (15)(52.7)(83.5)}{\sqrt{47,224 - (15)(52.7)^2}\sqrt{108,052 - (15)(83.5)^2}}$$

$$= 0.71$$

그 결과는 확실히 "대탈출"가설과 일치한다.

2. 行態 관계(behavioral relationships)에 관한 서술

앞 절에서는 두 변수 사이의 통계학적 관련성을 측정하는 두가지 수단을 소개하였다. 이를 배경으로 하여서 우리의 관심을 끄는 기초적인 주제를 이제 진행하여 보도록 하자. 이는 다름아닌 가설화된 경제적 관계에 관한 상세한 기술과 추정이다. 이러한 목적을 달성하기 위해서 소비함수의 추정 문제로 되돌아 가자. 소비함수는 1장에서 간략하게 검토한 것이다.

경제이론은 소비지출 C가 가처분소득 Y_d의 함수임을 시사한다. 가구의 가처분소득수준이 높을수록, 그 가구의 소비에 대한 지출수준이 높아진다. 더 나아가 이러한 관계를 선형 함수식으로 한다고 가정하면,

$$C_t = a + bY_{dt} \tag{2.23}$$

여기서 C_t는 소비지출의 t번째 값이고 Y_{dt}는 대응하는 가처분소득의 값이다. 예를 들어, t는 시기를 언급하는 것으로서 (2.23)의 경우에서는 t시기의 소비지출을 t시기의 가처분소득수준에 연결시키고 있다. 다르게는 t가 일정 시점에서의 개인들을 언급할 수 있다. 이 경우에 (2.23)은 t번째 사람의 소비지출을 그의 소득과 관련시킬 것이다.

여기서 우리가 구체화하고 있는 것은 어느 제안된 인과관계라는 것을 먼저 주의해야 한다. 이론은 소비지출수준이 가처분소득에 종속한다고 한다.

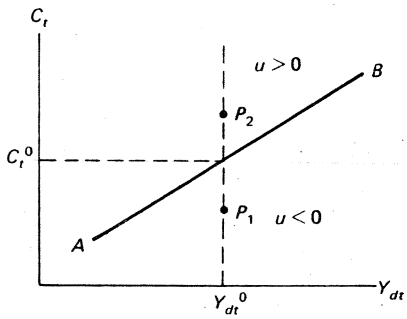
한 가구의 가처분소득이 증가함에 따라 그 가구는 증가된 소득의 일정 부분을 지출할 것으로 기대된다. 둘째로, (2.23)식은 C_t와 Y_{dt}의 정확한 관계를 분명히 하고 있다. 예를 들어 $a=100$이고 $b=0.9$라고 하면, (2.23)식은 $Y_{dt}=15{,}000$달러일 때 C_t가 정확히 13,600달러라고 보여 준다. 그러나 강조한 바 있듯이 자료를 볼 경우는 정확한 관계가 보이지 않는다. 전형적으로는 직선상에 정확하게 놓여있는 점들의 집단이 발견되는 것이 아니라, 그 대신에 산포도를 본다. 우리가 추정하고자 하는 소비와 가처분소득 사이의 관계는 정확한 것이 아니라 오히려 전형적인 (typical)관계이다. 유의해야 할 것은 다음과 같은 종류의 말이다. 즉, 가구의 가처분소득이 15,000달러이면, 평균적으로 그 가구의 소비에 관한 지출은 13,600달러가 될 것이다. 어느 특정한 경우에서나, 이제 우리가 논의할 많은 이유로 인하여 C_t가 그 전형적인 값으로부터 양 또는 음의 방향에서의 차이가 변한다. 이는 곧 우리의 단순한 소비함수를 다음과 같이 적어야함을 시사한다.

$$C_t = a + bY_{dt} + u_t \tag{2.24}$$

여기서 μ_t는 양이나 음의 값을 가지나 그 평균값은 0이어서 Y_{dt}로 주어진 값에 대응하는 C_t의 평균값은 $(a+bY_{dt})$이다. 예를 들어, C_t와 Y_{dt}사이의 평균 관계가 <그림 2.10>의 AB선으로 나타난다고 가정하자. 만약 $Y_{dt}=Y_{dt}°$이면, C_t의 평균은 $C_t°$가 될 것이다. 그러나 일반적으로 $Y_{dt}=Y_{dt}°$일지라도 C_t의 값은 어느 정도 $C_t°$와 괴리가 있다. 어떤 경우에는, P_1점과 같이 $C_t<C_t°$이고, 이는 $\mu_t<0$을 의미한다. 다른 경우로서 <그림 2.10>의 P_2와 같이 $C_t>C_t°$이면 $\mu_t>0$이다.

μ_t항은 계량경제학에서 매우 중요한 것으로 교란항[disturbance term :또는 오차항(error term)]이라고 불린다. 이 교란항은 경제관계가

정확한 것이 아니라 오히려 평균 행태의 유형을 표시하는 것임을 지적하는 방식이다. 이것은 우리가 왜 경제학에서 예외없는 정확한 관계를 발견하지 못하는가란 논점을 제기한다.* 왜 전형적으로 μ_i는 0과 차이가 있



〈그림 2.10〉

는가? 왜 이것이 진실인지에 관한 이유는 많이 있다.**

① 제외한 변수들

다시 소비자 지출의 보기를 고려하여 보자. 소비는 의심할 여지없이 가

---

* 경제학에는 우연하게 정확한 관계가 있지만, 이는 경제이론이 제시하는 행태의 관계가 아니다. 정확한 관계는 "회계 항등식"으로 알려진 것으로서, 정의에 의하여 참이 된다. 예를 들어, 경제학에서 잘 알려진 항등식은 대차대조표의 근본적인 것으로서 아래와 같다.

$$\text{자산} = \text{부채} + \text{자본}$$

이 관계는 언제나 정확하게 유지된다. 왜냐하면 자본이 정의되는 방식이 아래와 같기 때문이다.

$$\text{자본} = \text{자산} - \text{부채}$$

자본은 대차대조표의 균형을 유지하는 나머지 금액으로 정의되는 것이다.

** 이하의 논의는 J.Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), pp.10-11에서 옮긴 것이다.

처분소득이외의 변수 이외에 추가로 다른 많은 변수의 영향을 받는다. 예를 들어 C_t가 t시점의 소비지출이라고 한다면, 다음과 같이 가정할 수가 있을 것이다.

$$C_t = a + b_1 Y_{dt} + b_2 L_t + b_3 \dot{P}_t + b_4 R_t + \cdots \tag{2.25}$$

여기서 L_t는 t시점의 유동자산스톡(stock), $\dot{P}_t$는 t시기 동안에 일어난 가격의 변화율(%), R_t는 t시기중의 이자율 등과 같다.

다른 한편으로, 적어도 정상적인 상황에서는 Y_{dt}가 단연 C_t의 가장 중요한 결정변수이고, 다른 변수들의 효과는 매우 작아서 시간에 걸쳐서 말소되는 것으로 느껴질 것이다. 이러한 경우에 교란항은 이들 모든 생략된 항의 합을 나타낸다.*

$$u_t = (b_2 L_t + b_3 \dot{P}_t + b_4 R_t + \cdots) \tag{2.26}$$

요약하면, 교란항은 모든 적합한 요소가 평가되지 않기 때문에 제기될 것이다.

② 사람들의 예측하지 못할 행위

사람들, 특히 개인의 행태 유형은 완벽하게 예측할 수가 없다. 다른 한편으로 행위는 일반적으로 순수한 임의적 성격도 아니다. 이런 측면에서 (2.24)의 소비모형은 두가지 성분을 결합한 것으로 간주할 수 있다. 즉, 지출을 소득($a + bY_{dt}$)과 관계시키는 결정적인 성분과 예측 불가능한 (또는 "자유 의지") 성분 u_t이다. 이러한 틀안에서 교란항은 소비자가 전형적으로 소비하는 것보다 많거나 적게 하는 "돌발적인 충동", "생

* 이들 생략된 항이 "말소"되지 않는 경향을 가지는 경우에 대해서는 뒤에서 고려하도록 할 것이다.

각의 변화", 또는 다른 태도의 변화를 반영하거나 설명한다. 이러한 전형적인 양은 결정적인 성분, ($a+bY_{dt}$)에 의하여 주어진다.

관련된 분석틀에서, 교란항은 경제행위에 예측불가능한 사건이 미치는 효과를 반영하는 것으로도 생각될 수 있다. 예를 들어, 그 도시 외부로부터 예상하지 못한 친구가 한 명 찾아 왔다면, 주인은 그 친구의 저녁식사를 외식으로 해결함으로써 자신의 정상적인 예산을 초과할 것이다.

③ 개인들간의 다양한 행위

서로 다른 행위 때문에 특정한 가구들이 다른 가구들보다 더 높은 저축성향을 가진다는 것은 잘 알고 있다. 만일 어떤 시점에 다양한 가구의 지출을 설명하기 위하여 (2.24)를 사용한다면, 우리는 ($a+bY_{dt}$)를 가처분소득이 Y_{dt} 인 어느 가구의 전형적인 지출을 나타내는 것으로 취하고, 교란항 u_t를 그러한 평균과의 편차를 나타내는 것으로 할 수 있다. 즉, u_t는 저축에 대한 다양한 태도를 반영하는 것이다. 상대적으로 높은 저축성향을 가진 가구들은 음의 u_t 값과 대응하고, 낮은 저축성향을 가진 가구들은 양의 u_t 값과 연관될 것이다.

④ 측정의 오차

비록 $C_t = a + bY_{dt}$가 정확하더라도, C_t를 완전히 정확하게 측정할 수가 없을 것이다. 그러한 측정의 오차에 따른 결과로서, 실제로는 C_t와 다음과 같이 관련된 $\tilde{C}_t$를 관찰한 것이다.

$$\tilde{C}_t = C_t + u_t$$

여기서 u_t는 측정의 오차를 대표한다. 이러한 정식화에서, u_t가 양인지 음인지에 따라서 C_t를 과소 내지는 과대 평가하는 반면에, 만일 C_t의 측정을 반복해서 한다면, 그 측정치의 평균이 C_t가 될 것으로 기대된다

는 의미에서 오차가 말소되는 경향이 있음을 가정한다. 만일 $\tilde{C}_t = C_t + u_t$를 $C_t = a + bY_{dt}$에 대입하면, 다음을 얻는다.

$$\tilde{C}_t = a + bY_{dt} + u_t$$

다른 말로 하면, 소비지출과 가처분소득 사이의 관계가 정확하더라도, 측정된 지출과 소득 사이의 관계는 교란항을 엮어 넣는다.*

그러므로 평균 관계를 함수적인 관계로 특징화하고, 이것을 교란항 u_t를 포함하여 모형에서 지시할 것이다. 이러한 경제적 관계에 관하여 추가적으로 두가지 예를 들면 다음과 같다.

$$I_t = e + fR_t + u_t \tag{2.27}$$

과

$$Q_t = g + hL_t + u_t \tag{2.28}$$

여기서 I=투자, R=이자율, Q=산출량 수준, L= 총노동투입량을 나타낸다. 각각의 식에서 예를 들어 종속변수에 영향을 끼치는 추가적으로 중요한 변수들이 있음이 분명하다. 투자는 이자율과 마찬가지로 산출물에 대한 수요와 같은 여타 요인에 분명히 의존한다. 두번째 사례에도 노동과 같이 사용되는 여타 투입물들의 양에 따라서 산출량의 수준은 변동한다. 이러한 이유만으로도 변수들간의 관계에 있어서 일정한 부정확함이 기대된다.

3. 二變數回歸模型

좀전에 논의한 유형의 행태 관계를 뚜렷하게 지정하였다고 가정하자. 더

---

* 설명을 위해서 독립변수의 측정에서는 아무런 오차가 없다고 가정하자.

일반적인 용어로서 그 형식의 선형 관계를 다음과 같다고 한다.

$$Y_t = a + bX_t + u_t \quad , \quad t = 1, \ldots, n \tag{2.29}$$

여기서

Y_t =종속변수의 t 번째 관찰치

X_t =독립변수의 t 번째 관찰치

u_t =대응하는 교란항의 t 번째 값

그리고 a와 b는 미지의 모수

(2.29)식은 두개의 미지모수 a와 b를 갖는, Y_t, X_t와 u_t사이의 선형 관계이다. 이 관계가 모든 특정화된 t값, 즉 t = 1, 2, ……, n에서 유지된다고 가정한다. 이 t값들은 Y_t와 X_t에 관한 n개 관찰치에 대응하는 것이다. 교란항 u_t의 값은 "전형적인" 행위의 유형으로부터 벌어지는 편차를 반영하여 상이한 관찰치 간에 변동할 것으로 예견할 수 있다. Y_t, X_t와는 달리 교란항의 값들은 관찰 가능하다고 가정하지 않는다. 말이 나온 김에, (2.29)식에서 종속변수 Y_t와 독립변수 X_t는 때로 각각 피설명변수(regressand), 설명변수(regressor)라고도 언급된다.

우리의 첫번째 문제는 모수 a와 b의 값을 추정하여서 Y_t와 X_t사이의 관계를 양적으로 밝히는 것이다. 이를 위하여 독립변수 X_t와 교란항 u_t가 발생하는 방식에 관한 일종의 공식적인 가정을 맨 먼저 하여야만 한다. 실제로, 그러한 가정의 필요성, 특히 u_t에 관한 가정의 필요성은 명백하다. 예를 들어 Y_t는 X_t와 u_t 둘다에 종속하기 때문에, Y_t와 X_t사이의 어떠한 관계가 갖는 본성은 교란항 u_t의 특성에 의존함이 틀림없다.

가. 기본적인 가정

(1) 첫째로, 종속변수 X의 모든 값이 다 동일하지 않다. 적어도 X 값중의 하나는 다른 값과 달라야만 한다. 앞으로 보게 되겠지만, 이 조건이 만족되지 않으면 a와 b를 추정할 수가 없게 된다. 어느 정도 직관적으로 보아도 만약 X가 전혀 변하지 않는다면, 어떻게 Y가 X에 따라 변하는지를 관찰할 수가 없게 된다.

이러한 가정은 무엇이 X의 특정한 값들을 결정하는가라는 논점을 제기하는 것이다. 고전적인 가정은 실험자 자신이 X의 값들을 고르고서 Y의 대응하는 값 또는 결과된 값을 관찰하는 식이다. 예를 들어 (2.29) 식이 에이커당 부셀 (bushel) 로 측정된 옥수수 산출량 Y와 에이커당 파운드로 측정된 비료의 투입량 사이의 관계를 나타내는 것으로 하자. 이러한 관계를 조사하기 위해서 실험자는 첫번째 에이커의 토지를 $X=1$, 두번째 에이커를 $X=2$ 식으로 놓을 것이다 (예컨데, $X_1=1$, $X_2=2$).

경제학자들이 전형적으로 이상과 같이 다행스런 상황에 있지 않다는 것은 분명하다. 예를 들어 인플레이션율과 실업률의 관계에 대하여 조사하기를 원한다고 하자. 이 경우에 (2.29) 의 Y를 매년도의 물가상승률(%)로 정의하고 X를 그 해의 실업상태에 있는 노동력으로 정의할 수가 있다. 매년 실업률을 다르게 하여서 그 결과로 나타나는 인플레이션율을 관찰함으로써 Y와 X의 관계에 대한 조사를 진행할 수가 없음이 분명하다. 실업률은 전체 경제의 작용이 결정하는 것이고, 그러한 이유로 해서 실험적인 통제를 벗어나기 때문에 독립변수의 값을 고르고 체계적으로 변동시킬 수가 없다. 이러한 경우에는 X와 Y를 둘다 관찰해야만 한다. 그러므로 X의 값들을 발생시키는 기구가 무엇이든지 그 기구는 적어도 두개의 상이한 X값을 낳는다는 가정을 하고서 논의를 진행하기로 한다.

(2) 교란항 그 자체의 속성에 관한 일련의 세 가정을 다음과 같이 할 것이다.

2a. $E(u_t) = \mu_u = 0$

2b. $E(u_t - \mu_u)^2 = E(u_t^2) = \sigma_u^2$

2c. $s \neq t$일 때 u_t는 u_s와 독립적이어서 $E[(u_t - \mu_u)(u_s - \mu_u)] = \mathrm{cov}(u_t, u_s) = 0$

가정 2a는 모든 관찰치에 대해서 교란항의 기대값이 0임을 말한다. 간단한 설명으로서 모든 각각의 관찰치에 대해서 u_t의 값은 다음과 같이 결정된다. 우리에게 알려지지 않은 한 사람이 동전을 던져 올리고 있다. 만일 앞면이 나오면, 그는 $u_t = 1$이라 놓으나, 뒷면이 출현하면 $u_t = -1$로 놓는다. 이 경우에

$$E(u_t) = \tfrac{1}{2}(1) + \tfrac{1}{2}(-1) = 0$$

교란항은 평균값이 0이라고 가정하는 근본적인 이유는 아주 간단하다. (2.29)식에 구체화되어 있는 우리의 이론은 다양한 X값에 대응하는 Y의 평균 행태를 정확하게 묘사하고 있다고 가정한다. 즉, X값이 어떠하더라도 Y의 평균값은 $Y^m = (a + bX)$가 될 것으로 가정하는 것이다. 그러나 만약 $E(u_t) \neq 0$이라면 그렇지 못하게 된다. 예를 들어 (2.29)식에서 u_t가 일관적으로 양이었을 경우를 가정하자. 이는 X의 각 값에 대응하는 Y의 평균값이 $(a + bX)$를 능가함을 의미하여, 명백히 Y의 평균값이 $(a + bX)$라는 우리의 가정과 불일치하게 될 것이다.

대안적으로 (2.29)식은 교란항이 0의 평균을 갖지 않는 또 다른 식으로부터 도출될 것으로 간주할 수가 있다. 예를 들면, $E(u_t) = d \neq 0$이고 d가 상수라고 가정하자. 그러면 다음과 같이 정의할 수 있다.

$$v_t = u_t - d \tag{2.30}$$

또한 $E(v_t) = E(u_t) - d = d - d = 0$이 된다. (2.30)을 사용해서 $u_t = v_t + d$를 (2.29)에 대입하여 다음을 얻는다.

$$\begin{aligned} Y_t &= (a + d) + bX_t + v_t \\ &= a^* + bX_t + v_t \end{aligned} \tag{2.31}$$

여기서 $a^* = a + d$이고, 이미 지적하였듯이 $E(v_t) = 0$이다. 그러므로 (2.31)를 회귀모형으로 취할 수 있었다.

가정 $2b$는 교란항의 분산이 σ_u^2과 같은 상수이어서 t에 따라 체계적으로 변하지 않는다는 것을 말한다. 예를 들어 만일 소비지출과 가처분소득의 시간에 걸친 일련의 관찰치를 가지고 분석하고 있다면, 가정 $2b$는 교란항의 분산이 시간의 흐름에 따라서 더 커지거나 더 작아지게 되지 않는다는 것을 의미하는 것이다.* 또는 앞서의 단순한 보기에 따라서, 우리가 모르는 한 사람이 동전을 던져 올려서 앞면이나 뒷면이 나올 경우에 각각 $u_t = +t$ 또는 $u_t = -t$라고 놓는다고 하여도 가정에 위배되는 것이 발생한다. 이 경우에 u_t의 기대값은 여전히 0이지만,

$$E(u_t) = \tfrac{1}{2}(t) + \tfrac{1}{2}(-t) = 0$$

u_t의 분산은 각각의 연속적인 관찰치에 따라 더 커지게 되는 것이 분명하다.

이러한 가정의 근본적인 이유는 그 가정이 위배될 때 발생하는 추정문

---

* 대신에 만약 상이한 수준의 가처분소득을 갖는 가구들의 소비지출을 가리키는 예산 조사에 기초하고 있다면, 이러한 가정은 교란항의 분산 σ_u^2이 가처분소득의 수준에 따라 체계적으로 변하지 않는다는 것을 의미한다. 이 점에 대해서는 6장에서 더 보기로 한다.

제를 처리하는 기법과 함께 이 책의 뒤에서 공식적으로 전개되고 있다(6장을 보라). 지금 여기서는 만약 u_t의 분산이 우리의 모든 관찰치에 대하여 동일하지 않다면, 같은 의미로 모든 관찰치가 동등하게 믿을 만한 것이 되지 못함을 밝혀 둔다. 예를 들어, u_t의 분산이 두 특정한 관찰치에 대하여 0이고, 나머지 관찰치에 대해서는 어떤 양수라고 알고 있다고 가정하자. u_t의 평균은 0이기 때문에 이 두 관찰치는 1과 같은 확률로 $Y = a + bX$식을 만족시킨다. 따라서,

$$\begin{aligned} Y_1 &= a + bX_1 \\ Y_2 &= a + bX_2 \end{aligned} \tag{2.32}$$

이어서, (2.32)의 두 식을 풀어 a와 b를 구한다. 다른 말로 하면, 다른 모든 관찰치를 제외하고서 단지 처음 두 점만을 사용하여 a와 b를 추정할 수가 있다. 요약하면, 어떤 의미에서는 작은 분산에 대응하는 관찰치가 큰 분산에 대응하는 관찰치보다 더 귀중하다는 것이다. 현재의 단계에서는 모든 관찰치가 동일하게 중요한 것으로 보고자 하여 $2b$의 가정을 한 것이다.

가정 $2c$는 t번째 교란항의 값이 어느 다른 여타 변수(말하자면 s번째)의 값과는 독립적임을 말한다. 이러한 가정을 하는 이유는 우리가 현재 종속변수 Y_t에 작용하는 단 하나만의 체계적 또는 예측가능한 힘(즉, X_t)이 있는 모형을 구체화하기 위해서이다. 만약 교란항이 상호간에 관련되어 있다면, 이러한 경우가 되지 못함이 분명하다. 예를 들어, u_t가 바로 이전의 값, u_{t-1}과 負의 상관을 갖는다고 가정하자. 그러면 Y_t의 값은 X_t와 또한 u_{t-1}의 값에 체계적이면서 예측가능하게 종속된다. 왜냐하면, 적어도 부분적으로는 u_{t-1}이 u_t의 값을 결정하기 때문이다. 이러한 모형을 6장에서 고려할 것이지만, 우리의 논의는 한 관찰치에 대한 교란

항의 값이 다른 여타 관찰치에 대한 교란항의 값에 종속되지 않는다고 가정함으로써 보다 단순한 수준의 회귀분석부터 출발하기로 한다.

이상의 가정들을 가지고서 (2.29)식의 교란항을 평균값 0, 일정한 분산 σ_u^2 를 갖는 관찰 불가능한 확률변수로서 특성지을 수 있다. 또한 그것은 주어진 어떠한 경우에서라도 또 다른 경우의 값과 독립적이고 따라서 상관이 없는 값을 갖는 속성을 가진다.

(3) 우리의 마지막 가정은 u_t가 설명변수 X의 모든 n개 값과 독립적이라는 것이다. 따라서 cov (u_t, X_t) = 0 이다. 2a의 가정에 의하여 $E(u_t) = 0$이므로,

$$\begin{aligned}\text{cov}(u_t, X_t) &= E[(u_t - 0)(X_t - \mu_X)] = E(u_tX_t) - E(u_t\mu_X) \\ &= E(u_tX_t) - \mu_X E(u_t) = E(u_tX_t) = 0\end{aligned}$$

그러므로 그 가정은 $E(u_tX_t) = 0$임을 의미한다.

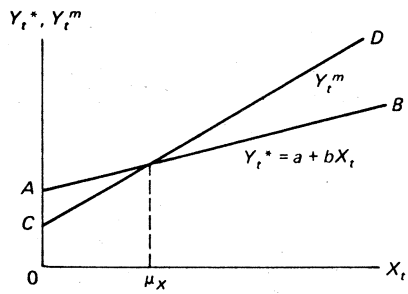
X_t와 u_t가 독립적이라는 가정이 필요한 근본적인 이유는,* 또한 cov (u_t, X_t) = 0 인 가정의 이유는 $E(u_t) = 0$이란 가정이 필요한 것과 유사하다. 만일 이들 가정의 어느 하나가 성립하지 않는다면, 특정한 X값에 대응하는 Y_t의 평균값, Y_t^m은 일반적으로 더 이상 ($a + bX$)가 되지 않는다. 예를 들어, X_t와 u_t가 正의 상관을 갖는 경우를 고려하여 보자. 正의 상관은 평균값보다 더 큰 u_t의 값〔$E(u_t) = 0$이므로 양이다〕이 평균값보다 더 큰 X값과 연관이 되는 경향이 있음을 의미한다.

---

* u_t가 X의 모든 n개 값과 독립적이라는 가정이 갖는 중요성은 기술적인 것으로서 뒤에서 자세히 보기로 한다.

유사하게, 평균값보다 작은 u_t의 값(음)은 평균값보다 작은 X값에 연관되는 경향이 있다. 이는 X_t의 큰 값들에만 대응하는 u_t의 평균이 양임을 의미하고, 역으로 X_t의 작은 값들에만 대응하는 u_t의 평균은 음이 될 것임을 시사한다. 이로부터 Y_t의 평균, Y_t^m은 X_t가 클 경우 ($a+bX_t$)를 초과하고 X_t가 작을 때는 ($a+bX_t$)에 못미친다는 것을 알게 된다.

이 문제를 <그림 2.11>로 설명하여 보겠다. AB선을 $Y_t^* = a + bX_t$ 를 그린 것으로 하자. 비슷하게 CD선을 X_t의 다양한 값에 대응하는 Y_t의 평균값, Y_t^m을 나타내는 것으로 하자. 그러면 교란항과 설명변수 사이의 正의 상관은 Y_t^m과 Y_t^* 사이의 관계가 <그림 2.11> 같이 나타나게 된다.*



〈그림 2.11〉

이것은 우리의 가정에 관한 논의를 완성시킨다. 다음과 같은 형식의 X와 Y 사이의 관계에 대한 특성화는

$$Y_t = a + bX_t + u_t$$

---

* 이론상 CD가 반드시 직선이 되지 않음에 대하여 약간의 양해를 구하고자 한다.

이제까지 논의한 가정들과 함께 우리의 기본적인 선형회귀모형을 구성한다. 다음으로 할 일은 우리의 가정을 가지고서 어떻게 a와 b의 추정치를 얻는 데에 사용할 것인가를 보는 것이다. 논의를 진행하는 가운데 각 가정의 정확한 기능이 무엇이며, 우리의 결과가 그 가정들에 어떻게 의존하는가를 더 명확하게 알 수 있게 된다.

4. 회귀식의 추정 : 대변수기법

우리가 사용하는 접근방식은 문헌상에 대변수(instrumental-variable) 추정으로 나와있는 데, 그 기법은 우리의 가정들을 기본적인 회귀모형부터 직접 X와 Y의 관찰된 표본값들에 부과하는 것이다.* 간략히 보겠지만, 이 기법은 a와 b의 추정량을 얻게 하여 준다. 이 접근이 갖는 매력은 우리가 회귀모형에서 정하는 가정들 각각의 특정한 중요성 또는 역할을 분명하게 보여줄 수 있는 점이다.

다시 기본적인 회귀식을 고려하여 보자.

$$Y_t = a + bX_t + u_t \ , \quad t = 1, \ldots, n \tag{2.29}$$

$E(u_t) = 0$이라는 가정때문에, X_t의 주어진 값에 대응하는 Y_t의 평균값은,

$$Y_t^m = a + bX_t \tag{2.33}$$

(2.33)식은 Y_t와 X_t 사이의 평균 관계로서 해석될 수 있을 것이다. (2.29)와 (2.33)으로부터,

---

* 우리가 사용하게 될 대변수 기법이 갖는 특정한 형식은 Arthur S. Goldberger, Topics in Regression Analysis (New York : Macmillan, 1968)에서 개발된 것이다.

$$Y_t = Y_t^m + u_t \tag{2.34}$$

(2.34)식은 Y_t를 그 평균 성분과 그 평균으로부터 편차를 갖게 한 항의 합으로 표현할 수 있음을 단지 말하고 있을 뿐이다. (2.34)를 다시 정리하면, 교란항을 다음과 같이 표현할 수가 있다.

$$u_t = Y_t - Y_t^m \tag{2.35}$$

이제 a와 b의 추정량, $\hat{a}$과 $\hat{b}$를 가졌다고 가정하자. (2.33)에 비추어 보면, Y_t의 평균값에 관한 추정량은,

$$\hat{Y}_t = \hat{a} + \hat{b}X_t \tag{2.36}$$

여기서 Y_t^m의 추정량에서 윗첨자를 생략함으로써 표기를 간단하게 하여 두었다. 비슷한 방식으로 (2.35)에 의하여 제시되었던 교란항의 추정량은 다음과 같이 될 것이다.

$$\hat{u}_t = Y_t - \hat{Y}_t \tag{2.37}$$

식(2.38)에서 미지모수, 즉 $\hat{a}$와 $\hat{b}$를 a와 b의 추정량으로 대체함으로써 교란항의 추정량을 얻을 수가 있다. 다시 (2.37)의 항들을 재정리하면 (2.34)에 대응하는 식을 산출할 수 있다.

$$\begin{aligned} Y_t &= \hat{Y}_t + \hat{u}_t \\ &= \hat{a} + \hat{b}X_t + \hat{u}_t \end{aligned} \tag{2.38}$$

(2.38)은 Y_t의 값을 a, b와 u_t의 추정량 (즉, $\hat{a}$, $\hat{b}$와 $\hat{u}_t$) 그리고 X_t의 값으로 표현한 것이다.

이제 다시 $\hat{a}$과 $\hat{b}$을 얻는 문제로 돌아가기로 하자. 회귀모형이 갖는 가

정중의 하나는 u_t가 0의 평균, $E(u_t)=0$을 갖는다는 것이다. 이는 어느 정도 직관적으로 다음의 사실을 알 수 있게 하여 준다. 즉, n개의 u_t값을 평균하면($\bar{u}=\sum u_t/n$), 그 평균은 "작은" 값을 가질 것이다. 더 공식적으로는 $E(u_t)=0$이 $E(\bar{u})=0$을 의미한다고 할 수 있다. 이 모든 사항은 만약 $\hat{u}_t$를 (2.37)로 정의하면, 다음과 같이 되는 것이 바람직할 것으로 보인다.

$$\left(\sum_{t=1}^{n}\frac{\hat{u}_t}{n}\right)=0 \tag{2.39}$$

또는 양변에 n을 곱해서

$$\sum_{t=1}^{n}\hat{u}_t=0 \tag{2.40}$$

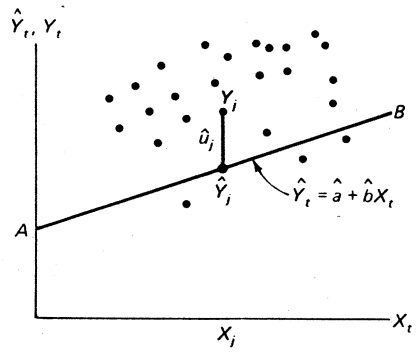
추정량 $\hat{u}_t$가 (2.39) 또는 (2.40)의 속성을 갖는 것이 좋을 것이다. 그 속성은 오차항에 관한 기본 가정중의 하나, 즉 $E(u_t)=0$과 대응하는 것이자 이 가정이 시사하는 것이기도 하다.

내용을 더 나아가기 전에 (2.40)의 중요성을 기하학적으로 해석하는 것이 도움을 줄 것이다. (2.38)로부터 만일, $\sum\hat{u}_t\neq 0$이면, $\sum Y_t\neq\sum\hat{Y}_t$임을 알았다. 예증을 위하여 $\sum\hat{u}_t=500$이라고 가정하자. 그러면 $\sum Y_t>\sum\hat{Y}_t$가 된다. 이제 <그림 2.12>를 보면, Y_t와 X_t에 관한 n개 관찰치는 흩어져 있는 점들로 나타나 있고, Y_t와 X_t 사이의 추정식은, 즉

$$\hat{Y}_t=\hat{a}+\hat{b}X_t$$

는 AB선이다. 점들은 일반적으로 AB선 위에 있다. 그 이유는 설명변수의 한 주어진 값, X_j에 대응하는 선의 높이가 $\hat{Y}_j$이기 때문이다. 그러나 이 높이는 대응하는 종속변수 Y_j보다 일반적으로 낮다. 왜냐하면 $\sum Y_t>$

$\sum \hat{Y}_t$ 이기 때문이다. 만일 $\sum \hat{u}_t$ 가 음이라면, 비슷한 논의로써 흩어져 있는 점들이 일반적으로 Y_t 와 X_t 사이에 추정된 관계 밑에 놓여 있을 것임이 분명하다. 따라서 $\sum \hat{u}_t = 0$ 이라는 (2.40)의 조건은 결국 점들이 추정한 선의 위나 아래에 없다는 것을 의미한다.



〈그림 2.12〉

(2.40)이 추정량 $\hat{a}$와 $\hat{b}$를 얻는 데 무슨 역할을 하는지가 이제 명백할 것이다. n개의 관찰치를 갖는 (2.38)을 모두 더하면,

$$\begin{aligned} \sum Y_t &= \sum \hat{Y}_t + \sum \hat{u}_t \\ &= n\hat{a} + \hat{b}\sum X_t \end{aligned} \tag{2.41}$$

왜냐하면 (2.40)에 의해서 $\sum \hat{u}_t = 0$ 이기 때문이다. (2.41)를 n으로 나누면,

$$\overline{Y} = \hat{a} + \hat{b}\overline{X} \tag{2.42}$$

여기서 $\overline{Y}$와 $\overline{X}$는 Y와 X의 표본 평균이다. $\overline{Y}$와 $\overline{X}$는 우리의 표본에서 알 수 있을 것이기 때문에, 두개의 미지수, 즉 $\hat{a}$과 $\hat{b}$가 남게 된다. 계

량경제학의 기술적인 전문 용어로는 (2.42) 〔또는 (2.41)〕를 "정규방정식" (normal equation)이라고 한다.

이 식의 해석을 진행하고 두번째 식을 전개하기에 앞서서, 정규방정식으로부터 무언가 더 직관적인 파생물을 끌어 내어 가지고 분석한다면 도움이 될 것이다. 다시 기본적인 회귀모형으로 되돌아 가자.

$$Y_t = a + bX_t + u_t$$

만약 X와 Y에 관해서 관찰한 n개의 모든 값을 윗식의 좌변과 우변에 걸쳐서 더하고, 다시 양변을 n으로 나누어 주면,

$$\frac{\sum Y_t}{n} = \frac{\sum a}{n} + \frac{\sum bX_t}{n} + \frac{\sum u_t}{n} \tag{2.43}$$

이는 다음과 같이 단순화된다.

$$\bar{Y} = a + b\bar{X} + \frac{\sum u_t}{n} \tag{2.44}$$

우리의 회귀모형에서 가정 $2a$로부터 다음을 알고 있다.

$$E(u_t) = 0$$

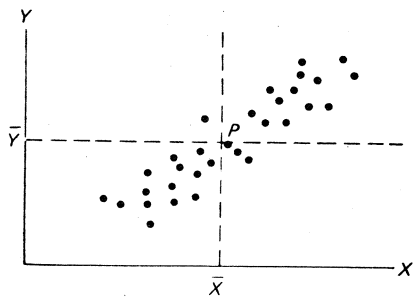
이는 (2.44)의 마지막 항의 기대값이 0이 될 것임을 의미한다. 그렇다고 $\sum u_t/n$이 0이 될 것이라고 하지는 않았다. 일반적으로 $\sum u_t/n$는 정확하게 0이 되지 않는다. 비록 표본 크기가 더 커짐에 따라 일정하게 주어진 한도로 편차가 생길 확률은 감소하지만, 0은 되지 않을 것이다.

대변수 기법은 본질적으로 (2.44)의 $\sum u_t/n$항을 무시하는 것과 결과적으로 같다. 왜냐하면, 기대값은 0이기 때문이다. 만약 이 항을 무시한다면 (즉, 0이라고 가정하면),

$$\bar{Y} = \hat{a} + \hat{b}\bar{X} \qquad (2.45)$$

윗식은 (2.42)와 동일하다. (2.44)에서 (2.45)로 가면서 a와 b가 추정량 $\hat{a}$과 $\hat{b}$로 대체된 것을 세심하게 보도록 하자. 그 이유는 정규방정식 (2.45)가 표현한 관계가 단지 $\sum u_i/n = 0$일 경우에만 (2.44)와 일치하기 때문이다. 오로지 이 조건이 지켜져야만 $\hat{a} = a$와 $\hat{b} = b$가 된다. 그러나 일반적으로 $\sum u_i/n$은 0이 정확하게 되지 않기 때문에, $\hat{a}$와 $\hat{b}$는 단지 a와 b의 추정량을 나타낼 뿐이다.

다음으로 이러한 정규방정식에서 볼 것은 정규방정식이 우리가 추정한 X와 Y의 관계에 대하여 무엇을 말하여 주는가이다. 정규방정식은 흩어져 있는 점들에 맞춘 선이 두 변수간의 표본 평균값을 통합한 점을 지나야만 한다는 것을 말한다. <그림 2.13>에서는 정규방정식은 점 P가 그 선 위에 있을 것을 지적하여 준다.



〈그림 2.13〉

이제 $E(u_i) = 0$라는 가정의 역할을 분명하게 알 수 있다. 이 가정은 우리가 흩어져 있는 점들을 맞출 선위에 한 점〔즉, $P(\bar{X}, \bar{Y})$〕을 위치할 수 있도록 한다. 두 점이 한 직선을 만들기 때문에, 만약 한 점

만 더 발견할 수가 있다면, 그 선을 나타내는 식을 세울 수 있고, X와 Y 사이에서 추정한 관계를 얻을 것이다. 이를 위해서 또 다른 가정을 이용하여야만 한다.

회귀모형에서 가정 (3)으로서 교란항 u_t가 X와 독립적이라고 하여서 그 결과 $\text{cov}(u_t, X_t) = 0$이었다. 또한, 이것이 다음을 의미한다고 보였다.

$$E(u_t X_t) = 0$$

이것은 직관적인 수준에서, 만일 u_t와 X_t의 관찰치로 이루어진 표본이 있다고 하면, 그 추정된 공분산,

$$\hat{\sigma}_{X,u} = \frac{\sum (u_t X_t)}{n}$$

는 거의 0과 같을 것이다. 왜냐하면, $E(u_t X_t) = 0$은 $E(\hat{\sigma}_{xu}) = 0$임을 의미하기 때문이다. 이는 우리가 $\hat{u}_t$에 부과할 수 있는 2차조건이

$$\frac{\sum (\hat{u}_t X_t)}{n} = 0 \tag{2.46}$$

또는 양변에 n을 곱하여

$$\sum (\hat{u}_t X_t) = 0 \tag{2.47}$$

다시 (2.38)식으로 돌아 가기로 하자.

$$Y_t = \hat{a} + \hat{b} X_t + \hat{u}_t \tag{2.38}$$

(2.38)의 양변을 X_t로 곱하면,

$$X_t Y_t = \hat{a} X_t + \hat{b} X_t^2 + \hat{u}_t X_t \tag{2.48}$$

X와 Y의 n개 관찰치를 (2.48)의 양변에서 모두 더하고서 n으로 나누어 주면,

$$\begin{aligned}\frac{\sum (X_t Y_t)}{n} &= \frac{\sum (\hat{a} X_t)}{n} + \frac{\sum (\hat{b} X_t^2)}{n} + \frac{\sum (\hat{u}_t X_t)}{n} \\ &= \hat{a}\bar{X} + \hat{b}\frac{\sum X_t^2}{n} + \frac{\sum (\hat{u}_t X_t)}{n}\end{aligned} \tag{2.49}$$

이제 표본의 값들에 대해서 $\sum(\hat{u}_t X_t) = 0$이라는 조건을 부과한다. 이 조건은 (2.49)의 마지막 항이 0과 같다는 것을 의미한다. 이는 다음과 같은 결과를 가져 온다.

$$\frac{\sum (X_t Y_t)}{n} = \hat{a}\bar{X} + \hat{b}\frac{\sum X_t^2}{n} \tag{2.50}$$

이제 X, Y의 관찰치들과 여전히 측정하여야 할 $\hat{a}$, $\hat{b}$ 사이의 이차적인 관계를 얻게 되었다. 이것이 2차정규방정식 (second normal equation) 이다.

두 정규방정식이 갖는 핵심적 중요성을 보면, 이러한 관계에 대한 직관적인 접근을 더 사용하는 것이 도움이 된다. 기본적인 회귀 관계부터 다시 한번 시작하면,

$$Y_t = a + bX_t + u_t$$

양변에 X_t를 곱하면,

$$X_t Y_t = aX_t + bX_t^2 + u_t X_t$$

다음에 모든 관찰치를 더하고 n으로 나누면,

$$\frac{\sum(Y_tX_t)}{n}=\frac{\sum(aX_t)}{n}+\frac{\sum(bX_t^2)}{n}+\frac{\sum(u_tX_t)}{n}$$
$$=a\overline{X}+\frac{b\sum X_t^2}{n}+\frac{\sum(u_tX_t)}{n} \tag{2.51}$$

전과 같이, 가정에 따라 (2.51)의 마지막 항이 갖는 기대값은 0이다. 그러므로 그 항을 그 값이 0이라고 가정하여 무시하기로 한다. 2차 정규방정식은 다음과 같다.

$$\frac{\sum(Y_tX_t)}{n}=\hat{a}\overline{X}+\hat{b}\frac{\sum X_t^2}{n} \tag{2.50}$$

모수 a와 b를 담고 있는 (2.51)에서 정규방정식(2.50)으로 가면서, a와 b을 $\hat{a}$과 $\hat{b}$로 대신한 것을 유의하기 바란다. 그것은 $\Sigma(u_tX_t)/n$이 일반적으로 정확하게 0이 되지 않고, 오로지 $\Sigma(u_tX_t)/n=0$인 경우에만 $\hat{a}=a$이고 $\hat{b}=b$이기 때문이다.

이제 두 식 (2.42), (2.50)과 두 미지수 $\hat{a}$과 $\hat{b}$를 가지고 있다. 추정량 $\hat{a}$와 $\hat{b}$를 풀 수 있는 위치에 서게 된 것이다. 풀기 위해서는 먼저 $\overline{X}$를 (2.42)에 곱하는 방법이 편리하다.

$$\overline{X}\overline{Y}=\hat{a}\overline{X}+\hat{b}\overline{X}^2 \tag{2.52}$$

(2.50)에서 (2.52)를 빼주면,

$$\frac{\sum(X_tY_t)}{n}-\overline{X}\overline{Y}=\hat{b}\left(\frac{\sum X_t^2}{n}-\overline{X}^2\right) \tag{2.53}$$

그 결과로 $\hat{a}$를 소거하고, 미지수가 $\hat{b}$ 하나인 한 개의 식을 갖게 된다. $\hat{b}$에 대해서 (2.53)을 풀면,

$$\begin{aligned} \hat{b} &= \frac{[\sum(X_t Y_t)/n] - \bar{X}\bar{Y}}{(\sum X_t^2/n) - \bar{X}^2} = \frac{\sum(X_t Y_t) - n\bar{X}\bar{Y}}{\sum X_t^2 - n\bar{X}^2} \\ &= \frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} \end{aligned} \tag{2.54}$$

일단 $\hat{b}$에 대해서 풀고 나면, (2.42)를 사용하여 간단하게 $\hat{a}$를 구할 수 있다.

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \tag{2.55}$$

이 4절은 이 책에서 가장 중요한 절이자 뒤에 나올 내용에 대하여 핵심적인 것이므로, 여기서 지금까지 전개했던 내용을 요약하고 간단한 수치로된 예를 드는 것이 유용할 것이다. 두 변수 사이의 가설화된 선형 관계로 시작한 바 있다. 이 관계는 정확한 것이 아니라 독립변수의 각 값에 대해서 종속변수가 어느 평균값을 중심으로 약간 변동하는 것을 허용하는 것이었다. 이러한 종속변수 값의 차이가 갖는 성격에 관하여 일련의 가정을 가지고서 상당히 상세하게 그 관계의 본성을 설명하였다. 이것이 이변수 또는 이변량 선형회귀모형으로 알려진 것을 이루고 있다.

우리의 문제는 이러한 관계가 갖는 모수들의 값을 추정하는 수단을 개발하는 것이다. 이를 위해서 대변수 기법을 선택하여, 회귀모형 그 자체의 가정이 시사하는 조건을 교란항의 추정량에 직접 부과하였다. 특히 $\hat{a}$과 $\hat{b}$을 구하기 위하여 $\sum \hat{u}_t/n = 0$과 $\sum(\hat{u}_t X_t)/n = 0$이란 조건을 부과한 바 있다. 이들 조건 각각은 하나의 정규방정식을 산출하였다. 달리 설

명하면, 각각의 조건은 관찰된 흩어져 있는 점들에 맞추어진 선위에 한 점이 위치할 수 있도록 함으로써, 그 두 점으로 추정된 관계를 풀 수가 있었다.

가. 보기

이제 이상의 기법을 사용하여 경제학적인 관계를 추정하기로 한다. <표 2.4>에는 미국에서의 1960～1969년 동안의 연간 소비와 가처분소득 수준이 나와 있다. 독자들은 아마도 앞에서의 소비와 가처분소득간의 관계에 관한 조사에서 대부분의 관심을 상이한 소득 수준에 따른 개별 가구의 소비 수준에 촛점을 맞추었음을 기억할 것이다. 횡단면분석(cross-sectional analysis)으로 알고 있는 방법을 사용하여 주어진 시점에서 가구 예산정보에 관한 표본을 가진 경우를 고려하였다. 그리고 어떻게 소비지출이 상이한 소득의 가구들 사이에서 변화하는가를 검토하기 시작하였다. 예를 들어 1970년의 소득과 소비에 관한 가구 자료를 연구한 것이었다. 따라서, 횡단면분석에는 실질적으로 시간을 불변으로 놓은 것이다.

대안적인 접근방식은 시계열분석(time-series analysis)을 채용하는 것인데, 이 분석을 가지고서 한 경제 단위의 행태 또는 경제 단위들의 총체적인 행태를 시간에 걸쳐서 검토한다. 예를 들면 한 경제의 총소비지출이 총가처분소득에 매년에 걸쳐서 어떻게 반응하는가를 조사할 수가 있을 것이다. 이것이 여기서 미국의 총체적인 자료를 사용하여 하려는 분석이다. 특히, 미국에서의 총소비와 가처분소득에 관한 10개의 관찰치(<표 2.4>)를 사용하여 소비지출에 대한 가처분소득 수준의 영향을 검토할 것이다. 먼저 다음과 같이 가정하고,

$$C_t = a + bY_{dt} + u_t$$

〈표 2.4〉 미국의 소비와 가처분소득

(경상 10억달러)

연 도	소 비 (C)	가처분소득 (Y_d)
1960	325	350
1961	335	364
1962	355	385
1963	375	405
1964	401	438
1965	433	473
1966	466	512
1967	492	547
1968	537	590
1969	576	630

출전 : Economic Report of the President (Washington, D.C. : U.S. Government Printing Office, Feb. 1970), pp.189, 195.

a와 b의 값을 추정하는데, 여기서 b는 한계소비성향으로 해석 가능하다. 따라서 다음을 계산해야만 한다.

$$\hat{b} = \frac{\sum (C_t - \bar{C})(Y_{dt} - \bar{Y}_d)}{\sum (Y_{dt} - \bar{Y}_d)^2}$$

와

$$\hat{a} = \bar{C} - \hat{b}\bar{Y}_d$$

필요한 계산은 <표 2.5>에 나와 있다. 그러므로 추정식은,

$$C = 13 + 0.89Y_d \tag{2.56}$$

추정 회귀선 AB는 10개의 흩어져 있는 점을 따라 <그림 2.14>와 같이 나타 난다.* 그 회귀선은 우리가 뒤에서 더 이야기할 문제인, C와 Y_d 사이의 연관 유형에 관한 좋은 근사치를 제공하는 것으로 보인다. 또한, 추정한 관계는 우리가 이론적으로 기대한 내용을 따르고 있는 것이 흥미롭다. 한계소비성향의 추정치는 0.89로서 양이며 0과 1 사이의 값을 갖고, 추정한 절편도 13으로서 양이다.

<표 2.5>에서 $\hat{a}$과 $\hat{b}$의 결정은 상당한 계산량을 요구한다는 점을 인식할 수 있을 것이다. 더하는 것(Summation)이 갖는 속성을 이용하여, 계산량을 어느 정도 줄이는 것이 가능하다. 특히, 다음에 유의할 필요가 있다.**

$$\begin{aligned}\sum (Y_t - \bar{Y})(X_t - \bar{X}) &= \sum (Y_t - \bar{Y})X_t - \sum (Y_t - \bar{Y})\bar{X} \\ &= \sum (Y_t - \bar{Y})X_t\end{aligned}$$

마찬가지로,

$$\sum (X_t - \bar{X})^2 = \sum (X_t - \bar{X})(X_t - \bar{X}) = \sum (X_t - \bar{X})X_t$$

계산을 간편하게 하기 위해서 이러한 관계를 이용하여 $\hat{b}$의 형식을 단순화 할 수 있다.

$$\hat{b} = \frac{\sum (Y_t - \bar{Y})(X_t - \bar{X})}{\sum (X_t - \bar{X})^2} = \frac{\sum (Y_t - \bar{Y})X_t}{\sum (X_t - \bar{X})X_t} \tag{2.57}$$

---

* 여기서는 <그림 2.14>의 CD선을 무시한다.

** 1장의 부록 A에서 명제 4를 참조하라.

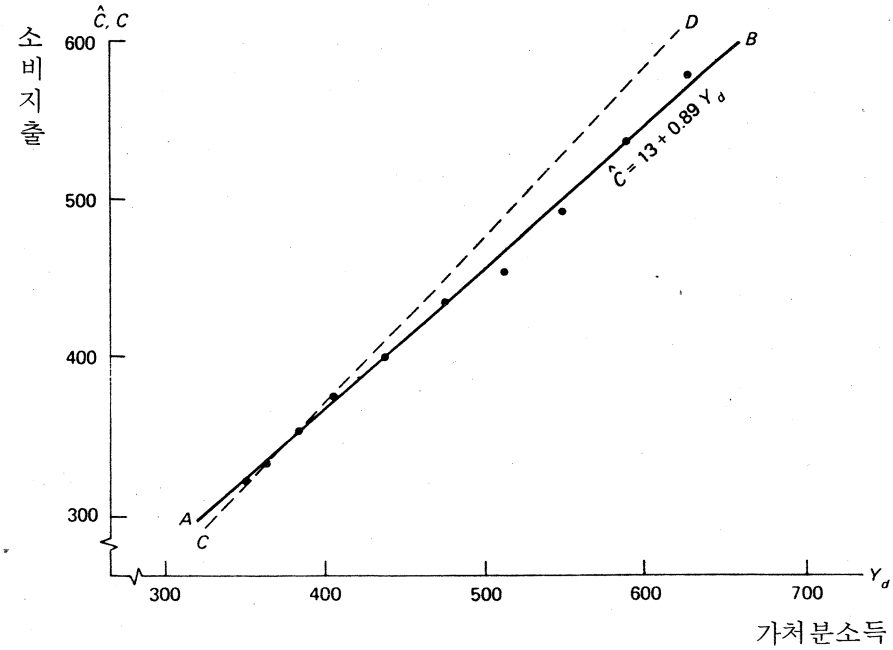
〈표 2.5 〉

(1) C_t	(2) Y_{dt}	(3) $(C_t - \bar{C})$	(4) $(Y_{dt} - \bar{Y}_d)$	(5) = (3) × (4) $[(C_t - \bar{C})(Y_{dt} - \bar{Y}_d)]$	(6) $(Y_{dt} - \bar{Y}_d)^2$
325	350	−105	−119	12,495	14,161
335	364	−95	−105	9,975	11,025
355	385	−75	−84	6,300	7,056
375	405	−55	−64	3,520	4,096
401	438	−29	−31	899	961
433	473	3	4	12	16
466	512	36	43	1,548	1,849
492	547	62	78	4,836	6,084
537	590	107	121	12,947	14,641
576	630	146	161	23,506	25,921

$$\sum C_t = 4{,}295 \qquad \bar{C} = 430$$
$$\sum Y_{dt} = 4{,}694 \qquad \bar{X}_d = 469$$

$$\sum (C_t - \bar{C})(Y_{dt} - \bar{Y}_d) = 76{,}038$$
$$\sum (Y_{dt} - \bar{Y}_d)^2 = 85{,}810$$

$$\hat{b} = \frac{76{,}038}{85{,}810} = 0.89$$
$$\hat{a} = \bar{C} - \hat{b}\bar{Y}_d = 430 - 0.89(469) = 13$$



〈그림 2.14 〉

$$= \frac{\sum (Y_t X_t) - \bar{Y} \sum X_t}{\sum X_t^2 - \bar{X} \sum X_t} = \frac{\sum (Y_t X_t) - n\bar{Y}\bar{X}}{\sum X_t^2 - n\bar{X}^2}$$

그러나 이상의 형태도, 특히 많은 관찰치가 있다면, $\hat{a}$과 $\hat{b}$의 결정에는 엄청난 계산 작업이 소요된다. 다행히도, 이러한 계산은 컴퓨터로 쉽게 할 수 있고, 컴퓨터로 하여금 이러한 일을 수행하고 $\hat{a}$과 $\hat{b}$의 값을 낼 수 있도록 지시하는 수많은 프로그램이 개발되어 있다.

나. 가정중의 하나에 관한 노트

$\hat{a}$과 $\hat{b}$가 갖는 속성을 설명하기에 앞서서 기본 가정중의 하나가 갖는 중요성을 보이기로 한다. 그 가정은 X_t가 적어도 두개의 다른 값을 취한다는 것이다. 이 점을 입증하기 위해서 이 가정이 X_t가 항상 X_0와 같은 특정한 값과 동일함으로써 위배된다고 가정하자. 그러면, $\hat{a}$과 $\hat{b}$를 결정하는 데에 사용되는 정규방정식인 식 (2.42)와 (2.50)은 다음과 같이 될 것이다.

$$\bar{Y} = \hat{a} + \hat{b}X_0 \tag{2.42A}$$

와

$$X_0\bar{Y} = \hat{a}X_0 + \hat{b}X_0^2 \tag{2.50A}$$

왜냐하면 $\sum X_t / n = X_0$, $\sum (X_t Y_t) / n = X_0$이기 때문이다. 만일 (2.50A)를 X_0로 나누면

$$\bar{Y} = \hat{a} + \hat{b}X_0$$

이며, 이것은 (2.42A)와 동일하다. 이 모든 사항은 우리가 오로지 한개의 식과 두개의 미지수, $\hat{a}$과 $\hat{b}$를 가지고 있음을 의미한다. 우리는 $\hat{a}$과

$\hat{b}$의 유일한 값을 구할 수 없으며, 따라서 a와 b를 추정할 수가 없음을 알게 된다.

이러한 이유로, 직관적인 수준에서 만일 X_t가 항상 X_0와 같다면, 기본적인 회귀모형

$$Y_t = a + bX_t + u_t \quad \text{는}$$

다음이 된다.

$$Y_t = a + bX_0 + u_t \tag{2.58}$$

이제 X_0는 상수이기 때문에, bX_0는 원래 상수항인 a와도 결합될 수 있어서, 모형은 다음과 같이 된다.

$$Y_t = A + u_t \tag{2.59}$$

여기서 $A = (a + bX_0)$이다. 우리의 회귀모형은 단지 하나의 상수항과 하나의 교란항을 갖는 것으로 축소된다.

만일 A를 추정하기 원한다면, 다시 대변수 기법으로 되돌아 가기로 한다. 자세히 하면 먼저 (2.59)로부터,

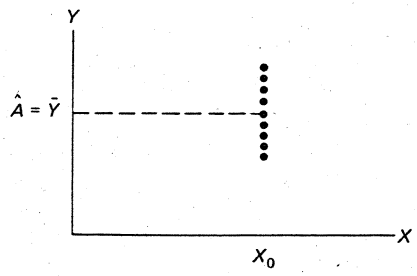
$$Y_t = \hat{A} + \hat{u}_t \tag{2.60}$$

그러면, $E(u_t) = 0$이란 가정이 다시 우리가 $\sum \hat{u}_t / n = 0$으로 놓을 것을 시사한다. 따라서 정규방정식은,

$$\frac{\sum Y_t}{n} = \hat{A} \tag{2.61}$$

다른 말로 표현하면, A의 추정량은 $\hat{A}=\bar{Y}$이 될 것이다.*

<그림 2.15 >로 나타내면, X_i가 X_0와 항상 같기 때문에, 산포도가 X_0 위에 수직으로 위치한 일련의 점들로 와해된 것을 볼 수 있다. 분명히 그러한 점의 집합은 X의 특정한 값 X_0에 대응하는 Y의 평균값 A를 추정할 수 있게 할 따름이다.



〈그림 2.15 〉

따라서, 만약 X_i의 값이 변하지 않으면, 대변수 추정 기법은 변수 X가 Y에 미치는 효과, 즉 b를 상수항과 분리하여 추정할 수 없게 한다는 사실을 발견한다. 이러한 경우에는 단지 결합한 효과 $A=(a+bX_0)$의 추정치를 얻을 수 밖에 없다. 다시 직관적으로 보면, X_i가 항상 같은 값을 가질 경우, X_i가 Y_i에 미치는 효과는 상수항과 "혼합"되거나, 분리가 불가능하게 된다. 부수적으로 다중회귀의 경우에 있어 이 문제의 일반화를 고려하고 있는 4장에서 논의를 계속하도록 한다.

5. $\hat{a}$과 $\hat{b}$의 속성

이제는 추정량 $\hat{a}$과 $\hat{b}$를 얻는 기법을 알게 되었다. 그러나 이 기법이

---

* 우리는 단지 한개의 모수, A만을 갖고 있기 때문에, 추정을 위해서도 오로지 하나의 정규방정식만 있으면 된다.

좋은 기법인지 아닌지에 대해서는 의문이 남아 있다. 분명 이들 모수의 추정량을 발생시키는 여타 방식들이 존재한다. 예를 들어 <그림 2.14>에 흩어져 있는 점 가운데서 아무 두 점을 취할 수 있으며, 그 정보를 가지고 소비와 가처분소득을 추정한 관계로 사용할 선을 그릴 수가 있는 것이다. 이런 방식은 우리가 식 (2.54)와 (2.55)에 도달하기 위하여 거쳤던 과정보다 훨씬 수월하다. 직관적으로 독자들은 아마도 우리가 사용했던 기법이 둘 중에서 더 좋은 것으로 느낄 것이다. 왜냐하면, 그 기법은 상당히 더 많은 정보를 체계적으로 이용하였기 때문이다. 우리가 <그림 2.14>에서 흩어져 있는 점에 맞추었던 선 AB는 관찰치가 제시하는 행태에 대해서 무리없이 대응하는 선으로 보인다. 만약 두 점만을 이용했다면, <그림 2.14>의 CD와 같은 선을 얻게 되는데, 이는 앞의 경우와 대조적으로 C와 Y_d사이의 전형적인 관계를 훨씬 나쁘게 나타내는 추정치인 것 같다.

이제 $\hat{a}$과 $\hat{b}$이 어떤 소망스러운 통계학적 속성을 가진다는 의미에서 "좋은" 추정량임을 보일 것이다. 특히, 다음의 사항을 보이게 된다.

(1) $\hat{a}$과 $\hat{b}$의 기대값은 각각 a와 b이다. 또한,

(2) $\hat{a}$과 $\hat{b}$의 분산은 "상대적으로" 작다.

결과적으로는 적어도 우리의 추정량이 무엇보다도 먼저 다른 추정량들에 비하여 오차의 폭을 작게 가지는 경향이 있음을 알게 될 것이다.

가. 不偏性*

맨 먼저 $\hat{a}$과 $\hat{b}$의 기대값이 사실상 a와 b임을 보일 것이다. 또는 다른 말로 한다면, 우리의 추정량이 불편적임을 보일 것이다. 그 증명에는

---

* 불편성에 관한 것은 J.Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), pp.18-20의 내용을 따른다.

가산이 갖는 5가지 속성을 이용한다.*

$$\sum (X_t - \bar{X}) = 0 \tag{2.62}$$

$$\sum (X_t + Y_t) = \sum X_t + \sum Y_t \tag{2.63}$$

$$\sum (X_t - \bar{X})(Y_t - \bar{Y}) = \sum (X_t - \bar{X})Y_t \tag{2.64}$$

$$\sum (X_t - \bar{X})(Y_t - \bar{Y}) = \sum (Y_t - \bar{Y})X_t \tag{2.65}$$

$$\sum (X_t - \bar{X})^2 = \sum (X_t - \bar{X})(X_t - \bar{X}) = \sum (X_t - \bar{X})X_t \tag{2.66}$$

식 (2.54) 에서 $\hat{b}$ 에 대하여 푼 결과를 갖고 있다.

$$\hat{b} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$$

이것을 (2.64) 를 사용하여 간단히 하면,

$$\hat{b} = \frac{\sum (X_t - \bar{X})Y_t}{\sum (X_t - \bar{X})^2} \tag{2.67}$$

이제 우리의 모형에서 $Y_t = a + bX_t + u_t$ 를 (2.67) 의 분자에 대입하면,

$$\hat{b} = \frac{\sum (X_t - \bar{X})(a + bX_t + u_t)}{\sum (X_t - \bar{X})^2} \tag{2.68}$$

(2.68) 의 분자를 전개하고, (2.63) 에서 제시한 명제를 사용하면, 다음을 얻는다.

---

* 이들 명제는 1장의 부록 A에서 공식적으로 증명하였다.

$$\hat{b} = \frac{a\sum(X_t - \bar{X}) + b\sum(X_t - \bar{X})X_t + \sum(X_t - \bar{X})u_t}{\sum(X_t - \bar{X})^2} \tag{2.69}$$

(2.69)를 다음과 같이 고쳐 쓰면,

$$\hat{b} = \frac{a\sum(X_t - \bar{X})}{\sum(X_t - \bar{X})^2} + \frac{b\sum(X_t - \bar{X})X_t}{\sum(X_t - \bar{X})^2} + \frac{\sum(X_t - \bar{X})u_t}{\sum(X_t - \bar{X})^2} \tag{2.70}$$

(2.62)로부터 윗식의 첫째 항이 0과 같음을 알 수 있다. (2.66)을 사용해서 두번째 항의 분모 형식을 $\sum(X_t - \bar{X})X_t$로 바꾸면, 두번째 항이 b와 같은 것에 불과함을 알 수 있다. 따라서,

$$\hat{b} = b + \frac{\sum(X_t - \bar{X})u_t}{\sum(X_t - \bar{X})^2} \tag{2.71}$$

뒤따르는 분석에서의 표기를 단순화하기 위해서, 다음과 같이 정하기로 하자.

$$A = \sum(X_t - \bar{X})^2 \tag{2.72}$$

$$w_t = (X_t - \bar{X}) \tag{2.73}$$

이 정의를 사용하면 (2.71)의 $\hat{b}$은 다음과 같이 된다.

$$\begin{aligned}\hat{b} &= b + \frac{\sum w_t u_t}{A} = b + \frac{w_1 u_1}{A} + \frac{w_2 u_2}{A} + \cdots + \frac{w_n u_n}{A} \\ &= b + \left(\frac{w_1}{A}\right)u_1 + \left(\frac{w_2}{A}\right)u_2 + \cdots + \left(\frac{w_n}{A}\right)u_n\end{aligned} \tag{2.74}$$

이제 $\hat{b}$이 불편추정량임을 보일 수 있는 입장이 되었다. 상세히 보면,

(2.74)로 부터

$$E(\hat{b}) = b + E\left[\left(\frac{w_1}{A}\right)u_1\right] + E\left[\left(\frac{w_2}{A}\right)u_2\right] + \cdots + E\left[\left(\frac{w_n}{A}\right)u_n\right] \qquad (2.75)$$

(w_1/A), ……, (w_n/A)항들은 설명변수 X_t의 n개 값에만 종속하고, 설명변수의 값들과 교란항의 값들은 독립적인 것으로 가정하였기 때문에,

$$E(\hat{b}) = b + E\left(\frac{w_1}{A}\right)E(u_1) + E\left(\frac{w_2}{A}\right)E(u_2) + \cdots + E\left(\frac{w_n}{A}\right)E(u_n) \qquad (2.76)$$

우리는 회귀모형에서 $E(u_t) = 0$임을 알고 있다. 따라서 첫째 항을 제외하고서는 전부가 0이어서,

$$E(\hat{b}) = b \qquad (2.77)$$

$\hat{b}$은 결국 b의 불편추정량이다.

다음에는 $\hat{a}$로 돌아가서, (2.55)로 부터 다음을 알고 있다.

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \qquad (2.55)$$

$Y_t = a + bX_t + u_t$이기 때문에,

$$\bar{Y} = a + b\bar{X} + \bar{u} \qquad (2.78)$$

(2.78)을 (2.55)에 대입하면,

$$\hat{a} = a + b\bar{X} + \bar{u} - \hat{b}\bar{X} \qquad (2.79)$$

$\hat{b}$를 대신하여 (2.74)를 대입하면,

$$\hat{a} = a + b\overline{X} + \bar{u} - b\overline{X} - \left(\frac{w_1\overline{X}}{A}\right)u_1 - \left(\frac{w_2\overline{X}}{A}\right)u_2 - \cdots - \left(\frac{w_n\overline{X}}{A}\right)u_n \qquad (2.80)$$

$b\overline{X}$와 $-b\overline{X}$는 상쇄되고 $\hat{a}$의 기대값을 취하면,

$$E(\hat{a}) = a + E(\bar{u}) - E\left(\frac{w_1\overline{X}}{A}\right)E(u_1) - E\left(\frac{w_2\overline{X}}{A}\right)E(u_2) - \cdots$$
$$- E\left(\frac{w_n\overline{X}}{A}\right)E(u_n) = a \qquad (2.81)$$

왜냐하면 $E(\bar{u}) = 0$, $E(u_i) = 0$이기 때문이다. 따라서 $\hat{a}$도 불편추정량이다.

나. $\hat{a}$과 $\hat{b}$의 분산 : 약간의 기초

$\hat{a}$과 $\hat{b}$의 분산에 관한 주제가 아직 남아 있다. 이제는 평균값이 대응하는 모수의 값과 같은 추정량을 만들어 내는 방식을 알고 있다. 분산에 관한 문제는 지금 $\hat{a}$과 $\hat{b}$이 각각의 평균값 a 그리고 b와 편차를 가지는 것이 기대될 수 있는 단계에까지 제기된다. 여기서 다른 기법들과 비교하여 지금까지의 절차가 상대적으로 작은 분산을 갖는 추정량을 산출할 수 있기를 바라게 될 것이다. $\hat{a}$과 $\hat{b}$의 분산에 대한 식을 도출하기에 앞서, 무엇보다도 먼저 왜 이들 추정량이 분산을 갖는가를 간략하게 논의하는 것이 도움을 줄 것으로 보인다.

우선, X와 Y에 관하여 가상적인 5개의 관찰치 표본이 있다고 하자.

	표본 1	표본 2
관찰치 1	(Y_{11}, X_1)	(Y_{12}, X_1)
관찰치 2	(Y_{21}, X_2)	(Y_{22}, X_2)
관찰치 3	(Y_{31}, X_3)	(Y_{32}, X_3)
관찰치 4	(Y_{41}, X_4)	(Y_{42}, X_4)
관찰치 5	(Y_{51}, X_5)	(Y_{52}, X_5)

Y의 하첨자에서 첫번째 숫자는 특정한 관찰치를 말하는 것이고, 두번째 숫자는 표본과 동일하다. 이 예에서는 두 표본에서 각 관찰치마다 X값이 동일하다고 가정한다. 그러한 X값에 대한 가정은 물론 Y값이 두 표본에서 같을 것이라고 의미하지는 않는다. 그 이유는 Y가 두가지 영향을 반영하고 있기 때문이다. 즉, (1) 평균 관계인 $Y^m = a + bX$에서 작용하는 X값의 영향 그리고 (2) 0의 평균값을 갖는 확률변수이자 가정에 의하여 변수값이 X와 독립적인 교란항, u_i의 존재이다.

두 표본 각각의 첫번째 관찰치를 보자. 교란항이 없다면, 매번의 경우 Y가 X_1의 영향만을 반영하기 때문에 Y_{11}이 Y_{12}와 같았을 것이다. 이 경우에 Y_{11}과 Y_{12}는 $Y = a + bX_1$의 값을 갖게 된다. 만일 이것이 모든 관찰치에 대해서 진실이라면, X와 Y값의 관찰한 집합은 두 표본에서 동일하고, $\hat{a}$과 $\hat{b}$에 관하여 계산한 값은 매번의 경우 a와 b의 실제 값과 같을 것이다.

그러나, 교란항의 존재는 Y의 관찰치가 X의 영향만을 반영한 값과 어느 정도 편차를 가질 것임을 의미한다.

특히,

$$Y_{11} = a + bX_1 + u_{11} \quad \text{그리고} \quad Y_{12} = a + bX_1 + u_{12}$$

여기서 일반적으로 $u_{11} \neq u_{12}$ 이다. 이것은 일반적으로 $Y_{11} \neq Y_{12}$ 를 의미한다. 다른 Y의 관찰치들에 대해서도 마찬가지이어서, 일반적으로 $Y_{i1} \neq Y_{i2}$ 이다. $\hat{a}$과 $\hat{b}$는 X와 Y의 관찰치로부터 직접 계산한 것이기 때문에, 일반적으로 $\hat{a}_1$과 $\hat{b}_1$은 각각 $\hat{a}_2$과 $\hat{b}_2$와 같지 않을 것이다. 여기서 하첨자는 $\hat{a}$과 $\hat{b}$의 값을 계산한 표본을 말한다. 그러므로, 교란항은 Y의 관찰치와 $\hat{a}$과 $\hat{b}$의 계산치 모두가 표본이 바뀜에 따라 변동하게 한다.

이제 이상의 내용을 일반화하자. 주어진 수의 X와 Y의 관찰치로 이루어진 P개의 표본이 있고, 여기서 설명변수 X값의 집합이 모든 표본에서 동일하다고 가정하자. $\hat{a}_i$과 $\hat{b}_i$을 i번째 표본에서 계산한 $\hat{a}$과 $\hat{b}$의 값으로 한다. Y값들이 표본이 달라짐에 따라 다를 것이기 때문에, $\hat{a}_i$과 $\hat{b}_i$의 값 또한 변동할 것으로 생각한다. 그러므로, 일반적인 조건 아래에서 P가 무한하면 (즉, 표본의 수가 무한 개이면), (2.82)에서의 (A)와 (B)의 총합은 각각 $\hat{a}$의 분산 $\sigma_{\hat{a}}^2$과 $\hat{b}$의 분산 $\sigma_{\hat{b}}^2$과 같게 될 확률이 1이 될 것이다.

$$\frac{\sum_{i=1}^{P} (\hat{a}_i - a)^2}{P} \quad \text{(A)}$$

$$\frac{\sum_{i=1}^{P} (\hat{b}_i - b)^2}{P} \quad \text{(B)} \qquad (2.82)$$

다음 절에서는 $\hat{a}$과 $\hat{b}$의 분산값을 구하는 식을 도출할 것이다. 여기서는 위에서 공식적으로 전개된 분산이 조건부 분산(conditional variance)이라는 점에 유의해야만 한다. 조건부라고 한 것은 이상의 설명들이 X값들의 집합이 여러 표본에 걸쳐 동일하다는 가정 아래에서 전개되었기 때문이다. 표본들에 걸쳐서 $\hat{a}$과 $\hat{b}$가 변동하는 것은 전적으로 상이한 표

본 간에 교란항의 값이 변동하기 때문이다. 누군가는 예견하였겠지만, (2.82)식의 값들은 부분적으로 바로 X의 공통적인 값의 무엇인가에 따른다. 실제로, 조건부 분산의 크기는 연구자로 하여금 자신의 추정량에 얼마만큼이나 불확실성이 따라 붙는가를 알게끔 한다. 그 추정량은 X의 특정한 관찰치 집합에 기초하는 것이다. 마지막으로, $E(\hat{a}) = a$와 $E(\hat{b}) = b$이기 때문에 P가 무한대라면, 일반적인 조건 아래에서 다음과 같을 확률은 1이다.

$$\sum_{i=1}^{P} \frac{\hat{a}_i}{P} = a \quad \text{와} \quad \sum_{i=1}^{P} \frac{\hat{b}_i}{P} = b \tag{2.83}$$

다. 추정량의 분산*

이제 $\hat{b}$의 분산에 관한 식을 먼저 도출하고 나서 $\hat{a}$의 분산을 구하는 식을 도출한다. $\hat{b}$에 대한 기본적인 식에서 시작하면,

$$\hat{b} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$$

앞에서 불편성을 증명하면서

$$\hat{b} = b + \frac{w_1 u_1}{A} + \frac{w_2 u_2}{A} + \cdots + \frac{w_n u_n}{A} \tag{2.74}$$

여기서

---

* 때때로 $\hat{a}$과 $\hat{b}$의 "분산"을 말하면서 분산으로 표기할 것이나, 독자들은 이것이 조건부 분산이라는 데에 유의해야 한다. 이 절에서의 결과를 유사하게 도출한 것은 다음을 보라. J.Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), pp.18-20.

$$w_t = (X_t - \bar{X}) \quad \text{이고} \quad A = \sum (X_t - \bar{X})^2$$

Var ($\hat{b}$)식을 도출하는 데 (2.74)를 이용하려면, 확률변수의 선형결합 합계가 갖는 분산에 대한 근본적인 관계를 이용할 필요가 있다. 자세히 보면 (그리고 2장의 부록에서 이 명제를 도출한다), 만약 확률변수 M이 있는데 다음과 같이 정의된다고 하자.

$$M = a_0 + a_1 Z_1 + a_2 Z_2 + \cdots + a_n Z_n \tag{2.84}$$

여기서 a들은 상수이며 Z들은 확률변수라 하고, Z들이 상관되지 않았다고 가정하자. 그러면,

$$\text{var}(M) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 \tag{2.85}$$

여기서 $\sigma_j^2 = \text{var}\ (\ Z_j\ )$.

조건부 분산이 갖는 중요성이 이제 명백하다. 특히 주어진 X값들의 집합에 대응하는 $\hat{b}$의 분산에 관심이 있다면, X의 n개 값을 단지 n개의 상수로 간주할 수가 있다. (2.74)로부터 w_t의 n개 값과 A의 값이 또한 상수로 간주될 수 있다. 이러한 조건아래에서 (2.74)는 교란항의 한 선형결합에 불과하게 된다. 가정에 의해서, 이들 교란항이 독립적이고 따라서 서로 상관되지 않기 때문에, 또한 교란항은 동일한 분산 σ_u^2을 갖기 때문에, (2.85)을 적용하여 다음과 같이 된다.

$$\begin{aligned} \text{var}(\hat{b}) = \sigma_b^2 &= \frac{w_1^2 \sigma_u^2}{A^2} + \frac{w_2^2 \sigma_u^2}{A^2} + \cdots + \frac{w_n^2 \sigma_u^2}{A^2} \\ &= \frac{\sigma_u^2}{A^2} \sum w_t^2 = \sigma_u^2 \frac{\sum (X_t - X)^2}{[\sum (X_t - X)^2]^2} \end{aligned} \tag{2.86}$$

결과는,

$$\text{var}(\hat{b}) = \sigma_{\hat{b}}^2 = \frac{\sigma_u^2}{\sum(X_t - \bar{X})^2} \tag{2.87}$$

(2.87)식은 어느 특정한 X값의 집합에 대응하는 $\hat{b}$의 분산을 나타낸다. b에 대한 추정량의 분산이 직접 교란항의 분산에 따라 변한다는 사실은 놀라운 일이 아니다. 어떠한 X값들의 집합에 대해서도, 교란항의 분산이 커지면 커질수록, $\hat{b}$의 분산이 더 커지게 된다. 어느 정도 직관적으로 보아도 기본 회귀모형에서의 불확실성이 커지면, 추정량에 대한 신뢰는 더 떨어지게 되는 것이다.

비슷한 방식으로 $\hat{a}$의 분산을 유도할 수 있다. $\hat{a}$에 관한 앞에서의 식은 다음과 같다.

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \tag{2.55}$$

$\bar{Y} = a + b\bar{X} + \bar{u}$임을 참조하여 $\hat{b}$에 (2.74)를 대입하면, 다음을 얻게 된다.

$$\begin{aligned}\hat{a} &= a + b\bar{X} + \bar{u} - b\bar{X} - \left(\frac{\bar{X}w_1}{A}\right)u_1 - \cdots - \left(\frac{\bar{X}w_n}{A}\right)u_n \\ &= a + \bar{u} - \left(\frac{\bar{X}w_1}{A}\right)u_1 - \cdots - \left(\frac{\bar{X}w_n}{A}\right)u_n\end{aligned} \tag{2.88}$$

$\bar{u}$를 다음과 같이 표현하면,

$$\bar{u} = \frac{u_1}{n} + \cdots + \frac{u_n}{n} \tag{2.89}$$

윗식을 (2.88)에 대입하여 정리하면

$$\hat{a} = a + \gamma_1 u_1 + \cdots + \gamma_n u_n \tag{2.90}$$

여기서 $\gamma_t = [(1/n) - \overline{X} w_t / A]$. X값에 관한 우리의 가정 아래에서는 $\hat{a}$도 교란항의 선형결합으로 축소된다는 점을 알게 된다. 그러므로 (2.85)를 응용하면, $\hat{a}$의 분산은 다음과 같다.

$$\begin{aligned} \text{var}(\hat{a}) = \sigma_{\hat{a}}^2 &= \gamma_1^2 \sigma_u^2 + \cdots + \gamma_n^2 \sigma_u^2 \\ &= \sigma_u^2 \sum_{t=1}^{n} \gamma_t^2 \end{aligned} \tag{2.91}$$

그런데,

$$\begin{aligned} \sum \gamma_t^2 &= \sum \left[\frac{1}{n^2} + \left(\frac{\overline{X}^2}{A^2} \right) w_t^2 - \left(2 \frac{\overline{X}}{nA} \right) w_t \right] \\ &= \frac{1}{n} + \left(\frac{\overline{X}^2}{A^2} \right) \sum w_t^2 - \left(\frac{2\overline{X}}{nA} \right) \sum w_t \end{aligned} \tag{2.92}$$

다음과 같은 이유로,

$$\sum w_t^2 = \sum (X_t - \overline{X})^2 = A \tag{2.93}$$

이고

$$\sum w_t = \sum (X_t - \overline{X}) = 0 \tag{2.94}$$

(2.91)에서 주어진 $\hat{a}$의 분산은 다음과 같이 표현할 수 있다.

$$\begin{aligned} \sigma_{\hat{a}}^2 &= \sigma_u^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{A} \right) \\ &= \sigma_u^2 \left(\frac{A + n\overline{X}^2}{nA} \right) \end{aligned} \tag{2.95}$$

마지막으로, 1장 부록 A에서

$$\sum (X_t - \bar{X})^2 = \sum X_t^2 - n\bar{X}^2$$

임을 상기하여 (2.95)에 직접 대입하면,

$$\sigma_{\hat{a}}^2 = \frac{\sigma_u^2 \sum X_t^2}{n \sum (X_t - \bar{X})^2} \tag{2.96}$$

$\sigma_{\hat{b}}^2$ 의 경우에서처럼, 주어진 어느 X_t 의 집합에 대해서 $\sigma_{\hat{a}}^2$ 의 값은 교란항의 분산 σ_u^2 에 따라 직접 변동한다. 이 중대한 시점에서 X_t 와 σ_u^2 의 값에 대한 가상적인 예를 가지고 $\hat{a}$과 $\hat{b}$의 분산을 실제로 계산하면 도움이 될 것이다. 여타 기본 정보로서 교란항의 분산 σ_u^2 의 분산이 10이라는 것을 알고 있다고 가정하자. 다음으로 <표 2.6>에서 보듯이 X와 Y의 관찰치 집합을 가진 것으로 하자. 이 경우에 다음의 계산 결과가 나온다.

〈표 2.6 〉

Y	X
8	3
12	6
14	10
15	12
15	14
18	15

$$\sum X_t^2 = 710, \quad \sum (X_t - \bar{X})^2 = 110, \quad n = 6$$

(2.87)과 (2.96)식을 이용하면, 위와 같이 주어진 X_t 값들의 집합에 대하여,

$$\text{var}(\hat{b}) = \frac{10}{110} = 0.09, \quad \text{var}(\hat{a}) = \frac{10(710)}{6(110)} = 10.8$$

라. 최소분산의 속성

이제 $\hat{a}$과 $\hat{b}$의 분산을 나타내는 식을 가지게 되었으며, 이 식은 뒤에서 가설을 검정하는 문제에 다다르게 되었을 때에 신뢰구간을 설정하는데에 유용함이 입증될 것이다. 그러나 먼저 이 분산이 다른 기법들에 의하여 a와 b를 추정할 경우보다 상대적으로 큰지 작은지를 아는 것이 좋다. 이런 측면에서 이 2장에서 얻어낸 $\hat{a}$과 $\hat{b}$보다 더 작은 분산을 갖는 a와 b의 선형, 불편추정량이 존재하지 않음을 보일 수가 있다. 선형 추정량이란 종속변수 Y의 값들로써 선형결합으로 표현할 수 있는 추정량을 의미한다. 예를 들면, (2.67)로부터 다음을 상기하는 것이다.

$$\begin{aligned} \hat{b} &= \frac{\sum (X_t - \bar{X}) Y_t}{\sum (X_t - \bar{X})^2} = \frac{\sum w_t Y_t}{A} \\ &= \left(\frac{w_1}{A}\right) Y_1 + \cdots + \left(\frac{w_n}{A}\right) Y_n \end{aligned} \tag{2.97}$$

(2.55)로 부터,

$$\begin{aligned} \hat{a} &= \bar{Y} - \bar{X}\hat{b} = \frac{\sum Y_t}{n} - \bar{X}\frac{\sum w_t Y_t}{A} \\ &= \sum \gamma_t Y_t = \gamma_1 Y_1 + \cdots + \gamma_n Y_n \end{aligned} \tag{2.98}$$

최소분산 명제의 증명은 개념적으로 어려운 것은 아니지만 상당한 정도로 길기 때문에, 이 장의 끝에 있는 부록에 두기로 하였다. 독자들이 그 증명을 학습하기를 권하지만, 만일 이 명제가 제시하는 내용을 그냥 믿는다고 하여도(적어도 당분간), 이하의 내용을 이해하는 데에는 어떤 어려움도 없을 것이다.

a와 b를 추정하는 기법뿐만 아니라, 이제는 그 기법이 좋은 기법이라는 것을 믿을 만한 근거도 알게 되었다. 첫째로, 그 기법은 모수의 불편추정량을 발생시키며, 둘째로 그 추정량은 a와 b의 모든 선형, 불편추정량중에서도 최소분산 추정량이다.

마. 분산 추정량

$\hat{a}$과 $\hat{b}$의 분산에 관해서 아직 마치지 못한 일이 하나 남아 있다. (2.89)와 (2.96)식에서 그 분산의 식을 알게 되었지만, 그 식은 회귀모형 안의 교란항이 갖는 분산 σ_u^2를 포함하고 있다. 문제는 σ_u^2이 a, b와 같이 일반적으로 미지수일 것이라는 점에 있다. 이는 $\hat{a}$와 $\hat{b}$의 분산값을 얻기 위해서는 먼저 σ_u^2를 추정해야만 한다는 것을 의미한다. 이제 σ_u^2의 추정량을 도출하는 문제를 고려할 것이다.

먼저 다음을 상기하도록 하자.

$$\sigma_u^2 = E(u_t - 0)^2 = E(u_t^2)$$

즉, 교란항의 분산은 그 자승의 평균값에 불과하다. 지금 우리는 기본 회귀모형으로부터 다음을 알고 있다.

$$u_t = Y_t - a - bX_t = Y_t - Y_t^m \tag{2.35}$$

a와 b값을 이미 알았다고 가정하자. 이 경우에 만약 X와 Y에 관한

n크기의 관찰치 표본을 갖고 있다면, (2.35)로부터 u_t에 관한 n개 값을 유도할 수 있다. 그러면, 단순히 표본에서의 u_t^2의 평균적인 값을 취함으로써 σ_u^2를 추정하는 것이 합리적일 것이다.

$$\frac{\sum u_t^2}{n} = \frac{\sum (Y_t - a - bX_t)^2}{n} \tag{2.99}$$

실제로는 이런 방법이 가능하지 않는데, 그 이유는 일반적으로 a와 b의 값을 알지 못하기 때문이다. 그러나, 그 분명한 절차는 (2.99)에서 a와 b를 $\hat{a}$과 $\hat{b}$으로 대체함으로써 σ_u^2의 추정량, 말하자면 $\hat{\sigma}_u^2$을 얻는 것이다. 그러면,

$$\hat{\sigma}_u^2 = \frac{\sum (Y_t - \hat{a} - \hat{b}X_t)^2}{n-2} = \frac{\sum (Y_t - \hat{Y}_t)^2}{n-2} = \frac{\sum \hat{u}_t^2}{n-2} \tag{2.100}$$

(2.100)의 분모가 n이 아니라 ($n-2$)임에 주의하자. 이는 (앞에서 논의하였듯이) 분자에서 ($n-2$)만의 자유도 밖에 가지지 못함을 가리킨다. 2개의 자유도는 추정량이 2개의 모수를 대체함으로써 상실한 것이다. 부수적으로 (여기서는 깊숙이 다루지 않겠지만) 이것은

$$E(\hat{\sigma}_u^2) = \sigma_u^2 \tag{2.101}$$

인 경우이다. $\hat{\sigma}_u^2$은 σ_u^2의 불편추정량인 것이다.

이제 교란항의 분산에 대한 추정량을 갖게 되었다. (2.87)과 (2.96)에 대응하여, 단지 σ_u^2를 $\hat{\sigma}_u^2$으로 대체함으로써 $\hat{a}$과 $\hat{b}$의 분산에 대한 추정량을 얻을 것이다. 특히, 분산의 추정량으로서 취하면,

$$\hat{\sigma}_{\hat{a}}^2 = \frac{\hat{\sigma}_u^2 \sum X^2}{n\sum (X_t - \bar{X})^2}$$

$$\hat{\sigma}_b^2 = \frac{\hat{\sigma}_u^2}{\sum (X_t - \bar{X})^2} \tag{2.102}$$

마지막으로 (2.101) 에서 $\hat{\sigma}_u^2$ 이 불편적이기 때문에 $\hat{\sigma}_{\hat{a}}^2$ 과 $\hat{\sigma}_{\hat{b}}^2$ 도 또한 불편추정량이다.

바. 보기

앞의 <표 2.5>에서 소비함수의 추정을 위하여 $\hat{a}$과 $\hat{b}$를 계산하는 데에 미국의 소비와 가처분소득에 관한 자료를 사용한 바 있다. 그 때 추정한 식을 상기하면,

$$C = 13 + 0.89 Y_d \tag{2.56}$$

이제 대응하는 분산에 관한 추정치를 계산할 수 있는 입장에 서있다. 먼저, <표 2.7>을 참조하여, $\hat{\sigma}_u^2$ 의 값을 계산하면,

$$\hat{\sigma}_u^2 = \frac{92}{10 - 2} = 11.5$$

또한 <표 2.5>로부터 다음을 알 수 있다.

$$\sum (Y_{dt} - \bar{Y}_d)^2 = 85{,}810$$

그래서,

$$\sum (Y_{dt}^2) = 2{,}289{,}172$$

따라서, 다음의 결과를 얻는다.

$$\hat{\sigma}_{\hat{a}}^2 = \frac{11.5(2{,}289{,}172)}{10(85{,}810)} = 31$$

$$\hat{\sigma}_b^2 = \frac{11.5}{85.810} = 0.0001$$

사. $\hat{a}$과 $\hat{b}$의 最小自乘 속성

추정량 $\hat{a}$과 $\hat{b}$에 대해서 지적하고 싶은 마지막 속성이 있다. X와 Y 사이의 관계를 추정하는 문제에 또 다르게 접근하는 방식은 가능한 한 흩어져 있는 점들에 가깝게 선을 맞추는 것일 것이다. 예를 들어, <그림 2.16>에서 선 AB가 가장 잘 흩어져 있는 점들에 가까이 갈 수 있도록 $\hat{a}$과 $\hat{b}$를 골랐다고 가정하자. 이런 측면에서 그 점 중에서 P_1과 같은 어떤 점을 고려하여 보자. 교란항 u_t 때문에 그 점은 일반적으로 정

〈표 2.7 〉

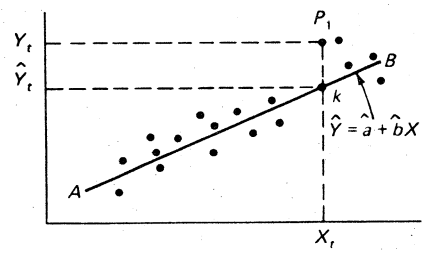
연 도	C	$\hat{C}$	$\hat{u} = C - \hat{C}^a$	$\hat{u}^2 = (C - \hat{C})^2$
1960	325	325	0	0
1961	335	337	− 2	4
1962	355	356	− 1	1
1963	375	373	2	4
1964	401	403	− 2	4
1965	433	434	− 1	1
1966	466	469	− 3	9
1967	492	500	− 8	64
1968	537	538	− 1	1
1969	576	574	2	4
				$\sum(C - \hat{C})^2 = 92$

a) 반올림관계로 $\hat{u}$의 합계는 0이 아니다.

확하게 선에 놓여 있지는 않을 것이고, 전형적으로는 AB의 위 또는 밑에 있을 것이다. P_1일 경우 P_1k선의 길이가 되는 점과 직선간의 수직거리는 X_t에 대응하는 Y의 관찰치인 Y_t와 Y를 계산한 값인 $\hat{Y}_t$ 사이의 차이를 나타 낸다. 여기서 $\hat{Y}_t$는 AB로 나타나는 추정 관계로부터 얻는다.

$$P_1k = (Y_t - \hat{Y}_t) = (Y_t - \hat{a} - \hat{b}X_t) \tag{2.103}$$

선 AB가 관찰한 점의 선으로부터의 편차를 최소화하기를 원한다고 가정하자. 단순하게 이들 편차의 합계를 최소화하는 데 따르는 한가지 어려움은 선 위의 점은 $(Y_t - \hat{Y}_t) > 0$인 반면에 선 밑의 점은 $(Y_t - \hat{Y}_t) < 0$인 것이다. 이는 선에서 상당히 멀리 떨어진 점들을 가지면서도,



〈그림 2.16〉

측정한 편차의 산술적 합계가 매우 작을 수가 있음(심지어 0도 됨)을 의미한다. 사실 이 합계는 선 AB를 가능한 한 높게 위치시킴으로써 최소화할 수 있었다. 그 이유는 이렇게 하면 $\sum(Y_t - \hat{Y}_t)$가 상당한 음의 값을 가지게 될 것이기 때문이다. 이러한 난점을 우회하는 방법은 이들 편차를 모두 자승하여(편차를 모두 양으로 만든다), 그 자승의 합계를 최소화하는 것이다. 이것이 a와 b의 추정량을 발생시키는 최소자승법(le-

ast-squares method) 이다. 즉, 산포도에서 다음과 같은 합계

$$S = \sum (Y_t - \hat{Y}_t)^2 = \sum (Y_t - \hat{a} - \hat{b}X_t)^2$$

가 최소화되는 선을 발견하는 방법이다.

미분을 이용하여 $\sum (Y_t - \hat{Y}_t)^2$ 을 최소화하는 $\hat{a}$과 $\hat{b}$의 값을 결정하는 것은 간단한 문제이다. 만일 독자들이 미분에 익숙하다면, 그 도출을 하여 보기 바란다. 이 장의 부록에 그 내용이 포함되어 있다. 독자들이 알아야 할 것은 a와 b의 최소자승 추정량을 계산할 경우 대변수 방식으로 도달한 결과와 정확하게 같은 결과를 얻는다는 점이다. 즉, 최소자승 추정량 또한 다음과 같다.

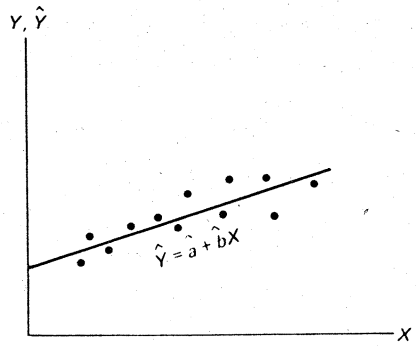
$$\hat{b} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

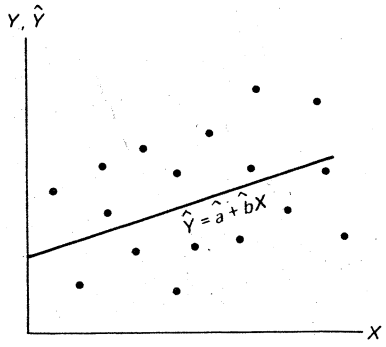
경제학 문헌에서 "최소자승"으로 추정하였음을 가리키는 식들을 자주 보게 되므로 윗식은 중요하다. 또한 윗식이 2장에서 전개하였던 추정 절차를 사용하여 얻어지는 식과 동일하다는 것도 깨달아야 한다. 마찬가지로 우리의 추정 절차도 Y의 계산한 값으로부터의 관찰치가 갖는 편차를 자승한 것의 합계를 최소화한 결과를 가져 온다.

6. 회귀모형이 갖는 설명력 (explanatory power) 의 측정

이제 두 변수 사이의 평균 관계라고 부르는 것을 추정하는 기법을 알았다. 즉, 회귀모형에서 모수 a와 b의 추정량을 얻을 수 있다. 그러나 아직은 이 관계가 얼마나 "빈틈이 없는지" (tight)를 측정하는 수단을 가지지 못했다. 예를 들어 <그림 2.17>과 <그림 2.18>에서는 Y와



〈그림 2.17〉



〈그림 2.18〉

X 사이의 추정한 관계가 동일하다. 그러나, 이 두개의 관계는 한가지 중요한 측면에서 다르다. 즉, 두번째 경우에서보다 첫번째 경우에서 흩어져 있는 점들이 선에 훨씬 더 가까운 것이다. 다른 말로 하면, 회귀선 주위로 관찰한 점들을 "더 빈틈없게 맞춘 것"을 얻은 것이다.

a와 b의 추정량에 더하여, X와 Y 사이의 관계가 갖는 이상의 측면을 측정하는 수단의 개발이 중요하다.* 어떤 뜻으로서는 진정 얼마나 모

* 실제로 이미 우리는 이것의 한가지 측정수단을 알고 있다. 교란항의 분산이 그것이다. 예를 들어, <그림 2.17>은 <그림 2.18>보다 교란항의 분산이 더 적음을 시사한다.

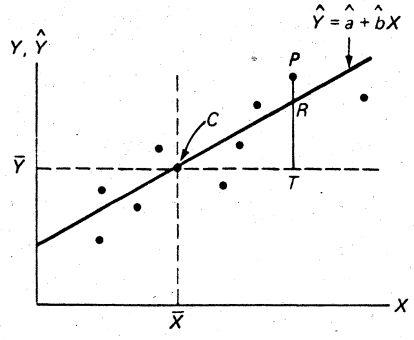
형이 좋은 것인가를 알기 원하는 것이다. 생각컨데, 우리는 이 모형을 가지고서 Y_t의 관찰치를 설명하려는 중이다. 만일 모형이 없다면, Y의 움직임을 설명할 수 없다. 이 경우에 할 수 있는 일은 기껏해야 Y_t의 예측치로 X_t값과 무관하게 $\overline{Y}$를 취하는 정도일 것이다. 문제는 우리의 모형이 이렇게 평균을 취하는 것보다 나은 것인지 그리고 만일 더 좋다면 얼마나 좋은 것인지 이다. 이러한 이유로 회귀모형이 갖는 설명력을 측정하는 수단을 이제부터 살펴보기로 한다. 즉, X와 Y 사이에 추정한 선형 관계가 Y의 변동을 얼마만큼이나 설명할 수 있는가에 관한 측정수단을 전개할 것이다.

가. 결정계수(coefficient of determination)

<그림 2.19>의 흩어져 있는 점들을 보면, 우리가 추정한 기법에 의하여 선 $\hat{Y} = \hat{a} + \hat{b}X$가 맞추어져 있다. 또한 그림에는 X와 Y의 표본 평균인 $\overline{X}$와 $\overline{Y}$가 표시되어 있다. (2.45)로부터 추정식의 한 속성이 다음과 같다는 점을 상기하자.

$$\overline{Y} = \hat{a} + \hat{b}\overline{X} \tag{2.45}$$

윗식은 회귀선이 <그림 2.19>의 점 C인 ($\overline{X}$, $\overline{Y}$)점을 통과한다는 것을 말하여 준다. 다음으로 P로 제시되는 관찰치를 보자. P에서의 Y값의 그 표본 평균값인 $\overline{Y}$과의 편차는 거리 PT이다. 그러나 이 편차중의 일부는 추정한 회귀식으로 "설명된다." 특히, 추정식은 편차중의 RT를 설명하고, PR을 "설명하지 않은 채" 남겨 둔다. 이러한 여러 거리를 다음과 같이 표현할 수 있다.



〈그림 2.19〉

$PT = (Y_t - \overline{Y}) = Y$의 표본 평균으로부터의 총편차

$RT = (\hat{Y}_t - \overline{Y}) = \overline{Y}$로부터의 Y_t의 설명된 편차

$PR = (Y_t - \hat{Y}_t) = \overline{Y}$로부터의 Y_t의 설명안된 편차

이상의 기초적인 지식을 가지고서 이제 회귀식이 갖는 설명력의 측정 수단을 보도록 하겠다. 맨 먼저 (2.38)식으로 부터 다음을 상기한다.

$$Y_t = \hat{Y}_t + \hat{u}_t \tag{2.38}$$

(2.38)의 양변을 모두 합하면,

$$\sum Y_t = \sum \hat{Y}_t + \sum \hat{u}_t \tag{2.104}$$

그러나, $\sum \hat{u}_t = 0$라는 조건을 부과하였기 때문에,

$$\sum Y_t = \sum \hat{Y}_t \tag{2.105}$$

이에 따라서 만일 양변을 n으로 나누면, 다음을 의미한다.

$$\bar{Y} = \bar{\hat{Y}} \tag{2.106}$$

이들 결과는 뒤에서 유용한 결과가 될 것이다. (2.38)로 이제 돌아가면, $Y_t = \hat{Y}_t + \hat{u}_t$ 인데, 양변을 자승한다.

$$Y_t^2 = \hat{Y}_t^2 + \hat{u}_t^2 + 2\hat{u}_t\hat{Y}_t \tag{2.107}$$

표본내 모든 값을 더하면

$$\sum Y_t^2 = \sum \hat{Y}_t^2 + \sum \hat{u}_t^2 + 2\sum(\hat{u}_t\hat{Y}_t) \tag{2.108}$$

여기서 다음을 알 수 있다.

$$\sum(\hat{u}_t\hat{Y}_t) = 0$$

그 이유는 우리의 추정 절차에서 다음과 같은 조건을 부과한 적이 있기 때문이다.

$$\sum(\hat{u}_t X_t) = 0 \quad \text{그리고} \quad \sum \hat{u}_t = 0$$

$\hat{Y}_t = \hat{a} + \hat{b}\,X_t$ 이므로

$$\sum(\hat{u}_t\hat{Y}_t) = \hat{a}\sum \hat{u}_t + \hat{b}\sum(\hat{u}_t X_t) = 0$$

윗식은 앞의 (2.108)의 마지막 항이 0 임을 의미하는데, 그 결과 (2.108)은 다음과 같이 간단히 된다.

$$\sum Y_t^2 = \sum \hat{Y}_t^2 + \sum \hat{u}_t^2 \qquad (2.109)$$

다음에 (2.109)의 양변에서 $n\overline{Y}^2$ 를 빼주면,

$$\sum Y_t^2 - n\overline{Y}^2 = (\sum \hat{Y}_t^2 - n\overline{Y}^2) + \sum \hat{u}_t^2 \qquad (2.110)$$

앞서의 결과인 $\overline{Y} = \overline{\hat{Y}}$ 을 상기하면, 이제 (2.110)을 다음의 형식으로 나타낼 수가 있다.

$$\sum (Y_t - \overline{Y})^2 = \sum (\hat{Y}_t - \overline{Y})^2 + \sum \hat{u}_t^2 \qquad (2.111)$$

식(2.111)은 곧 우리의 목적 달성에 극히 유용하다는 것이 입증될 것이고, 이 표현 가운데의 각 항이 의미하는 것을 세밀하게 검토하는 것이 중요한 일이다. 첫째로, (2.111)의 좌변은 Y_t의 자신의 표본 평균으로부터의 편차를 자승한 것의 총합이다. 이 항은 우리가 회귀식을 가지고 설명하려는 종속변수의 변동을 측정하는 것이다. 어느 정도 직관적으로 보면, 우리의 모형은 종속변수가 항상 불변이 안되는 이유를 설명할 수 있도록 요망된다. 특히, 우리의 변수 Y의 움직임이 변수 X의 움직임과 관련되거나 설명되기를 바라고 있다. 어쨌든, (2.111)의 좌변을 **총자승합** (total sum of squares ; TSS)이라고 부르자.

이제 (2.111)의 우변을 보자. $\hat{u}_t = (Y_t - \hat{Y}_t)$이므로, $\hat{u}_t$는 우리가 설명하는데에 실패한 것이다. 즉, $\hat{u}_t$은 회귀식 즉, $\hat{Y}_t = (\hat{a} + \hat{b} X_t)$ 으로 계산한 값으로부터 벌어진 Y_t의 관찰치가 갖는 편차이다. (2.111)의 마지막 항 $\Sigma \hat{u}_t^2$은 **오차** 또는 Y_t의 설명되지 못한 부분을 자승한 것의 합계이다. 이 합계를 **오차자승합** (error sum of squares : ESS)이라고

부른다. TSS와 ESS의 차이는 (2.111)에 따르면, $\Sigma(\hat{Y}_t - \bar{Y})^2$ 이고, 이 항은 분명하게 우리의 회귀모형이 설명하고 있는 총자승합의 일부를 나타내는 것임이 틀림없다. 이 항을 회귀자승합(regression sum of squares ; RSS , 즉 회귀모형이 설명한 자승합)이라고 부른다. (2.111)은 따라서 다음과 같다.

$$TSS = RSS + ESS \qquad (2.112)$$

이상의 관계를 〈그림 2.19 〉로 보는 것이 도움이 될 것이다. 점 P로 표시된 Y_t를 보면, PT가 Y_t의 그 표본 평균, $\bar{Y}$로부터의 편차를 가리킨다는 것을 앞에서 알았다. PR은 Y_t의 회귀선으로부터의 편차, 다른 말로 하면, 회귀선이 설명할 수 없는 $\bar{Y}$로부터의 Y_t의 변동의 일부이다. 또한 PT의 나머지 성분, 즉 RT는 Y_t의 회귀선이 설명하는 변동 부분이다. (2.111)에서 본 것은 표본의 각 점에서 PT, PR과 RT에 상당하는 거리를 고려한다면, RT에 상당하는 거리들의 자승합은 PR에 상당하는 거리들의 자승합에 RT에 상당하는 자승합을 더한 것과 동일하다.

요약하면, 다음과 같다.

$$TSS = \sum (Y_t - \bar{Y})^2 = \text{총자승합}$$

$$RSS = \sum (\hat{Y}_t - \bar{Y})^2 = \text{회귀 (또는 설명된) 자승합}$$

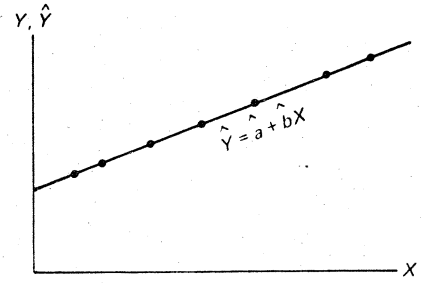
$$ESS = \sum \hat{u}_t^2 = \text{오차 (설명되지 않은) 자승합}$$

여기서 $TSS = RSS + ESS$, $RSS \geq 0$이고 $ESS \geq 0$이기 때문에, $TSS \geq RSS$ 그리고 $TSS \geq ESS$이다.

회귀식의 설명력을 측정하기 위하여 Y_t의 회귀식이 설명할 수 있는 변동 부분을 가리키는 측정수단이 요망된다. 그러한 측정수단의 하나는 다음과 같다.

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} \qquad (2.113)$$

여기서 R^2를 결정계수(coefficient of determination)이라 한다. 만일 회귀식이 Y_t의 모든 변동을 설명할 수 있다면(즉, 모든 t에 대하여 $\hat{Y}_t = Y_t$이면), $\hat{u}_t = 0$이어서 $ESS = 0$이다. 이 경우에 $RSS = TSS$이어서, $R^2 = 1$이다. $Y_t = \hat{Y}_t = \hat{a} + \hat{b}X_t$는 X_t의 완전한 선형결합이기때문에, Y_t와 X_t 사이의 산포도상의 모든 점은 한 직선에 놓여 있을 것이다. (즉, 교란항이 모두 0이 될 것이다.) 〈그림 2.20〉는 그러한 예를 묘사하고 있다.



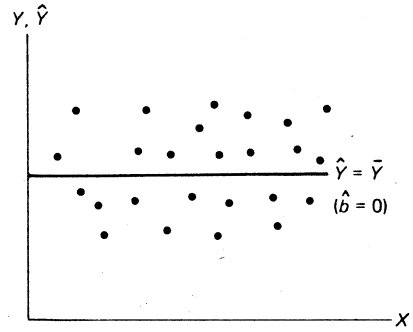
〈그림 2.20〉

다른 극단적인 예로서, 만일 회귀식이 아무것도 설명하지 못한다면, ESS는 그 최대값, 즉 $ESS = TSS$를 취할 것이다. $RSS = 0$이 되어 $R^2 = 0$임을 알 수 있다. 그러한 경우에는 $RSS = 0$이기 때문에, 모든 t에 대하여 $\hat{Y}_t = \overline{Y}$임이 틀림없다. 이는 곧 $\hat{b} = 0$을 의미한다. 이것을 보려

면, $\hat{a} = \overline{Y} - \hat{b}\,X$를 $\hat{Y}_t = \hat{a} + \hat{b}\,X_t$에 대입하여 다음을 얻는다.

$$\hat{Y}_t = \overline{Y} + \hat{b}(X_t - \overline{X}) \qquad (2.114)$$

모든 X_t의 값이 같지 않기 때문에 $X_t \neq \overline{X}$이어서, $\hat{Y}_t = \overline{Y}$는 $\hat{b} = 0$을 의미한다.



〈그림 2.21〉

이러한 경우에 계산한 또는 설명한 Y_t의 값, 즉 $\hat{Y}_t$가 변수 X_t의 값에 전혀 종속하지 않는다는 점에서 모형은 전적으로 부적절하다. 〈그림 2.21〉이 그러한 상황을 하나 그린 것이다.

〈그림 2.17〉과 〈그림 2.18〉에서 보인 것과 같은 더 전형적인 경우에서는 회귀식이 Y의 모든 변동이 아닌 일부 변동을 설명할 것이다. R^2은 0과 1사이에 있을 것이다. 회귀식이 Y의 변동을 더 설명할수록 (즉, 점들이 회귀선에 더 가까이 맞추어 질수록) R^2은 1에 더 가까이 간다. 그리고 X와 Y의 관계가 더 약해지면, R^2은 0에 더 가까워 진다. 따라서, R^2은 추정한 식이 설명할 수 있는 종속변수의 변동부분을 가리킨다. 결과적으로 예를 들면 $R^2 = 0.63$이란 추정한 관계가 종속변수의 변동을 63% "설명"할 수 있음을 말한다.

나. $R^2 = \hat{\rho}^2_{Y,\hat{Y}}$

이 장의 1절에서 두 변수의 선형 연관이 갖는 "강도"의 측정수단을 전개하였다. 이를 상관계수라고 부른 바 있다. 이 모수는 두 변수의 표준편차의 곱에 대한 두 변수의 공분산의 비율이라고 정의하였다. 아마도 상관계수를 회귀식에서의 관계가 갖는 강도를 측정하는 데도 사용할수 있을 것이다. 예를 들어 Y_t와 $\hat{Y}_t$사이의 상관계수, $\rho_{Y,\hat{Y}}$은 Y_t와 $\hat{Y}_t$이 얼마나 밀접하게 관련되었는가에 대한 측정수단이 될 것이고, 따라서 모형이 얼마나 Y_t의 값들을 "설명"할 수 있는가의 측정수단이 된다.

불행히도, $\rho_{Y,\hat{Y}}$는 일반적으로 잘 알지 못하는 것이다. 실제로는 추정하여야만 한다. 이 장의 1절에 있는 식(2.16)과 일치하는 $\rho_{Y,\hat{Y}}$의 추정량은 다음과 같을 것이다.

$$\hat{\rho}_{Y,\hat{Y}} = \frac{\sum (Y_t - \bar{Y})(\hat{Y}_t - \bar{Y})}{\sqrt{\sum (Y_t - \bar{Y})^2 \sum (\hat{Y}_t - \bar{Y})^2}} \qquad (2.115)$$

왜냐하면 $\bar{\hat{Y}} = \bar{Y}$이기 때문이다. 이제 $\hat{\rho}^2_{Y,\hat{Y}} = R^2$을 보이기로 한다. 즉, R^2통계량(statistic)은 단지 Y_t와 $\hat{Y}_t$ 사이의 상관계수에 불과하다. 그러므로, R^2과 $\hat{\rho}_{Y,\hat{Y}}$를 Y와 $\hat{Y}$ 사이 관계가 갖는 강도를 측정하는 대안적인 수단으로 볼 수가 없다.

먼저 (2.115)의 분자를 고려하자. 1장 부록 A로부터, 어느 두 변수를 Z_{1t}와 Z_{2t}라 하고 그 표본평균이 $\bar{Z}_1$과 $\bar{Z}_2$라 한다면, 다음과 같음을 상기하도록 하자.

$$\sum (Z_{1t} - \bar{Z}_1)(Z_{2t} - \bar{Z}_2) = \sum (Z_{1t} - \bar{Z}_1)Z_{2t}$$

(2.115)의 분자를 다음처럼 간단히 할 수 있다.

$$\sum (\hat{Y}_t - \bar{Y})Y_t$$

왜냐하면 $Y_t = \hat{Y}_t + \hat{u}_t$ 이라는 것을 알고 있기 때문에, 우리는

$$\sum (\hat{Y}_t - \bar{Y})Y_t = \sum (\hat{Y}_t - \bar{Y})(\hat{Y}_t + \hat{u}_t) = \sum (\hat{Y}_t - \bar{Y})\hat{Y}_t$$

로 쓸 수 있다.

그 이유는

$$\sum (\hat{Y}_t\hat{u}_t) = 0 \quad \text{그리고} \quad \sum (\hat{u}_t\bar{Y}) = \bar{Y} \sum \hat{u}_t = 0$$

이기 때문이다.

마지막으로 분자를 다음의 형식으로 놓을 수 있다.

$$\sum (\hat{Y}_t - \bar{Y})\hat{Y}_t = \sum (\hat{Y}_t - \bar{Y})(\hat{Y}_t - \bar{Y}) = \sum (\hat{Y}_t - \bar{Y})^2 = \text{RSS}$$

다음, $\hat{\rho}_{Y,\hat{Y}}$ 의 분모는

$$\sqrt{\sum (Y - \bar{Y})^2 \sum (\hat{Y}_t - \bar{Y})^2} = \sqrt{(\text{TSS})(\text{RSS})}$$

그러므로

$$\hat{\rho}_{Y,\hat{Y}} = \frac{\text{RSS}}{\sqrt{(\text{RSS})(\text{TSS})}} = \frac{\sqrt{\text{RSS}}}{\sqrt{\text{TSS}}} \qquad (2.116)$$

(2.116)으로부터, 다음을 알 수 있다.

$$R^2 = \hat{\rho}^2_{Y,\hat{Y}}$$

다. 보기

설명을 위해서 이 장의 앞에서 추정하였던 소비함수로 돌아가서 R^2 의 값을 찾는 것이 유용하다.

$$R^2 = \frac{\sum (\hat{Y}_t - \bar{Y})^2}{\sum (Y_t - \bar{Y})^2} = \frac{\sum (\hat{Y}_t - \bar{Y})\hat{Y}_t}{\sum (Y_t - \bar{Y})Y_t} = \frac{\sum \hat{Y}_t^2 - n\bar{Y}^2}{\sum Y_t^2 - n\bar{Y}^2}, \qquad (2.117)$$

앞서 보였듯이 $\bar{Y}=\bar{\hat{Y}}$ 이기 때문이다.

R^2 의 계산에서 한 걸음 더 나가는 것은 Y 의 계산값(즉, $\hat{Y}_t$)을 계산하는데 추정한 회귀식을 사용해야 한다는 것이다. 그 계산내용은 〈표 2.8 〉에 있다.

〈표 2.8 〉[a]

$$\hat{C} = 13 + 0.89\ Y_d$$

C	Y_d	C^2	$\hat{C}$	$\hat{C}^2$
325	350	105,625	325	105,625
335	364	112,225	337	113,569
355	385	126,025	356	126,736
375	405	140,625	373	139,129
401	438	160,801	403	162,409
433	473	187,489	434	188,356
466	512	217,156	469	219,961
492	547	242,064	500	250,000
537	590	288,369	538	289,444
576	630	331,776	574	329,476

$$\Sigma C^2 = 1{,}912{,}155$$

$$\Sigma \hat{C}^2 = 1{,}924{,}705$$

$$\bar{C}^2 = 184{,}900$$

$$n\bar{C}^2 = 1{,}849{,}000$$

$$R^2 = \frac{\Sigma \hat{C}_t{}^2 - n\bar{C}^2}{\Sigma C_t{}^2 - n\bar{C}^2} = 0.99$$

a) 회귀식에서 계수의 값이 단지 소숫점 둘째자리까지 반올림하여 사용하였기 때문에, 위의 숫자로 만든 R^2 의 값은 반올림에 따른 오차의 결과로 1을 약간 넘는다. 만일 계수를 충분한 자릿수까지 계산한다면, R^2 의 값은 0.99로 된다.

추정한 소비함수에서 R^2의 값은 0.99이고, 이 수치는 C와 Y_d 사이의 극도로 강한 연관을 가리키는 것이다. 즉, 추정한 회귀식은 C의 변동의 99%를 설명하고, 오직 1%만이 "설명안된 채" 남아 있다. 이는 일찌기 단지 〈그림 2.14〉에서 산포도와 회귀선으로 내린 시험적인 결론을 확고히 해 준다.

7. 實 例 : 비용함수의 추정

경제학 문헌의 실제 연구를 통해 이변수모형에 관한 서론의 결론을 맺기로 한다. 기업의 미시경제학 이론에 대한 핵심적인 관심사는 기업의 비용함수(cost function)이다. 특히, 거의 대부분의 교과서는 비용과 기업의 산출량 수준 사이의 관계를 길게 검토하고 있다. 일반적인 용어로,

$$C = f(Q)$$

여기서 C는 총비용, Q는 기업의 산출량이다. 이러한 형식에 대하여 보다 깊이 알기 위해서 일부 경제학자는 비용과 산출량에 관한 실제 자료를 가지고서 비용함수를 추정하는 데에 회귀분석을 사용하였다.

이러한 유형의 초기적이고 선구적인 노력은 실크 스타킹을 제조하는 공장의 비용함수를 다룬 조엘딘(Joel Dean)의 연구였다.* 딘은 양말류 제조공장의 스타킹 제조 비용과 산출량에 관한 月間 자료를 수집하였다. 〈그림 2.22〉는 散布圖로 그러한 자료를 그려 놓은 것이다. 그림에서 점들이 직선에 가깝게 뭉쳐있는 것으로 보아 선형함수가 공장의 비용함수를 잘 묘사하고 있음을 알 수 있다. 딘은 기본적인 이변수 회귀모형을 사

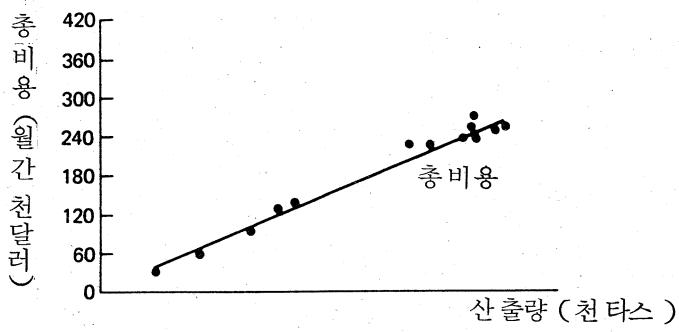
---

* Joel Dean, "Statistical Cost Functions of a Hosiery Mill," Journal of Business:Studies in Business Administration, Vol. XI, No.4 (July 1941), pp.1 ~ 116.

용하여 다음과 같이 가정하였다.

$$C_i = a + bQ_i + u_i \qquad (2.118)$$

여기서 C는 천달러 단위로 측정한 월간 총비용이고, Q는 천타스를



〈그림 2.22〉

출전 : Joel Dean, " Statistical Cost Functions of a Hosiery Mill, " Journal of Business : Studies in Business Administration, Vol. XI, No.4 (July 1941), p.101.

단위로 측정한 월간 산출량이다.* 또한 u는 교란항이다. 부수적으로 식(2.118)은 공장의 비용이 갖는 본성에 대해서 중요한 시사점을 갖고 있다. 그 식은 총비용함수이어서, 기대하는 월간한계비용은 월간 비용의 b단위이다. 우리의 측정 단위에 따라서, 이러한 사실은 만약 월간 산출량이 천 타스씩 증가하면, 월간 비용은 " b "천 달러씩 증가함을 의미한다. 이는 다시, 월간 산출량이 1 타스 증가하면 월간 비용이 b 달러 증가함을 의미한다. 모수 " a "는 월간 산출량이 0 일 경우 월간 비용이 " a "천 달러가 됨을 가리킨다. 이 " a " 천 달러가 기업의 월

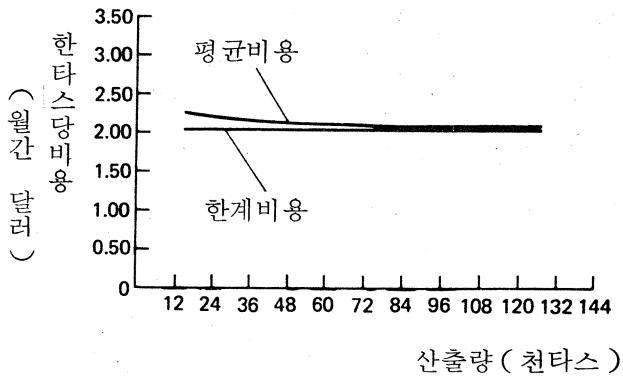
* 이러한 측정의 예로서, $C = 27$ 의 값은 27,000 달러에 상당한다. $Q = 17$ 의 값은 17,000 타스 또는 $(17{,}000 \times 12) = 204{,}000$ 켤레의 스타킹에 상당한다.

간 고정비용이다. 만약 a가 양이면, 공장의 평균 월간 비용은 산출량에 따라 감소한다. 왜냐하면 고정비용은 산출량의 단위가 매우 큰 수가 됨에 따라 덜어지기 때문이다.

딘은 이 장에서 전개한 대변수 기법과 동등한 최소자승법을 사용하여 회귀식 (2.118)을 추정하여, 다음의 결과를 얻었다.

$$C = 2.936 + 2.00Q \qquad (2.119)$$
$$R^2 = 0.95$$

산포도는 적합도가 밀접함을 보여 주었다. 결정계수가 0.95이어서 식 (2.119)는 총비용에서 관찰한 변동을 95% 설명한다. 더구나, 계수들의 추정치는 특정한 비용 정보를 제공한다. 즉, 계수들은 공장의 고정비용이 매월 2,936달러이고 스타킹 한 타스의 한계비용은 2 달러임을 가리킨다. 〈그림 2.23 〉은 이러한 비용함수를 그린 것이다. 이러한 정보는 비용 관계를 연구하는 경제학자뿐만 아니라 기업의 경영진에게도 상당히 귀중한 것이다.



〈그림 2.23 〉

출전 : Joel Dean, " Statistical Cost Functions of a Hosiery Mill," Journal of Business : Studies in Business Administration, Vol. XI, No.4 (July 1941), p.101.

부록. 세 命題의 증명

앞에서는 세가지 명제를 말로만 써보았다. 그 하나는 상관이 없는 확률변수의 합계가 갖는 분산이고, 또 하나는 대변수 추정량이 갖는 최소분산의 속성, 마지막 하나는 추정량이 갖는 최소 자승의 속성이다. 여기서는 이들 명제의 증명을 보이고자 할 것이다.

가. 확률변수의 합이 갖는 분산

X_1, X_2, …, X_n을 확률변수라 하고, a_1, a_2, …, a_n을 상수라 한다. 여기서 각 변수의 평균과 분산은 각각 μ_1, μ_2 …, μ_n과 σ_1^2, σ_2^2 …, σ_n^2이다. 다음에 아래를 정의한다.

$$Y = a_0 + a_1X_1 + \cdots + a_nX_n \qquad (2A.1)$$

Y는 확률변수 집합의 선형결합에 불과하기 때문에, 그 자신도 확률변수이다. 다음을 증명하기로 한다.

명제 I : X_1, X_2, …, X_n이 상관이 없는 변수라고 하고, Y의 분산을 σ_Y^2라 할 경우,

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \qquad (2A.2)$$

윗식은 만일 X가 서로간에 선형의 관계가 없다면, Y의 분산이 X 각각의 계수를 자승하여 곱한, X의 분산의 합과 같음을 말한다.

명제 I을 증명하려면, 먼저 Y의 평균 $E(Y) = \mu_Y$가 (2A.1)식의 우변이 갖는 평균과 같다는 것을 알아야 한다.

$$\begin{aligned}\mu_Y &= E(a_0) + E(a_1X_1) + \cdots + E(a_nX_n) \\ &= a_0 + a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n\end{aligned} \qquad (2A.3)$$

다음으로 정의에 의하여 Y의 분산은 $E(Y-\mu_Y)^2$이다. (2A.1)과 (2A.3)을 사용하면,

$$\sigma_Y^2 = E(Y-\mu_Y)^2 = E[a_1(X_1-\mu_1) + a_2(X_2-\mu_2) + \cdots + a_n(X_n-\mu_n)]^2 \quad (2A.4)$$

식 (2A.4)를 전개하면, 두 유형의 항—즉, 자승항과 서로 곱한 항—을 얻는다.

$$\sigma_Y^2 = E[a_1^2(X_1-\mu_1)^2 + a_2^2(X_2-\mu_2)^2 + \cdots + a_n^2(X_n-\mu_n)^2 + \cdots + 2a_3a_4(X_3-\mu_3)(X_4-\mu_4) + \cdots] \quad (2A.5)$$

(2A.5)의 마지막 항은 전형적으로 교차하여 곱한 준 항이다. 이러한 항은 n이 얼마나 큰 가에 따라서 엄청난 수가 될 것이다. 문제의 열쇠는 X가 가정에 의하여 서로 상관하지 않기 때문에 이들 교차하여 곱해 준 항 각각의 기대값이 0이 될 것이라는 점이다. (2A.5)의 교차하여 곱해 준 항을 예로 들어 보이면,

$$E[2a_3a_4(X_3-\mu_3)(X_4-\mu_4)] = 2a_3a_4E(X_3-\mu_3)(X_4-\mu_4) = 2a_3a_4\,\mathrm{cov}(X_3, X_4) = 0 \quad (2A.6)$$

윗식에서 Cov(X_3, X_4)는 X_3와 X_4 사이의 공분산이다. 교적항(cross-product term)의 기대값이 0과 같으므로, 식 (2A.5)는 아래와 같이 된다.

$$\begin{aligned}\sigma_Y^2 &= E[a_1^2(X_1-\mu_1)^2 + a_2^2(X_2-\mu_2)^2 + \cdots + a_n^2(X_n-\mu_n)^2] \\ &= a_1^2E(X_1-\mu_1)^2 + a_2^2E(X_2-\mu_2)^2 + \cdots + a_n^2E(X_n-\mu_n)^2 \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2\end{aligned} \quad (2A.7)$$

나. a와 b의 最小分散 추정량*

다음으로, 대변수 추정기법이 회귀모형의 계수에 대해서 불편측정량 $\hat{a}$과 $\hat{b}$을 만들어 내는 것을 보았다. 이제 대변수 추정량 $\hat{a}$과 $\hat{b}$이 a와 b의 모든 불편 선형 추정량중에서 가장 작은 분산을 가짐을 보이기로 한다. 이것은 문헌상에 가우스-마르코프 정리(Gauss-Markov Theorem)로 알려져 있다.

먼저 앞에서,

$$\hat{b} = \frac{\sum (X_t - \bar{X})Y_t}{\sum (X_t - \bar{X})^2} = \frac{\sum w_t Y_t}{A} = \sum Q_t Y_t \qquad (2A.8)$$

여기서 $Q_t = w_t / A$, $w_t = (X_t - \bar{X})$, $A = \Sigma (X_t - \bar{X})^2$. 마찬가지로 앞에서,

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = \sum \gamma_t Y_t \qquad (2A.9)$$

여기서 $\gamma_t = (1/n) - \bar{X}(w_t / A)$. 앞에서처럼 γ_t와 Q_t는 오로지 X_t의 값에만 종속한다. 그러므로, 만약 X_t의 값이 주어진다면, γ_t와 Q_t 값을 상수로 취급할 수 있다.

다음으로 가중치 Q_t가 갖는 두가지 속성을 설정할 필요가 있다. 먼저, 다음에 유의하자.

$$\sum Q_t = \sum \left(\frac{X_t - \bar{X}}{A} \right) = \frac{1}{A} \sum (X_t - \bar{X}) = 0 \qquad (2A.10)$$

---

* 이 절은 다음에 근거한다.
J. Johnston, Econometric Methods, 2nd ed.(New York:McGraw-Hill, 1972), pp. 18-23.

둘째로,

$$\sum (Q_t X_t) = \frac{1}{A} \sum (X_t - \bar{X}) X_t = \frac{1}{A} \sum (X_t - \bar{X})^2 = \left(\frac{1}{A}\right) A = 1 \qquad (2A.11)$$

왜냐하면, 다음을 기억하고 있기 때문이다.

$$\sum (X_t - \bar{X}) X_t = \sum (X_t - \bar{X})(X_t - \bar{X}) = \sum (X_t - \bar{X})^2 = A$$

이 교재에서는 이미 다음을 보였다.

$$\sigma_b^2 = \frac{\sigma_u^2}{\sum (X_t - \bar{X})^2} \qquad (2A.12)$$

(2A.12)를 다음과 같이 바꿀 수 있다.

$$\sigma_b^2 = \sigma_u^2 \sum Q_t^2 \qquad (2A.13)$$

왜냐하면,

$$\sum Q_t^2 = \frac{\sum w_t^2}{A^2} = \frac{\sum (X_t - \bar{X})^2}{A^2} = \frac{A}{A^2} = \frac{1}{A} = \frac{1}{\sum (X_t - \bar{X})^2}$$

이제 $\hat{b}^*$를 b의 어떤 다른 선형 추정량이라 하자. 그러면,

$$\begin{aligned} \hat{b}^* &= \sum (Q_t + v_t) Y_t \\ &= \hat{b} + \sum v_t Y_t \end{aligned} \qquad (2A.14)$$

여기서 (Q_t처럼) v_t는 X_t의 어떤 함수이지 Y_t의 함수가 아니다. 식 (2A.14)는 $\hat{b}^*$이 $\hat{b}$에 무엇인가를 더하여 준 것과 같다는 것을 말하여 주는 것에 불과하다. 여기서 무언가란 $\hat{b}^*$와 $\hat{b}$의 차이이다. $\sum Q_t = 0$과 $\sum (Q_t X_t) = 1$이므로 다음과 같이 된다.

$$\begin{aligned} \hat{b}^* &= \sum (Q_t + v_t)Y_t = \sum (Q_t + v_t)(a + bX_t + u_t) \\ &= a \sum (Q_t + v_t) + b \sum (Q_t + v_t)X_t + \sum (Q_t + v_t)u_t \\ &= a \sum v_t + b + b \sum (v_t X_t) + \sum (Q_t + v_t)u_t \end{aligned} \qquad (2A.15)$$

$\hat{b}^*$의 기대값을 취하고, Q_t와 v_t의 값이 X값에만 종속하기 때문에 상수임을 기억하면,

$$E(\hat{b}^*) = a \sum v_t + b + b \sum (v_t X_t) \qquad (2A.16)$$

왜냐하면 $E(u_t) = 0$ 이기 때문이다.

$\hat{b}^*$이 b의 불편측정량이라고 한단면, 다음이 틀림없다.

$$E(\hat{b}^*) = b$$

$\hat{b}^*$은 불편적이므로 식(2A.16)에서 다음과 같은 결과가 되는 것이 틀림없음을 시사한다.

$$\sum v_t = 0 \quad \text{그리고} \quad \sum (v_t X_t) = 0 \qquad (2A.17)$$

이러한 정보를 가지고 (2A.15)식의 마지막 행을 다음과 같이 다시 쓸 수 있다.

$$\begin{aligned} \hat{b}^* &= b + \sum (Q_t + v_t)u_t \\ &= b + (Q_1 + v_1)u_1 + (Q_2 + v_2)u_2 + \cdots + (Q_n + v_n)u_n \end{aligned} \qquad (2A.18)$$

이제 이 부록의 첫 절에서 얻은 명제 I을 $\hat{b}^*$의 분산에 관한 식을 얻는 데에 사용할 수 있다. $\hat{b}^*$은 u_t의 선형결합이고 이들 u_t는 상호독립이어서 상관하지 않기 때문에 $\hat{b}^*$의 분산은,

$$\sigma_{b^*}^2 = (Q_1 + v_1)^2\sigma_u^2 + \cdots + (Q_n + v_n)^2\sigma_u^2 \qquad (2A.19)$$
$$= \sigma_u^2 \sum (Q_t + v_t)^2$$
$$= \sigma_u^2[\sum Q_t^2 + \sum v_t^2 + 2\sum (Q_t v_t)]$$
$$= \sigma_u^2[\sum Q_t^2 + \sum v_t^2]$$

왜냐하면,

$$2\sum (Q_t v_t) = \frac{2\sum (X_t - \bar{X})v_t}{A} = \frac{2}{A}[\sum (X_t v_t) - \bar{X}\sum v_t] = 0 \qquad (2A.20)$$

(2A.13)과 (2A.19)로부터 다음을 즉각적으로 알 수 있다.

$$\sigma_{b^*}^2 = \sigma_u^2 \sum Q_t^2 + \sigma_u^2 \sum v_t^2 \qquad (2A.21)$$
$$= \sigma_b^2 + \sigma_u^2 \sum v_t^2$$

만일 일부라도 $v_t \neq 0$이면, (2A.21)의 마지막 항은 명백히 陽이기 때문에, 다음과 같은 결과를 얻는다.

$$\sigma_{b^*}^2 \geq \sigma_b^2 \qquad (2A.22)$$

$\sigma_{b^*}^2$은 오직 $\Sigma v_t^2 = 0$, 즉 모든 $v_t = 0$일 경우에만 b_b^2과 같음에 유의하자. 이는 물론 $\hat{b}^* \equiv \hat{b}$일 때만이다. 따라서 어느 다른 b의 불편 추정량도 대변수 추정량의 조건부 분산보다 더 큰 조건부 분산을 가진다는 것을 알았다. 동일한 일반적인 접근방식을 사용해서 이 명제가 a의 대변수 추정량에 대해서도 유지됨을 보일 수 있다. 이것은 흥미를 갖는 독자들에게 연습문제로 남겨 놓는다.

다. $\hat{a}$과 $\hat{b}$의 최소자승 속성

앞서 말한 바 있듯이, 최소자승 추정량은 $\hat{a}$과 $\hat{b}$에 대하여 다음의 합

계를 최소화함으로써 도출된다.

$$S = \sum (Y_t - \hat{Y}_t)^2 = \sum (Y_t - \hat{a} - \hat{b}X_t)^2 \qquad (2A.23)$$

(2A.23)을 $\hat{a}$과 $\hat{b}$에 대하여 편미분하여 그 결과를 0과 같다고 하면, 아래와 같다.

$$\frac{\partial S}{\partial \hat{a}} = 2\sum (Y_t - \hat{a} - \hat{b}X_t)(-1) = 0$$
$$\frac{\partial S}{\partial \hat{b}} = 2\sum (Y_t - \hat{a} - \hat{b}X_t)(-X_t) = 0 \qquad (2A.24)$$

식(2A.24)에 $(-\frac{1}{2})$을 곱하여 간단히 하면,

$$\sum Y_t = n\hat{a} + \hat{b}\sum X_t \quad \text{또는} \quad \bar{Y} = \hat{a} + \hat{b}\bar{X} \qquad (2A.25)$$

그리고

$$\sum (Y_t X_t) = \hat{a}\sum X_t + \hat{b}\sum X_t^2 \qquad (2A.26)$$

(2A.25)는 우리의 첫번째 정규방정식 (2.45)와 같고, (2A.26)은 n으로 나누고 나면 두번째 정규방정식 (2.50)과 같기 때문에, 최소자승 추정량과 우리의 대변수 추정량은 동일하다.

문 제

1. 절편의 추정량인 $\hat{a}$이 다음과 같이 표현될 수 있음을 보여라.

$$\hat{a} = \sum \left(\frac{1}{n} - \bar{X}W_t \right) Y_t$$

여기서

$$W_t = \frac{(X_t - \bar{X})}{\sum (X_t - \bar{X})^2}$$

2. 다음의 회귀모형을 고려하자. $Y_t = a + b X_t + u_t$, 여기서 X_t 와 Y_t 에 대한 관찰치는 다음과 같다.

X_t	Y_t
4	8
2	6
3	5
1	7
2	4

a, b 와 σ_u^2 을 추정하라.

3. 한 소비 분석가가 산포도의 점 (C_t, Y_t) 가 직선에 있지 않기 때문에 소비함수 $C_t = a + b Y_t$ 가 쓸모없다고 한다. 또한, 그는 때때로 Y_t 가 증가하는데 C_t 가 감소하는 것도 지적한다. 그는 C_t 가 Y_t 의 함수가 아니라고 결론을 내린다. 그의 주장을 평가하라.

4. $Y = 5 - 3X$ 라 하자. 상관계수 $\rho_{x,y} = \sigma_{x,y} / \sigma_x \sigma_y$ 가 -1.0 임을 보여라.

5. 변수 X_1, X_2, X_3 가 $\sigma_1^2 = 1.0$, $\sigma_2^2 = 3.0$ 그리고 $\sigma_3^2 = 5.0$ 의 분산을 갖는다고 하자. 이들 변수가 독립적이라고 한다.

$Y = 13 - 2 X_1 + 3 X_2 - 10 X_3$ 라 하자. Y 의 분산을 구하라.

6. 중류와 상류소득 가구가 도시의 세금이 주위 교외지역의 세금보다 많기 때문에 도시를 떠난다고 한다. 우리는 일정 시점에서 다수 도시의 관련 자료를 갖고 있다. 회귀모형으로 이러한 가설을 정식화하라 (한가지 방식 이상이 있다./).

7. 다음의 모형을 고려하자.

$$Y_t = a_1 + b_1 X_t + u_t \tag{1}$$

여기서 교란항 u_t 는 다음과 같은 식으로 설명변수에 종속한다.

$$u_t = a_2 + b_2 X_t + \varepsilon_t \tag{2}$$

여기서 ε_t 는 X_t 와 독립적인 교란항이고 또한 우리의 모든 표준적인 가정들을 만족시킨다. $b_2 > 0$ 이라고 가정한다. (1)의 b_1 은 X_t 의 Y_t 에 미치는 영향을 과소평가함을 보여라.

8. 문제 7에서 (2)식을 다음의 식으로 바꾸었다고 하자.

$$u_t = a_2 + b_2 X_t^2 + \varepsilon_t$$

X_t 에 Y_t 를 관련시키는 이변수 회귀모형의 가정중에서 어느 하나라도 위배될 것인가?

9. 어린이의 나이와 키 사이의 관계를 묘사하는 회귀모형을 새로이 만들자. 그러한 모형이 단점을 갖는지 아닌지를 논의하자.

10. 다음의 모형을 보자.

$$Y_t = a + bX_t + u_t$$

여기서 X_t 를 직접 관찰하지 못함으로써 측정상의 오차가 있다. 대신에 우리는 다음을 관찰하였다고 가정하기로 한다.

$$X_t^m = X_t + \varepsilon_t$$

여기서 ε_t 는 X_t 에 독립적인 교란항이고 0의 평균을 가지며, 우리의 여타 표준적인 가정을 모두 만족시킨다. 덧붙여서 ε_t 와 u_t 가 독립적이

라고 가정한다. 이것은 X_t^m과 u_t의 독립성을 의미한다.

a. X_t^m에 Y_t를 관련시키는 회귀모형을 새로 만들어라.

b. 이 모형의 기본 가정중에서 위배되는 것이 있는가?

제 3 장 회귀모형의 응용

앞장에서는 기본적인 二變數模型을 설명하고 그것의 推定量을 추정하는 技法을 살펴보았다. 관련되는 두 변수의 관찰치로 이루어진 표본을 사용하여서 이제는 回歸式의 a와 b의 추정량 그리고 그 추정량의 分散을 얻을 수 있다. 또한 추정한 式이 설명할 수 있는 종속변수의 변동 비율을 계산함으로써 두 변수 사이의 관계가 갖는 强度도 측정할 수가 있다. 이 章에서는 經濟학자들이 이 기법들을 경제행위에 대한 假說檢定과 豫測 또는 예상하는 데에 어떻게 사용하는가를 알아 보기로 한다.

1. 假說檢定과 信賴區間: 입문

2 장에서는 소비지출이 가처분소득 수준에 종속한다는 가설을 설명하고서, 이 관계의 線型을 추정하였다. 미국의 총체적인 시계열 자료를 이용하여서 한계소비성향에 대한 추정치가 陽이고 0과 1 사이의 값을 갖는다는 것을 알았다. a에 관한 추정치도 陽이었다. 그 결과는 소비함수에 대한 표준적인 케인즈주의자의 이론과 일치한다고 시사한 바 있다.

그러나 이상의 결과는 얼마나 믿을 만한가? 예를 들어 母數 a가 실제 陽인지를 얼마나 확신할 수 있는가? 한 예로서 만약 a의 추정치가 0.001이었다면 a가 0이 아니라 陽이라고 확신할 수 있었겠는가?

달리 앞서의 정보를 기초로 $b = 0.75$임을 믿을 만한 근거가 있었다고 가정하자. 우리의 b에 대한 추정치 즉 0.89는 $b = 0.75$라는 이전의 가설과 불일치하는 것인가? 요점은 추정치 0.89와 가설화한 $b = 0.75$ 사이의 不一致가 추정치를 얻기 위하여 사용한 標本의 크기로 인하여 생겼을 지도 모른다는 점이다. 비슷하게, 만약 미국의 성인 남자들의 평균 키가 5피트 10인치라고 하여도 어느 한 집단, 예를 들어 임의로 세 사람

을 골랐을 때의 평균 키가 정확하게 5피트 10인치라고 기대하지 않는다.

이상이 가설검정의 문제이다. 본질적으로 이들 문제에서는 모수의 추정치가 以前의 가설과 일치하는지의 여부에 관심을 가진다. 이와 밀접하게 관련된 문제가 신뢰구간의 문제이다. 예를 들어 b의 특정한 추정치 0.89가 있다고 하자. 이것을 點推定値(point estimate)라고 한다. 그러나 이 추정치를 해석하여야만 한다면, 아마도 한계소비성향이 0.89 "근처"라고 말할 것이다. 즉, b의 값이 정확히 0.89라고 기대하지 않는 것이다.

그러므로 점추정치에 대하여 일정한 확신을 가지고서 그 b값을 포함하고 있다고 느끼는 區間을 만드는 일이 필요하다. 아래에서 전개할 신뢰구간의 이론이 바로 그것이다. 이 이론은 우리로 하여금 區間推定値(interval estimate)를 만들어 내기 위해서 점추정치를 확장할 수 있게 한다. 그러한 구간이, 구간을 설정하는 방식 때문에 일정한 신뢰 수준을 가지고 관심 대상이 되는 모수가 그 구간에 있음을 기대할 수 있도록 하는 값의 範圍(range of values)이다. 예로써 신뢰구간에 대한 말은 다음과 같이 쓰인다. 즉, 확률 0.95를 가지고서 구간 $(\hat{b} \pm 0.07)$이 모수 b를 포함하고 있다. 마지막으로 다른 조건이 일정하다면(ceteris paribus) 확신 또는 신뢰를 더 할 수 있는 유일한 방법은 b를 포함하는 구간을 더 넓히는 것 뿐이라는 점을 보일 것이다. 만약 b가 $\hat{b}$의 0.07 내에 있다는 것을 95% 확신하고 있는데, b를 포함할 확률이 0.99일 구간을 원한다면, 새로운 구간은 $(\hat{b} - 0.07, \hat{b} + 0.07)$보다 더 넓어져야 한다. 예를 들면 99%의 신뢰구간은 $(\hat{b} - 0.10, \hat{b} + 0.10)$이 될 것이다.

가설검정과 신뢰구간의 문제는 다음과 같은 의미에서 상호 관련이 있다. 위에서 언급하였듯이 $b = 0.75$라는 가설을 검증한다고 가정하자. 우리는 아마도 b를 추정하고서 그 추정치가 얼마나 0.75에 가까운가를 볼 것이다. 만일 추정치와 0.75 사이의 차이가 매우 "작다"고 하면, 그 결과가 가

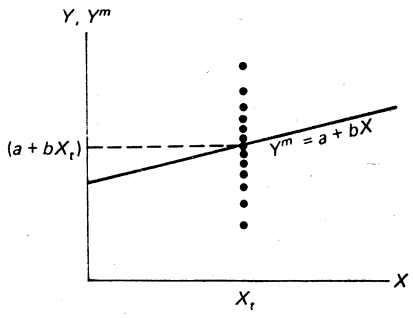
설을 지지한다고 느낄 것이다. 다른 한편으로 추정치가 0.75와 "크게" 다르다면, 가설이 관찰한 결과에 의하여 확인되지 않는다고 결론을 맺을 것이다. 그러면, $b \neq 0$이라는 것을 믿을 만한 좋은 근거를 갖게 된다. 이러한 종류의 의미있는 분석을 하기 위해서는, 반드시 母數의 가설화된 값과 추정치 사이의 차이가 "작은지" "큰지"를 구별할 수 있어야만 한다. 미리 보면, 문제가 되는 모수에 대하여 신뢰구간을 만들고 그 구간내에 가설화된 모수의 값이 놓여 있는지의 여부를 파악함으로써 차이가 큰지 작은지를 결정하는 것이다. 예를 들어, 구간 $(\hat{b} \pm 0.07)$이 b를 포함할 확률이 0.95일 경우, 자료를 기초로 추정한 $\hat{b}$이 0.89로 나오면 $b = 0.75$라는 가설을 95% 신뢰로 기각할 것이다. 구간의 폭이 모수의 값을 포함하고 있다는 신뢰와 관련이 있기 때문에, 추정절차에 따른 결과에 갖는 확신도는 신뢰구간과 연관된 확률 수준에 직접 종속한다.

가. 추가 가정

회귀모형의 모수들이 갖는 값에 대하여 가설을 검정하고 신뢰구간을 설정하기 위해서는 먼저 모형에서의 교란항 u_t가 갖는 속성을 더욱 확대해야만 한다. 2장에서는 교란항들 자체가 갖는 속성에 대하여 네 가지 가정을 한 적이 있다. 즉, 교란항들은 기대값이 0인 확률변수이고, 동일한 분산을 갖는다. 또한, 교란항들은 0의 공분산을 갖는다. 그리고 교란항들은 설명변수와 독립적이다. 여기서 한 조건을 추가로 도입하는 것이 유용하다. 즉, 교란항이 正規分布한다고 가정하는 것이다. 즉, 교란항의 密度 또는 확률함수는 正規分布曲線(normal curve)이다. 정규분포는 평균과 분산의 두 모수만을 가지지 때문에 정규분포하는 변수는 그 평균과 분산으로 완벽하게 지정된다. 교란항 u_t에 대한 가정을 다음과 같은 표시로 요약할 수 있다. 즉, $N(0, \sigma_u^2)$이다. 말로 표현하면, $N(0, \sigma_u^2)$은 평균이 0이고 그 분산이

σ_u^2 인 정규분포하는 변수를 가리킨다.*

교란항의 正規性에 대한 추가적인 가정으로서 〈그림 3.1〉과 같이 회귀모형의 기본 속성들을 그릴 수 있다. 어떤 주어진 X값, X_t에 대해서 Y의 평균값은 $(a + bX_t)$가 될 것이다. 그러나 교란항 때문에, Y는 그



〈그림 3.1〉

평균값으로 완전히 결정되지 않는다. X의 특정한 값 X_t에 대응하는 모든 Y에 대한 관찰을 반복하였더라면, Y의 관찰치 모두가 그 평균값인$(a + bX_t)$가 되었을 것으로 기대하지 않았을 것이다. 더구나 Y의 평균으로부터의 편차는 오로지 교란항이 일으키는 것이기 때문에, 만일 교란항이 정규분포한다면 그 편차가 정규분포하게 된다. 예를 들어 X_t에 대응하는 Y에 관하여 반복적인 관찰을 한 散布圖가 있다면, 그 산포도는 〈그림 3.1〉의 흩어져 있는 점들을 닮았을 것으로 기대한다. 여기서 "자료"(관찰치)는 Y의 평균값 $(a + bX_t)$에 멀어지기 보다는 더 밀집하여 있다. 이렇게 되는 이유는 正規密度曲線(normal density curve)이 평균(여기서 평

---

* $N(u_x, \sigma_x^2)$은 변수 X가 평균 u_x와 분산 σ_x^2를 갖고 정규분포를 하는 것을 가리키는 표준적인 표기이다. 여기서 N을 n과 혼동해서는 안된다. n은 표본의 크기를 가리키는 것이다.

균은 0이다)에서 멀어짐에 따라 차차 그 높이가 줄어들기 때문이다.*

이제 회귀모형의 모수에 대한 추정량 $\hat{a}$과 $\hat{b}$으로 되돌아 가자. 2장에서 이들 추정량의 분산을 찾는 과정에서 설명변수의 값이 주어졌을 때, $\hat{a}$와 $\hat{b}$는 모두 교란항 $u_1, u_2, \cdots, u_n$의 선형결합임을 보였다.** 통계학에는 정규분포변수의 선형결합은 그 자체가 정규분포한다는 定理가 있다. 따라서 주어진 X_t 값의 집합에 대해서 $\hat{a}$과 $\hat{b}$는 반드시 정규분포하게 된다. 2장에서 $\hat{a}$과 $\hat{b}$의 평균값이 각각 a와 b임을 알았고, (2.87)과 (2.96)식으로 그 분산을 나타내었다. 이들 결과에 u_t의 정규성이란 가정을 보태면, 다음과 같은 결론을 내릴 수 있다.

$$\hat{a} \quad \text{은} \quad N\left(a, \frac{\sigma_u^2 \sum X_t^2}{n\sum(X_t - \bar{X})^2}\right) \tag{3.1}$$

$$\hat{b} \quad \text{은} \quad N\left(b, \frac{\sigma_u^2}{\sum(X_t - \bar{X})^2}\right) \tag{3.2}$$

이상은, 추정량 $\hat{a}$과 $\hat{b}$이 정규분포한다면 a와 b에 관한 가설검정에 표준적인 통계학 기법을 사용할 수 있기 때문에, 특별히 유용한 결과이다. 기초통계학에서, 만일 정규분포하는 변수, v가 평균 μ_v 분산 σ_v^2 이라면 [즉, v가 $N(\mu_v, \sigma_v^2)$이라면],

$$Z = \frac{v - \mu_v}{\sigma_v} \tag{3.3}$$

가 표준 정규 변수이다. 다른 말로 하면, Z은 $N(0, 1)$이다. 標準正規分布表*** 로부터 Z의 값에 대한 특정 종류의 확률을 언급할 수가 있게 된 것이다. 그 표에서 예를 들어서 다음을 알 수 있다.

---

* 설명의 편의를 위해서 이 책에서는 교란항이 정규분포한다고 가정한다. 그러나 技術的으로는 이하의 많은 결과가 이러한 가정을 요구하는 것이 아니다.

** 식 (2.74)와 (2.90)을 보라.

*** 책 뒤의 통계표 1을 보라.

$$\text{Prob}\,(-1.65 \le Z \le 1.65) = 0.90$$
$$\text{Prob}\,(-1.96 \le Z \le 1.96) = 0.95 \qquad (3.4)$$
$$\text{Prob}\,(-2.58 \le Z \le 2.58) = 0.99$$

예로써, 이것은 만일 Z의 값이 임의로 고른 것이라면, Z의 값이 (−1.65)와 (+1.65)사이에 있는 확률이 0.90이라는 것을 의미한다. (3.4)의 첫번째 것은 다음을 의미하는 것이다.

$$\text{Prob}\,(Z \le -1.65) = 0.05 \qquad (3.5)$$

그리고

$$\text{Prob}\,(Z \ge 1.65) = 0.05 \qquad (3.6)$$

이렇게 되는 이유를 다시 상기하자. 즉, 正規分布曲線이 좌우 대칭이어서 곡선 아래 부분의 0.90이 ±1.65 사이에 놓여 있다면, 0.05는 반드시 좌측의 -1.65이하에 그리고 0.05는 우측의 +1.65 이상에 놓여 있기 마련이다.

이제 σ_u^2을 알고 있다고 가정하자, (3.1)과 (3.2)에서 주어진 $\hat{a}$과 $\hat{b}$의 분산을 $\sigma^2_{\hat{a}}$과 $\sigma^2_{\hat{b}}$이라고 표시하면,

$$\left(\frac{\hat{a}-a}{\sigma_{\hat{a}}}\right) \text{ 은 } N(0,1) \text{ 그리고} \left(\frac{\hat{b}-b}{\sigma_{\hat{b}}}\right) \text{ 은 } N(0,1) \qquad (3.7)$$

여기서 $\sigma_{\hat{a}}$과 $\sigma_{\hat{b}}$은 각각 $\hat{a}$과 $\hat{b}$의 標準偏差이다. (3.7)에 비추어 보면, (3.3)의 표준 정규 변수 Z에 대한 것과 마찬가지로 $(\hat{a}-a)/\sigma_{\hat{a}}$와 $(\hat{b}-b)/\sigma_{\hat{b}}$에 대해서도 동일한 말을 할 수 있다. 예를 들어,

$$\text{Prob}\left(-1.96 \le \frac{\hat{b}-b}{\sigma_{\hat{b}}} \le 1.96\right) = 0.95 \qquad (3.8)$$

나. $b \gtrless b_0$에 대한 $b = b_0$의 검정 ; σ_u를 알고 있을 경우

이제 신뢰구간을 구성할 수 있게 되었으며 또한 신뢰구간을 가설검정에 이용할 수 있는 위치에 있게 되었다. 예를 들어 $\hat{b}$에 대하여 (3.8)을 재정리 하면 다음을 얻는다.

$$\text{Prob}\,(\hat{b} - 1.96\sigma_{\hat{b}} \le b \le \hat{b} + 1.96\sigma_{\hat{b}}) = 0.95 \qquad (3.9)$$

식 (3.9)는 0.95의 확률을 가지고 구간

$$(\hat{b} - 1.96\sigma_{\hat{b}}, \hat{b} + 1.96\sigma_{\hat{b}}) \tag{3.10}$$

이 b의 값을 포함하며, 그 결과 (3.10)은 b에 대한 95%의 신뢰구간임을 말한다. 검정 절차를 제시하면 다음과 같다. 먼저 표본을 기초로 $\hat{b}$과 $\sigma_{\hat{b}}$을 계산한다. 그 계산이 마치면 $(\hat{b} - 1.96\,\sigma_{\hat{b}})$와 $(\hat{b} + 1.96\,\sigma_{\hat{b}})$를 계산한다. 마지막으로 가설화된 b의 값이 (3.10)으로 주어진 구간내에 있는지 여부를 본다. 만일 그 구간내에 없다면, 95%의 신뢰도를 가지고 가설을 기각한다. 만일 그 구간내에 있다면, 자료가 가설과 일치한다고 말하고 그 가설을 채택한다.

이상의 검정 절차가 절대 안전한 것은 아니다. 예를 들어 (3.9)는 $(\hat{b} \pm 1.96\,\sigma_{\hat{b}})$가 b를 포함하지 않을 5%의 기회가 있음을 의미하고 있다. 그러므로 비록 가설이 옳을지라도 b값($b = 0.75$)에 관한 가설을 기각할 수 있는 것이다. 좀더 공식적으로 보면, 가설이 사실상 옳을 때에 그 사전의 가설(통계학자들은 歸無假說이라 부른다)을 기각할 수 있다. 이러한 형식의 과오를 제1종 과오(Type 1 error)라고 부른다. 이러한 과오를 범할 확률은 대체로 α라고 표시하며 有意水準(level of significance)라고 한다. 부수적으로, α는 제1종 과오의 "크기"라고도 한다.

귀무가설(null hypothesis ; H_0)이 $b = 0.75$라고 다시 가정하자. 만약 $H_0 : b = 0.75$를 기각하고 더 이상의 정보가 없다면, 분명히 $b \neq 0.75$라 하고 끝나게 된다. 통계학 문헌에서는 이러한 진술을 (H_0에 대한) 對立假說(alternative hypothesis)라고 하고 보통 H_1으로 표시한다. 다른 말로 하여 고려하고 있는 가설을 전부 말하면 $H_0 : b = 0.75$, $H_1 : b \neq 0.75$가 될 것이다. 즉, 우리의 검정절차는 H_0이거나 H_1으로 이끌 것이다.

제1종 과오만이 우리가 범할 수 있는 과오가 아니다. 예를 들어, H_0가

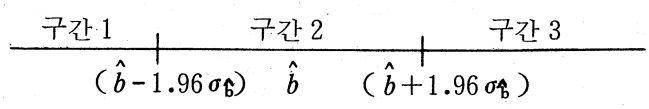
틀렸더라도(즉, H_1이 옳았을 때), H_0을 채택할 수 있다. 예로써, 신뢰구간이(0.86에서 0.92)라고 판명되었다고 가정하자. 더 나아가 $H_0 : b = 0.87$이나 b의 참된 값은 0.88이라고 가정하자. 그러면, 0.87이 신뢰구간내에 포함되기 때문에 $H_0 : b = 0.87$을 채택할 것이다. 그러나 이 경우에 $H_1 : b \neq 0.87$이 옳을 것이다. 신뢰구간은 한 값 이상을 포함하기 때문에 이같은 가능성은 항상 존재한다. 이러한 형식의 과오(귀무가설이 틀렸을 때도 귀무가설을 채택하는 과오)를 제2종 과오라고 한다. 제2종 과오를 범할 확률(이는 또한 제2종 과오의 "크기"라고도 한다)은 보통 β로 표시한다.

모수의 참된 값이 그 가설치와 크게 다를 경우에는 제2종 과오를 범할 확률이 낮아진다는 것을 알 수 있다. 예를 들어, $H_0 : b = 0.75$ 이나 실제로 $b = 0.98$이라 하면, b의 추정량 $\hat{b}$은 전형적으로(typically) 0.98에 "가까이" 있어서, 신뢰구간($\hat{b} \pm 1.96\ \sigma_{\hat{b}}$)은 보통 0.98 근방을(대체로) 중심으로 하는 값들을 포함하는 범위를 갖는다. 이러한 경우에 보통 $b = 0.75$라는 가설을 기각하는 것으로 끝을 맺게 된다. 다른 한편으로, b의 값이 0.751이라 하면, 신뢰구간은 전형적으로 가설화한 값 0.75를 포함하게 될 것이다. 그러므로 제2종 과오를 범할 기회가 매우 많을 것이다. 이는 母數의 가설화한 값이 그 진정한 값 근처에 있을 경우 제2종 과오를 더 많이 저지를 것임을 시사한다. 요컨데, 가설들 사이의 좋은 구별을 짓기는 어렵다.

다. 가설검정 : 한 해석

요약하면, 검정 절차는 본질적으로 모수의 가설치와 모수의 추정치 사이의 불일치를 기초로 歸無假說을 채택할 것인가 또는 기각할 것인가를 따지는 것이다. 이 절차를 보다 자세히 보기 위해서 〈그림 3.2〉의 구간 2와 같

이 b에 대한 95% 신뢰구간을 고려하여 보자. 만약, $\hat{b}$와 b의 가설화한



〈그림 3.2〉

값 b_0 ($b_0 = 0.75$라 하자) 사이의 불일치가 $1.96\,\sigma_{\hat{b}}$를 넘을 정도로 크다면, b_0는 구간 1 또는 구간 3에 있을 것이다. 그러한 경우에 우리는 95%의 신뢰를 갖고 가설 $b = b_0$를 기각한다. 따라서, "작은" 불일치에 비하여 "큰" 불일치라고 하는 결정 인자는 단지 추정량의 표준편차를 곱한 것, 즉 $1.96\,\sigma_{\hat{b}}$에 불과하다. 이것은 사리에 맞다. 예를 들어, $\sigma_{\hat{b}}$가 크다면 추정량의 정확성이 떨어져서, 우리는 귀무가설을 기각하기전에 추정치와 모수의 가설화 한 값 사이의 더 큰 불일치를 "관용"할 의사가 있는 것이다.

라. 採擇域과 棄却域

앞의 예에서 만일 다음과 같다면 귀무가설을 채택할 것임을 보았다.

$$|\hat{b} - b_0| < 1.96\sigma_{\hat{b}} \tag{3.11}$$

(3.11)의 項들을 조작하면 귀무가설의 채택이 $\hat{b}$값의 범위와 연관됨을 본다.

$$b_0 - 1.96\sigma_{\hat{b}} < \hat{b} < b_0 + 1.96\,\sigma_{\hat{b}} \tag{3.12}$$

(3.12)에서 정의한 $\hat{b}$값의 범위를 採擇域(acceptance region)이라 말한다. 일반적으로, 채택역은 귀무가설을 채택으로 이끄는 추정량 값의 범위이다. 역으로 귀무가설을 기각으로 이끄는 추정량 값의 범위를 棄却域(rejection region) 또는 때때로 臨界域(critical region)이라 한다. 앞의 예에서 임계역은 아래와 같을 것이다.

$$\begin{aligned} \hat{b} &> b_0 + 1.96\sigma_{\hat{b}} \\ \hat{b} &< b_0 - 1.96\sigma_{\hat{b}} \end{aligned} \tag{3.13}$$

가설검정을 위한 한 동등한 절차는 먼저 채택역과 기각역을 설정하고 나서, 표본을 기초로 어느 영역이 모수의 추정치를 포함하는가를 결정하는 것이다.

마. 신뢰구간 : 한 해석

진도를 나가기전에 식(3.9)에 관한 약간의 해석을 논의하고자 한다. 설명변수의 값이 주어져 있을 때, 확률로 일컬어 지는 확률변수는 $\hat{b}$이다. b와 $\sigma_{\hat{b}}$는 상수임에 유의하라. 이 둘은 확률변수가 아니다. 식(3.9)는 구간 $(\hat{b}-1.96\,\sigma_{\hat{b}},\ \hat{b}+1.96\,\sigma_{\hat{b}})$가 b를 포함할 확률이 0.95임을 말한다. 다른 말로 하면, 각각의 크기가 50인 다수의 표본(1000이라 하자)을 취할 수 있다고 하자. 더 나아가 이들 각각의 표본에서 X_t의 값들이 동일하다고 가정하자. 각 표본에 대해서 $\hat{b}$을 계산한다고 하자. 교란항이 표본에 따라 다를 것으로 기대하기 때문에, X_t의 값의 집합이 변하지 않더라도 Y_t의 값들이 표본마다 다를 것이다. $\hat{b}$은 Y_t들에 종속하기 때문에, 표본에 걸쳐서 변동할 것이다. 만일 각 표본마다 구간$(\hat{b}\pm 1.96\,\sigma_{\hat{b}})$을 계산하면, 본질적으로 1000개의 다른 구간을 갖게 될 것이다. (3.9)의 본질은 이러한 절차상으로 상수 b를 포함하는 구간이 $0.95(1000)=950$개가 됨을 기대하는 것이다.

바. 제1종과 제2종의 과오에 관한 약간의 논평

이상의 檢定은 95% 신뢰에 기초하였으므로 有意水準(level of significance) 5%를 가지게 된다. 그러나, 어떤 유의수준의 선택은 실험자가 갖는 판단의 자유에 맡겨져 있어서, 우리는 원하는 경우에 쉽게 다른 수준의 유의수준을 가지고 검정을 구성할 수 있다. 만일 $b=0$이 실제로 참인데, $b=0.75$라는 가설을 결코 기각하고 싶지 않다면, 제1종의 과오의 확률을 더 작게 $(\alpha<0.05)$ 갖는 검정을 구성할 수 있었다. 그러한 경우

에 선택하는 α는 연구자가 α를 자유롭게 선정할 수는 있지만 대개 $\alpha = 0.01$이다. 예를 들기 위하여, $\alpha = 0.01$을 갖는 검정을 구성하고자 한다고 가정하자. 그러면 식(3.8)에서 0.99수준의 확률을 선택하여 다음과 같은 구간으로 되어 버린다.

$$(\hat{b} \pm 2.58\sigma_{\hat{b}}) \tag{3.14}$$

다시 우리의 가설을 위의 구간이 가설화한 b의 값을 포함하는지 여부에 따라 채택 또는 기각할 것이다.

제1종 과오의 크기가 $\alpha = 0.01$과 연관된 (3.14)의 신뢰구간은 $\alpha = 0.05$의 크기를 갖는 제1종 과오인 구간(3.10)보다 크다. 만일 신뢰구간이 더 넓고, 여전히 그 중심이 점추정량 $\hat{b}$이라면, 우리가 제2종 과오를 범할 확률이 더 높음이 분명하다. 즉, H_0이 틀렸더라도 신뢰구간이 작을 때보다 클 경우에 더 채택할 가능성이 높다. 이는 제1종 과오의 가능성을 줄이는 만큼, 제2종 과오를 범할 가능성이 더 증가한다는 사실을 가리킨다. 이러한 相衝(trade-off)은 통계학 문헌에 잘 나와있다. 일반적으로 주어진 표본에서 제1종과 제2종 과오의 확률을 둘다 동시에 줄일 수가 없다. 간략히 말하면, 한 과오는 다른 과오의 희생을 치르고서야 줄일 수 있다. 이상의 논의에 비추어 볼 때, 檢定 목적의 신뢰구간을 구성하기 위해서는 먼저 제1종 또는 제2종의 과오를 범할 확률을 선택해야만 한다. 경제학자들은 $b = 0.75$와 같은 특정한 모수의 값을 특정화하는 가설을 검증하는 데에 $\alpha = 0.05$ 또는 $\alpha = 0.01$을 골라서 거의 변함없이 검정하고 있다. 전형적으로, 연관된 제2종의 과오에 대해서는 아무런 또는 명시적인 관심이 주어지지 않는다.

이것은 최선의 해결책이 아니다. 제1종 과오의 크기는(이것이 일단 결정되면 제2종의 과오를 결정한다) 조심스럽게 선정하여야 한다. 그러하지 않으면 어느 쪽의 과오나 바라지 않은 결과를 가져오는 행동을 취하게(또

는 결정을 하게) 될런지도 모른다. 그러므로, 각 유형의 과오와 연관되어 있는 손실을 생각해야 할 것이다. 이상적인 것은 어떤 방식으로 제1종 과오의 크기를 결정하는 데 따르는 손실의 중요성을 측정해야만 할 것이다. 예를 들어, 정부가 폭동의 원인에 관해서 연구하는 중이라고 가정하자. 이때 귀무가설은 특정한 정부 정책이 폭동을 일으키는 정세에 아무런 영향을 미치지 않는다라고 가정하자. 대립가설은 영향을 미친다는 것으로 가정하자. 이 경우에 제1종의 과오는 정부 정책이 실제로 영향을 미치지 않는데도 영향을 미치는 것으로 기대하게 이끄는 것이다. 이러한 과오가 가져오는 결과 중의 하나가 비효과적인 정책에 자금을 낭비하는 것이다. 다른 한편으로, 제2종의 과오는 정책이 폭동의 가능성을 줄이고 있는데도 정책이 기여하지 않는다고 결론을 맺게 한다. 이러한 과오의 귀결은 자금이 효과적인 정책에 쓰이지 않는 것이다. 분명히, 얼마나 각각의 과오가 중요한지에 대한 평가는 약간의 추가적인 가정에 달려있다. 예를 들어, 다수의 "反暴動政策"이 고려대상이 되지만 제한된 자금 사정으로 그 중의 하나만이 수행 가능한 경우를 가정하자. 그러면 제1종의 과오가 극히 중요하게 된다. 왜냐하면, 그 과오는 한정된 자금을 낭비하게 하고 폭동을 조성하는 분위기를 가져오기 때문이다. 다른 한편으로 제2종의 과오는 단지 연구자로 하여금 여타 대안적인 정책에 대하여 고려하도록 할뿐이다.* 그러므로, 이 경우에서 α는 매우 작은 값, 아마도 0.01보다도 작게 잡아야 할 것이다.

다른 한편으로 자금이 상대적으로 풍부하나, 극단적인 경우로서 검토 대상이 되는 정책중의 하나만이 효과적이라고 가정하자. 추가로, 자금은 쉽게 조달가능하므로 정부는 효과가 있을 만한 정책은 다 수행할 수 있다고 가정한다. 이 경우에 제1종의 과오에 기인하는 비효과적 정책에의 자금의 낭

---

* 여기서는 효과적인 다른 정책이 존재하고, "적절한" 기간 동안 검토될 수 있다고 가정하고 있다.

비는 그리 심각하지 않다. 유일한 손실은 오직 자금의 낭비뿐이다. 그러나 제 2 종의 과오는 효과적인 정책을 비효과적인 것으로 믿게 한다. 결과적으로 정부는 효과있는 정책을 기각하여서 사회 혼란이 발생할 것이다. 이 경우에 제 2 종 과오는 제 1 종 과오보다 훨씬 중요하다. 그러므로 이러한 조건 아래에서 제 1 종 과오의 적당한 "설정"은 $\alpha = 0.05$ 보다 α를 크게 놓는 것일 것이다. 예를 들어 $\alpha = 0.30$ 또는 그 이상을 보다 값비싼 제 2 종 과오의 크기를 줄이기 위하여 설정할 수 있을 것이다. 또한 마지막으로 다음을 지적하고 싶다. α를 적당하게 설정하는 문제에 대한 공식적인 해결책은 매우 복잡하고, 또한 그 해결책은 이 책의 범위를 넘어서 버린다는 것이다.* 그렇지만 독자들이 이와 관련된 주제를 적어도 일부나마 더 잘 이해하기를 바란다.

사. 가설 $b \gtrless 0$

실제로 경제학자들은 특정한 모수의 값을 갖는 가설을 보통 갖고 있지 못하다. 경제이론은 종종 두 변수가 正 또는 負의 관련을 갖고 있음을 시사할 뿐이다. 그러므로, 경제학자들이 관심을 갖는 가설은 종종 특정 모수의 부호만을 특성화한 것이다. 사실 어떤 경우에는 이론이 변수들은 관련한다는 것만 지적하고 있다. 즉, 그 관계가 負인지 正인지 자체가 實證的인 문제로 남아 있다. 예를 들어, 감자는 소비가 얼마나 가족의 소득 수준에 따라 변동할 것인가에 관심이 있다고 가정하자. 우리의 가설은

$$Q_t = a + bY_t + u_t \tag{3.15}$$

---

* 기초적인 지식을 주는 약간의 문헌은 Alexander H. Mood and Franklin A. Graybill, An Introduction to the Theory of Statistics (New York : McGraw-Hill, 1963)의 12장과 Arnold Zellner, Introduction to Bayesian Inferences in Econometrics (New York : Wiley, 1971)의 1, 2장을 보라.

여기서 Q_t는 t번째 가족이 소비하는 감자의 양이고, Y_t는 가구의 소득 수준이며, u_t는 이미 여러번 언급한바 있는 교란항이다. 이 경우에 b가 陽인지 陰인지 조차 불명확할 정도이다. 대부분의 상품에 대해서는 소득의 상승에 따라 소비량도 증가할 것으로 믿는다. 그러나 경제학자들이 "열등" 재라고 하는 일군의 상품은 소비가 소득에 반비례한다. 여기에 있는 생각은 가구의 소득이 증가함에 따라, 더 다양하고 값비싼 식단의 음식을 먹을 수 있기 때문이라는 것이며, 소득이 높은 가구일수록 감자를 상당한 정도의 다른 음식으로 대체하는 것을 발견한다고 하여도 놀라운 일이 아니다. 한편, 부유한 가구일수록 더 가난한 가구보다 사실상 더 많은 감자를 소비하는 것이 가능하다. 이 경우에 Q와 Y사이의 관계가 있다는 것을 알더라도, 그 관계의 부호조차 확신할 수 없을 것이다. 이는 우리가 검정하고자 하는 가설이, b의 부호에 대한 어떠한 사전적인 제약이 없는, $b \neq 0$ 임일 뿐이라는 점을 시사한다.

어떻게 $b \neq 0$이란 가설을 검정할 것인가? 예를 들어 단순히 95% 신뢰구간을 구성하여 그 구간이 $b \neq 0$이 아닌 어느 값을 포함하고 있는지 여부를 볼 수 없다. 만약 그렇게 한다면, 항상 가설을 채택할 것이다. 그 이유는 설정한 구간이 항상 $b \neq 0$인 어떤 값을 포함할 것이기 때문이다. 제2종 과오를 범할 확률은 여기서 1과 같다.

검정 절차가 가설이 기각될 수 없도록 추구된다면, 우리가 채택하는 가설에 대하여 명백히 어떠한 신뢰도 가질 수 없다. 이를 바로 잡기 위해서 경제학자들을 $b \neq 0$인 종류의 가설을 제2종 과오의 크기가 보통 0.05 또는 0.01의 작은 수와 같도록 하여 검정한다. 이는 단지 가설의 표지를 다시 붙임에 따라 쉽게 행하여 진다. 예를 들어, 귀무가설 $b \neq 0$ 에 대한 대립가설이 $b = 0$인 검정 문제가 있다. 그러므로 제2종 과오는 사실 $b = 0$일 때, 가설 $b \neq 0$을 채택하는 과오이다. 그러나 $b = 0$을 귀무가설,

$b \neq 0$을 대립가설로 간주하고 유의수준을 $\alpha = 0.05$와 같이 선택하기로 가정하자. 그러면 추정 절차는, 사실 $b = 0$일 때 $b = 0$을 기각할 확률이 $\alpha = 0.05$임을 시사한다. $b = 0$을 기각하는 것은 $b \neq 0$를 채택하는 것과 동등하므로, 검정 결과를 얻게 되는 것이다. 즉, 요구된 바와 같이, 사실 $b = 0$일 때 가설 $b \neq 0$을 채택할 확률이 0.05인 檢定을 짜맞춘 것이다. 예를 들어, 식(3.15)에서 만일 감자 소비량 Q_t와 가족의 소득 수준 Y_t가 아무런 관계가 없다면, 우리의 추정 절차가 이들 변수 사이에 ($b \neq 0$)인 관계가 있다고 믿게할 확률은 0.05인 것이다.

이상의 내용을 요약하기 위하여 특정한 母數 b가 0이 아니라는 가설을 검정하는 경우를 들어 보자. 여기서 우리는 다음과 같은 가설을 구성하자.

$$H_0: b = 0, \quad H_1: b \neq 0 \tag{3.16}$$

우리가 바라는 유의수준에 따라 α를 선정한다. 이미 언급하였듯이, 경제학자는 보통 α를 0.05 또는 0.01로 놓는다. 만일 $H_0 : b = 0$을 기각하면, 추정치가 0과 상당히 다르다라고 한다. 만일 $H_0 : b = 0$을 채택하면, 추정치가 0과 상당히 다르지 않다라고 한다. 후자의 경우는 사실상 변수들 사이에서 우리가 미리 설정한 유의수준에 부응하는 체계적인 관계를 발견할 수 없다고 말하는 것이다.

아. 가설 $b < 0$, $b > 0$

이상의 보기에서는 가설 $H_0 : b = b_0$와 $H_1 : b \neq b_0$를 검정하였으며, 여기서 b_0는 0일 수도 있었다. 이러한 검정을 兩側檢定(two-tailed test)이라고 한다. 즉, 대립가설 H_1은 b가 b_0보다 크나 작으나 둘중에 하나면 된다. 우리의 추정 절차에서는 H_0에 명시된 b값에 대하여 $\hat{b}$이 충분

히 큰 陽 또는 陰의 偏差를 가지게 되면 귀무가설을 기각하게 된다. 이와는 다르게 경제학자들은 자주 單側檢定(one-tailed test)에 관심을 가진다. 경제이론이 빈번히 변수들간의 부호를 제시하기 때문에, 이론으로 부터 도출한 가설들은 자주 $b > 0$ 또는 $b < 0$의 형식을 띠게된다. 다시 95%의 신뢰수준을 가지고 가설을 채택하고자 한다고 가정하자. 그뒤에, 검정의 목적을 달성하기 위하여 가설 $H_0 : b = 0$와 $H_1 : b > 0$ 또는 $H_1 : b < 0$을 설정한다. 여기서 제1종 과오를 범할 확률은 0.05이다.

이상의 단측검정의 실행은 간단하다. 예를 들어 다음의 가설을 보자.

$$H_0: b = 0, \qquad H_1: b > 0 \tag{3.17}$$

이들 가설은 b가 0과 같거나 0보다 크다는 것을 말한다. 따라서 b의 陰의 값에 관심을 가질 필요가 없다. 가설을 채택하는 데에 95%의 신뢰수준을 바란다고 가정하면, 추정 절차는 b값에 대해서 하한(lower bound)을 구성하여서 이 경계가 b값보다 작을 확률은 0.95라 하는 것이다. 이 하한은 95%의 오른쪽 끝이 열린(open-ended) 신뢰구간을 효과적으로 제공한다. 앞에서와 같이 하한은 추정량 $\hat{b}$의 값에 달려 있다. 우리의 추정 절차는 표본의 정보로부터 $\hat{b}$의 값을 결정하는 것이어서, 그 경계가 0보다 더 큰지의 여부를 결정하는 것이다. 만일 그 경계가 陽이면, 가설 $b = 0$을 기각한다. 만약 陽이 아니라면 가설을 채택하고 대립가설 $b > 0$을 기각한다.

하한은 다음과 같이 도출한다.

$$\text{Prob}\left(\frac{\hat{b} - b}{\sigma_{\hat{b}}} < 1.65\right) = 0.95 \tag{3.18}$$

여기서 1.65는 標準 正規分布曲線에 관한 값들의 표(통계표 1을 보라)

에서 찾은 것이다. 식(3.18)은 다음과 같이 다시 쓸 수 있다.

$$\text{Prob}\,(\hat{b} - 1.65\sigma_{\hat{b}} < b) = 0.95 \tag{3.19}$$

$\hat{b}$은 변수이고 $\sigma_{\hat{b}}$이나 b는 변수가 아님을 상기하면, 식(3.19)에서 하한 $(\hat{b} - 1.65\,\sigma_{\hat{b}})$은 모수 b의 값보다 작다는 것을 알게 된다. 따라서, 표본의 정보를 기초로 만일 $(\hat{b} - 1.65\,\sigma_{\hat{b}})$이 0보다 크다면 귀무가설 $b = 0$을 기각할 것이다. 다르게 보아서 만약 $(\hat{b} - 1.65\,\sigma_{\hat{b}}) \leq 0$이면, 가설 $b = 0$을 채택할 것이다. 이 후자의 경우에는 추정치가 0과 크게 다르지 않다고 말한다.

통계학에서는 $(\hat{b} - 1.65\,\sigma_{\hat{b}}) < b$와 같은 구간이 단측 신뢰구간으로 알려져 있다. 독자들은 만약 $H_1 : b < 0$에 대해서 $H_0 : b = 0$을 5%의 유의수준으로 검정하려면, 95%의 단측 신뢰구간을 다음과 같이 설정하여 끝내는 것을 보일 수가 있어야만 할 것이다.*

$$b < \hat{b} + 1.65\sigma_{\hat{b}} \tag{3.20}$$

이 경우에 $(\hat{b} + 1.65\,\sigma_{\hat{b}})$는 b값에 대한 상한(upper bound)을 나타낸다. 우리는 표본에서 $\hat{b}$을 평가하고서 $(\hat{b} + 1.65\,\sigma_{\hat{b}})$이 0보다 작은지 여부를 봄으로써 가설 $H_1 : b < 0$에 대하여 $H_0 : b = 0$을 검정할 것이다. 만일 0보다 작다면 H_0를 기각한다. 만약 상한이 0보다 크다면, H_0을 채택할 것이다. 마지막으로 논의를 b에 국한하여 행하였지만, 이제까지 전

---

* 힌트 : 이 경우에 상한을 바란다. 이 상한은 (3.18)과 매우 유사하여서 부등호의 방향만 반대로 된 것으로 얻을 수 있다. 특히, 다음에서 시작하면 된다.

$$\text{Prob}\left(\frac{\hat{b} - b}{\sigma_{\hat{b}}} > -1.65\right) = 0.95$$

개한 검정 절차는 모수 a도 적용된다. 독자들은 회귀식의 상수항이 갖는 값에 관심이 있다면, 앞에서와 똑같은 기법을 a에 대하여 단측 또는 양측 신뢰구간의 구성에 사용할 수 있다.

자. σ_u가 未知數일 때의 가설검정

앞에서 진행한 분석에서는 σ_u^2 을 따라서 $\sigma_{\hat{b}}$와 $\sigma_{\hat{a}}$도 알고 있다고 가정하였다. 그러나 보통 그렇지 못하기 때문에, 이제 이 가정을 제거할 단계가 되었다. 2장에서 보았듯이, σ_u^2 이 미지수일 경우 $\hat{a}$과 $\hat{b}$의 분산에 관한 추정량을 얻기 위해서는 σ_u^2 의 추정량 $\hat{\sigma}_u^2$ 을 사용해야만 한다. 분산 추정량은 다음과 같다.

$$\hat{\sigma}_{\hat{a}}^2 = \frac{\hat{\sigma}_u^2 \sum X_t^2}{n \sum (X_t - \bar{X})^2} \quad \text{그리고} \quad \hat{\sigma}_{\hat{b}}^2 = \frac{\hat{\sigma}_u^2}{\sum (X_t - \bar{X})^2} \tag{3.21}$$

여기서

$$\hat{\sigma}_u^2 = \frac{\sum (Y_t - \hat{a} - \hat{b}X_t)^2}{(n-2)} = \frac{\sum (Y_t - \hat{Y}_t)^2}{(n-2)}$$

이다.

이상의 변경으로 인하여 a와 b에 관한 가설의 검정(또는 신뢰구간의 설정)에서 더 이상 정규분포 곡선을 사용할 수가 없다. 대신에, 만일 (3.7)의 상대물(counterpart)을 만들면, 다음과 같다.

$$\frac{\hat{a} - a}{\hat{\sigma}_{\hat{a}}} \quad \text{그리고} \quad \frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \tag{3.22}$$

여기서 (3.22)의 둘 다는 $(n-2)$의 自由度를 갖는 t 분포로 나타낼 수가 있는 확률변수이다.* 신뢰구간의 경계를 결정하는 데에 정규분포 대신에 $(n-2)$ 自由度를 갖는 t 分布(통계표 2를 보라)를 사용한다는 점을 제외하면, 앞에서와 똑같은 절차를 따른다.

---

* t 분포는 정규분포를 상당히 닮은 것으로서, n이 커짐에 따라 그 극한에 이르면 정규분포에 접근한다.

차. 약간의 보기

t 분포에 기초한 가설검정을 설명하기 위하여 제 2 장에서 추정하였던 소비함수로 돌아 가자. 미국에서의 1960-1969년간 소비와 가처분소득 자료를 사용하여서 다음을 발견하였다.

$$\hat{a} = 13 \qquad \hat{\sigma}_{\hat{a}}^2 = 31 \qquad (\text{또는 } \hat{\sigma}_{\hat{a}} = 5.6)$$
$$\hat{b} = 0.89 \qquad \hat{\sigma}_{\hat{b}}^2 = 0.0001 \quad (\text{또는 } \hat{\sigma}_{\hat{b}} = 0.01)$$

가설 $a > 0$을 검정한다고 하고 가설의 채택에 95% 신뢰를 갖고자 한다고 가정하자.* 이 검정을 위하여 귀무가설 $H_0 : a = 0$과 대립가설 $H_1 : a > 0$을 설정하고, 유의수준을 0.05와 같다고 놓는다. 자유도 8 (즉, 10-2)을 갖는 t 분포를 통계표 2에서 찾아 보면, a에 대한 단측 신뢰구간의 하한은 다음과 같음을 알게 된다.**

$$a > (\hat{a} - t_{n-2;0.95}\hat{\sigma}_{\hat{a}}) = [13 - 1.86(5.6)] = 2.6$$

구간의 하한이 5%의 유의수준에서 0보다 크기 때문에 $H_1 : a > 0$을 채택한다. 즉, a의 값이 0보다 크다고 결론을 맺는다. 1%의 유의수준을 선정하였다면 다음과 같은 하한을 갖게 되었을 것임에 유의하라.

$$a > (\hat{a} - t_{n-2;0.99}\hat{\sigma}_{\hat{a}}) = [13 - 2.90(5.6)] = -3.2$$

따라서 신뢰구간은 ($a > -3.2$)이 될것이다. 그러므로 $H_0 : a = 0$을 채택하여 a가 0이라고 결론을 맺었을 것이다. 이는 제 1 종 과오의 크기에 대한 약간의 思考가 필요함을 시사한다. 왜냐하면, 검정의 결과는 분명히 제 1

---

* 주의 : 가설은 자료를 분석하고 추정치를 결정하기 앞서 설정해야만 한다. 그렇지 않으면 순환론의 愚를 범하게 된다.

** 여기서 사용한 표기법은 두루 쓰이는 것이다. 일반적으로, $t_{n-2;\gamma}$라는 수는 $n-2$의 자유도를 갖는 t 변수가 그 수보다 작을 확률이다.

종 과오의 크기에 달려 있기 때문이다.*

마지막으로, b에 대한 99% 신뢰구간을 단순히 兩側으로 구성하자. t 분포에 대한 통계표 2로부터 다음의 구간을 발견할 수 있다.

$$(\hat{b} \pm t_{n-2;0.995}\hat{\sigma}_b) = 0.89 \pm 3.36(0.01) = (0.89 \pm 0.03)$$

이 구간은 MPC의 값이 갖는 범위를 0.86에서 0.92까지 포함한다. 이로부터, $H_1 : b \neq 0.75$에 대한 귀무가설 $H_0 : b = 0.75$를 유의수준 1%로 검정하면, H_0을 기각하게 됨이 분명하다.

경제학자들이 전형적으로 자신의 회귀결과를 논문 또는 잡지의 기사에 보고하는 형식을 지적하여 두는 것도 유용할 것이다. 만일, 예를 들어, 독자들이 경제학 잡지나 책에서 이 책의 2장과 3장에서 추정하고 논의한 바 있는 소비함수를 우연히 보게 되면, 다음과 같음을 알게 될 것이다.

$$\begin{matrix} \hat{C} = 13 & + 0.89Y_d & n = 10 \\ (5.6) & (0.01) & R^2 = 0.99 \end{matrix} \qquad (3.23)$$

여기서 모수 추정치 아래의 괄호 안에 있는 숫자는 그 모수의 표준 오차에 대한 추정치(즉, $\hat{\sigma}_{\hat{a}}$과 $\hat{\sigma}_{\hat{b}}$)이다. 이러한 정보를 가지고 독자들은 각종 계수, 가설검정 등을 위한 신뢰구간을 쉽게 구성할 수 있다.

카. t 비율(t ratio) : 약식검정(a rule of thumb)

신뢰구간의 정확한 크기 또한 그로 인한 검정의 결과는 표본의 크기 n과 추정하는 모수의 數에 달려 있지만, 경제학자들은 추정한 회귀식을 볼 때 자주 사용하는 덜 엄격한 약식검정법을 갖고 있다. 예를 들어, 母數

---

* 다시금, 자료가 분석되기 앞서 제1종 과오의 크기를 결정해야 한다는 것을 강조한다. 그렇게 하지 않으면, 관심사가 되는 어떠한 가설이라도 단지 제1종 과오의 "적절한" 크기를 고르기만 하면 채택할 수 있기 때문이다.

推定値의 값이 대응하는 표준오차 추정치의 2배보다 크다면, 양측검정하의 5% 유의수준에서 모수 추정치가 0과 상당히 다르다고 보통 말할 수 있다. 즉, 양측의 대립가설에 대해서 모수가 0이라는 귀무가설을 고려한다면, 이상의 결과는 귀무가설의 기각을 가져온다는 것이다. 만일 모수 추정치가 추정한 표준오차의 3배 크기 이상이라면, 일반적으로 모수 추정치는 1%의 유의수준에서 0과 상당히 다르다고 할 것이다.

이러한 약식검정법들은 쉽게 합리화된다. 예를 들어, 대립가설 $b \neq 0$에 대해서 귀무가설 $b = 0$을 5%의 유의수준에서 검정한다고 하면, 다음의 구간을 검정의 기초로 삼을 것이다.

$$(\hat{b} \pm t_{n-2;0.975}\hat{\sigma}_{\hat{b}}) \tag{3.24}$$

만일 이 구간이 0을 포함하지 않는다면, 귀무가설을 기각할 것이다. 이제 $\hat{b}$는 陽일 수도 있고 陰일 수도 있으나, $t_{n-2;0.975}$와 $\hat{\sigma}_{\hat{b}}$은 항상 陽이다. 따라서 다음과 같다면 우리의 신뢰구간은 0을 포함하지 않는다.

$$(\hat{b} - t_{n-2;0.975}\hat{\sigma}_{\hat{b}}) > 0 \qquad (\hat{b} > 0) \tag{3.25}$$

또는

$$(\hat{b} + t_{n-2;0.975}\hat{\sigma}_{\hat{b}}) < 0 \qquad (\hat{b} < 0)$$

이상의 조건은 다음과 같이 다시 쓸 수 있다.

$$\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} > t_{n-2;0.975} \qquad (\hat{b} > 0) \tag{3.26}$$

또는

$$\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} < -t_{n-2;0.975} \qquad (\hat{b} < 0)$$

마지막으로 이상의 조건을 더 간결하게 다음과 같이 쓸 수 있다.

$$\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right| > t_{n-2;0.975} \tag{3.27}$$

그러므로 ($\hat{b}$ / $\hat{\sigma}_{\hat{b}}$)의 절대값이 t 분포에 따라 주어진 값 $t_{n-2;0.975}$ 를 능가하면, 귀무가설 $b=0$을 기각하여 버린다. 바꾸어 말하면, 5% 유의수준에서 대립가설 $b \neq 0$에 대한 가설 $b=0$을, 단지 비율 $\hat{b}$ / $\hat{\sigma}_{\hat{b}}$ 의 절대값이 $t_{n-2;0.975}$ 를 능가하는지 여부만을 관찰함으로써 검정할 수 있는 것이다. 이제는 경제학자들이 보통 크기가 적어도 $n=15$가 되는 표본을 가지고 작업을 한다는 것을 주지하게 된다. 만일 $n=15$이면, $t_{15-2;0.975} = t_{13;0.975} = 2.16$이다. 다른 한편으로 $n=\infty$이면, $t_{\infty;0.975} = 1.96$ 이다. 마지막으로 t 통계표를 잠시라도 보면 $13 < j < \infty$에 대한 $t_{j;0.975}$ 의 값은 2.16과 1.96 사이에 있음을 알 수 있다. 따라서, 그 약식검정 이란 만일 $\hat{b}$ / $\hat{\sigma}_{\hat{b}}$ 비율의 절대값이 2를 능가하면, 5% 유의수준에서 양측 대립가설에 대한 귀무가설 $b=0$을 기각하는 것이다. 예를 들어, $\hat{b}$ / $\hat{\sigma}_{\hat{b}}$ 비율이 3이면 t 통계표를 참조할 필요도 없다. 문헌에서는, 모수 추정량의 標準誤差에 대한 모수 추정량의 비율인 $\hat{b}$ / $\hat{\sigma}_{\hat{b}}$ 를 t 비율(t ratio)이라 한다. 1%의 유의수준을 원하면, $t_{15-2;0.995} = t_{13;0.995} = 3.01$과 $t_{\infty;0.995} = 2.58$을 찾는다. 이 경우에 대략적인 근사치로서 모수 추정치가 0과 상당히 다르다고 말하려면 그 전에 t 비율이 3을 능가하여야 한다.

이상과 매우 비슷한 방식으로 단측 가설검정의 "t 비율" 방식을 전개하자. 예를 들어, $b > 0$에 대하여 가설 $b=0$을 검정하고 있는데, 5% 유의수준에서 다음과 같다면,

$$(\hat{b} - t_{n-2;0.95}\hat{\sigma}_{\hat{b}}) > 0 \tag{3.28}$$

가설 $b=0$을 기각할 것이다. 마찬가지로 $b=0$에 대한 대립가설이 $b < 0$이고, 다음과 같다면,

$$(\hat{b} + t_{n-2;0.95}\hat{\sigma}_b) < 0 \tag{3.29}$$

가설 $b = 0$을 기각할 것이다.

이제 (3.28)과 (3.29)를 다음과 같이 다시 쓸 수 있다.

$$\frac{\hat{b}}{\hat{\sigma}_b} > t_{n-2;0.95} \tag{3.30}$$

그리고

$$\frac{\hat{b}}{\hat{\sigma}_b} < -t_{n-2;0.95} \tag{3.31}$$

필요한 모든 일은 다시 t 비율을 구하여 단측 대립가설 $b > 0$에 대응하는 것은 (3.30)의 $t_{n-2;0.95}$에 비교하고, 단측 대립가설 $b < 0$에 대응하는 것은 (3.31)의 $-t_{n-2;0.95}$에 비교하는 것으로 끝난다. 이 경우에 값의 범위는 $t_{13;0.95} = 1.771$, $t_{\infty;0.95} = 1.645$가 될 것이다. 물론, 기각하는 귀무가설을 위한 필요조건은 어느 경우에서나 다음과 같다.

$$\left|\frac{\hat{b}}{\hat{\sigma}_b}\right| > t_{n-2;0.95} \tag{3.32}$$

어떤 경우에서는 독자들이 t 비율을 얻기 위해 계수의 값을 추정한 표준 오차를 나누는 수고를 덜기 위하여, 저자가 이러한 나누기를 하여 그 비율(즉, t비율 자체의 표본값)을 그 계수 아래의 괄호 안에 적어 놓았다. 독자들은 괄호 안에 있는 숫자가 추정한 표준오차인지 또는 t 비율 인지를 가리키는 주를 세심하게 검사하여야 한다. 어쨌든, t 비율에 관한 이상의 어림짐작법은 가설검정을 대단히 쉽게 한다. t 비율은 자주 그 절대값이 3보다 크거나 1보다 작기 때문에, t 분포값의 표를 참조조차 하지 않아도 가설을 검정할 수 있을 때가 자주 있다.

2. 함수 형식의 문제

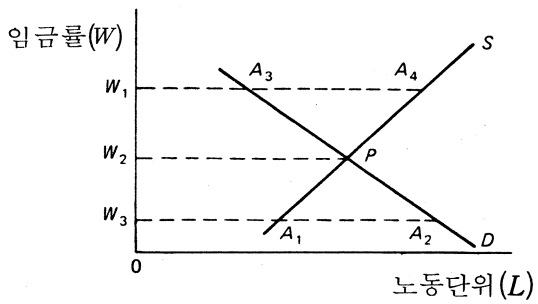
독자들은 1, 2장의 논의에서는 추정하고자 하는 관계의 형식을 線型으로 가정하였던 것을 아무런 의심없이 받아 들였다. 즉, 다음과 같은 형식을 가정한 것이다.

$$Y_t = a + bX_t + u_t$$

윗식은 분명히 가장 제한적인 조건이다. 경제이론 자체나 아니면 관찰한 점들의 흩어져 있는 상태는 두 변수 사이의 관계가 非線型이라는 것을 빈번하게 시사하고 있다. 그러나 어떻게 線型模型을 가지고 非線型 關係를 처리할 수 있겠는가?

가. 필립스곡선과 逆數變換

실제 경제 관계로 이상의 문제를 소개하는 것이 도움을 줄 것이다. 임금의 백분비 변화율($\dot{W}$)과 실업률(R) 사이의 비선형 관계의 전형적인 것으로 여겨지는 필립스곡선이 그 예이다. 노동의 수요와 공급이 있으며, 두가지 모두가 임금률에 종속하는 단순한 勞動市場模型을 보자. 그 모형은 〈그림 3.3〉에 설명되어 있는데, 여기서 수요와 공급이 만나서 均衡賃金率 W_2를 결정한다. 그러나, 임금률이 W_3에 있다면, 노동의 수요는 (A_2 A_1) 만큼 공



〈그림 3.3〉

급을 초과할 것이다. 이 경우에 노동의 陽의 초과수요가 존재한다고 하는데, 노동의 초과수요는 임금률의 수준을 올리는 압력을 가하는 경향이 있다. 역으로, $W=W_1$ 이면, 임금수준이 내려가도록 압력을 가하는 (A_4 A_3) 만큼의 陰의 초과수요(또는 초과공급)가 있다.

다음으로, 이상의 관계를 포착하기 위하여 단순한 動態的 調整機構를 가정하자. 즉, 한 시기에서 다음 시기로 넘어 갈 때의 임금률의 변화율($\dot{W}$)이 갖는 평균값이 직접 초과수요율에 비례한다고 가정한다. 이 가정은 합리적인 것으로 보인다. 고용주의 노동자에 대한 수요와 노동 공급 사이의 차이가 클수록, 우리는 임금수준에 대한 상승 압력이 더 커질 것으로 기대할 것이다. 따라서 다음과 같이 가정한다.

$$\dot{W}_t = \frac{(W_t - W_{t-1})}{W_{t-1}} = \alpha D_t^* + u_t \tag{3.33}$$

여기서 $D^* = (D_t - S_t) / S_t$는 t 시기에서의 초과 수요율이며, u_t는 교란항이다.

이상의 관계를 추정하기 위해서는 $\dot{W}_t$와 D_t^*에 관한 관찰치가 필요하다. 비록 $\dot{W}_t$에 관한 관찰치는 이용 가능하지만, 전형적으로는 D_t^*에 관한 관찰치(또는 자료)를 가지지 못한다. 만일 임금의 조정을 설명하는 조작 가능한 모형을 원한다면, 반드시 D_t^*와 관련된 어떤 변수를 찾아내서 그 변수를 일종의 代理變數 (proxy)로 사용할 수 있다. 초과수요율 D_t^*가 경제에서 실업률 R_t에 대한 체계적 관계를 가지고 있다고 보는 것은 설득력이 있을 것이다. 만일 실업률이 매우 낮다면, "수요와 공급이 꽉 물린" (tight) 노동시장의 전형적인 예로서, 실질적인 陽의 초과수요를 기대할 것이다. 그 반대도 마찬가지이다. 그러므로, R_t와 D_t^*가 負의 관련을 맺고 있다고 믿을 만한 근거가 있다. 이 관계를 가정하자.

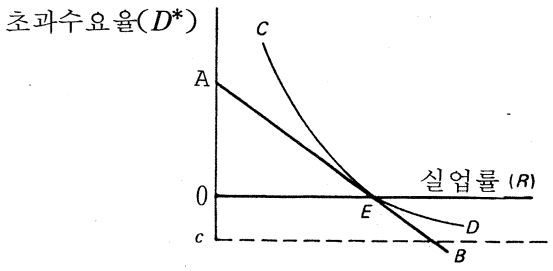
$$D_t^* = f(R_t) \tag{3.34}$$

(3.34)는 무슨 형식을 취할 것인가? 가장 간단한 가정은 그 관계가 線型이라는 것이다. 그래서,

$$D_t^* = e + gR_t \tag{3.35}$$

여기서 e와 g는 母數이고 $g < 0$이다. 그러나 약간만이라도 생각하면 (3.35)는 그 관계를 가장 잘 나타내는 형식이 아님을 깨닫게 하여 준다. 〈그림 3.4〉를 보자. 여기서 수직축은 초과수요율, 수평축은 실업률을 나타내고 있다. O점은 0의 초과수요를 보인다. D^*는 O점 위에서는 陽이고 아래에서는 陰이다. 0의 초과수요는 〈그림 3.3〉의 P에 대응하는데, 여기서 노동의 공급과 수요가 일치하고 결과적으로 임금수준의 변화에 대해서 아무런 압력이 존재하지 않는다. 그러나 약간의 摩擦的 失業은 존재한다. 즉, 動態的 經濟에서는 일부 사람이 한 직업에서 다른 직업으로 전직하는 과정에 있으나, 만일 초과수요가 0이라면 빈 자리의 수가 求職者의 수와 일치할 것이다. 그러므로 〈그림 3.4〉의 점 E는 마찰적 실업률(즉, 0의 초과수요에 대응하는 실업률)을 나타내며, 〈그림 3.3〉의 점 P에 대응한다.

다음으로 일련의 초과수요가 증가하는 시기(연속적으로 D^*의 값이 더 높아지는 시기)를 보자. 빈 자리의 수가 늘어나는 것은 실업한 사람이 직업을 찾는 데에 필요한 시간을 줄인다. 그러므로 D^*의 더 큰 값은 R의 더 작은 값에 수반될 것으로 기대한다. 그러나, D^*가 증가함에 따라 빈 자리



〈그림 3.4〉

의 수가 계속 증가함으로 R이 그것에 대응하여 동일한 양만큼 계속 감소하리라고 기대하지는 않는다. 이러한 이유의 하나는 실업률은 陰의 값을 취할수 없기 때문이다. 결과적으로 D^*와 R의 관계는 〈그림 3.4〉의 AB와 같이 線型關係일 수 없다. 이상의 내용은 D^*가 더 커질수록, 대응하는 R의 감소는 더 작아져야 한다는 점을 시사한다. 따라서 이들 변수 사이의 관계는 〈그림 3.4〉의 CD와 같이 非線型임이 틀림없고, 여기서 CD는 E점의 좌측에 대하여 윗쪽으로 구부러 진다. CD는 E점의 우측으로도 負의 기울기를 가지는데, 이것은 R의 더 큰값이 초과공급의 조건과 연관됨을 가리킨다. 설명을 위해서 CD 곡선이 E점의 오른쪽에 대하여 윗쪽으로 구부러 진다는 가정도 한다.

CD와 같은 곡선에 근사한 함수 형식의 하나는 다음과 같다.

$$D_t^* = c + d\left(\frac{1}{R_t}\right) \quad \text{여기서} \quad c < 0, d > 0 \tag{3.36}$$

여기서 D^*가 R에 반비례하여 변동한다고 가정한다. 윗식은 만약 d가 陽이면, 陰의 기울기를 갖지만 또한 윗쪽으로 구부러지는 非線型 曲線을 제공한다. 윗쪽으로 구부러진다는 것은 초과수요율 D^*가 증가함에 따라 D^*의 증가에 R의 단위당 감소가 더 작아진다는 것을 가리킨다. C가 陰이라고 가정하면, D^*는 R의 상대적으로 큰 값에 대해서 陰이 될 것이다.

이제 D^*의 대리변수 측정수단을 갖게 되었다. 왜냐하면 R_t에 대한 관찰치를 가졌기 때문이다. 만일 (3.36)을 賃金式(3.33)에 대입하면, 標準的인 필립스曲線을 얻는다.

$$\begin{aligned} \dot{W}_t &= \alpha D_t^* + u_t = \alpha\left[c + d\left(\frac{1}{R_t}\right)\right] + u_t \\ &= a + b\left(\frac{1}{R_t}\right) + u_t \end{aligned} \tag{3.37}$$

여기서 $a = \alpha c$이고 $b = \alpha d$이다. 식(3.37)은 임금의 변동율이 직접 실업률

의 逆數(reciprocal)에 따라 변동한다. $\dot{W}_t$와 R_t의 관찰치들로 이루어진 표본을 가지고 있다고 가정하자. 이러한 非線型 관계에서 어떻게 母數 a와 b를 추정할 수 있을까? 여기서 유의할 것은 c와 d를 추정하려 하지 않고 관찰 가능한 관계인 (3.37)의 母數 a와 b를 추정하려는 점이다.

비록 (3.37)이 $\dot{W}_t$와 $\dot{R}_t$ 사이의 비선형 관계이지만, $\dot{W}_t$와 $\dot{R}_t$의 逆數, 즉 $1/R_t$ 사이의 선형 관계로 해석할 수 있다. 그러므로 약간 표기를 바꾼다면 이제까지 線型模型을 위하여 개발한 推定技法을 (3.37)에 적용할 수 있다. 이를 더 명시적으로 나타내는 것으로서, 다음과 같은 새로운 변수를 가정한다.

$$Z_t = \frac{1}{R_t} \tag{3.38}$$

각각의 음이 아닌 R_t 값에 대해서는 〈표 3.1 〉의 예와 같이 대응하는 Z_t 값이 존재할 것이다. Z_t를 (3.37)의 $1/R_t$에 대입만 하면, 다음을 얻는다.

$$\dot{W}_t = a + bZ_t + u_t \tag{3.39}$$

〈표 3.1 〉 가상적인 관찰치 행렬

$\dot{W}_t$	R_t	Z_t
0.02	0.06	16.7
0.04	0.04	25.0
0.05	0.03	33.3
⋮	⋮	⋮

다른 말로 하면, 단순한 變換으로써 (3.37)의 비선형 관계를 (3.39)의 선형 형식으로 바꾼 것이다. 그러면 모수 a와 b의 값을 사용하는 데에

線型回歸模型과 $\dot{W}_t$, Z_t의 값을 사용할 수 있다. 예를 들어 2장에서 도출한 公式을 사용하면,

$$\hat{b} = \frac{\sum (Z_t - \bar{Z})\dot{W}_t}{\sum (Z_t - \bar{Z})^2}$$
$$\hat{a} = \bar{W} - \hat{b}\bar{Z} \tag{3.40}$$

위의 결과를 결합하면, 주어진 R_t 값에 대응하는 $\dot{W}_t$의 평균값의 推定量은,

$$\hat{\dot{W}}_t = \hat{a} + \hat{b}\left(\frac{1}{R_t}\right) \tag{3.41}$$

이상과 같이 하면, 예측 목적을 달성하기 위하여 식(3.41)을 사용할 수 있다. 예를 들어 실업률이 5%라면, 임금의 변동률은 다음과 같이 예측할 것이다.

$$\hat{\dot{W}} = \hat{a} + \hat{b}\left(\frac{1}{0.05}\right) = \hat{a} + 20\hat{b} \tag{3.42}$$

逆數變換은 두 변수 사이의 非線型關係를 線型關係로 변화시키는 데에 사용할 수 있는 多數의 變換중의 하나이다. 이 사실은 우리가 전개한 단순 선형모형이 애초에 보였던 것처럼 制限的인 모형이 결코 아니라는 점을 의미하는 것으로써 매우 중요한 것이다. 다양한 變換을 올바르게 사용함으로써, 상당히 종류가 많은 비선형 관계를 선형 형식으로 바꾸어 놓을 수 있다. 이것은 비선형 모형의 모수를 우리가 이미 전개한 방식을 사용하여 추정할 수 있게 하여 준다. 다음에는 계량경제학에서 널리 쓰이고 있는 다른 두 가지 변환을 보기로 한다.

나. 대수(logarithmic 또는 log)변환

다음의 生產模型의 모수를 추정하려 한다고 가정하자.

$$Q_t = aL_t^b e^{u_t} \tag{3.43}$$

여기서

Q_t = t 시기의 산출량 수준

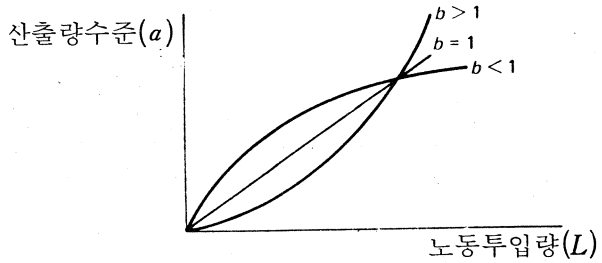
L_t = t 시기의 노동 투입량

e = 2.718에 근사한 상수항

u_t = t 시기의 교란항 그리고 a와 b는 추정하고자 하는 모수

이상의 모형에서는 본질적으로 노동의 유일한 생산요소라고 가정하고 있는데, 이 책의 뒤에서는 이러한 가정을 완화할 것이다.

당분간 (3.43)이 生産函數를 정확하게 묘사한 것으로 가정한다면, 특히 관심이 있거나 갖게 되는 가설은 모수 b에 대한 것이 될 것이다. 그 이유는 b가 전형적으로 규모에 대한 수확체감, 불변 또는 체증의 존재 여부를 가리키기 때문이다. 이들 경우는 〈그림 3.5〉에서 보이듯이, 각각 $b < 1$, $b = 1$, $b > 1$에 대응한다.



〈그림 3.5〉

(3.43)에서 즉각적으로 떠오르는 계량경제학적 문제는 그것이 非線型關係라는 점이다. 우리가 다시 필요로 하는 것은 이 비선형 관계를 앞에서 설정한 추정 기법을 적용할 수 있는 線型으로 변환하는 것이다.

우리가 찾는 변환은 對數變換이다. 예를 들어 (3.43)의 양변에 對數를 취하면,

$$\ln Q_t = \ln a + b \ln L_t + u_t \qquad (3.44)$$

여기서 ln은 변수의 "自然對數" (natural logarithm ; 즉 밑수가 $e \fallingdotseq$

2.718인 對數)이다. (3.44)는 변수의 對數도 線型임을 알 수 있어서 다음과 같은 對數變換을 할 수 있다.

$$Q_t^* = \ln Q_t, \quad a^* = \ln a \text{ 그리고 } L_t^* = \ln L_t \tag{3.45}$$

이상을 (3.44)에 대입하면,

$$Q_t^* = a^* + bL_t^* + u_t \tag{3.46}$$

이제 교란항 u_t에 대해서 표준적인 가정을 하면, (3.46)의 모수 a^*와 b^*는 대변수 기법(instrumental - variable technique)으로 추정할 수 있다는 것은 분명하다. 즉, Q_t와 L_t의 관찰치에 자연대수를 취하기만 하면, 다음을 계산하게 된다.

$$\hat{b} = \frac{\sum (L_t^* - \bar{L}^*)Q_t^*}{\sum (L_t^* - \bar{L}^*)^2} \equiv \frac{\sum (\ln L_t - \overline{\ln L}) \ln Q_t}{\sum (\ln L_t - \overline{\ln L})^2}$$
$$\hat{a}^* = \bar{Q}^* - \hat{b}\bar{L}^* \equiv \overline{\ln Q} - \hat{b}\,\overline{\ln L} \tag{3.47}$$

여기서 $\bar{L}^* = \sum L_t^*/n$ 그리고 $\bar{Q}^* = \sum Q_t^*/n$ 이다.

우리의 推定量이 不偏推定量임을 알기 때문에,

$$E(\hat{b}) = b \text{ 그리고 } E(\hat{a}^*) = a^* \tag{3.48}$$

방정식 (3.47)은 (3.43)에 있는 노동투입량 변수의 지수값의 불편추정량을 제공한다. 이 추정량으로부터 결과가 실제로 노동에 대한 수확체감인지를 알 수 있다.

또한 이상의 추정 절차는 a^*의 불편 추정량도 산출한다. 그러나 a가 생산함수에 나타나는 모수이므로, 모수 a에 관심이 있다. $a^* = \ln a$ 이므로 逆對數(antilog)를 취하면 $a = e^{a^*}$ 이다. a의 추정량은 다음과 같이 시사될 것이다.

$$\hat{a} = e^{\hat{a}^*} \tag{3.49}$$

그러나 $E(\hat{a}^*) = a^*$ 일지라도 $\hat{a}$은 a의 불편추정량이 아니다. 즉, $E(\hat{a}) \neq e^{E(\hat{a}^*)} = e^{a^*} = a$이다. 1장 부록 B에서 일반적으로 $E(e^{a^*})$와 같은 非線型函數의 기대값은 기대값의 함수 $[e^{E(\hat{a}^*)}]$와 같지 않음을 지적한 사실이 독자들에게 떠오를 것이다. 이것은 그 명제의 보기이다. 다행히도 $\hat{a}$가 a의 一致推定量임을 보일 수 있다.

요약하건데, 일반형의 함수를 갖는다면,

$$Y_t = aX_t^b e^{u_t} \tag{3.50}$$

윗식을 線型으로 놓는 데에 對數變數를 사용할 수 있다.

$$Y_t^* = a^* + bX_t^* + u_t \tag{3.51}$$

여기서 *는 대응하는 변수의 자연대수를 의미한다. 선형 추정 기법을 사용해서 (3.51)을 추정할 수 있는데, 이로부터 b의 불편추정량을 얻을 수 있다. 또한 逆對數를 취하여 偏倚를 갖지만 적어도 a의 一致推定量을 얻게 된다.

부수적으로, 對數形式은 경제 모형에서 아주 잘 쓰이는 函數形式이다. 왜냐하면, 그 기울기 계수가 독립변수에 대한 종속변수의 彈力度 (elasticity)로 해석할 수가 있기 때문이다. 예를 들어, (3.51) 또는 (3.50)에서 X_t에 대한 Y_t 평균의 탄력도는 b로 판명된다.* 그러므로, (3.51)과 같은 모형은 탄력도가 不變임을 의미한다.

---

* 예를 들면, (3.50)에서 주어진 X_t 값에 대해서 Y_t의 평균값은 $Y_t^m = aX_t^b E(e^{u_t})$이다. u_t와 X_t가 독립적이라는 우리의 가정은 $E(e^{u_t})$가 어떤 상수, C와 같다는 것을 의미하며, 일반적으로 1이 되지 않는다고 한다(즉, $E(e^{u_t}) \neq e^{E(u_t)} = e^0 = 1$). X_t에 대응하는 Y_t^m의 평균값을 $Y_t^m = a_1 X_t^b$로서 표현 할 수 있으며, 여기서 $a_1 = aC$ 이다. 다음으로 Y_t^m을 X_t에 대해서 微分하면 아래와 같다. (다음 페이지 계속)

다. 준로그(semilog)변환

成長率을 포함하는 模型을 定式化하는 데에는 준로그變換(semilog transformation)이 유용하다. 예를 들어, 일정 시기에 걸친 미국의 노동력 크기의 연간 평균 증가율을 추정하고자 한다고 가정하자. 다양한 임의의 사건으로 인한 미미한 변동은 있지만 이 기간에 걸쳐서 노동력이 일정한 불변의 성장율을 보였다고 한다면 미심쩍을 것이다. 그렇다면, 다음과 같은 관계를 가정하자.

$$L_t = a(1 + g)^t e^{u_t}, \qquad t = 1, 2, \ldots, n \tag{3.52}$$

여기서

$L_t = t$ 년도의 노동력 크기

a = 母數

$g = L_t$의 合成成長率(compound rate of growth)의 母數

u_t = 교란항

(3.52)에서 독립변수 t 그 자체는 指數로 나타난다. 대조적으로 (3.50)에서는 독립변수 X_t는 항상 b 乘된다. 그러나, (3.52)와 (3.50) 둘다는 곱셈으로 이루어져 있어서, 線型으로 관계를 변환하려면 對數를 취하여야 할 것으로 기대한다.

그러나 대수를 취하기전에, 시간에 걸친 성장 경로가 갖는 한 속성을 지적해야 한다. 〈표 3.2 〉에는 미국의 1956 - 1970 년에 걸친 민간 노동력

(앞 페이지 계속)

$$\frac{dY_t^m}{dX_t} = a_1 b X_t^{b-1} = \frac{bY_t^m}{X_t}$$

b에 대해서 풀면,

$$b = \frac{dY_t^m / Y_t^m}{dX_t / X_t}$$

이것은 X_t에 대한 Y_t^m의 彈力度이다.

크기의 자료가 있다. 이러한 자료를 〈그림 3.6〉에서와 같이 그리면, 곡선이 시간이 흐를수록 더 가파라짐을 발견한다. 이는 노동력이 근래에 들어 더 빨리 성장했다는 것을 시사한다. 그러나 이는 대체로 환상이다. 왜냐하면 주어진 백분비(percentage) 성장율이 해가 거듭될수록 노동력의 절대적 증가분을 늘리기 때문이다. 즉, 성장을 계산하는 해의 노동력 크기는 대체로 앞선 해의 크기보다 더 크기 때문이다(즉, $X > Y$이라면, X의 3%는 Y의 3%보다 크다).

〈표 3.2〉 미국 민간 노동력

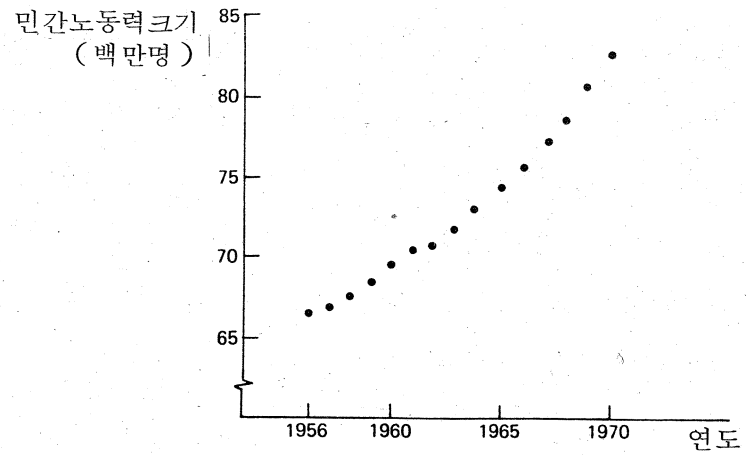
(백만 명)

연 도	L_t	연 도	L_t
1956	66.6	1964	73.1
1957	66.9	1965	74.5
1958	67.6	1966	75.8
1959	68.4	1967	77.3
1960	69.6	1968	78.7
1961	70.5	1969	80.7
1962	70.6	1970	82.7
1963	71.8		

출전 : Economic Report of the President (Washington,D.C:U.S Government Printing office, Feb. 1971), p.222.

이제 (3.52)의 양변에 對數를 취하면,

$$\ln L = \ln a + t \ln(1 + g) + u_t \qquad (3.53)$$



〈그림 3.6〉

다음과 같이 놓으면,

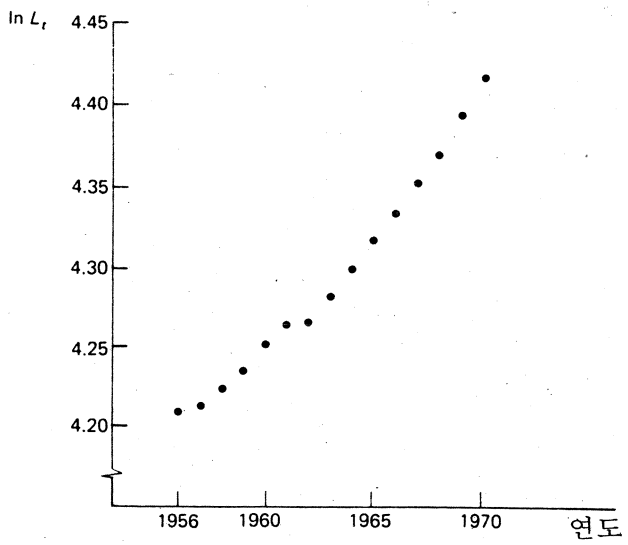
$$\begin{aligned} L_t^* &= \ln L_t \\ a^* &= \ln a \\ b^* &= \ln(1+g) \end{aligned} \tag{3.54}$$

다음의 식을 얻는다.

$$L_t^* = a^* + b^* t + u_t \tag{3.55}$$

윗식은 合成成長率이 L_t와 t 사이가 아니라 오히려 $\ln L_t$와 t 사이의 선형 관계임을 의미한다. 시간에 걸친 L_t값을 그리는 대신에 L_t의 대수값을 그리면, 〈그림 3.7〉과 같은 선형경로 주변의 관찰치를 갖게 된다.

(3.55)의 모수 a^*와 b^*를 추정하려면, 검토대상 시기의 매년도 $\ln L_t$와 t에 관한 관찰치가 있어야만 한다. 설명을 위하여 1956－1970년 시기에 걸친 미국 노동력 크기의 관찰치로 다시 돌아 가자. 각 년도의 L_t에 관한 관찰치는 즉각적으로 $\ln L_t$의 관찰치를 제공한다. t의 관찰치는 단순히 연도에 일련 번호를 부여하면 쉽게 얻을 수 있다. 즉, 1956



〈그림 3.7〉

년은 첫 해로서 $t=1$, 1970년은 15번째 해로서 $t=15$이다. 이상은 〈표3.3〉에 나와 있다.

〈표3.3〉의 정보로부터 다음을 쉽게 계산할 수 있다.

$$\hat{b}^* = \frac{\sum (t-\bar{t})L_t^*}{\sum (t-\bar{t})^2} = 0.0153$$

$$\hat{a}^* = \overline{L_t^*} - \hat{b}^*\bar{t} = 4.17$$

$\ln(1+g)=b^*$이기 때문에, $g=(e^{b^*}-1)$이다. 그러므로 성장률 g는 $\hat{g}=e^{\hat{b}^*}-1 \fallingdotseq 2.718^{(0.0153)}-1=0.016$으로 추정된다. 노동력의 연간 성장률은 1.6%로 나타난다. 앞 節에서 제시한 이유로, $\hat{g}$는 불편추정량은 아니지만 g의 일치추정량이다. 마찬가지로 a의 추정량 $\hat{a}=e^{\hat{a}^*}$은 편의가 있지만 일치추정량이다.

성장률을 추정하는 대안적인 기법(때때로 다른 문헌에 나와 있다)은 다음과 같이 설명할 수 있다. L_t의 최초 값인 1956년의 66.6백만 명과 최종 관찰치인 1970년의 82.7백만 명을 취하고 對數表를 이용하여 평

〈표 3.3 〉 미국 민간 노동력

(백만 명)

연도	L_t	$\ln L_t$	t
1956	66.6	4.199	1
1957	66.9	4.203	2
1958	67.6	4.214	3
1959	68.4	4.225	4
1960	69.6	4.243	5
1961	70.5	4.256	6
1962	70.6	4.257	7
1963	71.8	4.274	8
1964	73.1	4.292	9
1965	74.5	4.311	10
1966	75.8	4.328	11
1967	77.3	4.348	12
1968	78.7	4.366	13
1969	80.7	4.391	14
1970	82.7	4.415	15

균 연간 성장률을 계산한다. 즉, 66.6이 15년 뒤에 82.7이 되는 연간 성장률을 결정한다.

이 대안적인 절차는 추천할 만한 것이 아니다. 예를 들어, 이 절차는 〈그림 3.7 〉(t에 대한 $\ln L_t$의 도면)의 첫 점과 끝 점을 취해서 이 두 점을 잇는 선의 기울기를 계산하는 것과 같다. 다른 말로 하면, 흩어져 있는 점중에서 두 점에만 선을 맞춘 것이다. 이 절차는 표본 크기 n

$=2$일 때의 절차와 동일하다. $n > 2$ 크기의 모든 표본에 대해서 이 대안적인 절차는 우리의 절차보다 열등하다. 왜냐하면 전자는 모든 정보를 무시하기 때문이다.

라. 변환의 사용에 관하여 : 일반화

많은 경우에 事後的인 가설 그 자체는 관심의 대상이 되는 변수들 사이의 관계에 대하여 특정한 非線型을 시사할 것임을 보았다. 이러한 관계를 線型으로 놓고 단순히 선형 추정 절차를 적용하는 변환을 자주 발견할 수 있다. 이제 변환이 부리는 기교가 명백하다. 예를 들어 模型이 아래와 같고,

$$Y_t = a + bf(X_t) + u_t \tag{3.56}$$

여기서 $f(X_t)$는 X_t의 어떤 函數* 라면, $Z_t = f(X_t)$라고 정의하고, 따라서 Z_t에 Y_t를 관련시키는 線型模型을 가진다. (3.56)의 一般化는 아래와 같다.

$$g(Y_t) = a + bf(X_t) + u_t \tag{3.57}$$

여기서 f와 g는 두 개의 非線型일 수 있는 Y_t와 X_t의 함수이다. 이 경우에 Z_{1t}와 Z_{2t} 사이의 線型關係를 가지게 될 것이며, 여기서 $Z_{1t} = g(Y_t)$이고 $Z_{2t} = f(X_t)$이다. 마지막으로, 이러한 종류의 곱셈으로 이루어진 모형은

$$g(Y_t) = af(X_t)^a e^{u_t} \tag{3.58}$$

---

* $f(X_t)$는 어떠한 未知의 母數도 포함하지 않는다고 가정한다. 즉, X_t에 대한 관찰치를 갖는다면, $f(X_t)$에 관한 관찰치를 결정할 수가 있다.

Z^*_{1t} 와 Z^*_{2t} 에서 線型이 될 것이다.

여기서 $Z^*_{1t} = \ln g(Y_t)$ 와 $Z^*_{2t} = \ln f(X_t)$ 이다. 한 예로써, 독자들은 다음의 모형이 $\ln(1/Y_t)$ 와 $\ln X_t^2$ 에서 선형이라는 것을 스스로 확인해야 한다.*

$$\frac{1}{Y_t} = aX_t^{2\alpha}e^{u_t} \tag{3.59}$$

대부분의 경우에 경제이론은 관계의 정확한 형식에 대한 아무런 지침을 제공하지 않을 것이다. 예를들어 이론은 상품의 수요량은 그 가격에 負의 방향으로 변동할 것이라고 시사하지만 그 수요곡선이 線型인지, 對數型인지 아니면 더 복잡한 형식인지를 말하여 주지 않는다. 그러한 경우에 모형의 함수 형식은 종종 散布圖을 단순하게 검사함으로써 이루어 진다. 즉, 함수 형식은 흩어져 있는 점들의 윤곽이 갖는 유형을 확인하여 선택한다. 그러나, 산포도 접근방식은 오직 두 변수만 관련된 경우에 도움이 된다.** 대부분의 경제의 모형은 많은 변수를 포함하기 때문에, 이 문제에 대한 보다 심층적인 논의는 5장까지 미루어 두기로 한다.

---

* $X^{2\alpha}$ 는 $(X^2)^\alpha$ 로 쓸 수 있음.

** 또 다른, 그리고 매우 미묘한 문제가 또한 관련되있다. 만일 모형의 형식이 자료에 대한 첫번째 검사로 결정된다면, 循環性의 요소가 존재한다. 이론적으로는 먼저 모형을 특성화하고, 그 다음에 자료에 비추어 보아 모형을 검정해야 한다. 만일 자료를 검사하여 모형의 형식을 결정한다면, 아마도 적절한 절차는 새로운 자료의 집합을 가지고 그 형식을 사용하여 모형을 검정하는 것이다. 경제학자들은 종종 한 개보다 많은 표본을 얻을 수 없기 때문에, 이러한 순환성의 요소는 불행히도 간과할 수 있는 기회가 많지 않다.

3. 比例縮小 (scaling)와 측정 단위

回歸式의 실제 계산에서는 측정 단위(unit)를 지각있게 사용하는 것이 계산을 단순화하는 데서나, 어떤 경우에는 결과의 해석을 쉽게 하는 데에 모두 중요하다. 예를 들어, 〈표 3.4 〉의 총소비와 가처분소득의 관찰치를 보자.

〈표 3.4 〉

연도	가처분소득 (달러)	소비지출 (달러)
1960	350,000,000,000	325,000,000,000
1961	364,000,000,000	335,000,000,000
⋮		
1969	630,000,000,000	576,000,000,000

이들 C_t와 Y_d의 관찰치를 다음과 같은 소비함수를 추정하는 데에 사용하고자 한다고 가정하자.

$$C_t = a + bY_{dt} + u_t$$

C_t와 Y_t 각각의 관찰치 끝에 있는 9개의 0을 달고 다니는 것이 필요한가? 대답은 필요하지 않다는 것이다. 각 변수의 값을 적절한 단위로 측정하여 이 9개의 0을 간단히 제거함으로써 일감을 상당히 줄일 수 있다. 이 경우에서는 단위를 달러에서 10억 달러로 하는 것이다. 유사한 것으로는 천문학자가 인치가 아닌 光年으로 거리를 재는 것이 있다.

만일 변수를 10억 달러로 측정한다면,

〈표 3.4 A〉

연 도	가처분소득 (10 억달러)	소 비 지 출 (10 억달러)
1960	350	325
1961	364	335
⋮	⋮	⋮
1969	630	576

〈표 3.4 〉의 자료는 〈표 3.4 A 〉의 자료가 된다. 〈표 3.4 A 〉의 자료를 사용하면, 소비함수의 母數 a와 b를 쉽게 추정할 수 있다. 이제, 다른 연구자가 가처분소득과 소비지출을 1,000 억 달러로 측정하였다고 가정하자. 그의 관찰치 표는 〈표 3.4 B 〉와 같이 될 것이다. 이 자료를 사용하여 이 두번째 연구자도 소비함수의 모수 a와 b를 추정할 수 있다. 이제 문제는 〈표 3.4 A 〉를 기초로 추정한 모수와 〈표 3.4 B 〉를 기초로 추정한 모수 사이의 관계에 관한 것이다.

〈표 3.4 B 〉

연 도	가처분소득 (1,000 억달러)	소 비 지 출 (1,000 억달러)
1960	3.50	3.25
1961	3.64	3.35
⋮	⋮	⋮
1969	6.30	5.76

상이한 측정단위에 기초한 모수 推定値들 사이의 관계는 대응하는 모수들 사이의 관계와 아주 동일하다. 예를 들어 10 억 달러로 측정한 가처

분소득과 소비지출을 측정한다고 하고 다음의 모형을 세웠다고 가정한다.

$$c_t = a + by_{dt} + u_t \tag{3.60}$$

이 모형에서 母數 a는 가처분소득이 0일때의 (10억 달러로 측정한) 소비지출액이 될 것이다. b는 한계소비성향이다. 이제 (3.60)의 각 항을 100으로 나눈다고 가정하자. 다음의 식을 얻게 될 것이다.

$$C_t = A + bY_{dt} + U_t \tag{3.61}$$

여기서 $C_t = (\frac{1}{100})c_t$, $Y_{dt} = (\frac{1}{100})Y_{dt}$, 그리고 $U_t = (\frac{1}{100})u_t$이다.

식(3.61)은 1,000억 달러로 정의한 소비지출과 가처분소득에 관련된 소비함수이다. 이 식은 (3.60)에서 도출하였기 때문에, 그 식과 일치 하여야만 한다. 예를 들어, $Y_{dt} = 0$이면, $C_t = (A + U_t)$이고 여기에 100을 곱하면 $c_t = (a + u_t)$이다. 〈표 3.4 A〉의 자료를 사용하는 연구자는 실제로 (3.60)을 자신의 모형으로 생각하고 있는 반면에, 〈표 3.4〉의 자료를 사용하는 연구자는 (3.61)를 자신의 모형으로 생각한다. 이상의 내용은(3.60)을 (3.61)에 비교하였을 때, 〈표 3.4 A〉를 사용한 연구자가 얻은 절편의 추정치가 〈표 3.4 B〉의 자료로부터 도출된 절편의 추정치보다 100배가 더 큰 반면에, 한계소비성향은 둘 다 동일한 추정을 얻게 된다는 것을 말한다. 변수가 다르게 정의 되었다고 해서 이들 추정치가 불일치추정량이지는 않을 것으로 본다.

이들 관계의 증명은 간단하다. 먼저, $\bar{Y}_d = (\frac{1}{100})\bar{y}_d$ 그리고 $\bar{C} = (\frac{1}{100})\bar{c}$ 라고 한다. 그러면,

$$\hat{b}_1 = \frac{\sum(Y_{dt} - \bar{Y}_d)C_t}{\sum(Y_{dt} - \bar{Y}_d)^2} = \frac{\sum(y_{dt} - \bar{y}_d)c_t}{\sum(y_{dt} - \bar{y}_d)^2} = \hat{b}_2 \tag{3.62}$$

여기서 $\hat{b}_1$은 〈표 3.4 B〉와 (3.61)에 의한 b의 추정량이고, $\hat{b}_2$는〈표 3.4 A〉와 (3.60)에 대응하는 것이다. 절편항은,

$$\hat{A} = \bar{C} - \hat{b}\bar{Y}_d = \tfrac{1}{100}(\bar{c} - \hat{b}\bar{y}_d) = \tfrac{1}{100}\hat{a} \tag{3.63}$$

결과를 일반화하기로 하자. 다음의 모형을 고려하자.

$$y_t = a + bx_t + u_t \tag{3.64}$$

이제 다음과 같이 놓는다.

$$Y_t = s_1 y_t, \qquad X_t = s_2 x_t \tag{3.65}$$

여기서 s_1과 s_2는 상수 또는 比例要素(scale factor)이다. (3.65)를 (3.64)에 대입하면 Y_t와 X_t의 관계는 다음과 같다.

$$\begin{aligned} Y_t &= as_1 + \left(\frac{bs_1}{s_2}\right) X_t + s_1 u_t \\ &= A + BX_t + U_t \end{aligned} \tag{3.66}$$

여기서 $A = s_1 a$, $B = b s_1 / s_2$, 그리고 $U_t = s_1 u_t$이다. 그러므로, 만일 한 연구자가 먼저 y_t와 x_t를 (3.65)에서처럼 비례축소(즉, 불필요한 0을 제거하기 위하여)하여 (3.66)을 추정한 반면에, 다른 연구자는 (3.64)를 사용한다면, 그들의 대응하는 모수 추정치는 다음과 같이 주어질 것이다.

$$\hat{A} = s_1 \hat{a}, \qquad \hat{B} = \hat{b}\left(\frac{s_1}{s_2}\right) \tag{3.67}$$

(3.67)에 비추어 보면, 추정량의 分散들 사이의 관계는 다음과 같다.

$$\sigma_{\hat{A}}^2 = s_1^2 \sigma_{\hat{a}}^2, \qquad \sigma_B^2 = \left(\frac{s_1}{s_2}\right)^2 \sigma_b^2 \tag{3.68}$$

여기서는 증명하지 않더라도, 분산의 추정량 사이의 관계가 (3.68)에 있는 내용과 정확하게 동일하다는 것을 다음과 같이 보일 수 있다.

$$\hat{\sigma}_{\hat{A}}^2 = s_1^2 \hat{\sigma}_{\hat{a}}^2, \qquad \hat{\sigma}_B^2 = \left(\frac{s_1}{s_2}\right)^2 \hat{\sigma}_b^2 \tag{3.69}$$

比例要素가 주어져 있다면, 어느쪽의 연구에서 얻은 결과든지 다른 쪽의 연구의 결과로부터 직접 쉽게 도출될 수 있다.

이상의 比例縮小理論의 보기를 제시하기전에 명백하다고 할 수 있는 것을 지적하여 두어야 할 것같다. (3.64)의 모수 a 또는 b에 관한 가설이 있다면, (3.64)와 (3.66)의 어느 것으로라도 이 가설을 검정할 수가 있을 것이다. 예를 들어, 가설 $b = b^0$는 분명히 가설 $B = b^0(s_1/s_2)$ 와 동등하다. 이는 모수 a에 대해서도 비슷하게 적용된다. 비록 여기서 증명은 하지 않았지만, (3.64)로부터 도출된 $\hat{a}$과 $\hat{b}$이 검정된 a 또는 b에 관한 특정 가설은 오직 (3.66)으로 검정된 A 또는 B에 관해 대응하는 가설이 채택 또는 기각되어야만 역시 채택 또는 기각될 것이다.* 다른 말로 하면, $\hat{a}$과 $\hat{b}$에 대한 t비율이 $\hat{A}$과 $\hat{B}$에 대한 t비율과 동일하다.

가. 보기

이제 이상의 비례축소 원리를 간단히 응용한 예를 보기로 하자. 다시, 소비함수에 관심이 있다고 가정하자.

$$c_t = a + by_{dt} + u_t \tag{3.70}$$

더 나아가 c_t와 y_{dt}에 관한 자료를 수집하여 모수 a와 b를 추정하고서 마지막으로 가설 $b = b_0$를 검정한다고 하자. 이제 우리가 사용한 자

---

* 독자는 (3.69)에 비추어 보아 다음을 인지함으로써 이 점을 스스로 확인할 수 있다.

$$\frac{\hat{B} - B}{\hat{\sigma}_{\hat{B}}} \equiv \frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \quad \text{그리고} \quad \frac{\hat{A} - A}{\hat{\sigma}_{\hat{A}}} \equiv \frac{\hat{a} - a}{\hat{\sigma}_{\hat{a}}}$$

따라서, (3.66)에 기초한 A와 B의 신뢰구간은 (3.64)로부터 도출된 대응하는 신뢰구간을 단지 상하로 비례조정한 것에 지나지 않는다.

료가 부정확하다는 것을 알게 되었다고 가정하자. 즉, 자료를 수집한 방식 때문에, 가처분소득이 전부 10 % 높은 수치가 나왔다고 한다. 그러나, 소비지출에 관한 자료는 정확하다고 가정한다. 문제는 우리가 다시 정확한 자료를 가지고 연구를 새로이 반복해야 하는지의 여부이다.

y^*_{dt} 를 우리의 가처분소득 측정치라 하면, 앞에서의 측정 오차는 다음을 의미한다.

$$y_{dt}^* = (1.1)y_{dt} \tag{3.71}$$

여기서 y_{dt} 는 가처분소득의 진정한 값이다. (3.71)을 소비함수(3.70)에 대입하면, 다음을 얻는다.

$$\begin{aligned} c_t &= a + \left(\frac{b}{1.1}\right) y_{dt}^* + u_t \\ &= a + B y_{dt}^* + u_t \end{aligned} \tag{3.72}$$

여기서 $B = (b/1.1)$이다.

식(3.72)는 부정확한 자료의 사용과 관련된 모형이다. (3.67)에서의 결과를 사용하면, (3.72)로부터 a를 검정하는 어느 가설검정의 결과와 마찬가지로 절편 a의 추정치가 여전히 유용함을 알 수 있다. 그러나 한계소비성향의 추정치는 MPC의 불편추정치가 $\hat{b} = \hat{B}(1.1)$이기 때문에 낮게 나올 것이다. 우리는 이 낮은 추정치에 (1.1)을 곱하여 주어야 한다. 더구나, b 값에 관한 가설을 다시 검정해야만 한다. 이 再檢定은 가설 $b = b_0$가 가설 $B = b_0/1.1$임을 의미한다는 것을 일단 알고 있다면 상당히 쉬운 일이다. 원래 하였던 작업중에 다시 해야 하는 일은 거의 없다.

마지막으로 볼 것은 결과를 보고할 때 추정치를 의미있는 자릿수까지 마무리하는 것이 갖는 중요성이다. 이것은 식을 간단히 할 뿐만아니라 "겉치레"(spurious)의 정확성을 피하는 것이다. 독자들은 제 2 장에서 소비

함수를 추정하였을 때 단위를 세세하게 하지 않고 10억 달러로 한 자료를 사용했던 것을 기억할 것이다. 이 수치를 사용하여 다음의 회귀식을 추정하였다.

$$\hat{C} = 13 + 0.89Y_d$$

컴퓨터는 전형적으로 많은 자릿수까지 추정한 계수값을 가져오기 때문에 그 결과로 다음의 형식과 같이 결과를 보고할 수도 있었다.

$$\hat{C} = 13.186537 + 0.889632Y_d$$

그러나 윗식은 오히려 어리석은 행동의 결과이다. 기초 자료가 10억 달러에 아주 근접해야만 정확하기 때문에, 거의 1,000달러까지 내려가는 소비수준을 예측하려고 한다는 것은 이치에 어긋난다. 原資料를 기초로 하여서는 거의 불가능한 것이 분명한 정밀도를 갖는 것처럼 보이게 하는 환상을 낳게 하는 것이다. 결과적으로 독자들은 보고할 수치의 의미있는 자릿수를 기초 정보가 허용하는 정확성의 수준에 일치하도록 추구하는 데에 주의를 기울여야 한다.

4. 時差變數(lagged variable)의 이용

지금까지는 다음과 같은 형식의 관계를 고려하였다.

$$C_t = a + bY_{dt} + u_t$$

그리고

$$\dot{W}_t = a + b\left(\frac{1}{R_t}\right) + u_t$$

여기의 時系列分析(time-series analysis)에서 下添子 t는 時間을 말한다. 이상의 관계는 적어도 한가지 공통적인 중요한 속성을 갖고 있다. 즉, 종속변수의 값이 동일 시점에서 (또는 동일 시기에 걸쳐서)의 독립변수

의 값에 관련된다는 것이다. 예를 들면, 우리의 모형은 1950년의 소비지출이 1950년의 가처분소득에 종속하고 있다.

그러나, 경제학자들은 자주 변수들이 동일 시점에서 관련되있지 않은 모형을 다룬다. 예를 들어, 매월 말에 "봉급을 받거나" 소득을 수령하는 개인들의 집단의 소비지출을 설명하려 한다고 가정하자. 이들 개인은 이 소득의 일정 부분을 그 다음 달에 지출할 것으로 기대된다. 그러므로 어느 달의 지출이 그 전 달에 수령한 소득에 종속하는 시간에 걸친 逐次(sequence)가 있을 것이다. t가 한 달의 시기를 말하는 것으로 하면,

$$C_t = a + bY_{d(t-1)} + u_t \tag{3.73}$$

윗식은 예를 들어 다음을 가리키는 것이다.

$$C_{6월} = a + bY_{d\,5월} + u_{6월}$$

따라서 t시기의 소비는 $(t-1)$시기중의 받은 소득에 종속한다. 이를 자주 소비가 소득에 1시기 "뒤진다" 또는 C가 Y_d에 "1시기 시차를 갖고" 종속한다라고 말한다.

(3.73)과 같은 모형으로 이끄는 또다른 가정의 집합은 다음과 같다. 아래를 가정하자.

$$C_t^p = a + bY_{dt}^e \tag{3.74}$$

여기서,

C_t^p = 다가오는 t시기를 위하여 계획한 소비지출

Y_{dt}^e = 다가오는 t시기의 期待所得

간단히 하기 위하여 다음을 가정한다.

$$Y_{dt}^e = Y_{d(t-1)} \tag{3.75}$$

윗식은 사람들이 다가 오는 t시기의 소득이 현재 시기$(t-1)$의 소득

과 동일할 것으로 기대한다는 말이다. 또한 다음을 가정한다.

$$C_t = C_t^p + u_t \tag{3.76}$$

여기서 C_t는 t시기의 실제 소비지출이며, u_t는 교란항이다. 즉, 실제 지출은 계획한 지출과 다른 확률변수로서 평균이 0이어서 평균적으로는 실제 지출이 계획한 지출과 같다. 이 경우에 u_t는 예기치 못했던 의료비의 지출과 같이 앞서 내다보지 못한 사건의 실제 지출에 미치는 효과를 나타낸다. 어쨌든, C_t^p와 Y_{dt}^e를 (3.74)에 대입하면 다음을 얻는다.

$$C_t = a + bY_{d(t-1)} + u_t \tag{3.77}$$

윗식은 (3.73)과 동일하다.

비슷한 유형으로서 투자행위와 임금의 변동을 설명하는 時差關係도 유용할 것이다. 예를 들어 投資決定은 즉각적으로 이루어지지 않는다. 비록 투자결정이 즉각 이루어지더라도, 그 결정을 투자지출로 옮기는 데에는 시간이 걸린다. 이러한 이유로 다음을 가정한다.

$$I_{t-1}^d = a + br_{t-1} \tag{3.78}$$

$(t-1)$시기의 투자 결정 I^d는 같은 시기의 이자율 r에 달려 있다. 그러나, 다음을 가정하면,

$$I_t = I_{t-1}^d + u_t \tag{3.79}$$

투자지출은 1시기의 시차를 갖는 투자 결정에 종속한다. (3.79)를 I_{t-1}^d에 대입하여 종전에 소비에 대한 관계와 같은 관계를 얻는다. 즉,

$$I_t = a + br_{t-1} + u_t \tag{3.80}$$

1시기의 시차를 필립스곡선에 집어 넣어서 t시기의 임금의 백분율 변화

가 그전 시기의 노동에 대한 초과수요 수준(D^*_{t-1})에 종속한다는 것을 보이는 것을 연습문제로 남겨 둔다. 그 결과는,

$$\dot{W}_t = a + b\left(\frac{1}{R_{t-1}}\right) + u_t \tag{3.81}$$

윗식의 시차가 합당한 것으로 보이는가?

이제 時差變數를 포함하는 모형을 추정하는 문제가 우리의 모형을 가지고 처리할 수 있는지가 문제로 떠오른다. 그 대답은 할 수 있다는 것이다. 이를 보기 위하여, 다음의 모형을 고려하자.

$$Y_t = a + bX_{t-1} + u_t, \qquad t = 1, 2, \ldots, n \tag{3.82}$$

여기서 교란항 u_t에 대해서는 통상적인 가정을 하고 있다. 윗식은 Y가 1시기의 시차를 갖는 X에 종속한다고 말한다. Y_t와 X_t에 대한 n개의 관찰치를 가지고 있다고 가정하고, 이를 〈표 3.5〉에서처럼 배열한다.

〈표 3.5〉

Y	X
Y_1	X_1
Y_2	X_2
⋮	⋮
Y_n	X_n

Y_t는 X_t에 관련되지 않음을 유의하라. Y_t는 X_{t-1}에 종속한다. 이러한 이유로 Y값을 X의 이전 시기의 값과 〈표 3.6〉과 같이 짝지울 수 있다.

〈표 3.6 〉

Y	X
Y_1	X_0
Y_2	X_1
Y_3	X_2
⋮	⋮
Y_n	X_{n-1}

散布圖로 Y와 X값의 관찰치를 보이고자 하였다면, 그림상의 각 점은 Y값과 X의 이전 시기의 값을 나타내었을 것이다. 이들 점이 回歸線을 맞추는 대상이다. 〈표 3.5 〉에서 〈표 3.6 〉으로 옮기면서 관찰치 하나를 상실한 것에 유의하라. 우리는 X_0의 관찰치를 갖고 있지 않기 때문에 Y_1을 사용할 수 없고, Y_{n+1}을 알지 못하기 때문에 X_n을 사용할 수 없다. 1시기의 시차를 갖는 模型은 표본 크기를 1만큼 줄여 ($n-1$)이 되게 한다는 것을 알았다. 즉, 〈표 3.6 〉에서는 모형의 추정에 사용할 수 있는 관찰치쌍은 ($n-1$)개 밖에 없다.

이상의 시차 관계를 추정하기 위하여, 간단히 $Z_t = X_{t-1}$로 정의한다. 따라서, 어느 시기의 Z값은 그전 시기의 X값과 동일할 뿐이다. 따라서 기본 모형 (3.82)은 다음과 같이 쓸 수 있다.

$$Y_t = a + bZ_t + u_t, \qquad t = 2, \ldots, n \tag{3.83}$$

a와 b의 推定量은

$$\hat{b} = \frac{\sum_{t=2}^{n} (Z_t - \bar{Z})Y_t}{\sum_{t=2}^{n} (Z_t - \bar{Z})^2} \tag{3.84}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{Z}$$

여기서

$$\bar{Y} = \frac{\sum_{t=2}^{n} Y_t}{n-1}, \qquad \bar{Z} = \frac{\sum_{t=2}^{n} Z_t}{n-1}$$

이러한 계산에서 간단히 Y_1과 X_n을 버릴 수 있음에 유의하라.

가. 보기

이상의 절차를 설명하기 위하여 다시 2장에서 추정한 소비함수로 되돌아 가자.

$$C_t = a + bY_{dt} + u_t$$

그리고 이제 위의 함수를 1시기의 시차를 갖는 것으로 아래와 같이 하여 추정하자.

$$C_t = a + bY_{d(t-1)} + u_t \tag{3.85}$$

(3.85)에서는 어느 주어진 연도의 소비가 그전 해의 가처분소득 수준에 달려 있다고 가정하고 있다.

〈표 3.7〉

연도	소비 (10억달러)	연도	가처분소득 (10억달러)
1961	335	1960	350
1962	355	1961	364
1963	375	1962	385
1964	401	1963	405
1965	433	1964	438
1966	466	1965	473

연도	소 비 (10 억달러)	연도	가처분소득 (10 억달러)
1967	492	1966	512
1968	537	1967	547
1969	576	1968	590

다시 〈표 2.2 〉로 돌아 가서 그전 해의 가처분소득 관찰치를 소비의 관찰치와 짝지우면, 〈표 3.7 〉을 얻는다. 이제 10 개가 아닌 9 개의 관찰치가 있음을 유의하라. 우리의 추정 절차를 적용하면, 시차소비함수를 다음과 같이 얻게 된다.

$$\hat{C} = -20 + 0.98\,Y_{d(t-1)}, \qquad n = 9, \tag{3.86}$$
$$(0.2)\quad (44.3) \qquad\qquad R^2 = 0.99$$

여기서 괄호 안의 숫자들은 t 비율의 절대값이다. 2 장에서 추정한 소비함수처럼 식 (3.86) 은 높은 정도의 설명력을 갖고 있음을 나타내고 있다. 그러나, 시차 소비함수에서는 常數項이 95 % 신뢰수준에서 0 과 크게 다르지 않다 (t 비율도 0.2 에 불과하다). 더구나, MPC 의 추정한 값은 상당히 더 높다. 앞에서는 0.89 이었는 데에 비하여 지금은 0.98 이다. 결과적으로 시차소비함수에 기초한 政策 處方은 非時差의 경우와 다를 것이다. 이 두 모형을 識別할 수 있게 하는 技法의 필요성은 분명하다. 제 5 장에서 그러한 기법을 전개하기로 한다. 이에 덧붙여, 더 일반적인 時差關係模型을 다룰 것이다.

5 . 예측

이 節에서는 回歸分析의 두 가지 기본적인 응용에서의 두번째 것으로 들어간다. 이 章의 앞에서는 回歸 結果를 어떻게 경제 행위에 대한 假說을 檢定하는 데 사용할 수 있는가를 보였다. 이와 마찬가지로 중요한 것은 추

정한 回歸式을 어떤 사건이 경제 변수에 미치는 영향을 예측하거나 예상하는데 사용하는 것이다. 독자들은 제 1 장 서론에서 대안적인 규모의 조세삭감이 소비지출에 미치는 효과를 평가하여야 하는 경제자문가의 문제를 검토한 적이 있음을 상기할 것이다. 예를 들어, 만일 경제자문가가 새로운 조세삭감이 가처분소득을 얼마만큼 증가시킬지를 알고 있다면, 자신이 추정한 C와 Y_d의 관계를 이용하여 소비지출에 미치는 여타의 가능한 조세삭감이 갖는 효과를 예측할 수 있을 것으로 보인다. 이러한 방식에서 回歸式은 경제 정책이 가져올 것으로 보이는 효과를 평가하는 데 실제적인 도움을 주는 것이다. 回歸分析은 어떻게 經濟가 작용하는가를 量的으로 이해하고 또한 이어서 政策立案者가 이용할 수 있는 다양한 선택권이 갖는 效果를 예측하는데 도움을 준다.

이 점에서 예측의 문제를 보다 체계적으로 연구하고자 한다. 이미 익숙한 형식의 아래와 같은 관계가 있다고 하자.

$$Y_t = a + bX_t + u_t \tag{3.87}$$

최초로, 어떤 미래 시기 f에서 X값이 X_f가 될 것으로 알고 있다고 가정한다. 예를 들어, X가 가처분소득 수준이라면, X_f는 특정 크기의 조세삭감으로 인한 X값이라고 가정하는 것이다. 문제는 이 특정한 값 X_f에 대응할 Y의 값 Y_f를 예측하는 것이다. Y_f는 주어진 X값, 즉 X_f에 대응하는 Y의 미래값임에 유의하라.

예측의 문제에 대하여 첫번째로 알아 두어야 하는 것은 Y_f 자체가 確率變數라는 점이다. 예를 들어, 模型(3.87)에 따르면,

$$Y_f = a + bX_f + u_f \tag{3.88}$$

여기서 u_f는 미래 시기의 교란항이다. 이제 이미 늘 해왔듯이 u_f는 예측 불가능하다. u_f는 이전의 교란항 또는 종속변수 X의 값과 관련이 없

는 확률변수이다. a와 b를 알아서 $(a+bX_f)$를 계산할 수는 있지만 u_f의 예측 불가능한 영향으로 인하여, 여전히 Y_f를 완벽하게 예상할 수가 없다.

이에 덧붙여, 예측의 불확실성 또는 부정확성의 두번째 원인이 있을 것이다. 일반적으로 a와 b를 모르기 때문에, (3.88)에서 Y_f의 첫번째 成分, 즉 X_f에 대응하는 아래와 같은 Y_f의 평균을 추정하기 위해서는 a와 b의 추정치를 사용하여야만 한다.

$$Y_f^m = a + bX_f \tag{3.89}$$

요약하면, 일반적으로 예측에는 오차의 각기 다른 두 원인이 있다. 교란항 u_f의 예측 불가능한 효과와 모수 a와 b를 대신한 추정치의 사용이 그것이다.

(3.88)에서 X_f가 주어지면, Y의 대응하는 미래값 Y_f는 확률변수라는 것을 알았다. 이로써 Y_f의 點推定値 또는 예측만이 아니라 Y_f에 대한 신뢰구간의 설정도 요망된다. 즉, 우리의 예측이 갖음직한 정확성을 측정하는 수단을 원하는 것이다.

아래에서는 회귀 결과를 사용하여 예측도 하고 그 신뢰구간도 설정하는 技法을 설명하기로 한다. 논의는 2단계로 진행한다. 앞에서 지적하였듯이, 일반적으로 a와 b를 모르기 때문에 Y_f^m을 알 수 없다. 먼저 Y_f^m의 추정량과 그 추정량의 분산 추정량을 도출하는 문제부터 볼 것이다. 이를 배경으로 하는 두번째 단계는 Y_f 자체의 예측과 이와 연관된 신뢰구간의 결정이 될 것이다.

가. Y_f^m의 추정

(3.89)는 X_f에 대응하는 Y_f의 평균이다.

$$Y_f^m = a + bX_f \tag{3.89}$$

이제 추정량 $\hat{a}$과 $\hat{b}$이 이용 가능하다고 가정하자. 이 추정량은 $t = 1, 2, \cdots, n$ 시기에 대한 n 크기의 Y와 X의 標本에 기초하고 있으며, f는 미래 시기로서 $n < f$이다. 우리의 표준적인 가정 아래에서는 $\hat{a}$과 $\hat{b}$이 不偏推定量이다. 이 가정으로부터 식 $\hat{Y}_f = (\hat{a} + \hat{b}X_f)$를 Y_f^m의 불편 추정량으로 쓸 수 있다. 왜냐하면 주어진 X_f하에서 다음과 같기 때문이다.

$$\begin{aligned} E(\hat{Y}_f) &= E(\hat{a} + \hat{b}X_f) = E(\hat{a}) + [E(\hat{b})]X_f \\ &= a + bX_f = Y_f^m \end{aligned} \tag{3.90}$$

예를 들어 2장에서 추정한 소비함수로 되돌아 가서, 제안된 조세삭감이 5,000억 달러의 가처분소득 수준과 연관된다고 가정하자. 그러면 이러한 규모의 조세삭감과 연관된 평균 소비지출 수준은 다음과 같다.

$$\hat{C}_f^m = 13 + 0.89(500) = 13 + 445 = 458 \tag{3.91}$$

$\hat{Y}_f$는 X_f에 대응하는 Y_f의 평균에 관한 點推定量 Y_f^m임에 유의한다. 만일 Y_f^m에 대한 신뢰구간 또는 가설검정을 얻고자 한다면, $\hat{Y}_f$의 確率分布(또는 函數)가 필요하다. 이 분포는 통계학에서의 기본 定理로부터 쉽게 도출된다. 즉, 정규분포 변수의 線型結給은 그 자체가 정규분포하는 것이다. 교란항의 正規性에 관한 가정은 $\hat{a}$과 $\hat{b}$이 정규분포한다는 것을 의미한다는 사실을 상기하라. X_f가 주어지면, $\hat{Y}_f$는 단지 $\hat{a}$과 $\hat{b}$의 선형결합에 불과하므로, $\hat{Y}_f$도 반드시 정규분포하는 확률변수이다. $\hat{Y}_f$의 평균은 Y_f^m이다. 더 나아가 $\hat{Y}_f$의 분산이 아래와 같음을 보일 수 있다.*

---

* 이 節에서 다루는 식의 공식적인 전개는 J. Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), pp.38-43을 보라.

$$\sigma_{\hat{Y}_f}^2 = \sigma_u^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{t=1}^{n} (X_t - \bar{X})^2} \right] \tag{3.92}$$

여기서 $\bar{X} = \sum_{t=1}^{n} X_t / n$ 이고, X_1, …, X_n 은 추정량 $\hat{a}$과 $\hat{b}$이 기초하고 있는 X에 대한 관찰치이다. $\hat{Y}_f$ 는 $N(a + bX_f, \sigma_{\hat{Y}f}^2)$ 이다.

논의를 계속하기 앞서서, $\hat{Y}_f$의 분산이 $(X_f - \bar{X})$에 比例하는 것을 유의하라. 이는 특정한 값 X_f가 X에 대한 관찰치(이것은 추정량 $\hat{a}$과 $\hat{b}$를 구성하는데 사용된다)의 평균으로부터 멀어질수록, 추정량 $\hat{Y}_f$의 분산이 더 커지게 된다는 것을 말한다. 이는 상당히 합리적으로 보인다. 어느 정도 직관적으로 보면, 이것이 의미하는 것은 독립변수의 예측치가 관찰한 경험 내의 값들로부터 멀어질수록, 예측의 정확성에 대해서는 신뢰가 덜 가게 된다는 것이다. 근래의 지배적인 가처분소득 수준에 매우 근접한 가처분소득 수준과 연관되어 있는 소비지출의 평균 수준을 추정하는 데에는 상당한 신뢰감이 든다. 대조적으로 현재 소득 수준에 대략 2배가 되는 가처분소득 수준과 연관되어 있는 소비지출의 평균 수준에 관한 추정치에 대해서는 아마도 거의 믿지 않게 될 것이다.

마지막으로, 다음의 사항에 유의하여야 한다. 즉, 일반적으로 σ_u^2 은 未知數이므로 $\hat{Y}_f$의 분산도 알지 못하면서 추정하여야만 한다는 것이다. 제시할 추정량은 분명히 다음과 같다.

$$\hat{\sigma}_{\hat{Y}_f}^2 = \hat{\sigma}_u^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{t=1}^{n} (X_t - \bar{X})^2} \right] \tag{3.93}$$

위의 추정량은 앞에서 논의한 것처럼 불편추정량이다.

$$E(\hat{\sigma}_{\hat{Y}_f}^2) = \sigma_{\hat{Y}_f}^2 \tag{3.94}$$

이상의 내용은 만일 σ_u^2 을 알고 있다면 신뢰구간을 얻을 것이고, (3.

95)가 $N(0, 1)$라고 하였으므로 Y_f^m에 관한 가설을 검정할 수 있을 것이다.

$$\frac{(\hat{Y}_f - Y_f^m)}{\sigma_{\hat{Y}_f}} \tag{3.95}$$

σ_u^2을 모른다면, (3.96)이 自由度 $n-2$의 t變數임에 유념한다.

$$\frac{\hat{Y}_f - Y_f^m}{\hat{\sigma}_{\hat{Y}_f}} \tag{3.96}$$

예를 들어, σ_u^2을 모른다면 Y_f^m에 대한 95%의 신뢰구간은 다음과 같을 것이다.

$$(\hat{Y}_f \pm t_{n-2;0.975}\hat{\sigma}_{\hat{Y}_f}) \tag{3.97}$$

앞서의 설명으로 돌아가면, $Y_d = 500$에 대응하는 $\hat{C}_f$는 458이다. C_f^m에 대한 95%의 신뢰구간을 결정하면, 아래와 같다.

$$458 \pm 2.31\left[3.4\sqrt{\frac{1}{10} + \frac{(500-469)^2}{85{,}810}}\right] = 458 \pm 2$$

나. Y_f의 예측

이제 보통 핵심적인 관심사가 되는 논제로 돌아 가자. 즉, Y_f 자체의 예측과 그와 연관된 신뢰구간의 결정이 그것이다. 먼저 주어진 X_f 하에서 Y의 미래 수준 Y_f의 추정량(또는 예측치)은 Y_f^m의 추정량, 즉 $\hat{Y}_f = (\hat{a} + \hat{b}X_f)$와 동일하다. 그 이유는 Y_f의 예측 불가능한 교란 성분의 평균〔(3.88)을 보라〕이 0이기 때문이다. 다른 말로하면 단지 Y_f의 평균을 예측하는 수단으로 Y_f의 수준을 예측하는 것이다.

이 경우에 예측에 따른 오차는 다음과 같을 것이다.

$$e_f = Y_f - \hat{Y}_f \tag{3.98}$$

(3.98)의 e_f가 갖는 이름은 豫測誤差(forecast error)이다. 예측오차가 0의 평균을 갖는다는 가정에서 다음의 결과가 나온다.

$$\begin{aligned} E(e_f) &= E(Y_f) - E(\hat{Y}_f) \\ &= a + bX_f - a - bX_f = 0 \end{aligned} \tag{3.99}$$

u_f가 $u_1, \cdots, u_n$과 같이 평균 0과 분산 σ_u^2을 갖는 정규분포를 한다고 가정하자. 그러면 주어진 X_f에서 Y_f 또한 평균 $(a + bX_f)$와 분산 $\sigma_u{}^2$을 갖고 정규분포하는 것이 사실이다. (3.98)에서 예측오차 e_f는 정규분포 변수의 線型結合이므로($\hat{Y}_f$이 정규분포 변수임을 상기하라), e_f 또한 정규분포 변수임이 틀림없다.

이미 e_f의 평균을 0이라고 하였다. X_f가 주어졌을 때, Y_f와 $\hat{Y}_f$가 독립적이라는 것에 유의한 하면, e_f의 분산도 결정할 수 있다. 예를 들어, (3.88)로부터 Y_f가 종속하는 유일한 교란항이 u_f이다. 그러나, $\hat{a}$와 $\hat{b}$이 오직 결합(joint) 관찰치 $X_1, Y_1 ; X_2, Y_2 ; \cdots ; X_n, Y_n$으로서만 구성되기 때문에, $\hat{Y}_f$는 교란항 $u_1, u_2, \cdots, u_n$에 종속한다. 각각의 교란항이 여타의 모든 교란항과 독립적이라면, X_f가 주어졌을 때, Y_f와 $\hat{Y}_f$가 독립적일 것이다. (3.98)로부터 e_f가 두 독립적인 확률변수의 線型 合計이며 그 분산은(2장 부록을 보라) 아래와 같다.

$$\begin{aligned} \sigma_e{}^2 &= \sigma_{Y_f}{}^2 + \sigma_{\hat{Y}_f}{}^2 \\ &= \sigma_u{}^2 + \sigma_u{}^2 \left[\frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \\ &= \sigma_u{}^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \end{aligned} \tag{3.100}$$

요컨데, e_f는 $N(0, \sigma_e^2)$이다.

앞에서의 분석과 마찬가지로 σ_e^2의 불편추정량이 아래와 같게 될 것이다.

$$\hat{\sigma}_e^2 = \hat{\sigma}_u^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \tag{3.101}$$

(3.102)가 $N(0, 1)$이라는 것에 의하여

$$\frac{e_f}{\sigma_e} = \frac{Y_f - \hat{Y}_f}{\sigma_e} \tag{3.102}$$

또는 σ_u^2 을 모른다면 (3.103)이

$$\frac{e_f}{\hat{\sigma}_e} = \frac{Y_f - \hat{Y}_f}{\hat{\sigma}_e} \tag{3.103}$$

自由度 $n-2$의 t 변수라는 점에 의거하여 Y_f에 대한 신뢰구간(때때로 豫測區間이라고 불린다)을 구성한다. 예를 들어 σ_u^2 을 알지 못할 때, Y_f에 대한 95% 신뢰구간은 다음과 같이 될 것이다.

$$(\hat{Y}_f \pm t_{n-2;0.975}\hat{\sigma}_e) \tag{3.104}$$

(3.100)에서 $\sigma_e^2 > \sigma_{\hat{Y}f}^2$ 이고 (3.101)과 (3.93)에서 $\hat{\sigma}_e^2 > \hat{\sigma}_{\hat{Y}f}^2$ 임에 유의하라. 이들 결과는 Y_f에 대한 신뢰구간이 항상 $Y_f{}^m$에 대한 같은 수준의 신뢰구간보다 폭이 더 넓다는 점을 의미한다((3.104)를 (3.97)에 비교하라). 이는 자명하다. Y_f를 예측하는 데에는 두가지 어려운 점이 있다. 하나는 母數 a와 b를 모르는 것이고, 다른 하나는 교란항 u_f가 예측 불가능하다는 것이다. 다른 한편으로, Y_f^m을 예측하는 데에는 오직 단 하나의 어려움, 즉 a와 b를 모른다는 것밖에 없다.

마지막 설명으로서 추정한 소비함수로 되돌아가서(앞에서와 같이) $Y_d = 500$이 주어졌을 때 소비 수준에 대한 95% 신뢰구간을 계산한다. 그 구간이 아래와 같다.

$$(\hat{a} + \hat{b}Y_d) \pm t_{n-2;0.975}\hat{\sigma}_e$$
$$= 458 \pm 2.31\left[3.4\sqrt{1 + \frac{1}{10} + \frac{(500 - 469)^2}{85{,}810}}\right] = 458 \pm 8$$

앞서 지적한 바와 같이, C_f의 예측에 대한 신뢰구간(458 ± 8)은 C_f^m의 예측에 대한 신뢰구간 (458 ± 2)보다 폭이 더 넓다.

6. 보기 : 수요곡선의 추정

二變數 回歸模型에 대한 설명을 한 수요곡선의 추정을 예증하는 연습을 끝으로 결론짓기로 하자. 〈표 3.8 〉은 미국에서의 닭고기의 연간 판매량과 가격에 관한 실제 자료를 제시하고 있다. 즉 그 표는 1948년에서 1963년까지 미국의 매년 일인당 닭고기 소비량과 소비자가격지수로 디플레이트한 각년도의 가격을 가리키고 있다.

이 자료를 사용하여 닭고기의 수요곡선을 추정하자. 다음의 관계를 가정한다.

$$Q_t = aP_t^b e^{u_t} \tag{3.105}$$

여기서 Q_t는 파운드(pound)로 측정한 t시기(t번째 연도중의) 1인당 닭고기 소비량이며, P_t는 1파운드당 센트로 측정한 닭고기의 가격이고, u_t는 오차항이다. 그 측정의 예로써 $Q_t = 28.9$의 값은 t번째 연도중의 1인당 닭고기 소비량이 28.9파운드임을 의미한다. $P_t = 41.4$라는 값은 t번째 해의 일 파운드당 닭고기의 평균 가격이 41.4센트임을 의미한다. 〈표 3.8 〉에서 이 수치가 1959년도에 해당함을 알 수 있다.

교란항 u_t가 표준적인 모형이 갖는 가정들을 모두 만족시킨다고 가정하자. 그러면, 앞서 식(3.43)에 대한 설명에서 지적하였듯이, (3.105)와 같은 모형에서 모수 b는 彈力度로 해석할 수 있다. 이 경우에 b는 닭고기 가격이 1% 변함에 따라서 기대되는 일인당 연간 소비량의 百分率 변

〈표 3.8〉 닭고기의 1인당 소비량과 디플레이트한 가격 1948-1963[a]

연도	1인당 닭고기 소비량(파운드)	1파운드당 디플레이트한 가격(센트)
1948	18.3	75.4
1949	19.6	71.8
1950	20.6	68.0
1951	21.7	66.0
1952	22.1	65.0
1953	21.9	62.8
1954	22.8	56.4
1955	21.3	58.7
1956	24.4	50.4
1957	25.5	47.6
1958	28.2	45.8
1959	28.9	41.4
1960	28.2	41.4
1961	30.3	37.0
1962	30.2	38.6
1963	30.6	37.6

a) 가격은 소비자 물가지수로 디플레이트하였다. (1957 - 1959 = 100)

출전 : Frederick V. Waugh, Demand and Price Analysis - Some Examples from Agriculture (Washington, D.C. : U.S. Department of Agriculture, Technical Bulletin 1316, Nov. 1964), Table 5 - 1, p.39.

화를 가리킨다.

이 章의 앞에서 보았듯이 (3.105)를 다음의 線型으로 바꾸어 놓는 데에 對數變換을 사용할 수 있다.

$$\ln(Q_t) = A + b\ln(P_t) + u_t \tag{3.106}$$

여기서 $A = \ln(a)$이고, 對數는 밑수를 e를 가진다. 앞서 논의한 바와 같이 이것은 단지 두 변수의 自然對數를 취하고 종속변수의 對數 $\ln(Q_t)$를 독립변수의 對數 $\ln(P_t)$에 회귀시킴으로써 A와 b의 추정치, 즉 $\hat{A}$과 $\hat{b}$을 얻기 위한 것이다. 그러면 a의 (不偏)추정치는 $e^{\hat{A}}$로 얻을 수 있다. 〈표 3.8〉의 자료를 가지고 이상의 절차를 (3.106)에 적용하면 다음의 결과를 얻는다.

$$\ln(Q_t) = \underset{(0.13)}{5.87} - \underset{(0.03)}{0.68}\ln(P_t) \tag{3.107}$$

$$R^2 = 0.97$$

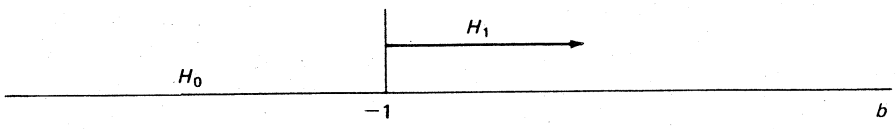
여기서 추정한 계수 밑의 괄호 안에 있는 숫자는 각각의 추정한 표준오차이다. 따라서, 닭고기에 대한 需要의 가격탄력도에 대한 추정치는 -0.68이다. 세자리까지의 a에 대한 추정치는 $e^{5.87} = 354$이다.

이제 (3.107)의 결과를 가지고서 3章에서 제시하였던 技法을 약간 설명하기로 한다. 예를 들어, 5%의 有意水準을 가지고서 닭고기의 수요가 닭고기의 가격과 관련되있다는 대립가설에 대하여 관련되있지 않다는 假說을 검증하는 것에 관심을 갖고 있다고 가정하자. 그렇다면, 歸無假說은 $H_0 : b = 0$이고 對立假說은 $H_1 : b \neq 0$이 될 것이다. (3.107)에서 b의 절대값을 취하여 이에 대응하는 추정한 표준오차로 나누면 (0.68/0.03), 20을 초과하여 매우 높은 t比率을 얻게 된다. 약식검정을 사용한다면, 가설 $H_0 : b = 0$이 (3.107)의 결과로 기각될 것이라는 즉각적인 결론을 내리게

된다.

두번째 설명으로서, 1%의 유의수준에서 수요가 가격에 彈力的이지 않다는 대립가설에 대하여 彈力的이라는 가설을 검정한다고 가정하자. 미시경제학의 원리로부터 수요의 가격탄력도가 -1보다 적다면(절대값으로는 1보다 크다), 그 수요는 가격에 대하여 탄력적이라고 말한다는 것을 상기하라. (3.105)의 수요모형에 따르면, 수요가 가격에 대하여 탄력적이라는 것은 $b < -1$의 어느 값에 대응한다. 따라서 귀무가설과 대립가설은 이 경우에 $H_0 : b < -1$과 $H_1 : b > -1$이 된다.

이 3장의 앞부분에서의 예증과는 달리 위의 귀무가설은 어떤 특정한 b의 값을 밝히지 않았다. 그럼에도 불구하고 이 가설은 신뢰구간 절차에 의하여 검정할 수 있다. 이 점을 알기 위하여 먼저 H_0가 지시하는 어떠한 가능한 b값보다 H_1의 b가 크다는 것을 시사하는 單側 對立假說이 H_1임에 유의하라. 이러한 상황은 〈그림 3.8〉에 나타나 있다. 이 章의 앞에서 논의한 것과 일치하도록 하한(lower bound ; LB) 하나만 갖는 單



〈그림 3.8〉

側 信賴區間을 구성한다. 이 신뢰구간은 $b \geq$ LB라는 형식을 가질 것이다. 만일 LB ≥ -1이라면 H_0를 기각한다. 만약 LB < -1이면, H_0를 채택한다. 이러한 검정 절차는 자동적으로 H_0를 채택하도록 이끌지 않아도 된다. 그렇게 되는 경우는 귀무가설이 $H_0 : b \leqq 0$인 형식이었을 때 생기는 난점에서 본 바 있다.

분석의 메카니즘은 간단하다. 통계표 2에서 $16 - 2 = 14$의 自由度를 갖

는 t 분포를 찾으면, 99 %의 單側 信賴區間은 다음과 같다.

$$b > (\hat{b} - t_{14;\,0.99}\hat{\sigma}_{\hat{b}}) = -0.68 - (2.624)(0.03) = -0.76$$

$-0.76 > -1$ 이므로 H_0 를 기각하고 닭고기에 대한 수요가 가격에 대하여 非彈力的이라는 대립가설을 채택한다.

다시 (3.107)의 결과를 보고서 推定한 式이 관찰치에 매우 잘 "맞고" 있다는 점에 유의하자. 그 식은 연간 1인당 닭고기 소비량을 관찰한 변동을 97 % 설명한다. 그러나 독자들은 아마도 적어도 두가지 이유로 인하여 약간의 불편을 느낄 것이다(또한 그래야만 한다). 첫째로, 그 式은 닭고기의 소비량 변동을 오직 가격 변화로만 설명하였다. 그러나 주지하다시피 1948년에서 1963년 기간에는 소비자의 所得이 상승하였다. 따라서 닭고기가 正常財(normal good)라면 소득의 상승이 닭고기의 구입 증대를 일부 설명할 것으로 기대할 것이다. 그러나 식(3.105)는 소득에 대한 언급을 생략하였다. 중요한 변수(이 경우에는 소득)를 생략하였기 때문에 닭고기 소비에 미친 가격의 효과뿐만 아니라 소득의 효과도 가격에 의한 것으로 돌려 버렸을지도 모른다. 이러한 이유로 우리의 추정치는 가격 변수가 소비자의 닭고기 구매에 미친 효과를 과대 평가한 것이다. 둘째로, 우리의 추정 절차에서는 市場의 供給 측면을 고려하지 않았다. 관찰한 가격과 수량은 수요의 영향만으로부터 나타난 것이 아니라 오히려 공급과 수요의 相互作用으로부터 초래된 것이다. 어떻게 해서든지 이 상호작용을 回歸模型에서 명시적으로 인지시켜야 한다. 이상의 문제들은 남은 章들에서 연구할 것이다.

문 제

1. I.Q 검사의 성적이 근년에 들어 향상되었으며, 그 평균이 이제 100 이상이라는 주장이 있다. I.Q 점수가 인구에 걸쳐 정규분포한다고 가정하자. 만일 100개의 검사 결과로 이루어진 표본에서 평균 점수가 110이고 추정한 분산이 4라고 할 때, 모집단의 평균 I.Q가 100를 넘는다는 가설을 5% 수준으로 검정하라.

2. 평균적으로 확률 표본의 크기가 30일 때가 표본 크기 20일 때보다 평균을 더 잘 추정할 수 있는 이유를 설명하라. 이 둘의 추정량은 모두 불편추정량인가?

3. 산업의 표준적인 1주당 노동시간이 40시간이라 하자. 노동시간이 40시간에서 벗어날 때마다 다시 40시간으로 되돌아 간다는 것이 우리의 가설이다. 이러한 가설을 정식화하는 한 방식은 $\triangle H_t = \beta + \alpha(40 - H_{t-1}) + u_t$이다. 여기서 $\triangle H_t = H_t - H_{t-1}$이기에, $H_t = (\beta + 40\alpha) + (1 - \alpha)H_{t-1} + u_t$가 된다. 미국의 분기별 자료를 가지고 다음의 회귀를 추정한다고 가정한다.

$$\hat{H}_t = \underset{(0.7)}{5} + \underset{(0.15)}{0.875H_{t-1}}, \quad R^2 = 0.98$$

여기서 계수 아래의 괄호에 있는 숫자는 추정한 표준 편차이다. 5%의 유의수준에서 노동시간의 변동이 40시간으로부터의 이탈에 종속한다는 가설을 채택하여야 하는가 또는 기각하여야 하는가? 그 이유는?

4. A씨가 남자의 평균키는 70인치라는 이론을 제시하고 있다. B씨는 A씨의 이론이 어떤 요소를 과대 평가하여서 평균을 과장하였다고 주장한다. 다음이 크기 4의 확률 표본의 결과라고 하자. 64, 74, 72, 62. 그 확률밀도함수는 $\sigma^2 = 4$의 정규분포한다고 가정한다. A씨의 이론을 5% 유의수준에서 검정하라.

5. 교란항이 정규분포한다는 가정이 갖는 중요성 또는 취지가 무엇인가?

6. 동부 사람들의 평균키가 67인치로 알려졌다고 하자. 동부의 외투 제조업자가 서부에 한 판매점을 열고자 하고, 서부 사람들은 평균적으로 더 크다고 믿을 만한 근거를 갖고 있다고 한다. 만약 그렇다면 그는 더 긴 외투를 생산하여야만 한다. 더 긴 외투를 생산하기 위해서는 기계설비를 상당히 그리고 막대한 비용을 들여 재장비하는 것이 필요하다고 가정하자. 또한 이 상대적인 키에 관한 가설은 서부에서 선정한 사람들의 확률 표본으로 검정한다고 한다. 제1종 오류와 제2종 오류의 귀결에 대하여 논하라.

7. 많은 경제 관계가 비선형일지라도 선형 회귀모형이 그렇게 제한적인 것만이 아닌 이유를 설명하라. 특히, 아래의 모형에 대하여 적절한 변환을 채택하여 관찰치 행렬을 도출하라.

$$Y_t = a + b\left(\frac{1}{1 - X_t}\right) + u_t$$

여기서 $n = 3$ 그리고 Y와 X에 대한 관찰치는 $Y_1 = 1$, $Y_2 = 10$, $Y_3 = 12$; $X_1 = 0$, $X_2 = 0.1$, $X_3 = 0.5$이다.

8. A씨가 소비함수를 추정하여 다음의 결과를 얻었다고 한다.

$$\hat{C} = \underset{(3.1)}{15} + \underset{(18.7)}{0.81 Y_d}, \qquad n = 19, \quad R^2 = 0.99$$

여기서 괄호 안의 숫자는 t 비율이다.

a. t 비율을 이용하여 Y_d가 통계적으로 유의한 변수라는 가설을 검정하라.

b. 모수 추정량의 표준 편차를 추정하라.

c. Y_d의 계수에 대한 95% 신뢰구간을 구성하라. 이 구간이 0을 포함하는가?

9. 복지수혜율(1개월기준의 가족당 지불 금액으로 측정)의 증가가 사람들이 노동을 여가로 대체함으로써 복지에 대한 수요를 증가시키는 상황을 고려하자. 또한 복지에 대한 수요의 증가가 다시 정치적 압력을 통하여 다음의 시기에 수혜율을 증가시킨다고 가정하자. 이 관계식을 두 개의 식으로된 모형으로 표현하라.

10. 어떤 州에 입지한 기업의 수가 그 주의 상대적인 조세율에 종속하는 상황을 고려하자. 또한, 비록 그 주에 거주자들에게 조세의 혜택이 있을지도 모르지만, 그 주에 입지한 기업수가 많을수록, 公害率은 더 상승한다고 가정한다. 기업의 입지와 공해의 관계를 회귀모형으로 표현하라.

11. 다음과 같은 표준적인 회귀모형을 고려한다.

$$Y_t = a + bX_t + u_t$$

X_t를 측정할 수 없다고 가정한다. 대신에 $X_t = 5 - 3X_t$를 관찰하였다고 하자. Y_t를 Z_t에 관련시키는 모형에서 모수는 무엇인가?

제 4 장 다중회귀분석

앞장에서는 두 변수사이의 통계적 관계를 살펴보았다. 예를 들어 소비지출의 경우, 다음의 모형을 가정하였던 것이다.

$$C_t = a + bY_{dt} + u_t \tag{4.1}$$

그리하여 a와 b의 값을 추정하고 이 관계에 대한 제 가설을 검정할 수 있게 해주는 기법을 개발하였던 것이다. 하지만 전형적으로 경제관계란 이보다 훨씬 더 복잡하기 마련이다. 곧 소비와 같은 어떤 특정변수의 값이 하나의 변수에 좌우되는 것이 아니라 일련의 독립변수들의 값에 좌우되는 것은 흔한 일이다.

예를 들어 t시점에서의 소비지출수준이 t기의 가처분소득뿐만 아니라 유동자산(liquid assets, A_t)* 과 전기의 가처분소득 $Y_{d(t-1)}$에도 좌우된다고 가정하자. 만일 유동자산에 대한 소비자의 보유액이 이례적으로 높다면(예를 들어 2차대전이 끝날 무렵 그랬던 것처럼), 소비지출이 정상적으로 당시의 소득수준과 관련된 것보다 훨씬 더 높을 것으로 예상될 것이다. 그러나 반대로 유동자산스톡이 이례적으로 적으면, 소비자들은 자산보유액을 보충하기 위해 어느 정도 지출을 줄이게 될 것이다. 과거의 소득 또한 현재의 소비자지출수준을 결정하는데 일정한 역할을 담당하기도 한다. 예를 들어 현시점에 이루어진 지출은 이전 시점의 소득을 반영하는 과거의 생활수준에 의해 야기될 수도 있다. 한가지 예로 다른 조건이 동일하다면 이전 시점의 소득이 높으면 높을수록 현재의 소비도 보다 높을 수 있다. 이러한 모든 것은 바로 다음과 같이 소비함수가 보다 복잡하게 됨

---

* 유동자산은 개인의 화폐보유고, 정기예금, 저축, 채권 등을 의미한다.

을 의미한다.

$$C_t = a + b_1 Y_{dt} + b_2 A_t + b_3 Y_{d(t-1)} + u_t \qquad (4.2)$$

이제 문제는 소비가 이러한 독립변수 각각에 어떻게 좌우되는지를 추정해 보는 것이 된다. 곧 달리 말하면 a, b_1, b_2, b_3의 값을 추정할 수있게 해주는 기법을 개발해야 하는 것이다. 그리고 어느 정도의 정확성을 위해 推定量의 分散을 설정할 수도 있을 것이다. 앞에서 이용한 자료를 직접 일반화하는 것으로서 임의적으로 모형의 변수에 대해 일련의 관찰치가 있다고 가정한다. 곧 예를 들면 (4.2)에 따라 〈표4.1〉에서와 같은 정보를 얻게 된다.

얼핏 보기에 문제는 이변수의 경우와는 성격이 어느 정도 상이하며 훨씬 더 복잡해 보인다. 특히 二變數回歸의 경우에 관찰치는 종속변수와 하나의 독립변수에 대한 것이었으며, 문제는 단지 전자가 후자와 함께 어떻게 변동하는가를 추정하는 것이었다. 이제 일련의 독립변수에 대한 관찰치가 존재하며, 상이한 모든 독립변수들의 영향을 가려내는 훨씬 어려워 보이는 작업에 직면하고 있다. 곧, 각 독립변수의 영향을 측정하기 위해서는 여타의 모든 독립변수들이 종속변수에 미치는 영향으로부터 그것을 어느 정도 분리해 내어야 하는 것이다.

보기에는 이렇게 복잡할지라도 多重回歸分析(multiple-regression analysis)(곧, 독립변수가 두가지 이상인 경우)은 二變量分析(bivariate analysis)의 직접적인 일반화라는 것을 처음에 강조한 바 있다. 처음에는 二變數(two-variable) 모형에서 가정한 교란항(disturbance term)의 특징과 관련하여 동일하게 가정한 다중회귀모형을 서술할 것이다. 다음에는 다시 교란항의 추정량에 대한 조건으로서 代變數技法(instrumental-variable technique)을 채택할 것이다. 이는 일련의 正規方程式(normal

〈표 4.1〉 관 찰 치

(단위 : 미화 1억달러)

시 간	소 비 (C_t)	가처분소득 (Y_{dt})	금융자산 (A_t)	전년도 가처분소득 ($Y_{d(t-1)}$)
1960	325.2	350.0	399.2	337.3
1961	335.2	364.4	424.6	350.0
1962	355.1	385.3	459.0	364.4
1963	375.0	404.6	495.4	385.3
1964	401.2	438.1	530.5	404.6
1965	432.8	473.2	573.1	438.1
1966	466.3	511.9	601.5	473.2
1967	492.1	546.3	650.4	511.9
1968	535.8	591.2	709.6	546.3
1969	577.5	631.6	731.6	591.2

출전 : Economic Report of the President (Washington, D.C.: U.S. Government Printing Office, Feb. 1971), pp.197, 204, 262.

equations)을 만들어낼 것이고, 그 解는 모형의 계수 각각에 대한 不偏推定量(unbiased estimators)을 제공하게 될 것이다. 그러므로 절차는 앞장에서와 거의 동일할 것이다. 곧 二變數의 경우를 이해한다면 多重回歸分析도 그리 어렵지는 않을 것이다.

1. 多重回歸模型

일반적으로 다중회귀모형은 다음과 같다.

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + \cdots + b_kX_{kt} + u_t, \quad t = 1, \ldots, n \qquad (4.3)$$

이 때,

Y_t = 종속변수의 t 번째 관찰치

X_{it} =독립변수가 k 개일 때, i 번째 독립변수(또는 설명변수)의 t 번째 관찰치 ($i = 1, \cdots, k$)

u_t = 교란항의 t 번째 값

b_i = i 번째 독립변수의 계수

회귀방정식에서 상수항을 나타낼 때, a 보다는 b_0를 이용함으로써 (4.3)의 모든 상수들을 b로 표현하는 것이 편리하다. 이제 多重回歸模型에 쓰이는 가정을 열거해 보기로 하자. 그런데 상당수는 二變數의 경우에 채택하였던 것과 동일하므로 언급할 필요가 거의 없을 것이다.

가. 교란항의 기대치 또는 평균치는 0이다 : $E(u_t) = 0$

나. 교란항의 분산은 상수이며 따라서 t 와는 독립적이다 :

$$E(u_t - 0)^2 = E(u_t)^2 = \sigma_u^2$$

다. 교란항의 값은 상호간에 독립적이며 이에 따라 어느 두 관찰치와 관련된 교란항, u_s와 u_t간의 共分散(covariance)은 0이다 :

$$\text{cov}(u_t, u_s) = E(u_t u_s) = 0$$

라. 교란항은 모든 설명변수의 값과 독립적이다. 보다 명확하게 하면, 모든 t와 s에 대해 u_t는 $X_{1s}, \cdots, X_{KS}$와 독립적임을 가정한다. 이에 따라 교란항, u_t와 회귀방정식 (4.3)의 각 독립변수간의 공분산은 0인 것이다. 이는 실험자가 독립변수의 값을 어떻게 정하든지, 또는 "경제"가 그것을 어떻게 정하든지 결코 그러한 특정치가 교란항의 값에

영향을 주지 못한다는 것이다. 보다 정형화하면 u_t 와 X_{it} 간의 공분산이 0이라는 조건은 다음과 같이 나타낼 수 있다.

$$\text{cov}(u_t, X_{it}) = E[u_t(X_{it} - \mu_{Xi})] = E(u_t X_{it}) - \mu_{Xi}E(u_t) = E(u_t X_{it}) = 0$$

(단, $E(X_{it}) = \mu_{Xi}$)

마. 설명변수간에는 完全多重共線性(perfect multicollinearity)이 존재하지 않는다. 곧 어떠한 독립변수도 여타 독립변수의 선형결합(linear combination)이 아닌 것이다. 예를 들어 다음과 같은 관계를 배제한다.

a. $X_{1t} = 3 - 2X_{2t} + 17X_{3t}$;
b. $X_{4t} = (X_{1t} + X_{2t} + X_{3t})/3$;
c. $X_{2t} = 3X_{8t}$

하지만 비선형관계(nonlinear relationships)는 배제하지 않는다. 곧, 가령 $X_{1t} = X_{2t}^2$, 또는 $X_3 = X_{5t} X_{6t}$ 이면 가정은 위배되지 않는 것이다.

앞의 네가지 가정은 二變數模型에서 알려진 것이며, 여기서의 설명도 앞에서와 동일하다. 이러한 가정을 세우는 이유가 모호하다면 제 2 장에 있는 二變數模型에 대한 토론으로 되돌아감으로써 기억을 되살려야 할 것이다.

여기에서 도입된 새로운 하나의 가정, 곧 가정 마는 실제로 제 2 장에서 행하였던 가정, 곧 二變數의 경우 독립변수 X_t는 적어도 두개의 상이한 값을 가져야 한다는 가정의 확장이다.

제 2 장에서는 二變數의 경우, 설명변수 X_t가 불변이어서 $X_t \equiv X_0$라면 단지 A라 불리는 하나의 母數(parameter)를 추정할 수 있을 뿐이며, 그값은 최초의 상수항 a와 X_t의 불변치에 의해 만들어진 상수항 bX_0에 좌우된다는 것을 보았다. 곧 $A = a. + b\ X_0$인 것이다. 간단하게 말한다면, 설명변수의 값이 변하지 않는다면, 그것이 Y에 미치는 영향은 최초의 상수항의

영향과 구분할 수 없을 것이다.

多重回歸의 경우에는 Y에 대한 X_{it}의 영향인 b_i가 상수항 b_0와 구분될 수 있을 뿐만 아니라 다른 모든 X의 영향과도 구분될 수 있어야만 할 것이다. 가정 마는 그러한 구분을 가능하게 하여 준다. 이는 2절에서 보게 될 것이다. 하지만 여기서는 가정 마가 위반되는 다음의 설명을 고려하여 보자.

X_{1t}는 변화하지만 항상 X_{2t}와 같다고 가정해 보자. 그러면 다른 조건이 동일할 때, X_{1t}가 한단위 증가하면 Y_t는 $(b_1 + b_2)$ 단위만큼 변화할 것이다. 왜냐하면 X_{1t}가 항상 X_{2t}와 같다면 X_2도 또한 한단위 증가할 것이기 때문이다. 이는 $X_{1t} = X_{2t}$라면 X_{1t}와 X_{2t}의 결합효과(combined effect), 곧 $(b_1 + b_2)$만이 추정될 수 있음을 뜻하는 것이다. Y에 대한 X_1의 영향을 Y에 대한 X_2의 영향으로부터 분리해 낼 수 있는 방법은 없는 것이다. 이는 다음과 같은 이유때문이다. 만일 $X_{1t} = X_{2t}$이면 기본모형(4.3)은 다음과 같이 다시 쓸 수 있게 된다.

$$Y_t = b_0 + BX_{1t} + b_3X_{3t} + \cdots + b_kX_{kt} + u_t \qquad (4.4)$$
$$(\text{단, } B = (b_1 + b_2))$$

곧, 두 설명변수가 항상 서로 같다면 이들 중 하나는 정보를 전혀 잃지 않고서도 모형에서 제거시킬 수 있는 것이다. 그러므로 단지 $(k-1)$개의 독립변수를 포함하는 모형으로 귀착될(또는 축소될) 수 있을 것이다. 곧, 이 모형은 원래의 두 설명변수의 "결합된(combined)" 영향인 하나의 모수(위에서는 B)를 포함하게 될 것이다. (4.4)는 또한 b_0, b_3, …, b_K도 추정할 수 없음을 뜻하는 것은 아님에 유의해야 한다. 제2절에서 이 문제를 보다 정식으로 다룰 것이다.

多重回歸模型은 본질적으로 二變數模型과 매우 유사하다. 곧 그것은 일련의 독립변수의 값이 하나의 종속변수의 값을 결정하는 線型函數關係(li-

near functional relationship)를 나타낸다. 이제 이러한 관계에 있는 모수의 값을 추정하는 방법을 개발해야만 할 것이다.

2. 대변수에 의한 추정

기억해야할 것은 다음과 같다. 곧, 代變數技法(instrumental-variable technique)은 회귀모형의 가정에 의해 주장된 일련의 조건을 교란항의 추정량에 부과하는 것을 포함한다는 것이다. 二變數의 경우 두가지 조건을 부과하였다.

조 건	관련된 가정
1) $\sum \frac{\hat{u}_t}{n} = 0$, 또는 $\sum \hat{u}_t = 0$	$E(u_t) = 0$
2) $\sum \frac{(X_t\hat{u}_t)}{n} = 0$, 또는 $\sum (X_t\hat{u}_t) = 0$	$E(X_t u_t) = 0$

이 두 조건을 이용하여 두개의 정규방정식을 만들었으며, 그에 따라 $\hat{a}$과 $\hat{b}$의 값을 구하였다. 여기에서도 동일한 접근법을 채택할 것이다. 하지만, 그렇게 하기 전에 몇가지 기본정의를 多重回歸의 틀로 확장시켜야만 한다. 예를 들면, 추정절차는 임의적으로 교란항의 추정량 $\hat{u}_t$에 좌우된다. 그러므로 多重回歸模型에 따라 $\hat{u}_t$을 정의해야 할 것이다.

만일 회귀모형이 (4.3)이라면 $X_{1t}, \cdots, X_{Kt}$와 관련된 Y_t의 평균치는 다음과 같다.

$$Y_t^m = b_0 + b_1 X_{1t} + \cdots + b_k X_{kt} \tag{4.5}$$

二變數의 경우에서처럼 Y_t는 평균과 교란항의 합으로 쓸 수 있다.

$$Y_t = Y_t^m + u_t \tag{4.6}$$

만일 $b_0, b_1, \cdots, b_K$를 알고 있다면, u_t의 값을 다음과 같이 도출할 수 있을 것이다.

$$u_t = Y_t - Y_t^m \tag{4.7}$$

방정식(4.6)과 (4.7)이 二變量回歸模型의 경우에 서로 일치한다는 것을 유의해야할 것이다.

$b_0, b_1, \cdots, b_K$의 추정량, 가령 $\hat{b}_0, \hat{b}_1, \cdots, \hat{b}_K$이 있다고 가정하여 보자. (4.5)에 비추어보면 Y_t의 평균치의 추정량은 다음과 같다.

$$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 X_{1t} + \cdots + \hat{b}_k X_{kt} \tag{4.8}$$

여기에서 위첨자 m은 표기를 간편화하기 위해 생략하였다. (4.7)에 의해 제시된 교란항의 추정량은 다음과 같게 될 것이다.

$$\begin{aligned} \hat{u}_t &= Y_t - \hat{Y}_t \\ &= Y_t - \hat{b}_0 - \hat{b}_1 X_{1t} - \cdots - \hat{b}_k X_{kt} \end{aligned} \tag{4.9}$$

(4.9)를 다음과 같이 다시 쓸 수 있을 것이다.

$$Y_t = \hat{Y}_t + \hat{u}_t \tag{4.10}$$

또는 더욱 완전하게 하면,

$$Y_t = \hat{b}_0 + \hat{b}_1 X_{1t} + \cdots + \hat{b}_k X_{kt} + \hat{u}_t \tag{4.11}$$

간단히 말하면, $\hat{Y}_t$과 $\hat{u}_t$에 대한 정의는 제2장의 이변수회귀모형의 그것과 아주 유사하다. 이제 정의는 충분하다. 그러면 추정으로 돌아가기로 하자.

가. 정규방정식

이미 언급한 것이지만 다중회귀의 경우에도 이변수의 경우와 같은 절차로 정규방정식을 만들 것이다. 단지 k개의 독립변수가 있다고 가정하면, 교란추정량(disturbance estimator)에 대해 부과된 조건이 $(k+1)$ 개일 것이다. 보다 명확하게 하면 다음과 같다.

조 건	가 정
1) $\sum \frac{\hat{u}_t}{n} = 0$, 또는 $\sum \hat{u}_t = 0$	$E(u_t) = 0$
2) $\sum \frac{(X_{1t}\hat{u}_t)}{n} = 0$, 또는 $\sum (X_{1t}\hat{u}_t) = 0$	$E(X_{1t}u_t) = 0$
3) $\sum \frac{(X_{2t}\hat{u}_t)}{n} = 0$, 또는 $\sum (X_{2t}\hat{u}_t) = 0$	$E(X_{2t}u_t) = 0$
⋮ ⋮	⋮
$k+1$) $\sum \frac{(X_{kt}\hat{u}_t)}{n} = 0$, 또는 $\sum (X_{kt}\hat{u}_t) = 0$	$E(X_{kt}u_t) = 0$

회귀모형에 (k+1)개의 조건과 (k+1)개의 모수 b_0, b_1, …, b_K가 나타남에 유의해야 할 것이다.

이제 이러한 조건들이 일련의 (k+1)개의 정규방정식을 어떻게 만들어내는지를 검토해 보기로 하자. 먼저 모두 n개의 관찰치 집합에 대한 방정식(4.11)의 합을 구하면 다음과 같다.

$$\sum Y_t = n\hat{b}_0 + \hat{b}_1 \sum X_{1t} + \hat{b}_2 \sum X_{2t} + \cdots + \hat{b}_k \sum X_{kt} + \sum \hat{u}_t \qquad (4.12)$$

만일 $\sum \hat{u}_t = 0$ 이라는 조건을 덧붙이면, 첫번째 정규방정식은 다음과 같이 도출될 것이다.

N1. $$\sum Y_t = n\hat{b}_0 + \hat{b}_1 \sum X_{1t} + \hat{b}_2 \sum X_{2t} + \cdots + \hat{b}_k \sum X_{kt}$$

(4.11)에 k개의 독립변수를 각각 곱하고 n개의 관찰치집합에 대해 합을 구하면 k개의 방정식을 더 얻게 된다. 자세하게 하면 다음과 같다.

$$\sum(X_{1t}Y_t) = \hat{b}_0 \sum X_{1t} + \hat{b}_1 \sum X_{1t}^2 + \cdots + \hat{b}_k \sum(X_{1t}X_{kt}) + \sum(X_{1t}\hat{u}_t)$$
$$\sum(X_{2t}Y_t) = \hat{b}_0 \sum X_{2t} + \hat{b}_1 \sum(X_{2t}X_{1t}) + \cdots + \hat{b}_k \sum(X_{2t}X_{kt}) + \sum(X_{2t}\hat{u}_t)$$

$$\vdots$$

$$\sum(X_{kt}Y_t) = \hat{b}_0 \sum X_{kt} + \hat{b}_1 \sum(X_{kt}X_{1t}) + \cdots + \hat{b}_k \sum X_{kt}^2 + \sum(X_{kt}\hat{u}_t)$$

만일 $\Sigma(X_{it}\hat{u}_t) = 0$ ($i = 1, 2, \cdots, k$)이라는 조건을 부과하면 이러한 방정식 각각의 마지막항은 없어지게 되며, 따라서 나머지 k개의 정규방정식을 얻게 된다.

N 2. $\sum(X_{1t}Y_t) = \hat{b}_0 \sum X_{1t} + \hat{b}_1 \sum X_{1t}^2 + \cdots + \hat{b}_k \sum(X_{1t}X_{kt})$

N 3. $\sum(X_{2t}Y_t) = \hat{b}_0 \sum X_{2t} + \hat{b}_1 \sum(X_{2t}X_{1t}) + \cdots + \hat{b}_k \sum(X_{2t}X_{kt})$

$$\vdots$$

N(k + 1). $\sum(X_{kt}Y_t) = \hat{b}_0 \sum X_{kt} + \hat{b}_1 \sum(X_{kt}X_{1t}) + \cdots + \hat{b}_k \sum X_{kt}^2$

위의 정규방정식의 합계는 단지 종속변수와 설명변수의 값에 의존함을 유의해야할 것이다. 일단 관찰치의 표본을 얻게 되면, 이와 같은 합계의 값을 계산할 수 있다. 결론적으로 말하면, 위의 정규방정식은 (k + 1)개의 미지수를 가진 (k + 1)개의 방정식의 집합으로 간주할 수 있다. 그 때 미지수는 $\hat{b}_0, \hat{b}_1, \cdots, \hat{b}_K$인 것이다. 일반적으로 위의 정규방정식의 집합에 대한 해를 간단하게 구함으로써 이러한 추정량의 값들을 결정할 수 있게 된다.

설명을 위해 독립변수가 2개인 다중회귀방정식을 고려하여 보자. 곧,

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t$$

위에서 서술한 절차에 따르면 다음의 세가지 정규방정식의 집합이 발생하게 될 것이다.

$$\sum Y_t = n\hat{b}_0 + \hat{b}_1 \sum X_{1t} + \hat{b}_2 \sum X_{2t}$$
$$\sum (X_{1t}Y_t) = \hat{b}_0 \sum X_{1t} + \hat{b}_1 \sum X_{1t}^2 + \hat{b}_2 \sum (X_{1t}X_{2t})$$
$$\sum (X_{2t}Y_t) = \hat{b}_0 \sum X_{2t} + \hat{b}_1 \sum (X_{2t}X_{1t}) + \hat{b}_2 \sum X_{2t}^2$$

Y_t, X_{1t} 와 X_{2t} 의 관찰치에 대한 계산이 다음과 같다고 하자.

$$n = 10 \qquad \sum X_{1t} = 2 \qquad \sum X_{2t} = 2$$
$$\sum X_{1t}^2 = 6 \qquad \sum (X_{1t}X_{2t}) = 1 \qquad \sum X_{2t}^2 = 4$$
$$\sum Y_t = 5 \qquad \sum (X_{1t}Y_t) = 6 \qquad \sum (X_{2t}Y_t) = 7$$

이 값들을 정규방정식에 대입하면 다음을 얻게 된다.

$$5 = 10\hat{b}_0 + 2\hat{b}_1 + 2\hat{b}_2$$
$$6 = 2\hat{b}_0 + 6\hat{b}_1 + \hat{b}_2$$
$$7 = 2\hat{b}_0 + \hat{b}_1 + 4\hat{b}_2$$

이 방정식의 집합을 풀면 다음을 얻게 되는 것이다.

$$\hat{b}_0 = 0.045, \quad \hat{b}_1 = 0.727, \quad \hat{b}_2 = 1.545$$

따라서 추정된 회귀방정식은 다음과 같게 될 것이다.

$$\hat{Y} = 0.045 + 0.727X_1 + 1.545X_2$$

이로부터 X_1 과 X_2 에 대한 값의 특정집합과 관련된 Y의 평균치의 추정량이 구해진다.

추측하는바 대로, 계수의 추정치를 결정하기 위한 일련의 정규방정식의 실제적인 해는 상대적으로 적은 수의 변수의 경우에도 천문학적인 숫자계산을 포함할 수 있다. 실제적으로 다중회귀분석을 이용하는데는 컴퓨터가 필

요하다. 그럼에도 불구하고 원칙적으로 계수의 추정치가 어떻게 결정되는지를 이해하는 것은 중요하다. 이를 통해 결과를 정확하게 해석할 수 있을 뿐만 아니라 뒤에 설명하게 되겠지만 문제점을 찾아낼 수 있는 것이다.

나. 完全多重共線性의 문제

하나(또는 2이상)의 설명변수가 여타 변수의 완전한(perfect) 선형결합이라면 추정절차는 실패하고 만다는 것을 앞의 가정 마에서 설명하였다. 이제 그 이유를 알아볼 차례이다. 예를 들어(4.3)의 X_K가 다음과 같다고 하자.

$$X_{kt} = c_0 + c_1 X_{1t} + c_2 X_{2t} + \cdots + c_{k-1} X_{(k-1)t} \tag{4.13}$$

여기에서 c_0, c_1, …, c_{K-1}은 상수이다. 곧, 우연적으로 이 상수중 어떤 것은 0일 수도 있는 것이다. 앞에서 (4.11)에 X_{Kt}를 곱하여 합한 다음 $\Sigma(X_{Kt}\hat{u}_t) = 0$으로 함으로써 ($k+1$)번째 정규방정식을 만들어 내었음을 상기해 보자. 만일 X_{Kt}가 (4.13)에 나타낸 것과 같다면, 단지 첫번째 정규방정식에 C_0를, 두번째에 C_1을 차례로 곱하여 합산함으로써 (k+1)번째 정규방정식을 도출해 낼 수 있다. 예를 들어 (4.13)에 비추어 볼 때, (k+1)번째 정규방정식의 좌변은 다음과 같이 표현할 수 있는 것이다.

$$\sum(Y_t X_{kt}) = c_0 \sum Y_t + c_1 \sum(Y_t X_{1t}) + \cdots + c_{k-1} \sum(Y_t X_{(k-1)t})$$

독자들은 세가지 독립변수를 가진 다중회귀모형의 예를 통해 실험해 보면 이를 이해하게 될 것이다.

그러나 이는 (k+1)번째 정규방정식이 독립적인 방정식이 아님을 뜻한다. 곧, 그것은 처음의 k개의 방정식의 선형결합인 것이다. 도출하고자 하는 모수추정량의 갯수는 (k+1)개이지만(곧, $\hat{b}_0$, $\hat{b}_1$, …,

$\hat{b}_K$), 독립적인 방정식은 단지 k개인 것이다. 결론적으로 말하면, 일반적으로 모수추정량에 대한 해를 유일하게 얻을 수 없는 것이다. 보다 직관적으로 말하면, 문제는 하나의 독립변수가 사실상 항상 여타 독립변수의 값의 가중합이므로 그 독립변수가 종속변수에 미친 영향을 여타 독립변수에 의한 영향으로 부터 분리해 낼 수 없다는 것이다.

하지만 그러한 경우, 그 변수들의 결합효과는 추정할 수 있다. 예를 들어 (4.13)을 (4.3)으로 치환하면 다음을 얻게 된다.

$$\begin{aligned} Y_t &= (b_0 + b_k c_0) + (b_1 + b_k c_1)X_{1t} + \cdots \\ &\quad + (b_{k-1} + b_k c_{k-1})X_{(k-1)t} + u_t \\ &= d_0 + d_1 X_{1t} + \cdots + d_{k-1}X_{(k-1)t} + u_t \end{aligned} \tag{4.14}$$

일반적으로 여기서 $d_i = (b_i + b_K c_i)$이다. 방정식 (4.14)에는 모수가 k개 있게 된다. 곧, d_0, d_1, …, d_{K-1} 이다. 만일 바로 (4.13)과 같은 관계가 존재한다면, 다섯번째 가정을 만족하는 설명변수는 k개인 것이다. 간략하게 말한다면, 추정량 $\hat{d}_0$, $\hat{d}_1$, …, $\hat{d}_{K-1}$ 을 포함하는 k개의 독립적인 정규방정식을 도출할 수 있으며, 그에 대한 유일한 해도 구할 수 있다.

$d_i = (b_i + b_K c_i)$이므로 d_i의 추정량을 b_i의 추정량으로 간주할 수는 없다. 곧, 일반적으로 Y에 대한 X_i의 영향을 추정하는 것은 불가능하다. 한가지 예외는 있다. 곧 $c_i = 0$인 경우가 그것이다. 예를 들어 $c_5 = 0$이면, $d_5 = b_5$이어서 d_5의 추정량을 b_5의 값의 추정량으로 간주할 수 있는 것이다. (4.13)을 보면 X_K의 값이 X_i의 값에 의존하지 않으면 특정 c_i는 0임을 알 수 있다.

일반적으로 X_{Kt}와 같이 여타 설명변수의 선형결합인 "退化 (degenerate)" 설명변수의 계수는 추정불가능하다. 그러나 회귀방정식에서 몇가지 다른 독립변수의 계수는 추정할 수도 있다. 예를 들면, 다음과 같다.

$X_{2t} = 3 - 17\,X_{1t} + 8\,X_{5t}$이면, $c_0 = 3$, $c_1 = -17$, $c_5 = 8$을 제외한 모든 c_i는 0이다. 결국 b_0, b_1, b_2와 b_5를 제외한 처음의 모든 b_i는 추정 가능한 것이다. 결론적으로 말하면, 독자들은 이 결과를 암기하기 보다는 이렇게 도출한 기법을 익혀야 할 것이다. 곧, 하나(또는 2이상)의 설명변수가 몇몇 여타 변수와 선형적으로 연관되어 있다면 간단하게 선형관계를 이용하여 "퇴화(degenerate)" 설명변수를 대체한 다음 축소된(reduced) 회귀모형의 모수를 추정하면 된다. 마지막으로 원래의 회귀모형의 모수를 축소된 회귀모형의 모수와 비교함으로써 원래의 회귀모형의 모수가 어떤 값으로 추정될 수 있는지를 결정하면 되는 것이다.

3. 추정량의 특성과 가설검정

가. 추정량에 대한 설명*

이장 여러 군데에서 강조하였듯이, Y에 대한 X_K의 영향처럼 어떤 특정한 독립변수의 영향을 추정하기 위해서는 여타 독립변수의 영향을 분리시키거나 설명하여야만 한다. 그럴때만 Y에 대한 X_K의 영향을 분리해낼 수 있는 것이다. 비록 직접적으로는 앞의 정규방정식의 집합으로부터 분명하지는 않지만, 이러한 것이 바로 추정절차의 한 과정인 것이다. 이 장의 부록에서 추정량 $\hat{b}_i$에 대한 식을 도출할 수 있는 대안적인 접근법을 이용할 것이다. 이 기법을 통해 독자들은 추정절차가 어떻게 상이한 독립변수들의 영향을 분리하는지를 명확하게 알게 될 것이다. 부록에 있는 추정량에 대한 식을 통해 이변수의 경우와 같이 이러한 추정량이 不偏推定量(unbiased estimators)임을 쉽게 알 수 있을 것이다.

하지만 여기서는 부록의 내용을 단지 간략하면서도 아주 직관적으로 검

---

* 이절은 어려운 내용이다. 식(4.19)에 대한 이해를 제외한 나머지 내용은 처음에는 읽지 않아도 좋다.

토해 볼 것이다. 방정식 (4.3)에서 다른 조건이 동일할 때 X_K의 한단위 변화가 Y에 미치는 영향을 추정한다고, 즉 달리 말하면 b_K의 값을 추정한다고 가정한다.

만일 X_K가 여타 X의 선형결합이면, b_K는 추정불가능함은 알려진 사실이다. 따라서 X_K는 그와 같은 선형결합이 아니라고 가정하여 보자. 하지만 이는 X_K가 여타 설명변수와 완전히 상관되어 있지 않음을 의미하는 것은 아니다. 가령, 소비함수에서 X_K는 유동자산, X_1은 가처분소득일 수도 있다. 그렇다고 한다면, 분명히 X_K와 X_1 사이에는 正의 相關(positive correlation)이 존재한다고 예상하게 될 것이다. 하지만 어떤 사람의 신장과 체중처럼 이 두 변수가 완전히 관련되어 있지 않아야 하는 것이다.*

회귀모형(4.3)에서 약간 일반화하면 X_K는 완전하지는 않지만 몇몇 또는 모든 설명변수와 관련되어 있을지도 모른다. 만일 그러할 경우, 이는 적어도 여타 설명변수의 값에 의해 X_K의 값을 설명할 수 있음을 뜻한다. X_k가 여타 설명변수와 관련되어 있는 선형회귀방정식을 써서 이러한 일을 시도하여 보기로 하자.

$$X_{kt} = c_0 + c_1X_{1t} + \cdots + c_{k-1}X_{(k-1)t} + v_{kt} \qquad (4.15)$$

여기에서 v_{Kt}는 교란항이다. 또한 대변수기법**에 의해 (4.15)의 모수 $c_0, c_1, \cdots, c_{K-1}$을 추정하여, 추정량으로서 $\hat{c}_0, \hat{c}_1, \cdots, \hat{c}_{K-1}$을 얻게 된다고 가정하자. 이 추정량을 이용하면, $X_1, \cdots, X_{K-1}$에 따른 X_K의 설명된 (또

---

* 어떤 의미에서 개인의 유동자산은 완전히 소득에 의해 결정되는 것은 아니다. 예를 들면, 두사람의 소득은 동일할지라도 유동자산은 다를 수 있다.

** 정규방정식은 다음에 의해 얻게 된다.

$$\Sigma\hat{v}_{Kt}=0\ ,\ \Sigma\ (\hat{v}_{Kt}X_{1t}) = 0\ ,\ \cdots,\ \Sigma(\hat{v}_{Kt}X_{K-1}) = 0$$

는 계산된)값은 다음과 같을 것이다.

$$\hat{X}_{kt} = \hat{c}_0 + \hat{c}_1 X_{1t} + \cdots + \hat{c}_{k-1} X_{(k-1)t} \tag{4.16}$$

X_{Kt}가 여타 설명변수의 완전한 선형결합이 아니라고 가정하였기 때문에 일반적으로 $X_{kt} \neq \hat{X}_{kt}$임을 알고 있다. 그러므로 $\hat{X}_{Kt}$를 다음과 같이 나타낼 수 있다.

$$X_{kt} = \hat{X}_{kt} + \hat{v}_{kt} \tag{4.17}$$

여기에서 $\hat{v}_{Kt}$는 여타 독립변수가 설명할 수 없는 X_{Kt}의 부분이다. 곧 $\hat{v}_{Kt} = X_{Kt} - \hat{X}_{Kt}$이다. 흔히 $\hat{v}_K$의 항은 X_K가 $X_1, \cdots, X_{K-1}$과 관련된 회귀의 잔차(residual)라고 한다.

$\hat{v}_{Kt}$는 b_K의 추정에 중요하다. 왜냐하면 그것은 어떤 의미에서 여타 설명변수와 독립적인 X_{Kt}의 부분을 나타내기 때문이다. 가령, (4.17)과 (4.16)에서 $\hat{v}_{Kt} = 0$이라면 완전다중공선성이 존재하며, 따라서 b_K는 추정불가능하다는 것을 알고 있다. 부록에서 보게 되겠지만 b_k의 추정량은 다음과 같이 나타낼 수 있다.

$$\hat{b}_k = \frac{\sum (\hat{v}_{kt} Y_t)}{\sum \hat{v}_{kt}^2} = \frac{\sum (X_{kt} - \hat{X}_{kt}) Y_t}{\sum (X_{kt} - \hat{X}_{kt})^2} \tag{4.18}$$

곧, $\hat{b}_K$에 대한 정규방정식의 해는 (4.18)과 같이 표현할 수 있는 것이다. 잔차항 $\hat{v}_K$의 값이 X_K를 통해서만 X_1의 값에 의존하게 된다는 사실을 유념해야 한다. 유사한 관계는 여타 추정량에도 유효하다. 부록에서 살펴보면 다음과 같다.

$$\hat{b}_i = \frac{\sum (\hat{v}_{it} Y_t)}{\sum \hat{v}_{it}^2} = \frac{\sum (X_{it} - \hat{X}_{it}) Y_t}{\sum (X_{it} - \hat{X}_{it})^2}, \qquad i = 1, \ldots, k \tag{4.19}$$

여기서 $\hat{v}_{it}$는 X_i의 회귀에서 여타 설명변수에 대한 잔차이다.* 요약하면, 추정량 $\hat{b}_i$는 단지 여타 설명변수와 완전다중공선성의 관계에 있지 않은 X_i의 부분, 곧 $\hat{v}_i$에만 달려 있다. 그 대신에, 추정량 $\hat{b}_i$는 단지 $\hat{v}_i$을 이용함으로써 구할 수 있다.

이제 $\hat{Y}_{ti}$은 X_{it}를 제외한 모든 설명변수와 완전다중공선성의 관계에 있는 Y_t의 부분이라고 하여 보자.** 그러면 $(Y_t - \hat{Y}_{ti})$은 이러한 설명변수와 다중공선성의 관계가 없는 Y_t의 부분이 될 것이다. 다음을 유념해야 할 것이다. (4.19)의 $\hat{v}_{it}$은 아래와 같다.

$$\sum \hat{v}_{it} = 0, \sum (\hat{v}_i X_{jt}) = 0, \quad j \neq i \tag{4.20}$$

이 때문에 다음이 만족되어야 한다.

$$\sum (\hat{v}_{it} \hat{Y}_{ti}) = 0 \tag{4.21}$$

---

* 이변수 경우의 $\hat{b}$에 대한 식과 (4.19)의 $\hat{b}_t$에 대한 식이 유사하다는 것에 유의하라. 제2장에서 이변수모형의 경우, 다음과 같다.

$$\hat{b} = \frac{\sum (X_t - \bar{X}) Y_t}{\sum (X_t - \bar{X})^2}$$

연습삼아 독자들은 이변량모형으로 부터 $\hat{b}$에 대한 식이 단지 (4.19)의 특별한 경우임을 논증해 낼 수 있다. 곧 이변량의 경우 $\hat{X}_t = \bar{X}$인 것이다(도움말 : 이변량의 경우, (4.15)에 따른 방정식은 $X_t = c + v_t$일 것이다).

** 이는 X_{it}를 제외한 모든 설명변수에 대해 Y_t를 회귀하여 그 예측치를 계산함으로써 구할 수 있다. 예를 들면, 다음의 회귀모형을 고려할 수 있다.

$$Y_t = \gamma_0 + \gamma_1 X_{1t} + \cdots + \gamma_{i-1} X_{(i-1)t} + \gamma_{i+1} X_{(i+1)t} + \cdots + \gamma_k X_{kt} + w_t$$

여기서 W_t는 교란항이다. 그러면 대변수기법을 이용하여 $\hat{\gamma}_0, \hat{\gamma}_1, \cdots, \hat{\gamma}_{(i-1)}, \hat{\gamma}_{(i+1)}, \cdots, \hat{\gamma}_k$을 얻게 되며, 이에 따라 다음을 얻게 된다.

$$\hat{Y}_{ti} = \hat{\gamma}_0 + \hat{\gamma}_1 X_{1t} + \cdots + \hat{\gamma}_{i-1} X_{(i-1)t} + \hat{\gamma}_{i+1} X_{(i+1)t} + \cdots + \hat{\gamma}_k X_{kt}$$

(4.21)로부터 다음을 얻게 된다. 곧, $Y_t = \hat{Y}_{ti} + (Y_t - \hat{Y}_{ti})$이므로 (4.19)의 추정량 $\hat{b}_i$은 다음과 같이 나타낼 수 있을 것이다.

$$\hat{b}_i = \frac{\sum \hat{v}_{it}(Y_t - \hat{Y}_{ti})}{\sum \hat{v}_{it}^2} = \frac{\sum (X_{it} - \hat{X}_{it})(Y_t - \hat{Y}_{ti})}{\sum (X_{it} - \hat{X}_{it})^2} \tag{4.22}$$

이제 지금까지의 절차가 여러 설명변수들이 종속변수 Y_t에 미치는 영향을 어떻게 분리해 내는지가 분명하게 되었다. 모든 여타 설명변수의 선형적인 영향이 X_{it}와 Y_t 두 변수에서 제거된 후, (4.22)의 추정량 $\hat{b}_i$은 단지 그 두변수에 의존하게 되는 것이다. "여타 변수들"의 영향을 제거함으로써 절차는 다변량의 경우를 이변량의 경우로 축소 시킨다는 것을 직관하게 될 것이다.

나. 추정량의 분산

이 장 부록에 있는 (4.19)의 도출을 통해 진행하기로 하자. 특히, 보다 정식으로 다루게 되면, 우선 $\hat{b}_1$이 b_1의 불편추정량임을 보일 수 있다.(곧, $E(\hat{b}_i) = b_i$). 둘째로 $\hat{b}_i$의 조건분산(conditional variances)에 대한 식을 개발할 수 있다. 보다 명확하게 하면, 다음을 보이게 된다.

$$\text{var}(\hat{b}_i) = \sigma_{b_i}^2 = \frac{\sigma_u^2}{\sum \hat{v}_{it}^2}, \qquad i = 1, 2, \ldots, k \tag{4.23}$$

여기에서 σ_u^2은 처음의 다중회귀방정식(4.3)의 교란항의 분산이다.*

다. 信賴區間과 가설검정 : 서론

이변수의 경우와 마찬가지로 교란항이 정규분포를 한다고 계속 가정한다

---

* 다시 한번 이변수경우의 $\hat{b}_1$의 분산에 대한 식과 다중회귀방정식의 $\hat{b}_i$의 경우가 유사함에 유의하라. 그 차이는 이변수모형의 $\hat{b}$의 분산에 대한 식의 분모에서 $\Sigma (X_t - \bar{X})^2$의 항이 $\Sigma \hat{v}_{it}^2 = \Sigma (X_{it} - \hat{X}_{it})^2$으로 대체되는 것이다.

면 설명변수의 값이 주어졌을 때, 각 추정량 $\hat{b}_i$은 정규분포를 한다는 것이 사실일 것이다. 이는 각 $\hat{b}_i$이 교란항의 선형결합이며, 따라서 제3장에서 검토해 보았듯이 정규변수(normal variables)의 선형결합 또한 정규분포를 하기 때문인 것이다. 결과를 기호로 나타내면 다음과 같다.*

$$\hat{b}_i \sim N(b_i, \sigma_{\hat{b}_i}^2), \qquad i = 0, 1, \ldots, k \tag{4.24}$$

여기에서

$$\sigma_{\hat{b}_i}^2 = \frac{\sigma_u^2}{\sum \hat{v}_{it}^2}$$

남아있는 문제점은 일반적으로 σ_u^2이 알려져 있지 않기 때문에 추정량의 분산을 결정할 수 없다는 것이다. 따라서 관찰치표본으로부터 σ_u^2의 추정량을 세울 필요가 있다. σ_u^2이 실제로는 u_t^2의 평균, 곧 $\sigma_u^2 = \mathrm{E}[u_t^2]$이므로, b_i의 값을 알고 있다면, σ_u^2의 추정량으로서 표본평균을 취하는 것은 유의미할 것이다. 표본평균은,

$$\frac{\sum u_t^2}{n} = \frac{\sum (Y_t - b_0 - b_1 X_{1t} - \cdots - b_k X_{kt})^2}{n} \tag{4.25}$$

불행하게도 계수의 값은 모르지만, 그에 대한 추정량은 있다. (4.25)의

---

* b_0에 관한 설명변수는 $X_{0t} \equiv 1$ ($t = 1, 2, \cdots, n$). 그러므로 $\hat{v}_{0t}$은 다른 모든 설명변수 $X_{1t}, \cdots, X_{Kt}$에 대한 X_{01}의 회귀에서 잔차가 될 것이다. 곧,

$$X_{0t} = e_1 X_{1t} + e_2 X_{2t} + \cdots + e_k X_{kt} + v_{0t}$$

이 회귀방정식은 상수항이 종속변수이기 때문에 상수항을 포함하지 않음에 유의하라. 대조적으로 (4.15)와 같이 여타 v_{it}를 정의하는 회귀방정식은 상수항을 포함한다. 왜냐하면 하나의 설명변수는 종속변수이기 때문이다. 간단하게 말하면 X_{0t}는 또다른 설명변수로 간주될 수 있는 것이다.

b_i 각각을 그 추정량으로 대체하여 교란항의 분산에 대한 추정량으로서 얻을 수 있게 된다.

$$\hat{\sigma}_u^2 = \frac{\sum_{t=1}^{n}(Y_t - \hat{b}_0 - \hat{b}_1 X_{1t} - \cdots - \hat{b}_k X_{kt})^2}{n-(k+1)} \tag{4.26}$$

(4.26)의 분모가 〔$n-(k+1)$〕임에 유의해야할 것이다. 곧, 이는 (k+1)개의 모수추정에서 결과하는 분자의 (k+1)개의 자유도의 상실을 반영하는 것이다. 이변수의 경우에서처럼 $E(\hat{\sigma}_u^2) = \hat{\sigma}_u^2$ 인 것은 물론이다.

(4.26)의 추정량은 불편추정량이다. (4.26)을 이용하여 각 $\hat{b}_i$의 분산을 다음과 같이 추정할 수 있다.

$$\hat{\sigma}_{b_i}^2 = \frac{\hat{\sigma}_u^2}{\sum \hat{v}_i^2} \tag{4.27}$$

라. 신뢰구간과 가설검정

이제 개별계수에 대한 信賴區間을 설정할 차례이다. 제3장의 결과를 이용하면 $\hat{b}_i \sim N(b_i, \sigma_{\hat{b}i}^2)$이므로 다음을 얻게 되는 것이다.

$$\frac{\hat{b}_i - b_i}{\sigma_{\hat{b}_i}} \sim N(0,1)$$

그러면, 바로 이변수의 경우에서처럼 신뢰구간을 설정할 수 있으며, σ_u^2와 $\sigma_{\hat{b}i}^2$을 알고 있다면, 정규곡선(normal curve)에 의해 가설을 검정할 수 있다. 예를 들어 정규곡선에 기초한 경우 b_i에 대한 95%의 신뢰구간은 다음과 같게 됨을 독자 스스로 입증해야 한다.

$$\hat{b}_i \pm 1.96\sigma_{\hat{b}_i}$$

이변수의 경우에서 처럼 σ_u^2은 전형적으로 알려져 있지 않으며, 따라서 추정해야만 하는 것이다. 결국, 일반적으로 신뢰구간을 설정하거나 또는 t분포에 의해 가설을 검정하기 위한 t비율을 만들게 된다. 이를 위해 다

음을 주목하게 된다.

$$\frac{\hat{b}_i - b_i}{\hat{\sigma}_{\hat{b}_i}} \tag{4.28}$$

(4.28)은 $(n-k-1)$의 자유도를 가진 t 변수이다. (4.28)의 t 변수의 자유도가 항상 (4.26)의 분산추정량의 분모와 동일하다는 것에 유의해야 한다. 이변수의 경우 $k=1$임을 생각해 보자. 그러면 추정할 모수는 2개이므로 $(n-2)$의 자유도를 가진 t 분포를 얻게 된다.

(4.28)을 이용하면, 개별계수들에 관한 가설을 검정하기 위해 제3장에서 개발하였던 동일한 절차를 채택할 수 있게 되는 것이다. 예를 들어 다음의 모형이 있다고 가정하여 보자.

$$Y_t = b_0 + b_1 X_{1t} + \cdots + b_9 X_{9t} + u_t, \qquad t = 1, \ldots, 25 \tag{4.29}$$

그리고 0.05의 제1종의 과오(Type 1 error)* 로 다음의 가설을 검정한다고 하여 보자.

$$H_0: b_3 = 0,$$
$$H_1: b_3 \neq 0$$

여기에서 관찰치 표본의 수는 25이다.

이 문제에서 $n=25$, $k=9$이므로 다음과 같다.

$$\frac{\hat{b}_3 - b_3}{\hat{\sigma}_{\hat{b}_3}}$$

이는 $(25-9-1=15)$의 자유도를 가진 t 변수인 것이다. t 분포에 대

---

* 제1종의 과오(Type 1 error)란 眞을 僞로 판단함으로써 범한 과오를 말한다. 위의 경우 귀무가설을 기각할 확률이 0.05인 것이다(역자주).

한 통계표 2 에서 b_3에 대한 95 퍼센트의 신뢰구간은 다음과 같다.

$$(\hat{b}_3 \pm 2.131\hat{\sigma}_{\hat{b}_3}) \tag{4.30}$$

이제 제 3 장에서와 마찬가지로 표본을 이용하여, $\hat{b}_3$과 $\hat{\sigma}^2_{\hat{b}3}$을 계산하고, 이 값들을 (4.30)에 치환한 다음, 그로부터 결과하는 구간이 귀무가설의 구간을 포함하는지의 여부를 알아보면 된다. 그렇다고 한다면 H_0를 채택할 것이고 그렇지 않다면 기각시킬 것이다. 대신 간단하게 t 비율을 계산하고 略式檢定(rule of thumb)에 의해 그것의 절대값과 2 를 비교할 수도 있다. 물론 이러한 경우 臨界値(cut-off value)는 2.131 인 것이다. 여하튼 (4.28)과 같은 결과가 존재하는 경우, 가설검정과 신뢰구간에 관한 문제는 이변량회귀의 경우와 마찬가지로 다중회귀의 경우에서도 동일한 방식으로 해결되는 것은 명약관화한 사실이다.

4. 다중결정계수

앞절에서 다중회귀모형의 개별계수의 추정과 그에 대한 가설검정을 위한 기법을 구축하였다. 이제 회귀방정식 전체의 설명력(explanatory power)에 대한 논의가 남아 있다. 모든 독립변수를 통틀어 생각하면 어느 정도 만큼의 종속변수의 변동량을 설명할 수 있는가?

제 2 장에서 이변수의 경우에 대해 그러한 척도 R^2 을 개발하였다. R^2 은 추정된 회귀방정식이 설명할 수 있는 Y의 변동량을 가리키는 0 과 1 사이의 값을 가짐을 상기하여야 할 것이다. 유사한 절차를 여기에서도 따르면, 동일한 설명력을 가진 R^2에 대한 동일식이 다중회귀분석에도 적용될 수 있음을 보일 것이다.

다음의 기본적인 다중회귀모형을 고려하여 보기로 하자.

$$Y_t = b_0 + b_1X_{1t} + \cdots + b_kX_{kt} + u_t \tag{4.31}$$

(4.31)을 대변수추정기법에 의해 추정하면 Y_t 를 다음과 같이 나타낼수 있음을 이미 보았다.

$$Y_t = \hat{Y}_t + \hat{u}_t \tag{4.32}$$

여기에서

$$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 X_{1t} + \cdots + \hat{b}_k X_{kt} \tag{4.33}$$

이고, $\hat{u}_t$ 은 다음과 같다.

$$\sum \hat{u}_t = 0, \qquad \sum (\hat{u}_t X_{it}) = 0, \qquad i = 1, \ldots, k \tag{4.34}$$

이제 표본에 대해 (4.32)를 합산하면 다음을 얻게 된다.

$$\sum Y_t = \sum \hat{Y}_t + \sum \hat{u}_t \tag{4.35}$$

$\sum \hat{u}_t = 0$ 이므로 이변량회귀의 경우와 마찬가지로 다음과 같게 된다.

$$\sum Y_t = \sum \hat{Y}_t \tag{4.36}$$

(4.36)을 n 으로 나누면 Y_t 의 표본평균은 $\hat{Y}_t$ 의 표본평균과 같다는 사실이 나타나게 된다.

$$\bar{Y} = \bar{\hat{Y}} \tag{4.37}$$

(4.32)로 돌아가서 방정식의 양변을 제곱하여 보자.

$$Y_t^2 = \hat{Y}_t^2 + \hat{u}_t^2 + 2\hat{Y}_t \hat{u}_t \tag{4.38}$$

표본전체에 대해 합산하면 다음을 얻게 된다.

$$\sum Y_t^2 = \sum \hat{Y}_t^2 + \sum \hat{u}_t^2 + 2\sum (\hat{Y}_t \hat{u}_t) \tag{4.39}$$

(4.39)의 마지막항은 0과 같다. 다음을 보기로 하자.

$$\begin{aligned}\sum (\hat{u}_t \hat{Y}_t) &= \sum \hat{u}_t (\hat{b}_0 + \hat{b}_1 X_{1t} + \cdots + \hat{b}_k X_{kt}) \\ &= \hat{b}_0 \sum \hat{u}_t + \hat{b}_1 \sum (\hat{u}_t X_{1t}) + \cdots + \hat{b}_k \sum (\hat{u}_t X_{kt}) = 0\end{aligned}$$

이는 (4.34)에 의해 도출된다. 따라서 방정식 (4.39)는 다음과 같이 간단하게 정리된다.

$$\sum Y_t^2 = \sum \hat{Y}_t^2 + \sum \hat{u}_t^2 \tag{4.40}$$

다음으로 (4.40)의 양변에서 $n\bar{Y}^2$을 빼기로 한다.

$$\sum Y_t^2 - n\bar{Y}^2 = (\sum \hat{Y}_t^2 - n\bar{Y}^2) + \sum \hat{u}_t^2 \tag{4.41}$$

(4.37)에서 $\bar{Y} = Y$임을 상기하면 (4.41)을 다음과 같이 나타낼 수 있다.*

$$\sum (Y_t - \bar{Y})^2 = \sum (\hat{Y}_t - \bar{Y})^2 + \sum \hat{u}_t^2 \tag{4.42}$$

그런데 이것은 제2장에서 도출된 이변수의 경우에 대한 해당방정식과 일치하는 것이다.

이러한 관계를 다음과 같이 표현하였음을 기억해야할 것이다.

$$\text{TSS} = \text{RSS} + \text{ESS} \tag{4.43}$$

여기에서 $\text{TSS} = \Sigma (Y_t - \bar{Y})^2$, $\text{RSS} = \Sigma (\hat{Y}_t - \bar{Y})^2$, 그리고 $\text{ESS} = \Sigma \hat{u}_t^2$이다. 總自乘合(total sum of squares, TSS)은 종속변수의 표본평균에 대한

---

* 여기서 사용하는 전제는 제1장의 부록A에 의한 것으로 어떠한 변수 Z_t에 대해 $\Sigma (Z_t - \bar{Z})^2 = \Sigma Z_t^2 - n\bar{Z}^2$이다.

변동량으로 이는 회귀방정식으로써 설명하려는 것이다. 곧, 회귀모형은 왜 Y_t는 상수가 아닌지를 설명해 줄 수 있을 것이다. 방정식이 설명할 수 없는 부분은 誤差自乘合(error sum of squares, ESS)이다. 곧, TSS와 ESS의 차이는 회귀방정식이 설명할 수 있는 부분, 곧 回歸自乘合(regression sum of squares, RSS)임에 틀림없다.

이변수의 경우와 마찬가지로 관찰치 Y의 변동량중에서 추정된 회귀방정식이 설명할 수 있는 부분을 회귀방정식의 설명력의 척도로 이용한다. 따라서 다음을 얻게 된다.

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\text{RSS}}{\text{TSS}} = \frac{\sum (\hat{Y}_t - \bar{Y})^2}{\sum (Y_t - \bar{Y})^2} \tag{4.44}$$

여기에서 R^2은 다중결정계수(coefficient of multiple determination)라 한다.

돌이켜 보면 모든 경우 Y의 계산치가 그 관찰치와 같다면 곧 $Y_t = \hat{Y}_t$이면, 각 $\hat{u}_t$은 0일 것이고, 따라서 ESS = 0일 것이다. 그러므로 RSS = TSS이고, R^2은 1이라는 극대값이 될 것이다. 추정된 방정식이 종속변수의 변동량을 전혀 설명하지 못하는 다른 극단적인 경우, ESS는 ESS = TSS라 할 수 있을 만큼 가능한 한 크게 되어 RSS = 0이자 R^2 = 0이 된다. R^2이 1에 가까울수록, 추정된 회귀방정식의 설명력은 더욱 커지는 것이다. 마지막으로 그에 대한 증명을 서술하지는 않을지라도(왜냐하면 그것은 이변수의 경우와 동일하므로), R이 Y_t와 $\hat{Y}_t$사이의 상관계수의 추정량에 다름없다는 사실을 보일 수 있음도 지적해 두고자 한다.

5. 다중회귀분석 : 두가지 實例

지금까지 다중회귀분석의 원리를 전개하였으므로 이제 두가지 실제적인 다중회귀연구를 검토함으로써 사용되는 기법을 관찰하고, 이러한 연구의 결과

를 해석하여 보기로 한다.

가. 多變量소비함수

이장 처음에 소비지출수준은 현재의 소득수준 뿐만 아니라 여러가지 변수에 의존한다고 하였다. 사실상 경제학자들은 소비에 관한 많은 이론을 정립하였으며, 이러한 이론에 대한 광범위한 계량경제학적 검토를 수행하였다. 예를 들면, 알버트 안도(Albert Ando)와 프랑코 모딜리아니 (Franco Modigliani)는 平生消費理論을 가정하였는데, 그 이론에서 어떤 개인의 현재 소비수준은 자신의 미래, 곧 평생의 소득흐름(stream of income)에 달려있다고 그들은 주장하였다.* 이 가설에 대한 한가지 간단한 시행분석으로서 그들은 다음을 주장하였다.

$$C_t = b_0 + b_1 Y_{dt} + b_2 A_{t-1} + u_t \qquad (4.45)$$

여기에서 현재의 가처분노동소득(disposable labor income) Y_{dt}는 노동서비스로부터 얻는 예상소득(expected income)에 대한 代理變數(proxy variable)로 쓰이며, $(t-1)$ 시점말기의 소비자의 순자산은 예상비노동재산소득(expected nonlabor property income)이다. 곧, 소비자들은 A_{t-1}을 가지고 t시점에 참여하는데, 그들은 그로부터 임대소득(rental income) 또는 이자소득(interest income)을 얻는다. 이에 따라 안도와 모딜리아니는 다중회귀분석을 써서 (4.45)의 계수의 값을 추정하려 하였다. 1929～1959년동안 (1941～1945년의 전시기간은 제외)의 소비지출, 가처분노동소득과 순자산에 대해 경상 1억달러로 측정한 연도별 자료를 이용하여

---

* The 'Life Cycle' Hypothesis of Saving : Aggregate Implications, and Tests," American Economic Review, 53(March 1963), pp.55～84.

그들은 이러한 변수들에 대해 25개의 관찰치를 수집하였다. 이러한 시계열자료를 써서 그들은 다중회귀기법을 이용하여 방정식을 추정한 결과, 다음의 사실을 알게 되었다.

$$\hat{C}_t = \underset{(1.46)}{5.33} + \underset{(0.047)}{0.767}Y_{dt} + \underset{(0.010)}{0.047}A_{t-1} \qquad N = 25, \quad R^2 = 0.999 \tag{4.46}$$

(주 : 추정계수하단의 괄호안 숫자는 각각의 표준오차임)

안도와 모딜리아니의 결과에 첫번째 주목하게 되는 것은 추정계수의 부호가 예상한대로 陽이며(positive), 가처분노동소득의 추정된 한계소비성향(MPC)이 0과 1사이에 있다는 것이다. 그러므로 추정방정식은 소비행위에 대해 가정하였던 특징과 일치하는 것으로 나타나게 되었던 것이다.

다음으로 (4.46)에서 주목해야할 것은 세 경우 모두에서 모수추정치가 각각의 표준오차의 추정치의 세배를 넘는다는 것이다. t비율(t ratios)과 관련된 略式檢定(rules of thumb)에 따르면, 이는 다음을 의미한다. 곧, 귀무가설 $b_0 = 0$, $b_1 = 0$, 또는 $b_3 = 0$중 어느 것이나 단측검정 또는 양측검정에 대해 고려해볼 경우 (4.46)의 결과에 기초하여 통상의 5%또는 1%의 유의수준하에서 기각되는 것이다.

더욱 자세한 설명으로서 $\hat{b}_1$, 곧 가처분노동소득의 MPC가 0.5와 같다는 가설을 양측검정하여 보기로 하자.

$$H_0 : b_1 = 0.5,$$
$$H_1 : b_1 \neq 0.5$$

자유도 22와 5%의 유의수준의 경우, 통계표2에서 보면 t의 임계치(critical value)가 2.07이다. 그리하여 (4.46)으로부터 계산된 t비율은 다음과 같다.

$$\frac{0.767 - 0.5}{0.047} = \frac{0.267}{0.047} = 5.7 > 2.07$$

(4.46)의 결과에 기초하여 5%의 유의수준하에서 귀무가설은 기각될 것이다.* 회귀분석에 대한 연습을 좀 더 하고자 한다면 실례를 통해 풍부히 검토해 볼 수 있는 것이다. 예를 들면 위의 경우에 얼핏 보아도 $\hat{b}_1 = 0.767$은 가설치 (hypothesized value) $b_1 = 0.5$에서 나온 표준오차의 두배 (2×0.047)이상임이 분명한 것이다. 따라서 가설 $b_1 = 0.5$는 복잡한 계산없이도 기각될 수 있는 것이다.

마지막으로 안도와 모딜리아니가 정식화한 방정식이 상당한 설명력(explanatory power)을 가지고 있음에 유의해야 한다. $R^2 = 0.999$로 두 독립변수는 소비에 대한 관찰치의 변동량의 99% 이상을 설명할 수 있는 것이다. 그들의 이론에 대한 연산형태(operational form)를 받아들이는 경우, 이러한 실증적인 결과들은 소비행위에 관한 그들의 견해와 일치하는 것으로 나타난다.

이러한 결과들을 설명하는데는 분석결과가 여타 이론에도 부합될 수 있다는 것을 인정하는 것이 중요하다. 실증적인 결과들이 그 이론에 부합된다고 할 수 있는 반면에 그렇다고 해서 소비에 관한 여타 모든 이론이 부정확하다고는 할 수 없다. 이는 경제학에서 어떤 특정문제에 대한 실증작업이란 어떤 가설에 적합한 또는 부합되지 않는 증거가 축적되는 연속적인 과정이기 때문인 것이다. 곧 어떤 가설에 대한 실증적인 지지는 그 가설에 부합되는, 그리고 그에 못지않게 경쟁적인 가설에 부합되지 않는 분석결과의 정도에 달려있는 것이다.

나. 도시 조세의 연구

이 장의 결론을 맺기 위해 이번에는 횡단면자료를 이용하여 또다른 다중회귀연구를 검토할 것이다. 도시문제가 위급한 오늘날 대부분의 대도시의

---

* 분명히 H_0는 1%의 유의수준하에서도 기각될 것이다.

시장들은 예산상승을 감당해낼 수 있는 추가적인 수입의 원천을 걱정스럽게 구하고 있다. 문제는 대체로 도시의 경제에 심각한 손상을 입히지 않은 채 새로운 조세를 부과하거나 현존의 조세에 대한 세율을 올리는 것 중의 하나가 된다. 특히 도시에서의 높은 수준의 세율은 중간계층의 교외로의 이전을 조장하며, 도시의 재화와 용역에 대한 구매력을 떨어뜨린다는 사실은 많은 관찰자들이 주장하고 있다. 미국의 많은 도시중에서 주요 수입원 중의 하나가 소매세(tax on retail sales)이다. 곧 그러한 조세는 도시의 소매를 줄이는 결과를 초래하였으며, 이에 따라 도시의 경제를 쇠퇴시키는 원인이 되었다는 주장이 있다.

이것은 사실인가, 그리고 만일 그러하다면 그것은 양적인 의미에서 아주 중요한 것인가? 이는 대답하기 곤란한 문제이지만 계량경제학적 분석을 이용하여 그 문제에 도달하는 방법에 관해 몇가지 생각하여 보기로 하자. 사실상 교외보다 도심에서의 세율이 높다는 것이 도시에서의 판매액을 줄이게 된다면, 다른 조건이 동일한 경우(other things equal) 어떤 도시의 판매세율과 교외의 판매세율과의 격차(differential)가 크면 클수록 도시의 1인당 소매액수준은 낮아질 것이다. 만일 1인당 소매액과 이러한 소매세격차의 크기를 결정할 수 있는 도시에 대한 하나의 표본을 얻을 수 있다면, 조세격차변수에 대한 판매액변수의 회귀방정식을 구할 수 있으며, 이에 따라 계수추정치의 부호와 크기를 검토하고 적절한 검정을 수행할 수 있을 것이다.

그러한 접근법이 충분히 이치에 닿는 반면에 이러한 방법에는 용이하지 않은 느낌을 가져다 주는 한가지 문제가 내재한다. 가설은 다른 조건이 동일한 경우 보다 높은 판매세격차가 보다 낮은 판매액수준과 관련된다는 것이지만, 도시의 표본내에서는 분명히 다른 조건이 동일하지 않은 것이다. 어떤 특정 도시의 소매액수준은 분명히 조세이외에도 많은 여타의 중요한 변

수들, 곧 소득수준과 교외인구의 상대적인 크기와 같은 변수들에 의존하는 것이다. 예를 들어 도시 거주민이 부유할수록 1인당 판매액수준이 높을 것임은 당연하다. 게다가 교외인구가 많을수록 잠재적인 도시의 구매자는 많을 것이고, 도시의 소매액도 높을 것이다. 관념상으로는 여기에서 하고자 하는 것은 판매세격차의 영향을 분리해낼 수 있도록 도시판매액에 대한 이러한 여타 결정요인의 영향을 검토하는 것이다.

이에 따라 다중회귀분석이 사용되어야 할 것이다. 특히 조세변수 뿐만 아니라 도시판매액에 대한 여타 중요한 결정요인을 포함하는 일련의 독립변수에 대한 판매액변수의 회귀방정식을 구할 수 있다. 그와 같은 연구는 바로 존 마이크셀(John Mikesell)에 의해 이루어졌다.*

마이크셀은 다음과 같이 가정하였다.

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} + b_4X_{4t} + u_t \qquad (4.47)$$

여기에서

Y_t = t번째 도시의 1인당 소매액

X_{1t} = t번째 도시에 대한 판매세격차

X_{2t} = t번째 도시의 1인당 소득

X_{3t} = t번째 도시에 대한 도시인구와 전체 수도권인구의 비율

X_{4t} = t번째 도시의 면적

판매세격차(sales-tax differential) 변수는 다음과 같이 정의된다.

$$X_{1t} = \frac{1 + t_c}{1 + t_s}$$

---

* "Central Cities and Sales Tax Rate Differentials : The Border City Problem," National Tax Journal 23 (June 1970), pp.206 ~ 213.

여기에서 t_c는 그 도시의 판매세율이며, t_s는 그 도시주위의 교외의 평균판매세율이다. t_s가 불변인 채 t번째 도시의 판매세율의 증가는 X_{1t}를 증가시킬 것이고, 그러므로 방정식 (4.47)에 따라 1인당 도시판매액의 감소가 초래될 것이다. 왜냐하면 $b_1 < 0$ 가정되고 있기 때문이다.

마이크셀은 미국의 173개 도시와 각각의 교외를 대상으로 이러한 모든 변수에 대한 자료를 수집할 수 있었다. 그리하여 그는 이러한 자료를 이용하여 다중회귀기법에 의해 (4.47)을 추정한 결과 다음을 알게 되었다.

$$\hat{Y} = \underset{}{4.5} - \underset{(2.94)}{7.44X_1} + \underset{(0.10)}{0.43X_2} - \underset{(0.04)}{0.11X_3} - \underset{(0.02)}{0.08X_4} \qquad N = 173, \quad R^2 = 0.26 \qquad (4.48)$$

여기에서 괄호안의 숫자는 관련된 추정표준오차이다. 직관적으로 조세변수 X_1의 계수는 예상된 負의 부호를 갖고 있으며 해당 표준오차의 크기의 두배를 넘는다는 것에 주목하게 된다. 그것은 t비율이 2를 넘는다는 것을 의미한다. 이는 다음과 같이 주장된다. 곧 5%의 유의수준하에서 귀무가설 $b_1 = 0$을 기각하고 대립가설 $b_1 < 0$을 채택할 수 있다는 의미에서 사실상 도시의 판매액과 판매세격차사이에는 負의 관계가 있다는 것이다. 더군다나 b_1의 추정치는 전형적으로 그 영향의 크기가 상당히 클 수 있음을 나타내는데 유의하여야 할 것이다. 곧, 도시와 교외의 판매세간의 격차가 1% 증가하면, 평균적으로(on average) 중심도시의 소매액이 약 7% 감소하게 되는 것이다.*

---

* 방정식 (4.47)은 마이크셀의 결과의 아주 간단한 예이다. 사실상 그는 다음과 같은 형태의 곱셈관계를 가정하였다.

$$Y_t = b_0 X_{1t}^{b_1} X_{2t}^{b_2} X_{3t}^{b_3} X_{4t}^{b_4} e^{u_t}$$

그리하여 로그를 취하면 다음을 얻는다.

$$\ln Y_t = \ln b_0 + b_1 \ln X_{1t} + b_2 \ln X_{2t} + b_3 \ln X_{3t} + b_4 \ln X_{4t} + u_t$$

다음장에서 이런 종류의 곱셈관계를 검토할 것인데, 여기에서는 그것이 앞 장에서의 로그변환시도의 간단한 일반화라는 사실에 유의하고자 한다.

또한 이러한 결과들은 조세변수에 대한 도시판매액의 단순회귀(simple regression)보다 훨씬 더 설득력이 있다는 사실을 인식하여야 하는 것이다. 이러한 다중회귀의 결과로 마이크셀은 판매액에 대한 1인당 소득, 도시-교외인구비율과 도시면적의 영향을 분명하게 설명하였던 것이다. 바꿔 말하면 도시의 판매액에 대한 조세변수의 영향은 설명된 이러한 여타 변수들의 영향과 함께 측정된 것이라 할 수 있다. 그러므로 마이크셀의 결과는 판매세격차가 중심도시의 소매액의 상당한 감소를 초래한다는 전제와 일치하는 것이다.

부록. 추정량의 특성

이 부록의 목적은 회귀계수에 대한 추정량과 그 분산을 보다 엄격하게 개발하여 이러한 추정량이 불편추정량임을 증명하기 위한 것이다. 더구나 본문에서 직관적으로 보았듯이 사용하게 될 접근법은 다중회귀분석이 상이한 독립변수의 영향들을 선별할 수 있는 방법에 몇가지 추가적인 통찰력을 제공할 것이다.

본문에서 어떤 회귀모형의 독립변수중 적어도 몇가지는 비록 완전하지는 않을지라도 몇가지 또는 모든 여타 변수와 관련될 수 있다고 하였다.

본문에서처럼 회귀모형을 통해 여타의 변수에 의하여 마지막 독립변수를 설명할 수 있다고 한다.

$$X_{kt} = c_0 + c_1 X_{1t} + \cdots + c_{k-1} X_{(k-1)t} + v_{kt} \qquad (4A.1)$$

대변수기법에 의해 (4A.1)의 모수를 추정한다고 하여보자. 이를 위해 $\Sigma \hat{v}_{Kt}=0$, $\Sigma(\hat{v}_{Kt} X_{1t})=0$, $\cdots$, $\Sigma(\hat{v}_{Kt} X_{(K-1)t})=0$으로 놓음으로써 일련의 정규방정식을 얻게 될 것이다.

$$\begin{aligned}
\sum X_{kt} &= n\hat{c}_0 + \hat{c}_1 \sum X_{1t} + \cdots + \hat{c}_{k-1} \sum X_{(k-1)t} \\
\sum (X_{kt} X_{1t}) &= \hat{c}_0 \sum X_{1t} + \hat{c}_1 \sum X_{1t}^2 + \cdots + \hat{c}_{k-1} \sum (X_{1t} X_{(k-1)t}) \qquad (4A.2) \\
&\vdots \\
\sum (X_{kt} X_{(k-1)t}) &= \hat{c}_0 \sum X_{(k-1)t} + \hat{c}_1 \sum (X_{1t} X_{(k-1)t}) + \cdots \\
&\quad + \hat{c}_{(k-1)} \sum X_{(k-1)t}^2
\end{aligned}$$

(4A.2)에서 미지수가 k개, 곧, $\hat{c}_0$, $\hat{c}_1$, …, $\hat{c}_{K-1}$ 가 있으며, 주어진 가정하에서 일반적으로 미지수의 해를 구할 수 있도록 방정식이 k개 있음에 유의해야 할 것이다. 이에 따라 다음을 설정할 수 있게 된다.

$$\hat{X}_{kt} = \hat{c}_0 + \hat{c}_1 X_{1t} + \cdots + \hat{c}_{k-1} X_{(k-1)t} \qquad (4A.3)$$

v_{Kt}의 추정량은 다음과 같을 것이다.

$$\hat{v}_{kt} = X_{kt} - \hat{X}_{kt} \qquad (4A.4)$$

(4A.4)의 항을 재배열하면 다음을 얻게 된다.

$$X_{kt} = \hat{X}_{kt} + \hat{v}_{kt} \qquad (4A.5)$$

지금까지 한 것에 대해 유의해 보자. X_K의 값이 두 요소로 분리되었다. 첫번째 부분 $\hat{X}_{Kt}$은 여타 설명변수와 직접적으로 관련된 X_K의 부분이다. 곧, 그것은 어떤 의미에서 여타 X들에 의해 설명될 수 있는 X_K의 변동량의 부분인 것이다. 그와 대조적으로 $\hat{v}_{Kt}$은 여타 독립변수에 의해 설명될 수 없는 X_K의 부분이다. 만일, 모든 t에 대해 $\hat{v}_{Kt} = 0$이라면, 완전 다중공선성이 존재하며 〔(4A.5)와 (4A.3)을 볼 것〕, b_K는 추정불가능하다는 사실을 유의해야 한다. 이러한 이유로 $\hat{v}_{Kt}$는 b_K의 추정에 중대한 역할을 담당한다고 생각해 볼 수 있다.

이를 기억해두면서 $\hat{b}_1$을 결정하는 정규방정식이 (k+1)개 있음을 이장의 본문으로부터 되새겨 보도록 하자. 그중 하나가 다음과 같다.

$$\sum (Y_t X_{kt}) = \hat{b}_0 \sum X_{kt} + \hat{b}_1 \sum (X_{1t} X_{kt}) + \cdots + \hat{b}_k \sum X_{kt}^2 \qquad (4A.6)$$

X_{Kt} 는 $\hat{X}_{Kt}$ 와 $\hat{v}_{Kt}$ 의 합으로 표시할 수 있음을 (4A.5)로부터 알고 있다. 또한 $i = 1, 2, \cdots, (k-1)$ 에 대해 $\Sigma \hat{v}_{Kt} = 0$, $\Sigma (\hat{v}_{Kt} X_{it}) = 0$ 임도 알고 있다. 그리고 그에 따라 $\hat{X}_{Kt}$ 이 $X_{1t}, X_{2t}, \cdots X_{(K-1)t}$ 의 선형결합이므로 $\Sigma (\hat{v}_{Kt} \hat{X}_{Kt}) = 0$ 임도 알고 있다〔(4A.3)을 볼 것〕. 만일 (4A.6) 에서 X_{Kt} 를 $(\hat{X}_{Kt} + \hat{v}_{kt})$ 으로 치환하여 0 과 같은 항을 모두 없애면, 다음을 얻게 된다.

$$\sum (Y_t \hat{X}_{kt}) + \sum (Y_t \hat{v}_{kt}) = \hat{b}_0 \sum \hat{X}_{kt} + \hat{b}_1 \sum (X_{1t} \hat{X}_{kt}) + \cdots + \hat{b}_k \sum \hat{X}_{kt}^2 + \hat{b}_k \sum \hat{v}_{kt}^2 \qquad (4A.7)$$

(4A.7)의 우변에 있는 처음의 $(k+1)$ 항의 합은 단지 $\Sigma (Y_t \hat{X}_{Kt})$ 와 같다.

$$\sum (Y_t \hat{X}_{kt}) = \hat{b}_0 \sum \hat{X}_{kt} + \hat{b}_1 \sum (X_{1t} \hat{X}_{kt}) + \cdots + \hat{b}_k \sum \hat{X}_{kt}^2 \qquad (4A.8)$$

이를 간략하게 보기 위해 (4A.8)의 좌변의 합에 다음의 Y_t 를 대입한다.

$$Y_t = \hat{b}_0 + \hat{b}_1 X_{1t} + \cdots + \hat{b}_k X_{kt} + \hat{u}_t$$

그리고 다음을 주목하기로 한다.

$$\sum (\hat{u}_t \hat{X}_{kt}) = \hat{c}_0 \sum \hat{u}_t + \hat{c}_1 \sum (\hat{u}_t X_{1t}) + \cdots + \hat{c}_{(k-1)} \sum (\hat{u}_t X_{(k-1)t}) = 0$$

이와 같이 하는 이유는 (4A.7)에 (4A.8)을 대입하면 다음의 결과를 얻기 때문이다.

$$\sum (Y_t \hat{v}_{kt}) = \hat{b}_k \sum \hat{v}_{kt}^2 \qquad (4A.9)$$

그러면 $\hat{b}_K$ 의 해는 다음과 같다.

$$\hat{b}_k = \frac{\sum (Y_t \hat{v}_{kt})}{\sum \hat{v}_{kt}^2} \qquad (4A.10)$$

간략하게 말하면, $\hat{b}_K$은 단지 Y_t와 $\hat{v}_{Kt}$에 의존한다. 그러데 여기에서 $\hat{v}_{Kt}$은 X_{Kt}의 "독립적인 (independent)" 변동량을 표시한다.

(4A.10)의 일반화는 간단하다. 일반적으로 다음과 같다.

$$\hat{b}_i = \frac{\sum (Y_t \hat{v}_{it})}{\sum \hat{v}_{it}^2} \qquad (4A.11)$$

여기에서 $\hat{v}_{it}$은 여타 모든 X에 대한 X_{it}의 회귀의 殘差(residual)이다. 곧, 각 b_i의 추정량은 Y_t의 값, 그리고 관련 설명변수X_{it}의 "독립적인" 변동량을 표시하는 항인 $\hat{v}_{it}$의 값에 의해 표현할 수 있는 것이다.

가. 불편추정량(unbiased ēstimators)

앞의 절에서 다음을 알고 있다. 곧,

$$\hat{b}_i = \frac{\sum (Y_t \hat{v}_{it})}{\sum \hat{v}_{it}^2} \qquad (4A.11)$$

$\hat{b}_i$가 불편추정량을 보이기 위해서 본문의 (4.3)의 기본모형을 다음과 같이 다시 쓰기로 한다.

$$Y_t = b_0 + b_1 X_{1t} + \cdots + b_i(\hat{X}_{it} + \hat{v}_{it}) + \cdots + b_k X_{kt} + u_t \qquad (4A.12)$$

여기에서 $(\hat{X}_{it} + \hat{v}_{it})$은 X_{it}를 치환한 것이다. 다음으로 (4A.12)에 $\hat{v}_{it}$을 곱하고, n개의 관찰치에 대해 합산한 다음, (4A.11)의 $\Sigma (Y_t \hat{v}_{it})$에 대입하면, 다음을 얻게 된다.

$$\hat{b}_i = b_i + \frac{\sum (\hat{v}_{it} u_t)}{\sum \hat{v}_{it}^2} \qquad (4A.13)$$

(4A.13)을 얻기 위해 정규방정식 $\Sigma \hat{v}_{it} = 0$과 $\Sigma (\hat{v}_{it} X_{jt}) = 0$ ($j \neq i$)을 사용하였으며 이에 따라 $\Sigma (\hat{v}_{it} \hat{X}_{it}) = 0$를 도출, 사용하였다. 유의해야

할 것은 $\hat{v}_{it}$ 은 u_t의 값에 좌우되지 않는다는 것, 그리고 $\hat{v}_{it}$ 은 단지 X의 주어진 값에만 의존한다는 것이다. 어떤 주어진 X의 집합에 대해서도 u_t 가 X와는 독립적이라는 것을 가정하고 있기 때문에 $\hat{v}_{it}$ 은 또한 주어진 것이 될 것이며, 따라서 다음과 같게 된다.

$$\begin{aligned} E(\hat{b}_i) &= E(b_i) + E\left[\frac{\sum(\hat{v}_{it}u_t)}{\sum \hat{v}_{it}^2}\right] \\ &= b_i + \left(\frac{\hat{v}_{i1}}{\sum \hat{v}_{it}^2}\right)E(u_1) + \cdots + \left(\frac{\hat{v}_{in}}{\sum \hat{v}_{it}^2}\right)E(u_n) \\ &= b_i \end{aligned} \qquad (4A.14)$$

따라서 b_i 의 추정량은 불편추정량이다.

나. 추정량의 분산

이제 (4A.13)으로부터 조건분산(conditional variance)을 도출하는 것은 간단한 일이다. 특히 (4A.13)을 확장시키면 다음을 얻게 된다.

$$\hat{b}_i = b_i + \left(\frac{\hat{v}_{i1}}{\sum \hat{v}_{it}^2}\right)u_1 + \left(\frac{\hat{v}_{i2}}{\sum \hat{v}_{it}^2}\right)u_2 + \cdots + \left(\frac{\hat{v}_{in}}{\sum \hat{v}_{it}^2}\right)u_n \qquad (4A.15)$$

$M_{it} = \hat{v}_{it} / \sum \hat{v}_{it}^2$ 이라 두면 다음을 얻게 된다.

$$\hat{b}_i = b_i + M_{i1}u_1 + \cdots + M_{in}u_n \qquad (4A.16)$$

곧, $\hat{b}_i$ 는 교란항의 선형결합인 것이다. 상관되지 않은 확률변수의 선형합(linear sum)의 분산에 대해 제2장의 식을 이용하면 다음을 얻게 된다.

$$\text{var}(\hat{b}_i) = M_{i1}^2\sigma_u^2 + M_{i2}^2\sigma_u^2 + \cdots + M_{in}^2\sigma_u^2 \qquad (4A.17)$$

$A = \sum \hat{v}_{it}^2$ 이라 하자. 그러면 $M_{it}^2 = v_{it}^2/A^2$ 이다. 이를 이용하면 (4A.17)은 다음과 같이 다시 쓸 수 있다.

$$\text{var}(\hat{b}_i) = \frac{\sigma_u^2}{A^2}(\hat{v}_{i1}^2 + \hat{v}_{i2}^2 + \cdots + \hat{v}_{in}^2)$$

$$= \frac{\sigma_u^2(\sum \hat{v}_{it}^2)}{A^2} \qquad (4A.18)$$

$$= \frac{\sigma_u^2}{A^2} A = \frac{\sigma_u^2}{A}$$

간략하게 하면 다음과 같을 것이다.

$$\text{var}(\hat{b}_i) = \frac{\sigma_u^2}{\sum \hat{v}_{it}^2}, \qquad (i = 1, \ldots, k) \qquad (4A.19)$$

문 제

1. 다음의 회귀모형을 생각해 보자.

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + u_t$$

a. 이 모형의 바탕을 이루고 있는 표준적인 가정을 열거하라.

b. 정규방정식을 쓰고, 각 방정식과 관련된 특정한 가정을 약술하라.

c. 표본에서 $n = 100$, $\Sigma X_1 = \Sigma X_2 = \Sigma X_1 X_2 = 0$, $\Sigma Y = 10$, $\Sigma(YX_1) = 30$, $\Sigma(YX_2) = 20$, $\Sigma X_1^2 = 35$, $\Sigma X_2^2 = 3$을 가정한다. a_0, a_1과 a_2를 추정하라.

2. 모형 $Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + a_3(X_{1t} - X_{2t}) + a_4 X_{1t} X_{2t} + \varepsilon_t$를 생각해 보자. 표준적인 가정하에서 어떤 모수가 추정불가능한가? 그 이유는?

3. 다음의 회귀모형을 생각해 본다. 곧,

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t$$

여기에서 X_{1t}, X_{2t}, Y_t에 대한 관찰치는 다음과 같다.

X_{1t}	X_{2t}	Y_t
1	2	7
2	1	8

1	3	5
3	1	6
1	2	4

정규방정식을 쓰라.

4. 중위소득가계와 상위소득가계는 상대적으로 높은 세율, 높은 범죄율, 높은 거주비용때문에, 그리고 그들이 "보다 넓은 공간"을 원하기 때문에 도시를 떠난다고들 한다. 그와 같은 가설을 검정하기 위해 이용할 수 있는 회귀모형을 설정하라. 이러한 가설을 검정하기 위해 횡단면자료와 시계열자료의 상대적인 장점을 설명하라.

5. 다음을 가정한다.

$$D_{1t} = a_0 + a_1 P_{1t} + a_2 P_{2t} + \cdots + a_k P_{kt} + b\bar{P}_t + cY_t + u_{1t}$$

여기에서 D_{1t}는 상품1에 대한 수요, P_{1t}는 상품1의 가격, P_{2t}, $\cdots P_{kt}$는 여타 재화 (k−1)개의 가격, $\bar{P}_t = \sum_{i=1}^{K} (P_{it})/\,k$는 전체가격수준, 그리고 Y_t는 소득이다. 간단하게 말해서 상품1에 대한 수요는 그 상품의 가격, 여타 재화의 가격, 전체가격수준과 소득에 좌우된다 라고 일반적으로 믿고 있다.

a. 모형의 추정에서 어떤 문제가 발생할 것으로 보이는가?

b. 위의 모형에서 어떠한 모수를 추정할 수 있는가? 논증하시오.

6. 다음의 모형을 생각해 보자.

$$Y_t = b_0 + b_1 X_t + b_2 X_t^2 + \varepsilon_t$$

a. 위의 문제는 완전다중공선성을 갖고 있는가?

b. 다음의 관찰치가 존재한다고 하자.

Y	−1	−1	2	4
X	0	1	2	5

정규방정식을 구하라.

제 5 장 다중회귀분석의 추가적인 기법

앞 장에서는 二變數回歸模型에 대한 推定技法을 독립변수가 다수인 경우로 일반화하였다. 이 장에서는 多重回歸分析에서 이용할 수 있는 몇가지 추가적인 기법을 검토할 것이다. 보다 분명하게 말하면, 우선 時差變數(lagged variables)에 대한 앞에서의 분석을 다중회귀의 경우로 확장시킬 것이다. 그리고 이와 관련하여 시차가 주어진 관계(lagged relationships)의 여러 형태를 추정하기 위한 세가지 방법을 개발할 것이다. 둘째로, 假變數("dummy" variables)라는 개념을 도입할 것이다. 그런데 그러한 변수는 경제관계에 관여하는 몇가지 질적인 영향을 설명할 수 있게 하여준다. 예를 들어 가변수를 이용하면, 인종적 요인 또는 종교적 요인과 같은 것들이 어떤 유형의 행동에 미치는 영향을 설명할 수 있다. 이 기법에 의해 상투적인 양적인 의미에서는 정상적으로 측정할 수 없는 변수들을 분석에 포함시킬 수 있을 것이다. 마지막으로 함수형태에 관한 문제로 돌아가서 상이한 많은 유형의 관계를 추정하기 위해 다중회귀분석을 어떻게 이용할 수 있는지를 보게 될 것이다.

1. 時差關係의 추정

이미 제 3 장에서 종속변수가 이전 시점의 독립변수의 값에 좌우되는 방정식을 추정하기 위해 이변수모형을 이용할 수 있음을 보았다. 예를 들면, 어떤 특정시점의 소비지출이 이전 시점의 가치분소득수준에 좌우되는 경우를 고려하였음을 상기하여 보자. 곧, 다음과 같은 형태일 것이다.

$$C_t = a + bY_{d(t-1)} + u_t \qquad (5.1)$$

하지만 왜 현재의 소비가 바로 이전 시점의 가처분소득에만 좌우되는지에 대한 필연적인 이유는 없다. 곧, 현재의 시점 뿐만 아니라 이전 시점들의 소득수준 또한 소비지출에 어느 정도 영향을 줄 수도 있는 것이다. 만일 이것이 사실이라면, 다음과 같은 형태의 관계가 존재할 것이다.

$$C_t = a + b_0 Y_{dt} + b_1 Y_{d(t-1)} + \cdots + b_k Y_{d(t-k)} + u_t \qquad (5.2)$$

이러한 유형의 관계를 分布時差(distributed lag)라 한다. 이는 어떤 시점에서든지 종속변수의 값은 독립변수의 과거치(past values)의 가중합(weighted sum)에 의존함을 의미한다. 어느 정도 직관적으로 종속변수(이 경우에는 C_t)가 독립변수(Y_{dt})의 현재값에 "느리게(sluggish)" 적응하는 것으로 생각해 볼 수 있을 것이다. 왜냐하면, 독립변수의 과거치(시차가 주어진 값)로부터 형성된 "관성(inertia)" 때문이다.

그러한 C_t 와 Y_{dt} 사이의 관계를 초래하게 될지도 모르는 정형적인 모형의 한가지 예란 t시점의 소비지출이 $(t+1)$시점의 예상소득(expected income)에 의존한다고 가정되어 다시 예상소득은 이전 시점들의 소득수준의 가중합으로서 결정되는 경우이다. 예를 들어, 다음과 같이 가정하자.

$$C_t = a + bY^e_{d(t+1)} + u_t \qquad (5.3)$$

여기에서 $Y^e_{d(t+1)}$ 는 $(t+1)$시점의 예상소득이다. 예상소득은 단지 현재와 과거소득의 가중합이라고 가정해 보자.

$$Y^e_{d(t+1)} = \alpha_0 Y_{dt} + \alpha_1 Y_{d(t-1)} + \cdots + \alpha_k Y_{d(t-k)} \qquad (5.4)$$

(5.4)를 (5.3)에 대입하면 다음을 얻게 된다.

$$C_t = a + b_0 Y_{dt} + b_1 Y_{d(t-1)} + \cdots + b_k Y_{d(t-k)} + u_t \qquad (5.5)$$

여기에서 $b_0 = b\alpha_0$, $b_1 = b\alpha_1$ 등이다. 그러므로 C_t는 Y_{dt}에 느리게 반응한다. 왜냐하면 Y_{dt}는 $Y^e_{d(t+1)}$를 결정하는데 단지 하나의 요인이기 때문이다.

여하튼 원칙적으로는 직접 다중회귀기법을 이용하여 방정식(5.2)을 간단하게 추정할 수 있을 것이다. 현재의 소비지출은 현재의 소득과 이전 9년간의 소득에 의존한다는 가설을 세워보자. 그러면 다음의 방정식을 얻게 될 것이다.

$$C_t = a + b_0 Y_{dt} + b_1 Y_{d(t-1)} + \cdots + b_9 Y_{d(t-9)} + u_t \qquad (5.6)$$

(5.6)을 추정하기 위해서는 〈표5.1〉에 표시된 형태의 자료가 있으면 된다. 각 행이 결합관찰치(joint observation)인 〈표5.1〉의 항목에 해당하는 소비와 가처분소득에 대한 자료를 이용하면서 다중회귀기법을 사용함으로써 $\hat{a}$, $\hat{b}_0$, $\hat{b}_1$, ···, $\hat{b}_9$을 얻을 수 있게 된다. 곧, 〈표5.1〉에 표현된 자료가 존재한다고 가정하면, 간단하게 $Y_{d(t-1)}$을 X_{1t}로, $Y_{d(t-2)}$를 X_{2t}로, ···, $Y_{d(t-9)}$를 X_{9t}로 다시 정의할 수 있는 것이다. 그리고 이에 따라 소비함수 (5.6)이 마치 다음과 같은 형태의 일반적인 (ordinary) 다중회귀방정식인 것처럼 그것을 추정할 수 있을 것이다.

$$C_t = a + b_0 Y_{dt} + b_1 X_{1t} + b_2 X_{2t} + \cdots + b_9 X_{9t} + u_t \qquad (5.7)$$

X_{1t}, ···, X_{9t}에 관한 관찰치는 〈표5.1〉에 들어있는 것이다.*

---

* 예를 들어 $X_{2(1951)} = Y_{d(1949)}$임을 독자들은 알아야 할 것이다.

〈표 5.1〉

C_t	Y_{dt}	$Y_{d(t-1)}$	$Y_{d(t-2)}$	···	$Y_{d(t-9)}$
C_{1950}	$Y_{d(1950)}$	$Y_{d(1949)}$	$Y_{d(1948)}$	···	$Y_{d(1941)}$
C_{1951}	$Y_{d(1951)}$	$Y_{d(1950)}$	$Y_{d(1949)}$	···	$Y_{d(1942)}$
⋮	⋮	⋮	⋮	⋮	⋮
C_{1970}	$Y_{d(1970)}$	$Y_{d(1969)}$	$Y_{d(1968)}$	···	$Y_{d(1961)}$

이는 어떤 경우에는 추정할 수 있는 적절한 형태의 관계일 수 있는 반면, 몇가지 문제점을 드러내고 있다. 1시점의 시차를 2변수관계에 도입하였을 때 하나의 관찰치를 상실하였음을 제 3장에서 상기해 보아야 할 것이다. 현재의 경우 문제는 더욱 어렵게 되어 있다. 왜냐하면 방정식(5.2)에 포함되고 있는 가처분소득의 추가적인 시차치(additional lagged value) 각각에 대해 관찰치를 또다시 상실하기 때문이다. 가령 1960년부터 1969년까지의 10년동안의 소비와 가처분소득에 대한 관찰치가 있으며, 소비가 지난 9년간의 소득과 현재의 소득에 좌우되는 방정식(5.6)을 추정하기로 한다고 해보자. 이러한 경우, 10개의 관찰치중에서 9개를 잃어버리게 되는 것이다. 곧, (5.6)에서 나타나는 모든 변수에 대한 관찰치가 존재하는 연도는 유일하게 1969년 뿐일 것이다.

$$C_{1969} = a + b_0 Y_{d(1969)} + b_1 Y_{d(1968)} + \cdots + b_9 Y_{d(1960)} \tag{5.8}$$

1969년 이전의 소비에 대한 관계는 이용할 수 없는 1960년 이전의 가처분소득에 대한 자료를 필요로 하는 것이다. 결과적으로 회귀방정식을 추정할 수 있을 만큼 충분한 관찰치를 얻지 못하게 되는 것이다.*

* 단지 한개의 결합관찰치가 있다면, 하나의 독립적인 정규방정식만이 있게 됨을 독자들은 증명할 수 있어야만 할 것이다. 왜냐하면 모든 정규방정식은 첫번째 정규방정식에 비례하기 때문이다. 예를 들면, 방정식(5.7)과 (5.8)이 관련하여 두번째 정규방정식의 비례요인은 $\overline{Y}_{d(1969)}$ 일 것이다. 그리고 세번째 방정식의 (다음 페이지 계속)

(5.6)과 같은 모형의 추정과 관련된 또다른 한가지 문제는 k가 "크면(large), (그것은 항상 $k \geq 5$를 의미하도록 쓰인 것이다.)" 추정할 모수가 많다는 것이다. 게다가 이러한 모수들은 서로 현저히 관련된 변수에 대응하게 될 것이다. 예상할 수 있듯이(그리고 다음장에서 정식으로 보게 되듯이) 이것은 종속변수에 대한 상이한 독립변수의 영향들을 분리시키기 어렵게 한다. 다시 말하면, 이러한 조건아래에서 회귀모수추정량의 분산은 커지게 될 것이다.

이들은 분포시차분석과 관련된 두가지 기본문제이다. 첫째, 관찰치가 시차로 인해 상실된다. 그리고 둘째, 모수가 너무 많은 나머지 추정의 신뢰성이 없어지는 경우가 가끔 있다. 이러한 문제를 처리하기 위해 경제학자들은 시차부여(lagging)로 인해 상실되는 관찰치의 수를 줄이고 (또는) 추정할 모수의 수를 줄이는 분포시차에 대한 모형을 개발하였다. 이제 그러한 두가지 모형을 서술하기로 한다.

가. 코익(Koyck) 시차

그 첫번째 모형이 코익시차(Koyck lag)이다.* 이 모형은 분포시차관계

(앞 페이지 계속) 비례요인은 $Y_{d(1968)}$일 것이다. 〔도움말 : (5.7) 또는 (5.6)에 대한 유일한 관찰치는 t=1의 경우에 대한 것으로, 여기서 1시점은 1969년을 가리킨다.〕 정규방정식은 다음의 조건에서 구하게 된다.

$$\sum_{t=1}^{1} \hat{u}_t = \hat{u}_1 = 0, \sum_{t=1}^{1} (\hat{u}_t Y_{dt}) = \hat{u}_1 Y_{d1} = 0, \sum_{t=1}^{1} (\hat{u}_t X_{1t}) = \hat{u}_1 X_{11} = 0, \cdots \sum_{t=1}^{1} (\hat{u}_t X_{9t}) = \hat{u}_1 X_{91} = 0$$

그러므로 두번째 정규방정식은 첫번째 방정식에 Y_{d1}을 곱한 것이고, 세번째는 첫번째에 X_{11}을 곱한 것이 된다. 이 결과는 어떤 회귀모형이 k개의 모수를 가지고 있을때, 모수를 추정하기 위해서는 적어도 k개의 결합관찰치가 있어야 함을 일컫는 보다 일반화된 결과의 특별한 경우인 것이다.

* L.M.Koyck, Distributed Lags and Investment Analysis(Amsterdam: North Holland, 1954).

의 모수에 관해 한가지 가정을 세우고 있는데, 그것에 의해 시차관계는보다 단순한 형태로 변환될 수 있게 된다. 따라서 그 결과로 시차가 보다 적어지거나 추정할 모수가 보다 적어지는 것이다. 불행하게도 이러한 "보다 간단한" 형태도 흔히 간과되는 심각한 복잡성을 또한 갖고 있는 것으로 판명된다. 코익시차모형이 널리 이용되고 있기 때문에 그것을 설명하고 나서 그 결과에 대해 지적하기로 한다. 게다가 이에 대한 설명은 이후의 토론을 위한 중요한 수단으로 이용될 것이다.

비록 소비가 이전 연도의 가처분소득에 의존할지라도 보다 오랜 과거의 소득의 영향은 보다 최근년도의 소득의 영향보다는 적다는 가설을 세워보기로 하자. 보다 명확하게 하면, 현재의 소비는 가처분소득의 현재수준과 과거수준의 (오차항을 포함한) 가중합이며, 그 가중치는 보다 오랜 시점에 대해 연속적으로 감소한다고 가정하자. 코익시차의 공식화는 이러한 가중치들이 기하적으로 감소함을 가정한다.

예를 들면, λ를 0과 1사이에 있는 상수라고 하자. 그러면 (5.2)와 관련하여 코익공식은 다음과 같을 것이다.

$$b_i = \lambda^i b_0 \qquad i = 1, 2, \ldots, k \tag{5.9}$$

(5.9)를 (5.2)에 대입하면 다음을 얻게 된다.

$$C_t = a + b_0 Y_{dt} + (b_0\lambda) Y_{d(t-1)} + (b_0\lambda^2) Y_{d(t-2)} + \cdots + (b_0\lambda^k) Y_{d(t-k)} + u_t \tag{5.10}$$

방정식(5.10)은 소비가 현재의 소득수준과 과거의 소득수준에 의존하지만, λ의 멱(power)이 높아질수록 λ가 계속 작아지기 때문에 과거로 되돌아감에 따라 점차 작아지게 되는 것을 말한다.

이제 코익시차 공식의 의미를 보기로 한다. (5.10)에 1시점의 시차를 준 다음, λ를 곱하면 다음과 같다.

$$\lambda C_{t-1} = \lambda a + (\lambda b_0) Y_{d(t-1)} + (\lambda^2 b_0) Y_{d(t-2)} + \cdots + (\lambda^{k+1} b_0) Y_{d(t-k-1)} + \lambda u_{t-1} \quad (5.11)$$

만일 이제 (5.10)에서 (5.11)을 빼면, 다음과 같을 것이다.

$$C_t - \lambda C_{t-1} = (a - \lambda a) + b_0 Y_{dt} - \lambda^{k+1} b_0 Y_{d(t-k-1)} + (u_t - \lambda u_{t-1}) \quad (5.12)$$

(5.12)의 항을 다시 정리하면 다음을 얻게 되는 것이다.

$$C_t = (a - \lambda a) + b_0 Y_{dt} + \lambda C_{t-1} - (\lambda^{k+1} b_0) Y_{d(t-k-1)} + (u_t - \lambda u_{t-1}) \quad (5.13)$$

이제 k가 크다고 (시차가 주어진 해당년도의 수가 크다고) 가정하면 (5.13)의 거의 마지막항인 $(\lambda^{k+1} b_0) Y_{d(t-k-1)}$ 은 작을 것이다. 따라서 근사하게 이 항은 0과 같다고 하기로 한다.*

또한 표기를 간단하게 하기위해 $a^* = (a - \lambda a)$라고 하자. 그러면 (5.13)은 다음과 같이 간단하게 될 것이다.

$$C_t = a^* + b_0 Y_{dt} + \lambda C_{t-1} + (u_t - \lambda u_{t-1}) \quad (5.14)$$

이제 다음과 같다고 하자.

$$v_t = (u_t - \lambda u_{t-1}) \quad (5.15)$$

v_t는 단지 교란항에 의존하기 때문에 v_t 자체를 교란항으로 간주하는 것은 합리적이다. v_t가 교란항의 특성에 관한 회귀모형의 모든 가정을 만족한다는 가정(이는 불행하게도 여러 문헌에서 흔히 자행된다)을 잠깐동안 세워보자. 이러한 경우, (5.14)는 다음과 같이 표현할 수 있는것이다.

$$C_t = a^* + b_0 Y_{dt} + \lambda C_{t-1} + v_t \quad (5.16)$$

---

* 물론 $Y_{d(t-k-1)}$가 유한인 반면, $k \to \infty$이라고 가정하면 이 항은 0이 될 것이다. 이는 여러 문헌에서 행해지는 통상의 가정이다.

여기에서 a^*는 상수이고

$$E(v_t) = 0$$
$$E(v_t^2) = \sigma_v^2$$
$$E(v_t v_{t-i}) = 0$$
$$E(v_t Y_{dt}) = E(v_t C_{t-1}) = 0$$

(5.16)이 (5.2)를 얼마나 간단하게 나타내고 있는지에 주목해야 할 것이다. 현재의 소비에 대한 가처분소득의 모든 시차치들의 영향이 하나의 단일항 속에 포함된다. 곧 소비의 값 그 자체가 1시점만큼 시차가 주어진 것이다. 소득의 시차치의 계수대신 λ의 값을 추정하기만 하면 된다. 다시 말하면, 코익시차모형이 v_t에 관한 가정을 받아들인다면, (5.2)와 같은 모형의 모수는 추정에서 단 하나의 관찰치만을 잃게 되는 (5.16)과 같은 모형에 의해 추정될 수 있는 것이다.

추정절차에 관하여 보다 명확하게 하기 위해서는 (5.16)을 모수 a^*, b_0와 λ를 포함하는 다중회귀모형으로 간주하게 될 것이다. 代變數技法을 직접 적용시키면, 이러한 모수의 추정량, 곧, $\hat{a}^*$, $\hat{b}_0$와 $\hat{\lambda}$을 얻게 된다. 이것으로 (5.2)의 모든 모수의 추정량을 구할 수 있을 것이다. 가령 시차가 무한이라고(k가 무한이라고) 가정하면 다음과 같다.

$$\hat{b}_i = (\hat{\lambda}^i)\hat{b}_0, \qquad i = 1, 2, \ldots \tag{5.17}$$

그리고 $a^* = (a - \lambda a)$ 이므로 $a = [a^*/(1 - \lambda)]$ 일 것이고, 따라서 a의 추정량은 다음과 같을 것이다.

$$\hat{a} = \frac{\hat{a}^*}{1 - \hat{\lambda}} \tag{5.18}$$

결과를 보다 일반적으로 표현하면, (k가 무한이라는 가정을 포함한) 코익시차 가정으로 다음 형태의 모형을 보다 간단한 형태로 만들 수 있게

된다.

$$Y_t = a + b_0X_t + b_1X_{t-1} + \cdots + u_t \qquad (5.19)$$

그 형태는 다음과 같다.

$$Y_t = a^* + b_0X_t + \lambda Y_{t-1} + v_t \qquad (5.20)$$

여기에서는 단지 세개의 모수, 곧 a^*, b_0와 λ를 추정하기만 하면 되는 것이다. (5.19)와 (5.20)의 모수사이의 관계는 다음과 같다.

$$a = \frac{a^*}{1-\lambda}, \quad b_i = \lambda^i b_0, \quad i = 1, 2, \cdots \qquad (5.21)$$

분명히 모형을 단지 하나의 변수의 시차치를 포함하는 것으로 변환시킴으로써 코익시차의 공식화는 단지 하나의 결합관찰치만을 상실하게 될 것이다. 이제 코익모형의 매력은 분명해 진다.

하지만 위에서 말하였던 것처럼 코익모형에도 몇가지 곤란한 문제가 있다. 먼저, 추정할 방정식을 도출하는데 교란항 $v_t = (u_t - \lambda u_{t-1})$이 정상적으로 교란항에 대해 부여하는 모든 조건을 만족한다라고 단순히 가정하였음을 생각해야 한다. 불행히도 이것은 일반적으로 사실이 아니다. 만일 원래의 방정식(5.19)에서 u_t가 회귀모형의 가정을 만족한다면 (5.20)에서의 v_t는 일반적으로 그렇치 않다. 보다 명확하게 하면 v_t의 값은 서로간에 0이 아닌 상관을 가지고 있으며, 또한 하나의 독립변수-종속변수 Y의 시차치와도 상관을 갖고 있는 것이다. 예를 들어 $v_t = (u_t - \lambda u_{t-1})$이고 $v_{t-1} = (u_{t-1} - \lambda u_{t-2})$이면, v_t와 v_{t-1}은 공통된 항, 곧 u_{t-1}을 갖고 있기 때문에 서로 독립적이지 못하다. 진정으로 u_t에 관한 통상의 가정아래에서 다음과 같게 된다.

$$
\begin{aligned}
E(v_t v_{t-1}) &= E[(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2})] \\
&= E(u_t u_{t-1} - u_t \lambda u_{t-2} - \lambda u_{t-1}^2 + \lambda^2 u_{t-1} u_{t-2}) \\
&= -\lambda \sigma_u^2 \neq 0
\end{aligned} \tag{5.22}
$$

이제 방정식 (5.20) 에서의 교란항의 연속적인 값들사이의 共分散 (covariance) 이 0 이 아님을 알게 되었다. 마찬가지로 다음의 사실을 독자들이 보일 수 있도록 과제로 남겨두기로 한다.*

$$
E(v_t Y_{t-1}) \neq 0 \tag{5.23}
$$

요약해보면, 회귀모형의 모든 가정을 만족하는 (5.19) 의 형태의 방정식에서 출발하면 이 방정식의 코익변환 (Koyck transformation) 은 일반적으로 이러한 가정중의 몇가지에 대한 위배로 귀결된다.** 더구나 이러한 위배의 결과는 심각한 것이다. 가령 이후의 장에서 보게 되듯이 (5.23) 은 b, b_i 와 λ 의 추정량이 불편추정량이면서 일치추정량이 아님을 의미한다.***

---

* 도움말 : (5.20) 에서 Y_t 는 v_t 에 직접 의존하므로 Y_{t-1} 은 직접 v_{t-1} 에 의존할 것이다. $Y_{t-1} = a^* + b_0 X_{t-1} + \lambda Y_{t-2} + v_{t-1}$ 이다. 그러므로 v_t 와 v_{t-1} 은 독립적이지 않기 때문에 Y_{t-1} 은 분명히 v_t 와 관련되어 있을 것이다. (5.23) 을 수식으로 논증하기 위해서는 Y_{t-1} 에 v_t 를 곱하여 기대치를 구하면 된다. 이 때 $E(Y_{t-2}\, v_t) = 0$ 임에 유의하라

** 제 6 장과 제 7 장에서 회귀모형의 가정에 대한 위배를 처리하는 기법을 개발할 것이다.

*** 이는 회귀모형 (5.20) 의 세번째 정규방정식이 $\Sigma(\hat{v}_t\ Y_{t-1}) = 0$ 이라는 조건에 기초하고 있기 때문이다. 그런데 그 조건은 더 이상 (5.23) 과 일치하지 않는다. 보다 자세한 것은 이후에 설명하기로 한다.

위에서 개관한 추정문제에 덧붙여 말하면, 코익공식이 과거 시점의 영향은 어떤 특수한 방식으로 연속적으로 감소한다는 것을 가정하고 있다는 점에서 그것은 아주 제한적이다. 이는 반드시 그렇지만은 않을 것임이 분명하다. 예를 들어 습관의 영향때문에 바로 이전 시점의 소득수준이 현재 시점의 소득보다는 소비에 더욱 큰 영향을 미치게 되는 것도 사실일 수 있는 것이다. 곧, 그러한 관계는 분명히 코익시차공식에는 부합되지 않는다. 그러므로 코익시차보다 신축적이면서도 회귀모형의 가정에 대한 위배로 귀결되지 않는 분포시차모형(distributed lag model)이 극히 유용할 것이다.

나. 알몬(Almon) 시차*

코익시차절차에 의해 초래된 위배사항이 이후 개발될 어느정도 보다 일반적인 모형의 구조에서 처리될 수 있다라고 할지라도, 그 해결이 간단한 것은 아니다. 이러한 어려움은 알몬시차기법을 통해 피할 수 있다. 여기서 강조해야 할 것은 이 방법이 코익공식처럼 시차변수의 존재로 인해 상실되는 관찰치의 수를 줄이는 것은 아니라는 것이다. 하지만 알몬의 방법은 추정할 모수의 수를 축소시킨다. 더군다나 그것은 코익절차에 비해 두가지 색다른 장점을 갖고 있다. 첫째, 회귀모형의 가정중 어떠한 것도 위배하지 않는다. 둘째, 앞으로 보게 되듯이 허용할 수 있는 시차구조라는면

---

* Shirley Almon, "The Distributed Lag Between Capital Appropriations and Expenditures," Econometrica, 33(Jan, 1965), pp. 178-196. 이 절에서의 논의는 페어(Ray Fair)와 야페(Dwight Jaffee)의 알몬기법에 대한 설명을 따른다. Ray Fair and Dwight Jaffee, "A Note on the Estimation of Polynomial Distributed Lags", Econometric Research Memorandum No. 120, Princeton University, Feb. 1971.

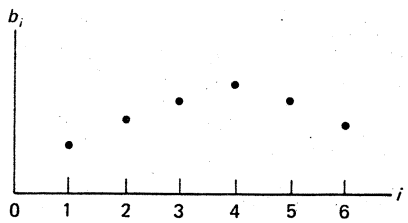
에서 코익의 방법보다 훨씬 신축적인 것이다.

시차관계에 대한 일반식으로 돌아가서 추정하고자 하는 모형이 다음의 형태라고 가정해 보자.

$$Y_t = a + b_0X_t + b_1X_{t-1} + \cdots + b_kX_{t-k} + u_t \qquad (5.24)$$

여기에서 교란항 u_t는 일반적인 가정을 만족한다. 앞에서와 같이 보다 오랜 시점과 관련된 X의 계수는 보다 가까운 시점과 관련된 계수보다 작다는 것을 예상할 수 있을 것이다. 다른 한편, 몇가지를 고려하는 경우, (5.24)와 같은 모형에서는 이것이 사실일 필요는 없게 된다. 곧 어떤 경우에 b는 실제로 처음에는 증가하다가 (곧, $b_1 > b_0$) 점차 줄어들기 시작할 수도 있는 것이다. 가령, 말하자면 인식에서의 시차, 정보수집에서의 지연, 또는 의사결정에 포함되는 시간소요 때문에 투자지출과 같은 변수는 실제로 수요조건이 오늘 어떠한가 보다는 이전 시점에 어떠하였는가에 보다 반응이 높을 수도 있는 것이다. 간단하게 말하면 상이한 가정하에서 (5.24)와 같은 모형에서 b의 값의 유형이 상이하다는 것을 예상할 수 있다.

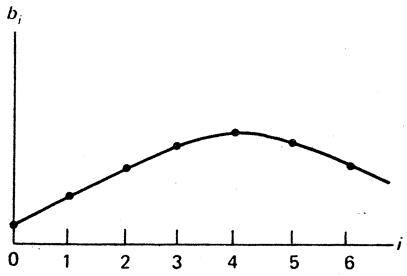
코익의 기법과는 달리 알몬의 기법은 그러한 b의 고정된 관계를 가정하지 않는다. 그 대신 b의 유형이 어떠하든지간에 그 유형은 다항식으로 표현될 수 있다고 가정한다. 예를 들어 b가 처음에 증가하다 감소한다고 예상하면, 그 유형은 어느 정도 〈그림 5.1〉에서와 같을 것이다.



〈그림 5.1〉

〈그림 5.2〉에서처럼 〈그림 5.1〉의 점들을 통과하는 부드러운 곡선을 그릴 수 있다고 가정하여 보자. 이제 〈그림 5.2〉의 곡선을 수식으로 나타내 보기로 하자. 수학에서는 일반적인 조건하에서 곡선은 다항식으로 근사화시킬 수 있다는 정리가 있다. 다항식의 차수* 를 결정하는 법칙은 차수가 최소한 곡선의 굴곡점의 수보다는 많아야 한다는 것이다. 이 법칙을 〈그림 5.2 〉에 적용하면 곡선을 2차다항식으로 근사화시킬 수 있다. 명확하게 하면 알몬의 기법을 이용하여 다음을 가정하기로 한다.

$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 \tag{5.25}$$



〈그림 5.2 〉

여기에서 α_0, α_1 과 α_2 는 결정될 상수이다. 만일 방정식 (5.25) 가 〈그림 5.2 〉의 곡선과 유사하다면, 다음과 같아야 한다.

$$\begin{aligned} b_0 &= \alpha_0 \qquad (i = 0) \\ b_1 &= \alpha_0 + \alpha_1 + \alpha_2 \qquad (i = 1) \\ b_2 &= \alpha_0 + 2\alpha_1 + 4\alpha_2 \qquad (i = 2) \\ &\vdots \\ b_k &= \alpha_0 + k\alpha_1 + k^2\alpha_2 \qquad (i = k) \end{aligned} \tag{5.26}$$

---

* "다항식의 차수(degree of a polynomial)"는 변수의 최고의 제곱수를 말한다는 것을 대수학에서 상기해야할 것이다. 그러므로 $Y = b_0 + b_1 X + b_2 X^2$ 는 이차다항식이며, $Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3$ 은 3차다항식인 것이다.

(5.26)의 각 b_i에 대한 식은 단지 i가 특정계수의 하첨자의 값과 같다고 놓음으로써 (5.25)에서 직접 도출된다.

방정식(5.25)는 시차가중치(lagged weight)의 값 b_i가 시차 그 자체의 길이 i와 관련된다는 점에서 처음에는 약간 이상해 보일 것이다. 실제로 이미 코익기법에서 유사한 관계가 있었던 것이다. 가정은 다음과 같다.

$$b_i = \lambda^i b_0 \tag{5.27}$$

방정식(5.27)에서는 b_i는 다시 i와 관련된다. (5.27)과 (5.25)의 유일한 차이란 방정식의 형태인 것이다.

알몬기법을 실행하기전에 그것이 어느 정도 신축적인지를 간단하게 논증해보기로 하자. b가 〈그림 5.3〉와 같은 유형을 따른다고 생각하였다고 한다. 그러면 간단하게 다음을 가정하게 될 것이다.

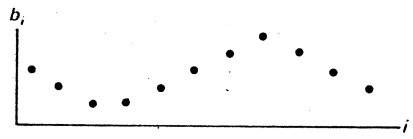
$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \alpha_3 i^3 \tag{5.28}$$

이는 다음을 의미한다.

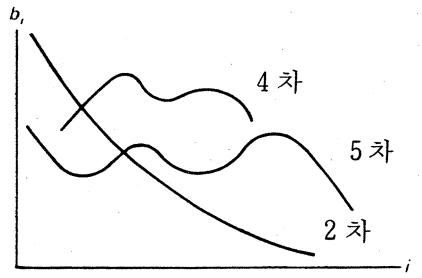
$$\begin{aligned} b_0 &= \alpha_0 \\ b_1 &= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 \\ b_2 &= \alpha_0 + 2\alpha_1 + 4\alpha_2 + 8\alpha_3 \\ &\vdots \\ b_k &= \alpha_0 + k\alpha_1 + k^2\alpha_2 + k^3\alpha_3 \end{aligned} \tag{5.29}$$

보다 일반적으로, 알몬의 방법을 이용하기 위해 해야할 일이란 b에 대해 가정한 유형에서의 굴곡점의 수를 세어 b를 i의 다항식으로 나타내는 것이다. 여기에서 다항식의 차수는 곡선에서의 굴곡점의 수에 1을 더한 것이다. 〈그림 5.4〉에서는 관련다항식의 차수에 따라 가능한 시차유형

의 수를 나타내고 있다.



〈그림 5.3〉



〈그림 5.4〉

이제 시차관계를 추정하기 위해 알몬의 기법을 어떻게 이용해야 할지를 알아보기로 한다. 앞에서 나타낸 일반식으로 돌아가기로 하자.

$$Y_t = a + b_0X_t + b_1X_{t-1} + \cdots + b_kX_{t-k} + u_t \qquad (5.24)$$

경제이론에 의하면 시차형태를 나타내는데는 2차다항식이 적당하다고 가정한다. 그러면 다음의 형태를 갖게 될 것이다.

$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 \qquad (5.30)$$

만일 (5.24)의 b를 (5.30)의 식으로 치환하면 다음과 같게 된다.

$$\begin{aligned} Y_t = {} & a + \alpha_0X_t + (\alpha_0 + \alpha_1 + \alpha_2)X_{t-1} \\ & + (\alpha_0 + 2\alpha_1 + 4\alpha_2)X_{t-2} + \cdots + (\alpha_0 + k\alpha_1 + k^2\alpha_2)X_{t-k} + u_t \end{aligned} \qquad (5.31)$$

(5.31)의 항을 정리하면 다음과 같다.

$$Y_t = a + \alpha_0\left(\sum_{i=0}^{k} X_{t-i}\right) + \alpha_1\left(\sum_{i=1}^{k} iX_{t-i}\right) + \alpha_2\left(\sum_{i=1}^{k} i^2X_{t-i}\right) + u_t \quad (5.32)$$

이제 다음과 같이 정의함으로써 표기를 간단하게 하여보자.

$$Z_{1t} = \sum_{i=0}^{k} X_{t-i}, \quad Z_{2t} = \sum_{i=1}^{k} iX_{t-i} \quad \text{그리고} \quad Z_{3t} = \sum_{i=1}^{k} i^2X_{t-i} \quad (5.33)$$

그러면 (5.32)를 다음과 같이 다시 쓸 수 있을 것이다.

$$Y_t = a + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + u_t \quad (5.34)$$

방정식(5.34)는 Y_t를 Z_{1t}, Z_{2t}, Z_{3t}와 관련시키는 통상다중회귀모형(ordinary multiple-regression model)이다. 표준적인 추정기법을 이용하여 a, α_0, α_1과 α_2의 추정량을 쉽게 만들어낼 수 있는 것이다. $\hat{a}$, $\hat{\alpha}_0$, $\hat{\alpha}_1$과 $\hat{\alpha}_2$을 이러한 추정량이라고 해보자. 그러면 (5.30)에서 b의 추정량은 다음과 같음을 볼 수 있을 것이다.

$$\begin{aligned} \hat{b}_0 &= \hat{\alpha}_0 \\ \hat{b}_1 &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 \\ \hat{b}_2 &= \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2 \\ &\vdots \\ \hat{b}_k &= \hat{\alpha}_0 + k\hat{\alpha}_1 + k^2\hat{\alpha}_2 \end{aligned} \quad (5.35)$$

유의해야할 것은 이 기법을 이용하여 단지 세 모수 α_0, α_1, α_2에 대한 추정량을 구함으로써 k개의 모수 $b_0, \cdots, b_k$에 대한 추정량을 구할 수 있다는 것이다. 이제 b가 α보다 많은 어떠한 상황에서라도 알몬기법으로부터 구한 b의 추정량은 다중회귀기법을 직접 (5.24)에 적용시켜 구한 b의 직접적인(direct) 추정량보다 양호하다는(보다 작은 분산을 가지고

있다는 의미에서)것을 알 수 있을 것이다.* 불행하게도 알몬기법에 의해 구한 추정량 $\hat{b}_0, \cdots, \hat{b}_k$ 의 분산에 대한 간단한 식을 표시할 수 없다. 하지만 실제로 각 회귀계수의 값과 관련한 통상의 통계적 검정을 실시할 수 있도록 컴퓨터는 이러한 분산의 추정치(estimates)를 제공해 줄 것이다.

알몬기법의 일반화와 변형은 간단하다. 가령 방정식(5.24)가 또다른 변수를 포함할 수 있도록 확장시켰다고 하자

$$Y_t = a + b_0X_t + b_1X_{t-1} + \cdots + b_kX_{t-k} + cW_t + u_t \qquad (5.36)$$

여기에서 W_t 는 또다른 하나의 독립변수이다. 만일 b가 (5.30)과 같은 관계를 따른다고 다시 가정한다면, 바로 동일한 단계를 따라 cW_t 항의 존재라는 한가지 예외를 제외하고는 (5.36)을 (5.34)와 동일한 방정식으로 축소시킬 수 있다. 곧,

$$Y_t = a + \alpha_0Z_{1t} + \alpha_1Z_{2t} + \alpha_2Z_{3t} + cW_t + u_t \qquad (5.37)$$

곧, 추가적인 변수를 포함시켜도 분석에는 아무런 영향이 없는 것이다. 사실상, 심지어 알몬의 방법을 동일한 방정식내에서 시차가 주어진 몇가지 독립변수 각각에 적용시킬 수도 있다.

알몬시차절차의 이용과 관련하여 좀더 상세하게 관찰해 보기로 한다. 첫째, 이용자가 b의 값에 대해 몇가지 "終點(endpoint)"제약을 부과하고자 할 수도 있다. 예를 들면, b_0 또는 b_k가(아니면 둘다) 0과 같다고 지정할 수도 있다. 가령 정보수집의 지연때문에 현재 시점의 독립변수

---

* 누구든 예상할 수 있듯이, 이러한 결과는 (5.30)과 같이 b와 α사이의 가정된(assumed) 관계가 사실이라는 가정에 기초하고 있는 것이다.

의 값이 현재의 행동(곧, 현재 시점의 종속변수의 값)에 영향을 미치지 않는다고 생각할 수 있기 때문이다. 곧, 방정식(5.24)에서 $b_0 = 0$인 것이다. 이러한 경우 종속변수Y는 단지 (5.24)에서의 X의 시차치에만 달려있다. 다른 한편, 의사결정방법 때문에 k 또는 그 이상의 시차가 주어진 X의 값이 Y에 영향을 주지 않는다고 생각할 수도 있다. 따라서 방정식(5.24)에서 $b_k = 0$인 것이다.

가령 $b_k = 0$ 와 같은 정보를 모형에 통합하는 한가지 방법은 단지 기본방정식(5.24)에서 X_{t-k}를 빼고 이전과 마찬가지로 진행하는 것이다. 하지만 이는 일반적으로 행해지는 방법인 것은 아니다. 실제로, $b_0 = 0$ 이거나 $b_k = 0$, 또는 둘다 성립한다 라는 정보는 (5.30)과 같은 기본가정을 이용하여 α에 대한 조건으로 바꾸며, 그에 따라 방정식을 추정하는 것이다. 그에 대한 증명은 비록 이 책의 범위를 벗어나는 것이지만 이러한 어느 정도 간접적인 접근법을 채택하게 된다.

왜냐하면 몇가지 가정하에서 결과적으로 생기는 불편추정량의 분산이 X_k와 X_{t-k}를 간단히 빠뜨리는 직접적인 접근법의 경우보다 작기 때문이다. 여기에 흥미가 있는 독자들을 위해 이 장의 부록A는 이 문제를 보다 자세하게 다루고 있다. 곧, 알몬의 기법이 어떻게 이러한 여러 조건을 통합할 수 있는지를 보여주고 있지만, 직접적인 접근법을 채택하는 실제의 이유도 제시하고 있다. 이는 부수적일지라도 어느 정도 중요성은 띠고 있다. 왜냐하면 알몬시차에 대한 대부분의 컴퓨터프로그램이 사용자들에게 종점제약을 지정하도록 요구하기 때문이다.

둘째, 분명히 유의해야할 것은 연구자들이 회귀모형의 시차의 길이 (곧, k의 값)와 b의 일반유형을 알고 있으며 따라서 다항식의 종류가 결정될 수 있는 것처럼 이 절의 내용이 서술되었다는 것이다. 실제로는 k도 b의 유형도 전혀 알지 못할 수도 있는 것이다. 그러한 상황에서는 다음

의 절차를 권하고자 한다. 첫번째, b의 어떠한 "합리적인" 유형도 포괄할 수 있을 정도로 높은 다항식의 차수, 가령 d를 선택하라. 대부분의 경우 3차 또는 4차다항식이면 충분하다. 그리고 나서 연구대상의 관계에 부합된다고 믿는 최대의 "합리적인" 시차가 k*라고 가정하라. 한가지 예로서 분기별 자료가 이용되고 있으면, k*는 3년과 4년의 시차에 해당되는 12 또는 16이면 된다. 만일 월별 자료가 이용되면, 아마도 k*=36을 선택하면 될 것이다. 여하튼 일단 d를 선택하였으면, k=d, d+1, …, k*에 대해 연구대상의 관계를 추정하라. 여기에서 d는 다항식의 차수이다. 그런데 k≥d인 시차만을 고려하기로 한다. 왜냐하면 시차의 길이가 적어도 d만큼 긴것으로 (α가 있는 만큼 b가 있는 것으로) 가정하고 있기 때문이다.* k의 여러 값에 대응하는 모든 회귀방정식은 동일한 자료로 추정해야 한다. 한가지 유의해야 할 것은 이것은 회귀방정식을 추정하기 위해 처음의 k*개의 관찰치를 버리고 나머지 (n−k*) 의 관찰치만을 이용하도록 요구한다는 것이다. 이에 따라 상이한 방정식들에 대한 R^2−통계량을 비교할 수 있게 된다. 왜냐하면 그것들은 동일한 자료에 모두 기초하고 있기 때문이다. 그러므로 R^2−통계량을 극대화하는 것을 k의 값으로 선택해야 하는 것이다. 몇가지 가정하에서 이러한 절차가 k와 회귀모수의 추정량을 일치추정량이 되도록 하는 것을 볼 수 있다.

다. 보기

간단하게 알몬시차구조를 추정하기 위한 절차를 설명하기 위해, 소비함수로 돌아가 보기로 하자. 〈표 5.2〉에는 제 2 장에서 실례의 소비방정식을

---

* 알몬시차절차의 목적이 추정할 모수의 수를 줄이기 위한 것임을 상기해야할 것이다. k>d이면, 이는 발생하지 않는다.

추정하기 위해 사용된 1960－69년의 소비지출과 가처분소득에 대한 열개의 관찰치를 재생하기로 한다.(표2.2를 보라) 추정을 위해 다시 한번 이 자료를 이용할 것이다. 하지만 이 경우에는 소비지출이 분포시차를 가진 채 가처분소득에 의존하는 것으로 가정한다. 보다 명확하게 표현하면, 소비가 금년과 이전 4년간의 가처분소득에 의존하는 것으로 가정한다. 또한 시차가 2차다항식의 형태를 가진 것으로 가정하기로 한다. 이에 따라 소비방정식은 다음과 같다.

$$C_t = a + b_0 Y_{dt} + b_1 Y_{d(t-1)} + b_2 Y_{d(t-2)} + b_3 Y_{d(t-3)} + b_4 Y_{d(t-4)} + u_t \quad (5.38)$$

여기에서 $b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$ 이다.

알몬형태의 소비방정식을 구하기 위해 다음을 계산하여야 한다.

$$Z_{1t} = \sum_{i=0}^{4} Y_{d(t-i)}, \quad Z_{2t} = \sum_{i=1}^{4} i Y_{d(t-i)} \quad \text{그리고} \quad Z_{3t} = \sum_{i=1}^{4} i^2 Y_{d(t-i)}$$

종속변수의 값에 따른 Z의 값은 〈표5.3〉에 나타나 있다. 유념해야할 것은 4시점의 시차를 도입한 결과로 4개의 관찰치를 "잃어버림" 으로써 〈표5.3〉이 단지 6년의 자료를 포함하고 있다는 것이다. 계산을 설명하기 위해 다음을 연산함으로써 $Z_{3(1969)}$의 값을 구해 보기로 한다.

$$\begin{aligned} Z_{3(1969)} &= Y_{d(1968)} + 4Y_{d(1967)} + 9Y_{d(1966)} + 16Y_{d(1965)} \\ &= 590 + 2{,}188 + 4{,}608 + 7{,}568 = 14{,}954 \end{aligned}$$

〈표5.3〉에서 새로 만든 자료를 이용함으로써 다음의 방정식을 추정하기 위한 표준절차를 사용할 수 있다.

$$C_t = a + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + u_t \quad (5.39)$$

추정방정식은 다음과 같다.

〈표 5.2〉 미국의 소비와 가처분소득

(경상 1억달러)

연 도	소비 (C)	가처분소득 (Yd)
1960	325	350
1961	335	364
1962	355	385
1963	375	405
1964	401	438
1965	433	473
1966	466	512
1967	492	547
1968	537	590
1969	576	630

출전 : Economic Report of the President (Washington, D.C. : U.S. Government Printing Office, Feb. 1970), pp.189, 195.

〈표 5.3〉

연 도	C_t	Z_{1t}	Z_{2t}	Z_{3t}
1964	401	1,942	3,667	10,821
1965	433	2,065	3,859	11,347
1966	466	2,213	4,104	12,030
1967	492	2,375	4,392	12,826
1968	537	2,560	4,742	13,860
1969	576	2,752	5,112	14,954

$$\hat{C}_t = -43.5 + 1.02Z_{1t} - 1.44Z_{2t} + 0.35Z_{3t} \qquad (5.40)$$

α_i에 대한 추정치를 이용하여, b_i에 대한 추정치를 계산할 수 있다.

$$\hat{b}_0 = \hat{\alpha}_0 = 1.02$$
$$\hat{b}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 = 1.02 - 1.44 + 0.35 = -0.07$$
$$\hat{b}_2 = \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2 = 1.02 + 2(-1.44) + 4(0.35) = -0.46$$
$$\hat{b}_3 = \hat{\alpha}_0 + 3\hat{\alpha}_1 + 9\hat{\alpha}_2 = 1.02 + 3(-1.44) + 9(0.35) = -0.15$$
$$\hat{b}_4 = \alpha_0 + 4\hat{\alpha}_1 + 16\hat{\alpha}_2 = 1.02 + 4(-1.44) + 16(0.35) = 0.86$$

이에 따라 4시점의 알몬시차를 가진 추정소비방정식은 다음과 같다.*

$$\underset{(3.3)}{\hat{C}_t = -43.5} + \underset{(5.3)}{1.02Y_{dt}} - \underset{(0.9)}{0.07Y_{d(t-1)}} - \underset{(2.8)}{0.46Y_{d(t-2)}} - \underset{(1.9)}{0.15Y_{d(t-3)}} + \underset{(4.3)}{0.86Y_{d(t-4)}} \qquad R^2 = 0.99 \qquad (5.41)$$

계수의 추정치와 더불어 t비율과 결정계수의 절대치가 포함되어 있다. 이 추가적인 정보는 알몬시차옵션을 가진 대부분의 컴퓨터프로그램에 의해 제공된다.

2. 가 (Dummy) 변수의 이용

지금까지는 가처분소득수준 또는 임금률의 변화율과 같이 양적인 면에서 측정할 수 있는 변수들만을 다루었다. 하지만 매우 중요한 어떤 변수들이 질적인 성격을 갖고 있다고 생각할 때가 있다. 예를 들면, 총소비지출수준이 가처분소득뿐만 아니라 그 나라가 전시인지 평화시인지의 여부에도 의

* 이 경우 추정방정식이 우리의 기대에 아주 잘 들어맞는 것은 아니다. 소득변수의 시차치 계수의 부호를 볼 때, 소비함수의 공식화에 대해 보다 많이 생각해 보아야 할 것이다.

존한다라고 생각해 볼 수도 있다. 전쟁기간동안 도덕적 설득과 사실상의 통제가 흔히 소비재화의 이용을 제한함으로써 어떤 가처분소득수준에 대해 평화시보다 낮은 소비수준을 예상할 수 있는 것이다. 그러나 평화시 또는 전쟁시에 대한 변수를 회귀방정식에 어떻게 도입할 수 있는가?*

이 문제에 대한 한가지 접근법은 두가지 별도의 소비함수를 추정하는 것이다. 곧 전시의 자료를 이용하여 "전시 소비함수"를 추정하고, 평화시의 자료를 이용하여 "평화시 소비함수"를 추정하는 것이다. 그에 따라 두가지 상이한 소비방정식을 구하게 될 것이다. 하지만 어떤 가정들을 취하고자 한다면, 단 하나의 방정식에 대한 추정을 포함하는 보다 효율적인 절차가 있다.

전시통제가 가처분소득에 대한 限界消費性向을 변경하지는 않지만, 그 대신 平均消費性向을 줄이는 것으로 가정하기로 하자. 곧, 〈그림 5.5〉에 의하면, 전시기간동안의 소비함수가 평화시와 동일한 기울기를 갖지만, 보다 낮은 절편(또는 보다 작은 상수항)을 갖는 것으로 가정하는 것이다. 이러한 가정에 의하면 하나의 회귀방정식에 의해 전시와 평화시의 소비함수를 둘다 표현할 수 있다.

그것은 다음과 같다.

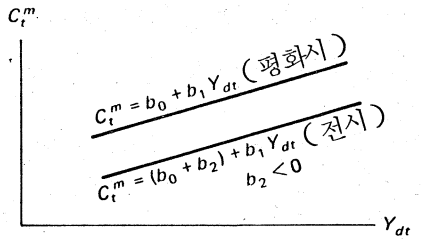
$$C_t = b_0 + b_1 Y_{dt} + b_2 D_t + u_t, \qquad t = 1, \ldots, n \tag{5.42}$$

여기에서 $D_t = 0$ (평화시의 경우)

$D_t = 1$ (전시의 경우)

---

* 그에 따른 논의에 대한 보다 진전된 취급은 다음의 책을 볼 것. Arthur S. Goldberger, Economic Theory(New York:Wiley, 1964), pp.218-227.



〈그림 5.5〉

평화시에는, 곧, $D_t = 0$일 때 방정식(5.42)는 다음과 같아짐을 의미한다.

$$C_t = b_0 + b_1 Y_{dt} + u_t \tag{5.43}$$

전시기간동안에는 곧 $D_t = 1$ 일 때는 다음과 같다.

$$C_t = (b_0 + b_2) + b_1 Y_{dt} + u_t \tag{5.44}$$

여기에서 $b_2 < 0$일 것이다.

시점이 $t = 5$에서 $t = 9$의 시점은 전시이며, 그 나머지는 평화시인 것으로 가정하자. 그러면 방정식(5.42)에 따라 〈표 5.4〉에서와 같은 일련의 자료를 갖게 될 것이다. 이러한 자료를 이용하여, 표준적인 다중회귀기법을 쓰면 방정식(5.42)의 계수의 값을 추정할 수 있다.

사실상 이 과정을 수행한 결과 다음과 같은 방정식을 얻게 되었다라고 가정해 보자

$$\hat{C}_t = 40 + 0.9 Y_{dt} - 30 D_t \tag{5.45}$$

여기에서 가령 D_t변수와 관련된 t 비율은 (5.42)의 모수 b_2가 0이 아니게끔 충분한 크기이다. 그러면 전쟁은 소비지출에 대해 상당히 부정적인 영향을 준다라고 결론을 내리게 될 것이다. 추정소비함수는 다음과 같을 것이다.

〈표 5.4 〉

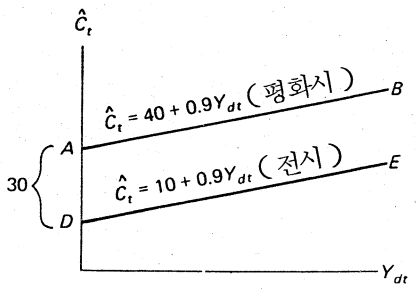
t	소 비 지 출	가 처 분 소 득	D
1	C_1	Y_{d1}	0
⋮	⋮	⋮	⋮
4	C_4	Y_{d4}	0
5	C_5	Y_{d5}	1
6	C_6	Y_{d6}	1
7	C_7	Y_{d7}	1
8	C_8	Y_{d8}	1
9	C_9	Y_{d9}	1
10	C_{10}	Y_{d10}	0
⋮	⋮	⋮	⋮
n	Cn	Y_{dn}	0

$$\hat{C}_t = 40 + 0.9Y_{dt} \qquad \text{(평화시의 경우)} \qquad (5.46)$$

$$\hat{C}_t = 10 + 0.9Y_{dt} \qquad \text{(전시의 경우)} \qquad (5.47)$$

만일 소비지출이 1억달러의 단위로 측정된다면, (5.46)과 (5.47)의 비교를 통해 해당 소득수준에 대해 소비지출이 전시동안 30억달러정도 적다는 것을 알게 된다. 〈그림 5.6 〉은 이 함수들을 나타내고 있다. 그림에서 전시소비함수 DE 는 평화시소비함수 AB 와 동일한 기울기를 가지지만, AB 보다 30정도 적은 수직절편을 가진 직선임을 볼 수 있는 것이다.

그 대신 전시의 조건이 소비방정식에서 한계소비성향을 감소시키지만, 상



〈그림 5.6〉

수항을 감소시키지 않는다고 가정할 수도 있는 것이다.* 이러한 두번째의 경우, 두 기간을 포함하는 회귀방정식은 다음과 같게 될 것이다.

$$C_t = b_0 + b_1 Y_{dt} + b_2(Y_{dt}D_t) + u_t \tag{5.48}$$

여기에서 다시

$D_t = 0$ (평화시의 경우)

$D_t = 1$ (전시의 경우)

이 방정식은 평화시에 다음과 같음을 의미한다.

$$C_t = b_0 + b_1 Y_{dt} + u_t \tag{5.49}$$

왜냐하면 $D_t = 0$이기 때문이다. 전시기간동안은 그 반면에 다음과 같다.

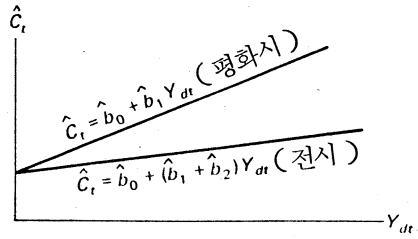
$$C_t = b_0 + (b_1 + b_2)Y_{dt} + u_t \tag{5.50}$$

여기에서 $b_2 < 0$ 임을 예상하게 될 것이다. 앞에서처럼, 방정식(5.48)을 추정하기 위해 〈표 5.4〉에서와 같은 자료를 이용할 수 있을 것이다. 그에

---

* 실제로 채택되었던 전시통제의 형태를 연구함으로써 절편이 이동하는지, 아니면 그 대신 MPC가 이동하는지 등을 결정하게 될 것이다.

로 말하면 가변수기법은 가장 효과적인 방법으로 모수이동(parameter shifts)에 관한 이전의 지식뿐만 아니라 이용가능한 모든 표본정보를 이용한다.



〈그림 5.7〉

방정식을 추정할 수 있을 만큼 관찰치의 수가 충분하다면, 부수적으로 회귀방정식에서 독립변수로 쓰고 싶은 만큼의 가변수를 이용할 수 있다. 예를 들면, 상이한 가구의 소비행위를 설명하고자 한다고 가정하자. 가구의 소비지출수준은 가처분소득뿐만 아니라 그 가구의 특성, 곧 어린이의 존재여부, 가구단위가 주택을 소유하였는지의 여부, 인종, 가장의 연령 등에 의존한다고 생각해 볼 수 있는 것이다. 만일 가구의 어떤 표본에 대한 모든 정보를 얻을 수 있다면 예를 들어 다음과 같은 방정식을 추정할 수 있을 것이다.

$$C_t = b_0 + b_1 Y_{dt} + b_2 F_t + b_3 H_t + b_4 R_t + b_5 A_t + u_t \qquad (5.51)$$

여기에서

C_t = t번째 가구의 소비지출

Y_{dt} = t번째 가구의 가처분소득

$$F_t = \begin{cases} 1 & \text{(어린이가 있는 경우)} \\ 0 & \text{(어린이가 없는 경우)} \end{cases}$$

$$H_t = \begin{cases} 1 & \text{(주택을 소유한 경우)} \\ 0 & \text{(주택을 소유 하지 않는 경우)} \end{cases}$$

따라 추정된 관계는 〈그림 5.7〉에서 나타낸 곡선과 유사하게 될 것이다. 곧 그림에서 전시소비함수가 추정된 평화시의 소비함수보다 작은 기울기를 갖지만 동일한 수직절편을 갖는다.

위의 방정식에서 D와 같은 변수는 "假變數(dummy variable)"라 한다. 그것은 어떤 조건이 유지되면 1의 값을 갖게 되며, 그렇지 않은 경우 0의 값을 갖게 된다. 가변수의 이용은 회귀분석을 극도로 강력하게 확장한 것임을 볼 수 있듯이 그것에 의해 양적인 단위로 측정할 수 없는 (때때로 근본적으로 중요한)변수를 포괄하기 위해 분석범위를 확장시킬 수 있게 된다. 가변수를 이용함으로써 종속변수의 값에 영향을 주는 중요한 질적 요인들의 효과를 계산할 수 있는 것이다.

다시 위의 첫번째 예를 생각해 보자. 그 예에서 전시기간동안 절편은 변하였지만, MPC는 변하지 않은 것으로 가정하였다. 가변수접근법 대신 두 개의 소비함수, 곧 전시에 대한 함수와 평화시에 대한 함수를 구하였다면, 셋이상인 네개의 모수를 추정하게 될 것이다. 전시와 평화시에 MPC가 동일하다는 가정하에서는 그 결과 하나의 모수에 대한 두개의 추정치가 있게 된다. 그에 따라 문제는 MPC의 유일한 "最良(best)"추정치를 구하기 위해 두 추정치를 이용하는 것이다. 그러한 경우 아마도 두 추정치의 "평균을 구하는" 어떤 기법을 개발해야 할 것이다.

만일 두 추정치를 "적절한(optimal)"방법으로 결합하게 된다면, 그 결과는 가변수절차에 의해 만들어진 것과 일치할(identical) 것임을 보일 수 있다. 보다 정식화하면, $\hat{b}_{1p}$을 평화시의 방정식과 자료에 기초한 MPC의 추정량이라 하고, $\hat{b}_{1w}$을 전시의 방정식과 자료로부터 도출한 추정량이라 하자, 표준적인 가정하에서 이러한 추정량들은 불편추정량일 것이다. 만일 이러한 추정량들을 그 결과가 가능한 한 최소 분산을 가진 MPC의 불편추정량이 되도록 하는 방법으로 결합한다면, 그에 따른 추정량은 가변수기법에 의해 만들어진 추정량과 일치할 것이다. 어느 정도 직관적으

$$R_t = \begin{cases} 1 & (\text{백인인 경우}) \\ 0 & (\text{백인이 아닌 경우}) \end{cases}$$

$$A_t = \begin{cases} 1 & (\text{가장이 50세 이상인 경우}) \\ 0 & (\text{50세 미만인 경우}) \end{cases}$$

u_t = 오차항

가. 實例

가변수의 적용범위는 실제로 제한이 없다. 아주 상이한 종류의 문제를 포함하고 있는 또다른 한가지 예가 필자중의 한사람*에 의한 최근의 연구이다. 그 연구주제는 지방의 행정제도의 형식적 특성이 국가재정의 분권화정도에 구조적으로 영향을 미치는가하는 것이다. 또는 간단히 말하면, 제도 그 자체가 전체로서의 공공부문에서 중앙정부의 재정행위가 상대적으로 차지하는 정도를 결정하는데 중요한가 이다.

이미 (인구규모와 1인당 소득수준과 같은) 여타 변수들의 중요성을 확인하였다면, 절차는 그 지방이 "연방(federal)" 헌법을 가지고 있다면(곧, 어느정도의 정치적 자율성을 지방정부에 보장하는 헌법을 가지고 있다면) 1의 값을, 연방헌법이 없는 경우(곧, 지방정부의 권한범위를 중앙정부가 결정하는 경우) 0의 값을 가지는 가변수를 도입하는 것이다. 53개 지역의 표본에 대한 횡단면자료를 이용한 결과, 추정방정식은 다음과 같았다.

$$\hat{G} = \underset{(12.1)}{96} - \underset{(1.3)}{1.21 \ln P} - \underset{(2.3)}{0.004Y} - \underset{(5.5)}{0.6Z} - \underset{(4.7)}{15.9F} \qquad N = 53 \quad R^2 = 0.65 \qquad (5.52)$$

(추정계수 아래의 괄호안 숫자는 t비율의 절대치)

여기에서

---

* W.E. Oates, Fiscal Federalism, (New York : Harcourt Brace Jovanovich, 1972), chap.5.

G=총공공수입에 대한 중앙정부의 지분(퍼센트)

lnP=인구규모의 자연로그(1천명단위)

Y=미국달러로 표시한 1인당소득(1965)

Z=총공공수입의 백분율로 표시한 사회보장기부금

$$F=\begin{cases}1 & \text{(연방헌법이 있는 정부)}\\ 0 & \text{(연방헌법이 없는 정부)}\end{cases}$$

방정식(5.52)의 결과는 연방헌법의 존재가 재정의 분권화정도를 증가시키는데 기여한다는 가설과 분명히 일치하는 것이다. 가변수F의 계수는 음의 부호를 가지며, 4를 넘는 t비율을 가진다. 따라서 보통의 5퍼센트 유의수준하에서 G와 F사이에 아무런 관계가 없다는 귀무가설을 쉽게 기각할 수 있다. 계수의 크기는 인구규모, 소득 등의 영향을 고려한 후 연방지역을 가진 중앙정부가 평균적으로 연방헌법이 없는 지방을 가진 중앙정부보다 총공공수입중 약 16퍼센트 적게 수취함을 의미하고 있다. 그러므로 행정제도에 의해 공식적으로 부과된 제한은 재정행위의 분권화정도를 결정하는데 상당히 중요한 것으로 나타난다.

나. 약간의 추가결과

가변수는 또한 행태에서 계절적인 차이와 지역적인 차이를 분리해 내는데 매우 유용하다는 것이 증명되었다. 자동차 판매수준은 가을에 새로운 모형이 소개된 결과로, 또는 기상조건에 달려있는 여러 농작물의 산출량에 따라 분명히 구조적으로 계절과 함께 변화한다. 만일 이러한 변수들에 대해 분기별 또는 월별자료를 이용하고 있다면 그러한 효과를 고려하기 위해 여러 계절과 관련된 가변수를 도입할 수 있다. 그와 마찬가지로 행태에서 지역적 차이를 예상하게 되는 경우, 여러 지역에 대한 가변수를 도입함으로써 이를 고려할 수 있다.

이를 행하는 방법을 보기 위해(그리고 피해야할 함정을 지적하기 위해) 다음의 예를 고려해 보자. 어떤 특정 재화의 생산량과 그 생산에 대한 노동의 투입량 사이의 관계를 이해하고자 한다고 가정하자. 여기에서 어떤 구조적인 계절의 힘이 작용하고 있다고 생각하는 것이다. 그에 따라 다음을 가정할 수 있을 것이다.

$$Q_t = b_0 + b_1 L_t + b_2 S_t + b_3 H_t + b_4 F_t + b_5 W_t + u_t \qquad (5.53)$$

여기에서

Q_t = t번째 분기의 산출량단위

L_t = 노동투입량단위

$$S_t = \begin{cases} 1 & (4월 \sim 6월간의 \ 분기일 \ 때) \\ 0 & (그 \ 이외의 \ 분기일 \ 때) \end{cases}$$

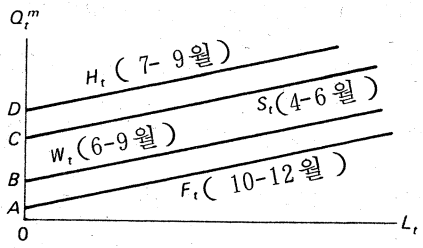
$$H_t = \begin{cases} 1 & (7 \sim 9월간의 \ 분기일 \ 때) \\ 0 & (그 \ 이외의 \ 분기일 \ 때) \end{cases}$$

$$F_t = \begin{cases} 1 & (10 \sim 12월간의 \ 분기일 \ 때) \\ 0 & (그 \ 이외의 \ 분기일 \ 때) \end{cases}$$

$$W_t = \begin{cases} 1 & (1 \sim 3월간의 \ 분기일 \ 때) \\ 0 & (그 \ 이외의 \ 분기일 \ 때) \end{cases}$$

유의해야 할 것은 각 분기에 대해 가변수를 집어 넣었다는 것이다.〈그림 5.8〉에 의하면 모형은 산출량의 평균수준 Q_t^m이 $(b_0 + b_1 L_t)$과 같다는 것을 의미한다. 여기에서 b_1은 계절에 의존하는 추가적인 양(아마 음수일 것이다.)을 더한 직선의 공통기울기이다.

방정식(5.53)을 추정하려 한다고 가정하자. 현재의 형태에서는 추정할 수 없음을 알게 된다. 왜냐하면 설명변수사이에 완전다중공선성이 존재하기 때문이다. 곧, 독립변수는 완전하게 그리고 선형적으로 관련되지 않는다는 가정이 위배된 것이다. 분명하게 하면 다음과 같다.



〈그림 5.8 〉

$$S_t + H_t + A_t + W_t \equiv 1 \tag{5.54}$$

어떤 분기에 하나의 변수는 1의 값을, 여타 변수는 0의 값을 갖는다는 사실을 알고 있다. 곧, 그 합은 항상 1이어야 하는 것이다. 앞장에서 상기해야 할 것은 완전다중공선성이 존재할 때 유일하게 계수를 추정할 수 없다는 것이다. 왜냐하면 정규방정식의 수가 충분하지 않기 때문이다.

하지만 간단하게 방정식으로부터 하나의 가변수를 빼고, 약간의 해석을 수정함으로써 이 문제를 쉽게 해결할 수 있다. 가령, 방정식(5.53) 으로부터 W_t를 제거하여 다음을 얻는다고 가정하자

$$Q_t = b_0 + b_1 L_t + b_2 S_t + b_3 H_t + b_4 F_t + u_t \tag{5.55}$$

이제 설명변수간의 선형종속(linear dependence)을 제거하였으면, 이에 따라 방정식(5.55)의 다섯개 모수를 추정할 수 있게 될 것이다. 겨울, 곧 1～3월동안 S_t, H_t와 F_t는 0일 것이고, 따라서 방정식의 "상수"항은 b_0일 것이다. 곧 〈그림 5.8〉에서 b_0는 수직절편 OB에 해당할 것이다. 마찬가지로 다시 방정식(5.55)와 관련하여 각 가변수의 계수는 해당계절의 영향이 겨울의 영향과는 어떻게 상이한지를 알게 된다. 가령 S_t가 1의 값을 갖는 봄의 분기(4～6월)동안 다음과 같게 된다.

$$Q_t = (b_0 + b_2) + b_1 L_t + u_t \tag{5.56}$$

〈그림 5.8〉에서 4～6월 분기와 관련한 곡선의 수직절편, 곧 OC 는 (b_0 $+b_2$)와 같다. 그러므로 b_2 는 춘기의 영향이 동기의 영향과 다른 방향과 크기를 가리킨다. 가령 〈그림 5.8〉에서 b_2 는 양수임을 예상하게 될 것이다. 마찬가지로 〈그림 5.8〉로부터 $b_3 > 0$ 이고 $b_4 < 0$ 임을 예상하게 될 것이다.

이제는 복습할 단계이다. 사계절이 있어서 방정식의 수준이 각 계절에 따라 변한다고 한다면, 하나의 계절을 표준으로 선택하여 그와 관련된 여타 계절의 영향을 고려할 것이다. 위의 예에서 겨울을 표준으로 선택하였던 것이다. 만일 방정식에 S_t 를 빼고 H_t, F_t 와 W_t 를 포함하였다면, 봄을 표준으로 선택한 것이다. 이러한 경우 b_0 는 봄 동안의 방정식의 높이를 나타낼 것이다. 어떤 의미에서 이러한 모든 것은 다음을 의미하는 것으로 요약해 볼 수 있다. 곧, 사계절이 있고, 방정식의 수직절편이 각 계절에 따라 변한다고 생각하면, 이러한 방정식절편을 나타내는 4개의 모수를 포함하는 하나의 방정식을 구해야할 것이다. 상수항이 그러한 하나의 변수이므로 단지 3개의 가변수가 필요하다. 유의할 것은 처음의 방정식(5.53) 의 모수를 추정할 수 없다는 것이다. 왜냐하면 방정식이 다섯개의 "절편"모수, 곧 b_0, b_2, b_3, b_4, b_5 를 포함하지만 4계절과 관련하여 단지 절편이 4개이기 때문이다. 하나의 가변수가 남아도는 것이다. 일반화하면 다음과 같다. 지역적인 변동 때문에 k개의 상이한 방정식을 생각한다면 단지 (k -1)개의 가변수만이 필요한 것이다. 이 논의를 독자 스스로 복습하기 위해서는 이 절 처음에 이용하였던 소비방정식으로 돌아가서 방정식(5.42) 에 다음과 같도록 놓으면 어떠한 일이 일어날 것인지를 생각해 보아야 할 것이다. 곧, 전시에 대한 가변수로 다음을 넣고,

$$W = \begin{cases} 1 & (\text{전시기간}) \\ 0 & (\text{평화기간}) \end{cases}$$

평화시에 대한 두번째의 가변수를 넣는 것이다.

$$P = \begin{cases} 1 & (\text{평화기간}) \\ 0 & (\text{전시기간}) \end{cases}$$

분명하게 말하면 W와 P 둘다 포함되면, 방정식을 추정할 수 없음을 보일 수 있다는 것이다. 두번째로 전시와 평화시 둘다를 포괄하는 하나의 방정식을 W나 P하나로 구할 수 있음을 확신해야 할 것이다.

3. 함수형태에 대한 보론

제 3 장에서 비선형관계를 선형형태로 바꿀 수 있게 함으로써 표준적인 선형회귀모형을 이용할 수 있도록 하여주는 몇가지 변환(transformation)을 실험하였다. 이러한 변환의 이용은 다중회귀의 경우로 쉽게 일반화하게 된다. 여하튼 이에 대해 자세히 검토하지는 않을 것이다. 오히려 한가지 특별한 예인 로그변환(logarithmic transformation)을 간단히 선택하여 이변수모형에 대한 분석을 다중회귀의 경우로 확장시킬 것이다. 그러므로 앞에서 이변수정식에 제한하지 않았던 몇가지 유용한 변환을 추가적으로 개발할 것이다.

가. 로그변환의 일반화

제 3 장에서 다음과 같은 형태의 단순 생산관계를 검토하였음을 상기해야 할 것이다.

$$Q_t = aL_t^b e^{u_t} \tag{5.57}$$

여기에서 L_t, 곧 t시점에 사용된 노동량은 Q_t의 생산에서 유일한 변수요인(variable factor)이었던 것이다. (5.57)에 로그를 취함으로써 이 생산함수를 다음과 같이 나타낼 수 있다.

$$\ln Q_t = \ln a + b \ln L_t + u_t \tag{5.58}$$

그러면 (5.58)은 하나의 선형형태로 변환된 것이다.

$$Q_t^* = a^* + bL_t^* + u_t \tag{5.59}$$

여기에서

$$Q_t^* = \ln Q_t$$
$$a^* = \ln a$$
$$L_t^* = \ln L_t$$

이다.

이러한 형태에서 (5.59)의 모수의 불편추정량 $\hat{a}^*$와 $\hat{b}$을 얻기 위해 정규적인 추정절차를 이용하였으며 그에 따라 $e^{\hat{a}^*}$을 a의 불편추정량으로, 그리고 적어도 일치추정량으로 간주하였다.

하지만 (5.57)의 명백한 한계란 단지 하나의 생산가변요소를 포함하고 있다는 것이다. 재화와 용역은 전형적으로 많은 종류의 투입물을 이용함으로써 생산된다. 이에 따라 보다 현실적인 생산함수라면 여러 요인들의 가변적인 양을 고려해야만 하는 것이다. 이러한 내용을 수행하는 (5.57)의 일반화된 형태를 생각해 보기로 하자.

$$Q_t = b_0 F_{1t}^{b_1} F_{2t}^{b_2} \cdots F_{kt}^{b_k} e^{u_t} \tag{5.60}$$

여기에서 각각의 F_t는 t시점동안 사용된 특정생산요소의 양을 나타낸다. 예를 들면 F_{1t}는 t시점에 사용된 노동량, F_{2t}는 자본량, F_{3t}는 토지면적 등을 나타낸다. 이러한 특정형태의 생산관계는 콥-더글라스(Cobb-Douglas) 생산함수로 알려져 있다.*

---

* 콥-더글라스생산함수는 흥미롭고 편리한 성질을 (다음 페이지 계속)

알고자 하는 것은 이 관계에서의 b의 값으로 이에 의해 여러 투입물의 양이 변함에 따라 산출량이 어떻게 변하는지를 알 수 있게 된다. 가령 어떤 특정상품의 생산이 규모에 대한 수확체증(increasing returns to Scale)현상을 보이는지의 여부에 관심이 있다고 하자. 이는 다음을 의미한다. 다른 조건이 동일할 때(other things equal)**, 모든 생산요소의 투입량을 두배로 한다면, 산출량은 두배 이상이 될 것인가? 만일 그렇다고 한다면, 이는 규모에 대한 수확체증현상인 것이다. 하지만 만일 산출량이 정확히 두배이면 규모에 대한 수확불변현상이 존재하는 것이며, 산출량이 두배보다 적다면, 규모에 대한 수확체감현상이 존재하는 것이다. 이것은 콥-다글라스 생산함수에서 단지 $\Sigma_{k=1}^{i} b_i$를 구함으로써 쉽게 결정된다. 이러한 것을 알기 위해 한가지 간단한 예를 들고, 일반화는 연습으로 남겨두기로 한다. 단지 노동과 자본을 사용하여 생산하는 하나의 상품이 있다고 가정하자. 그러면 다음과 같게 된다.

$$Q_t = b_0 L_t^{b_1} K_t^{b_2} e^{u_t} \tag{5.61}$$

여기에서 L_t와 K_t는 t시점동안 Q의 생산에 각각 사용된 노동량과 자본량을 나타낸다. 다음으로 노동과 자본의 투입량을 두배로 하였다고 가정해보자. 그리고 Q_t'를 그에 따른 산출량수준이라 하자. 그러면 다음과 같을 것이다.

(앞 페이지 계속) 많이 가지고 있다. 이 성질로 인해 경제분석에서 그 생산함수는 대단히 중요하게 된다. 이러한 성질에 관한 논의에 대해서는 다음을 볼것.
James M.Henderson and Richard E.Quandt, Microeconomic Theory, 2nd ed. (New York:McGraw-Hill, 1971), pp. 79-85.

** 이 "다른 조건이 동일할 때"라는 조건은 교란항과 관련이 있다. 곧 현재의 논의에서 투입량이 변할 때 교란항은 변하지 않는다고 가정하고 있는 것이다.

$$Q_t' = b_0(2L_t)^{b_1}(2K_t)^{b_2}e^{u_t} = b_0(2^{b_1})(L_t^{b_1})(2^{b_2})(K_t^{b_2})e^{u_t}$$
$$= 2^{(b_1+b_2)}b_0L_t^{b_1}K_t^{b_2}e^{u_t} = 2^{(b_1+b_2)}Q_t \qquad (5.62)$$

(5.62)의 마지막 방정식으로부터 알 수 있는 것은 $(b_1 + b_2) > 1$이면 산출량이 두배 이상이며 규모에 대한 수확체증인 것이다. 그리고 만일 $(b_1 + b_2) = 1$이면, 산출량이 정확히 두배이며, 규모에 대한 수확불변이 존재한다. 마지막으로 $(b_1 + b_2) < 1$이면, 산출량은 두배 미만이 되며, 이는 규모에 대한 수확체감을 의미한다. 보다 일반화하면, 규모에 대한 수확체증, 불변 및 감소의 경우는 각각 $\Sigma_{i=1}^{k} b_i$가 1을 초과하거나 1과 같거나 아니면 1보다 작은 경우에 해당된다는 것을 보일 수 있다.*

실증적으로 문제는 어떤 특정상품에 대한 생산관계의 특성을 결정하기 위해 b의 값을 추정하는 것이다. (5.60)을 추정가능한 형태로 만들기 위해 로그변환을 이용하게 된다. 곧 (5.60)에 로그를 취하면, 다음을 얻게된다.

$$\ln Q_t = \ln b_0 + b_1 \ln F_{1t} + b_2 \ln F_{2t} + \cdots + b_k \ln F_{kt} + u_t \qquad (5.63)$$

여기서 아래와 같이 정의하기로 한다.

$$Q_t^* = \ln Q_t$$
$$b_0^* = \ln b_0$$
$$F_{it}^* = \ln F_{it}$$

---

* 콥-더글라스생산함수(5.60)의 또다른 유용한 특성은 각각의 b_i가 요소 i에 대한 산출탄력성으로 해석될 수도 있다. 곧, F_i가 1퍼센트 증가하고 다른 모든 투입량이 불변인 채 그대로 있다면, 산출량Q는 b_i퍼센트 증가할 것이다. 하지만 유의해야 할 것은 어떠한 F_i도 $F_i = 0$이면 산출량Q가 또한 0이라는 의미에서 각 요소가 생산과정에서 필수불가결하다는 것이다.

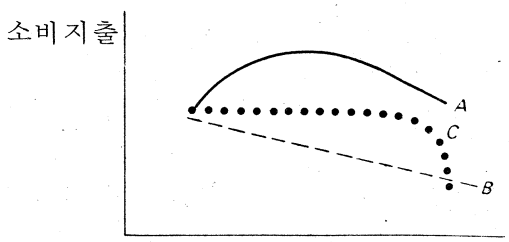
이를 (5.63)에 대입하면, 다음을 얻게 되는 것이다.

$$Q_t^* = b_0^* + b_1 F_{1t}^* + b_2 F_{2t}^* + \cdots + b_k F_{kt}^* + u_t \qquad (5.64)$$

(5.64)에서 정의한 변수들에 대해 표준기법을 이용하면, (5.64)의 모수의 불편추정량 $\hat{b_0}^*, \hat{b_1}, \cdots \hat{b_K}$을 구할 수 있게 된다. 이런 식으로 각각의 생산요소에 대한 산출탄력성의 불편추정량을 구할 수 있다. 이변수의 경우에서 처럼 b_0의 불편, 일치추정량은 $\hat{b_0} = e^{\hat{bo}*}$이 될 것이다. 마지막으로 이 장의 부록B에서 설명한 결과를 이용하면, 규모에 대한 수확불변이 존재하는지의 여부에 대한 가설을 추정할 수 있다.

나. 多項式형태의 독립변수

경제학자들은 실제로 가끔 경제변수들의 관계가 어떠한 형태를 띠게 될 것인지를 확신하지 못한 채 그것이 비선형일 가능성을 고려하고자 할 때가 있다. 가령 나이가 개인의 소비지출에 미치는 영향을 고려하기로 하자. 개인의 나이와 경험의 정도가 증가함에 따라 상이한 행동에 대한 지식이 소비재와 용역에 대한 지출수준의 확대를 초래할 가능성이 있는 것이다. 하지만 어떤 나이를 넘어서게 되면 개인은 "감퇴(slow-down)하기"시작하여 실제로 소비지출수준이 줄어들 수도 있다(소득과 같은 다른 관련 변수들이 불변인 채 그대로 있을 때). 소비지출과 연령사이의 그와 같은 관계는 〈그림 5.9〉에서 굵은선 A로 묘사되고 있다.

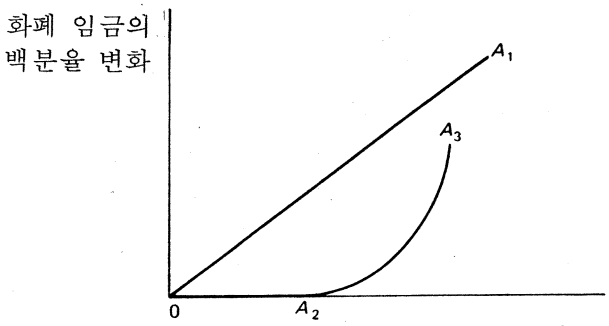


〈그림 5.9〉

다른 한편, 개인의 나이에 따라 안정(security)에 대한 수요, 그에 따라 저축이 증가하고, 그 결과 소비지출이 지속적으로 감소하는 경우도 있을 수 있다. 그러한 관계는 그림에서 점선B로 나타난다. 다시 이 예에서 소비자의 지출이 처음에는 단지 매우 느리게 감소하지만 나이가 많아짐에 따라 감소율이 증가하기 시작할 수도 있는 것이다. 이 마지막 관계는 〈그림 5.9〉에서 점선C로 표현되고 있다.

어떠한 비선형관계도 그럴듯 해보이는 변수들에 대한 또다른 예로서, 생계비용(cost of Living)의 변화가 화폐임금변화에 의해 측정되는 임금조정분에 미치는 영향을 고려하기로 하자. 물론 생계비용의 변화분은 임금의 변화분에 완전히 반영될 수도 있다. 곧, 다른 조건이 동일할 때, 만일 지난해 동안 생계비가 X%만큼 상승하였다면, 임금은 X%만큼 상향조정될 것이다.*

다른 한편, 생계비의 "작은(small)"변화는 인식되지 않을 것이고, 따라서 그와 관련된 임금증가를 나타내지 않을 것임을 알 수 있다. 이러한 경우 단지 생계비의 "커다란(large)"변화만이 임금조정에 반영된다고 가정할 수 있는 것이다. 이러한 여러 가능성들은 〈그림 5.10〉에 각각 곡선 OA_1과 OA_2A_3로 나타나고 있다.



〈그림 5.10〉

* 생산성의 상승을 반영하기 위해 이러한 어느정도의 임금증가를 추가할 수도 있을 것이다.

이러한 예와 관련하여, 변수들간의 관계의 형태가 명확하지 않을 때, 그 관계의 추정과 검정에 대한 문제에 주목해보기로 한다. 이장 뒷절에서 추가적인 예를 들어봄으로써 결론을 일반화할 것이다.

종속변수Y가 독립변수X와 불확실하게 관련되어 있다고 가정함으로써 (논의의 단순화를 위해 이변수의 경우를 이용한다) 시작하기로 한다. 이 가정은 다음과 같이 표현될 것이다.

$$Y_t = f(X_t) + u_t \tag{5.65}$$

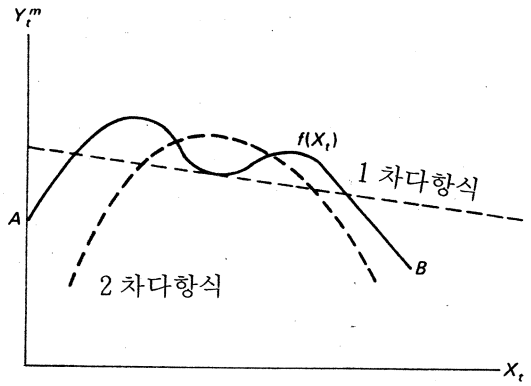
여기에서 u_t는 교란항이다. 방정식 (5.65)가 단지 나타내는 것은 Y의 t번째 값, 곧 Y_t가 X의 t번째 값, 곧 X_t와 교란항 u_t에 의존한다는 것이다. 방정식 (5.65)에서 $f(X_t)$의 형태를 알지 못하기 때문에 Y_t와 X_t간의 관계를 추정하기 위해서는 분명히 $f(X_t)$가 무엇인지를 알아내거나 아니면 그 대신 그에 근사한 어떤 것을 차용해야만 할 것이다. 알몬시차기법과 관련하여 설명, 사용했던 한가지 정리를 다시 생각하여 봄으로써 후자의 접근법을 취하기로 한다. 명확하게 말하면 그 정리란 일반적인 조건하에서 어떤 함수(또는 곡선)는 다항식에 의해 어떤 정확한 차수(degree)에 근사화할 수 있다는 것이다. 만일 그렇다고 한다면, (5.65)의 미지의 함수에 이 정리를 적용할 수 있게 되고 따라서 다음을 얻게 된다.

$$f(X_t) \doteq a_0 + a_1 X_t + a_2 X_t^2 + \cdots + a_k X_t^k \tag{5.66}$$

일반적으로 정확도를 높이기 위해서는 다항식의 차수(k)를 높여야만 한다. 어떤 의미에서 이는 알몬시차기법에서 다항식과 관련한 논의를 따

르고 있다. 그 절에서 다항식의 그래프위의 굴곡점의 수는 기껏해야 다항식의 차수보다 하나 적다는 것을 지적하였다. 이로부터 내릴 수 있는 결론은 고차다항식은 저차다항식보다 형태에서 보다 굴곡이 많다는 것이다. 이는 다음을 의미한다. 곧, 보다 근사한 형태를 얻으려면, 미지의 함수의 형태와 밀접해질 수 있을 만큼 충분하게 굴곡성이 높도록 고차다항식이 필요한 것이다. 전형적으로 함수형태를 보다 복잡하게 근사화하려면, 다항식의 차수는 보다 높아야만 하는 것이다.

이는〈그림 5.11 〉에서 미지의 함수 $f(X_t)$에 의해 표현되고 있다. 여기에서 미지의 함수 $f(X_t)$는 굵은 선 AB가 나타내는 형태를 띤다고 가정하기로 한다. 이제 1차다항식(k=1)을 이용하면 그 결과로 나타나는 직선에 의해 비록 빈약할지라도 이 곡선은 근사화 되는 것이다. 만일 그 대신 2차다항식(k = 2)을 이용하면 좀더 근사화될 것이다. k = 3 등을 취함으로써 훨씬 더 개선될 수 있는 것이다.



〈 그림 5.11 〉

Y_t 와 X_t 사이의 관계의 일반적인 형태에 대한 몇가지 가설이 있다고 가정하자. 게다가 이러한 가정중에서 가장 복잡한 것이 의미하는 것은 k = k_m 의 값이 (5.66)에서의 근사형에 "적당한"것이라는 사실임을 가정하자.* "적당하다(adequate)"라는 의미는 (5.66)에서 "근사하게 동

* 대부분의 경제에의 적용에는 $k_m = 3$ 의 값이 합리적이다.

일한 (approximately-equal -to) " 이라는 기호, 곧 ≒이 거의 정확성을 잃지 않고서도 등호로 대체할 수 있다는 것이다. 아울러 유의해야 할 것은 k_m차 다항식이 가장 복잡한 가설에 의한 형태에 적당한 근사형이라고 한다면 그것은 또한 앞으로 고려할 보다 간단한 모든 형태에도 적당한 근사라는 것이다.

k_m에 관한 이러한 가설하에 (5.66)에서 $k = k_m$을 대입하고 이것을 다시 (5.65)에 대입하면 다음을 얻게 된다.

$$Y_t = a_0 + a_1 X_t + a_2 X_t^2 + \cdots + a_{k_m} X^{k_m} + u_t \tag{5.67}$$

방정식 (5.67)을 다음의 형태와 같은 치환에 의해 표준모형으로 바꿀 수 있게 된다.

$$Z_{it} = X_t^i, \quad i = 1, \ldots, k_m \tag{5.68}$$

곧, (5.68)을 (5.67)에 대입하면, 다음을 얻게 된다.

$$Y_t = a_0 + a_1 Z_{1t} + a_2 Z_{2t} + \cdots + a_{k_m} Z_{k_m t} + u_t \tag{5.69}$$

여기에서 이는 표준형태인 것이다.

$\hat{a}_0, \hat{a}_1, \cdots, \hat{a}_{Km}$을 (5.69)의 모수의 추정량이라 하자. 그러면 Y_t와 X_t 사이의 추정된 관계 (estimated relationship)는 다음과 같다.

$$\hat{Y}_t = \hat{a}_0 + \hat{a}_1 X_t + \cdots + \hat{a}_{k_m} X_t^{k_m} \tag{5.70}$$

왜냐하면 $f(X_t)$의 추정량이 다음과 같이 주어졌기 때문이다.

$$\widehat{f(X_t)} = \hat{a}_0 + \hat{a}_1 X_1 + \cdots + \hat{a}_{k_m} X_t^{k_m} \tag{5.71}$$

이제 변수Y_t가 변수X_t에 의존하는지의 여부를 검정하는 문제에 대해

생각해 보기로 하자. 얼핏 보기에 이 가설은 (5.69)에서 단지 $a_1 = 0$, $a_2 = 0, \cdots, a_{k_m} = 0$이라는 가설을 차례로 검정함으로써 검정할 수 있는 듯해 보인다. 아마 이러한 귀무가설중 어느 것도 기각되어버리면 Y_t와 X_t는 상당히 관련되어 있다라는 결론을 내리게 될 것이다. 거꾸로 이러한 귀무가설 전부를 채택한다면 결론은 Y_t와 X_t가 그다지 관련되어 있지 않다는 것이다.

불운하게도 이러한 방식으로 Y_t와 X_t 사이의 관계에 대한 이러한 가설들을 검정할 수는 없다. 왜냐하면 이른바 "구성의 오류(fallacy of composition)"때문이다. 곧, 이 경우 가설은 하나 이상의 변수와 관련되어 있기 때문이다. 특별히 검정하고자 하는 가설은 다음과 같다.

$$H_0: a_1 = a_2 = \cdots = a_{k_m} = 0 \tag{5.72}$$

이 장의 부록B에서 (5.72)와 같은 가설을 검정하는 절차를 개발할 것이다. 하지만 여기에서 지적할 것은 $a_1 = 0$, $a_2 = 0$, …, $a_k = 0$이라는 가설이 어떤 유의수준, 가령 5%하에서 차례로 검정되어 채택된다면, (5.72)의 가설은 여전히 5%의 유의수준하에서 기각될 수도 있는 것이다. 바꾸어 말하면, 가령 (5.72)에서의 k_m과 같이 하나 이상의 모수와 관련된 가설은 일반적으로 그것을 각각 하나의 모수와 관련된 k_m의 가설로 나누고 어떤 유의수준하에서 차례로 가설을 검정함으로써 적절하게 추정될 수 있는 것은 아닌 것이다.

위의 추정기법을 회귀모형이 추가적인 설명변수를 포함하는 경우로 확장시키는 것은 간단하다. 가령 모형을 다음과 같다고 하여 보자.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + f(X_{3t}) + u_t \tag{5.73}$$

여기에서 u_t는 교란항이며, $f(X_{3t})$의 특정형태는 불확실하다고 가정한

다. 그리하여 위의 절차를 따를 때, $f(X_{3t})$에 관한 가설이 의미하는것이 $k = k_m$은 적절한 다항식근사화(adequate polynomial approximation)를 나타낸다는 것이라면, 다음을 가정하게 될 것이다.

$$f(X_{3t}) = a_0 + a_1 X_{3t} + \cdots + a_{k_m} X_{3t}^{k_m} \tag{5.74}$$

(5.74)를 (5.73)에 대입하면 다음을 얻게 된다.

$$Y_t = A + b_1 X_{1t} + b_2 X_{2t} + a_1 X_{3t} + \cdots + a_{k_m} X_{3t}^{k_m} + u_t \tag{5.75}$$

여기에서 $A = a_0 + b_0$이다. 다시 다음과 같이 표기하자.

$$Z_{it} = X_{3t}^{i} \qquad i = 1, 2, \ldots, k_m \tag{5.76}$$

그러면 방정식(5.75)를 다음과 같이 쓸 수 있다.

$$Y_t = A + b_1 X_{1t} + b_2 X_{2t} + a_1 Z_{1t} + \cdots + a_{k_m} Z_{k_m t} + u_t \tag{5.77}$$

이는 표준형인 것이다. 이에 따라 다항식근사화에 의거하면 (5.77)의 모수의 불편추정량 $\hat{A}$, $\hat{b}_1$, $\hat{b}_2$, $\hat{a}_1$, …, $\hat{a}_{km}$을 얻을 수 있다.

유의해야 할 것은 이 경우에 a_0와 b_0 각각의 추정량을 구할 수 없다는 것이다. 왜냐하면 (5.77)에서는 단지 그것들의 합 A만이 나타나기 때문이다. 위의 간단한 형태와는 달리 $f(X_{3t})$을 단지 부가상수(additive constant)까지만 추정할 수 있을 뿐이다. 바꾸어 말하면 단지 $f(X_{3t})$의 변수부분 가령 $f_v(X_{3t})$만을 추정할 수 있는 것이다.

$$\widehat{f_v(X_{3t})} = \hat{a}_1 X_{3t} + \cdots + \hat{a}_{k_m} X_{3t}^{k_m} \tag{5.78}$$

하지만 정상적으로 $f(X_{3t})$에서 중요한 부분은 변수요소이다. 왜냐하면 이 요소가 Y_t가 X_{3t}와 함께 변화하는 형태를 나타내기 때문이다.

이 단계에서 위의 기법을 일반적으로 설명변수의 수가 어떠한 경우에도 확장시킬 수 있음은 분명하다. 또한 이러한 하나 이상의 설명변수가 불분명한 형태로 모형에 들어 있는 경우를 포괄하기 위해 이러한 절차를 확장시킬 수 있다는 것도 명백하다.*

다. 함수형태의 결합

함수형태에 대한 논의를 한꺼번에 하기 위해 동일한 회귀방정식에 여러가지 상이한 변환을 이용하는 것은 완전히 사리에 맞는 것임을 강조해 두고자 한다. 사실상 앞의 몇가지 예에서 한가지 또는 그 이상의 변수가 로그형태나 아니면 역수형태로 나타나는 반면, 다른 변수들은 어떤 유형의 변환에도 따르지 않는다는 사실을 볼 수 있었다. 한가지 예로 다음과 같은 필립스곡선관계의 보다 복잡한 변형을 생각해보자.

$$\dot{W}_t = b_0 + b_1\left(\frac{1}{R_t}\right) + b_2\pi_{(t-1)} + b_3\dot{P}_t + b_4\dot{P}_t^2 + b_5 \ln G_t + u_t \qquad (5.79)$$

여기에서

$\dot{W}_t$ = t 시점동안의 임금의 변화분(%)

R_t = t 시점의 실업률

$\pi_{(t-1)}$ = ($t-1$)시점의 기업이윤율

$\dot{P}_t$ = t 시점의 가격변화분(%)

lnGt = t 시점의 노동력의 증가율의 자연로그

u_t = 교란항

유의해야 할 것은 동일한 방정식에 역수변환, 로그변환, 시차관계, 그리고

---

* 가령 다음과 같은 형태의 모형을 생각해 보기로 한다.

$$Y_t = a_0 + f_1(X_{1t}) + f_2(X_{2t}) + a_1 X_{3t} + u_t$$

하나의 독립변수에 대한 다항식형태가 이용된다는 것이다. 여기에서는 변환기법을 논증하기 위해 (5.79)를 고려하고 있는 것이다. 하지만 실제로 각각의 변환 이면에는 "이유(reasons)"(경제적 가설)가 존재하는 것이다. 가령 $\dot{W}$와 R사이의 선형관계는 (R이 음수를 가정하는 것으로 제한되기 때문에) 아주 진실되어 보이는 것은 아니며, 역수변환이 의미있는 것임을 제3장에서 보았던 것이다. 이윤변수(π_{t-1})에 대해서는 임금협상에 참가한 노동조합이 체결과정에서의 한 요소로서(정상적으로 이전(preceding) 회계시점에 대해 입수할 수 있는) 이윤율을 이용하는 것으로 예상할 수 있다. 만일 전기의 이윤율이 비정상적으로 높다면 노조는 그에 따라 보다 높은 임금증가를 요구하는 것이 정당하다고 생각할 것이다. 마찬가지로 가격변수의 경우 다항식형태를 생각해볼 수 있다. 왜냐하면 "작은(small)"가격변화가 대체로 주목받지 못하는 반면 커다란 가격변화는 임금증가요구를 촉발한다면(그림 5.10을 볼 것), 임금조정과 가격조정사이의 관계는 선형일 수 없기 때문이다. 마지막으로 노동력증가는 노동공급의 증가와 관련되어 있기 때문에 그것은 노동자의 협상지위에 영향을 미칠 수도 있는 것이다. 이러한 점에서 노동력증가에 관한 변수는 그 변화의 크기가 현재의 노동공급수준과 관련하여 측정될 수 있도록 전형적으로 (절대수보다는) 백분율증가임을 제3장에서 보았던 것이다. 설명을 위해 (5.79)에서 변수 G_t의 경우 로그변환을 채택하기로 한다. 이러한 모든 것이 의미하는 바는 함수형태의 선정이 기본적으로 시행착오의 과정은 아니라는 것이다. 곧 함수관계에 관해 가장 적절한 형태를 선정하는데 보유하는 이론적, 제도적 정보가 어떠한 것이든지 그것을 이용해야 한다는 것이다.

임금방정식으로 돌아가서 단지 다음과 같은 일련의 변환을 이용함으로써

(5.79)를 선형형태로 만들 수 있다.*

$$Z_{1t} = \left(\frac{1}{R_t}\right)$$
$$Z_{2t} = \pi_{(t-1)}$$
$$Z_{3t} = \dot{P}_t$$
$$Z_{4t} = \dot{P}_t^2$$
$$Z_{5t} = \ln G_t$$

이러한 변환들을 (5.79)에 대입하면, 방정식의 선형형태를 얻게 된다.

$$\dot{W}_t = b_0 + b_1 Z_{1t} + b_2 Z_{2t} + b_3 Z_{3t} + b_4 Z_{4t} + b_5 Z_{5t} + u_t \qquad (5.80)$$

R, π, $\dot{P}$과 G의 관찰치로부터 Z의 값을 쉽게 계산할 수 있으며, 따라서 Z의 값과 표준추정절차를 이용하여 $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$, $\hat{b}_3$, $\hat{b}_4$과 $\hat{b}_5$를 계산할 수 있는 것이다.

이 모든 것이 의미하는 바는 다중회귀추정에 내재하는 유연성의 범위가 넓다는 것이다. 계량경제학자들은, 특히 회귀분석 초창기에 관계의 선형형태에 대한 그들의 기본적인 신뢰를 비판할 때가 가끔 있었다. 하지만 상상력을 구사하여 변환을 현명하게 이용함으로써 회귀분석모형은 매우 복잡한 함수형태를 보다 폭넓게 처리할 수 있다는 사실을 이제 알 수 있는 것이다.

4. 實例: 화폐에 대한 수요

화폐금융론의 주요논제는 화폐수요에 대한 이론과 측정이다.** 사실상 이

---

* 실제로 변환 $Z_{3t} = \dot{P}_t$는 불필요하며, 단지 표기의 일관성을 기하기 위한 것이다.

** 상세한 논의에 대해서는 다음을 볼 것.
David E.Laidler, The Demand for Money:Theories and Evidence, 2nd ed.(New York:Dun-Donnelley,1977).

논제에 많은 것이 달려있다. 왜냐하면 전체 경제행위에 영향을 미치는데 재정, 금융정책이 지니는 잠재적 유효성은 화폐수요함수의 형태와 그것의 모수의 값에 좌우되기 때문이다. 경제이론에 의하면 실질화폐잔고(Real Money Balances, 구매력이 불변이게끔 전체가격수준에 대해 조정된 명목화폐잔고)에 대한 수요는 적어도 세가지 종류의 변수, 곧 소득, 채권에 대한 이자율(또는 여타 금융자산에 대한 수익률), 그리고 아마도 순자산에 의존할 것이다. 간략하게 말하면, 소득이 증가함에 따라 사람들은 거래를 목적으로 보다 많은 화폐잔고를 수요하게 될 것이다. 그리고 이자율이 상승함에 따라 그들은 화폐보유고를 감소시키고자 할 것이다. 왜냐하면 (적어도 최근까지 이자도 벌지 않는) 현금잔고(cash balances) 보유의 기회비용이 증가하기 때문이다. 그리고 사람들의 순자산이 증가함에 따라 자신의 증가된 자산을 유지시킬 수 있는 형태로서 현금잔고를 증가시키려 할 것이다. 다음과 같이 설정함으로써 이 모든 것을 요약해 볼 수 있다.

$$M_d = f(Y, r, W) \tag{5.81}$$

여기에서

M_d = 화폐잔고의 수요

Y = 실질소득

r = 이자율

W = 자산(또는 순자산)

여기에서 이자율의 부분효과는 負(negative)이며, 소득과 자산변수의 부분효과는 正(positive)임을 예상하게 된다.

많은 경제학자들은 (5.81)에 관한 방정식을 추정하기 위해 많은 계량경제학적인 작업을 행하였다. 이러한 효과들은 이 장에서 검토한 여러가지 변환을 포함하는 몇가지 함수형태로 표현되었던 것이다. 또한 그것들

은 흔히 어떤 변수들의 시차치(lagged values)도 포함하고 있다. 한가지 예로 마틴 브론펜브레너(Martin Bronfenbrenner)와 토마스 메이어(Thomas Mayer)의 연구결과를 서술해보기로 한다.* 고들의 출발점은 다음과 같은 곱셈형태의 화폐수요함수이다.

$$M_{dt} = b_0 Y_t^{b_1} r_t^{b_2} W_t^{b_3} M_{d(t-1)}^{b_4} e^{u_t} \tag{5.82}$$

여기에서 u_t는 교란항이며 다른 모든 변수들은 (5.81)과 관련하여 정의된 것이다. 브론펜브레너와 메이어는 (5.82)의 양변에 로그를 취해 다음과 같은 표준적인 선형형태를 구하였다.

$$\ln M_{dt} = \ln b_0 + b_1 \ln Y_t + b_2 \ln r_t + b_3 \ln W_t + b_4 \ln M_{d(t-1)} + u_t \tag{5.83}$$

설명을 위해 u_t는 표준적인 모든 가정을 만족한다고 가정하자. 그러면 (5.83)은 다음과 같이 쓰게 된다.

$$(\ln M_{dt} - b_4 \ln M_{d(t-1)}) = B + b_1 \ln Y_t + b_2 \ln r_t + b_3 \ln W_t + u_t \tag{5.84}$$

여기에서 $B = \ln b_0$이다. 이러한 형태에서 (5.83)과 같은 모형은 종속변수 $\ln M_{dt}$의 현재치(present value)와 요인 b_4를 곱한 그것의 시차치(lagged value)의 차이(difference)를 설명하는 것으로 이해할 수 있음을 알게 된다. 가령 화폐수요의 로그가 예를들어 3% 증가한 추세선(trend line) 근처에서 임의적으로 단순히 변동하고 있음을 인식하였다고 하자. 이러한 경우 (5.83)이나 (5.84)에서의 계수의 값에 관한 기대는 $b_4 = 1.03$이며, $B = b_1 = b_2 = b_3 = 0$일 것이다. 곧 화폐수요가 소득, 이자율과 자산변수에 반응하지 않는다고 생각할 것이다.

브론펜브레너와 메이어는 미국의 1919~1956년동안의 연도별 자료를 이

---

* Martin Bronfenbrenner and Thomas Mayer, "Liquidity Functions in the American Economy", Econometrica, 28(1960), pp. 810-834.

용하여 최소자승법(least-squares technique)에 의해 방정식(5.83)을 추정한 결과 다음을 얻게 되었다.

$$\widehat{\ln M} = \underset{(.003)}{0.11} + \underset{(.09)}{0.34}\ln Y - \underset{(.01)}{0.09}\ln r - \underset{(.08)}{0.12}\ln W + \underset{(.06)}{0.72}\ln M_{t-1} \qquad (5.85)$$

$$R^2 = 0.99$$

여기에서 모수추정치 하단의 괄호안 숫자는 추정표준편차이다. 결과를 설명하기 위해 보면, 소득, 이자율, 자산과 시차가 주어진(lagged) 화폐변수 각각에 대해 계산한 t비율은 3.8, 9, 1.5, 12에 가깝다. 약식검정(rule of thumb)을 이용하면 이러한 결과는 다음을 의미한다. 곧 5%의 유의수준하에서 귀무가설을 $H_0 : b_3 = 0$으로, 대립가설을 $H_1 : b_3 \neq 0$로 하면, H_0가 채택될 것이다. 그와는 대조적으로 5%의 유의수준하에서 $i = 1, 2, 4$일 때, 귀무가설 $H_0 : b_i = 0$, 대립가설 $H_1 : b_i \neq 0$의 경우는 어떤 것을 고려하더라도 H_0는 기각된다.

귀무가설 $H_0 : b_2 = 0$은 특히 화폐경제학자들의 관심사이다. 이 가설의 기각은 보다 높은 이자율에서 사람들은 채권과 이자를 낳는 여타 금융자산들로부터 얻을 수 있는 보다 높은 수익을 도모하기 위해 화폐형태의 자산은 보다 작은 부분만을 보유할 것이라는 그들의 신념을 합리화하여 줄 것이다. 이 결과의 한가지 중요한 의미는 재정정책이 총수요에 어느 정도 영향을 미친다는 것이다. 만일 화폐수요가 이자율에 반응을 보이지 않는다면, 재정정책은 경제의 전체지출에 아무런 순효과(net effect)를 주지 않고서도 개인지출의 변화분을 단지 상쇄시킬 수 있을 것이다.*

---

* 이점에 대해서는 다음을 볼 것.
Laidler, op. cit., chap.2. 간단하게 말하면, 화폐수요가 전체적으로 이자율에 영향을 받지 않으면, LM곡선은 수직일 것이다. 따라서 조세경감이나 정부지출의 증대는 단지 이자율을 끌어올릴 것이며, 동일한 양의 민간지출을 "구축할(crowd out)"것이다. 이러한 경우에 명목 GNP수준은 단지 화폐공급의 크기에만 의존하게 된다.

부록A. 알몬시차 기법에서의 종점제약

이 부록은 時差構造(lag structure)를 추정하기 위한 알몬기법에서의 終點制約(endpoint restrictions)의 이용을 검토하고 있다. 본문에서 언급한 바와 같이 경제학자는 가끔 b의 유형뿐만 아니라 b_0 또는 b_k 아니면 그 둘 모두의 정확한 값을 알고 있다고 느낄때가 있다. 덧붙여 말하면 이 값은 전형적으로 0이다. 만일 이러한 모수들의 값을 알고 있다면 이 정보를 추정절차속에 포함시켜야 할 것이다.

가령 b의 유형이 〈그림5A.1〉의 곡선 A로 표시된 것과 유사하다고 생각했다 하자. 이런 경우 b는 마지막의 b가 0이 될 때까지 처음에는 감소하였다가 다음에는 증가하다가 결국 감소하게 된다. 이러한 유형에서 종점제약, 곧 $b_k = 0$을 부과할 수 있는 것이다. 다른 한편, b의 유형이 〈그림5A.1〉의 곡선 B로 표시된 것과 같다고 생각하였다 하자. 이런 경우에는 두개의 종점제약, 곧 $b_0 = b_k = 0$가 있을 수 있는 것이다.

일반적으로 실험자는 이러한 종점제약중에서 어느 하나를 부과하거나 아니면 둘다 부과할 수 있는 것이다. 그 절차란 현재의 기법을 간단하게 일반화시킨 것이다. 다시 방정식 (5.24)를 고려하면서 $k = 10$이라고 (방정식이 10개의 시차를 가지고 있다고) 가정하자. 곧 다음과 같을 것이다.

$$Y_t = a + b_0X_t + b_1X_{t-1} + \cdots + b_{10}X_{t-10} + u_t \qquad \text{(5A.1)}$$

먼저 종점제약을 무시한 채, 가정된 b의 유형이 3차다항식이라고 가정하자.

$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \alpha_3 i^3, \quad i = 0, \ldots, 10 \qquad \text{(5A.2)}$$

그러면 (5A.2)에서 주어진 것과 같은 각각의 b에 대한 식을 (5A.1)에 대입하면 다음과 같을 것이다.

$$Y_t = a + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + \alpha_3 Z_{4t} + u_t \qquad (5A.3)$$

여기에서

$$Z_{1t} = \sum_{i=0}^{10} X_{t-i} \quad , \quad Z_{2t} = \sum_{i=1}^{10} i X_{t-i}$$
$$Z_{3t} = \sum_{i=1}^{10} i^2 X_{t-i} \quad , \quad Z_{4t} = \sum_{i=1}^{10} i^3 X_{t-i}$$

이제 종점조건을 부과해 보기로 하자. 특히 $b_{10} = 0$이라고 생각한다고 하자. 그러면 (5A.2)로부터 다음을 얻게 될 것이다.

$$b_{10} = \alpha_0 + 10\alpha_1 + 100\alpha_2 + 1000\alpha_3 = 0 \qquad (5A.4)$$

(5A.4)로부터 다음의 사실이 분명하다. 곧, 조건 $b_{10} = 0$을 부과하면 다음을 얻게 되는 것이다.

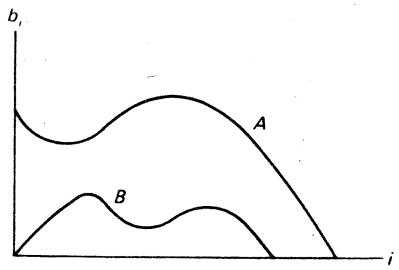
$$\alpha_0 = -10\alpha_1 - 100\alpha_2 - 1000\alpha_3 \qquad (5A.5)$$

곧, 조건 $b_{10} = 0$은 α들사이의 관계에 대한 제약을 의미하는 것이다. 본질적으로 이것이 그 해이다. 간단하게 (5A.3)으로 돌아가서 α_0 대신 (5A.3)으로 치환하여 보자. 그러면 다음을 얻게 된다.

$$Y_t = a + \alpha_1 Q_{1t} + \alpha_2 Q_{2t} + \alpha_3 Q_{3t} + u_t \qquad (5A.6)$$

여기에서

$$Q_{1t} = Z_{2t} - 10Z_{1t}$$
$$Q_{2t} = Z_{3t} - 100Z_{1t}$$
$$Q_{3t} = Z_{4t} - 1000Z_{1t}$$



〈그림 5A.1〉

방정식(5A.6)은 표준형에 속하며, 따라서 α_1, α_2, α_3는 표준적인 다중회귀기법을 써서 추정할 수 있는 것이다. $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$이 추정량이라고 가정하자. 그러면 (5A.5)로부터 α_0에 대한 추정량은 다음과 같을 것이다.

$$\hat{\alpha}_0 = -10\hat{\alpha}_1 - 100\hat{\alpha}_2 - 1000\hat{\alpha}_3 \qquad \textbf{(5A.7)}$$

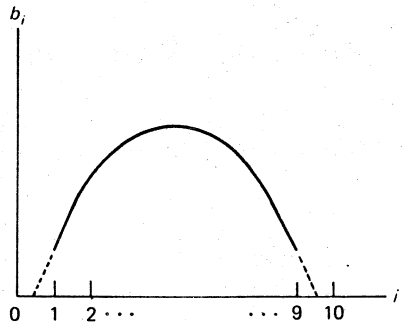
마지막으로 b의 추정량은 (5A.2)로부터 도출될 것이다.

$$\begin{aligned} \hat{b}_0 &= \hat{\alpha}_0 \\ \hat{b}_1 &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 \\ &\vdots \\ \hat{b}_9 &= \hat{\alpha}_0 + 9\hat{\alpha}_1 + 81\hat{\alpha}_2 + 729\hat{\alpha}_3 \\ \hat{b}_{10} &= 0 \end{aligned} \qquad \textbf{(5A.8)}$$

간단히 말해서, 종점제약 $b_{10} = 0$을 부과하게 되면 α_0를 치환할 수 있게 되고, 따라서 그것을 표준적인 알몬회귀모형(5A.3)으로부터 소거해 낼 수 있는 것이다. $b_0 = 0$와 $b_{10} = 0$ 둘다를 부과하게 되면, 회귀모형 (5A.3)에서 α_0과 α_1 둘을 α_2와 α_3의 식으로 치환할 수 있음을 보이는 것은 독자들의 연습으로 남겨두기로 한다. 이에 따라 α_0과 α_1는

실제로 추정하게 될 방정식에서 사라질 것이다.

종점제약을 부과하는 이러한 간접적인 방법은 (단지 분석에서 X_t와 X_{t-k}를 빼버리는) 직접적인 방법에 의해 산출된 분산 보다도 더욱 작은 분산을 갖는 불편추정량을 얻게 된다는 것을 본문에서 언급하였다. 하지만 이는 어떤 특정한 가정하에서만 타당하다. 그 가정이란 종점모수 (endpoint parameters)가 여타의 0이 아닌 모수(nonzero parameters)와 동일한 곡선상에 있다는 것이다. 예를 들어, 종점제약이 $b_0 = b_{10} = 0$이라면 $i = 0$과 $i = 10$일 때, 다항식의 값은 0임을 가정해야만 한다. 이는 위의 예(5A.4)에서 $b_{10} = 0$에 관해 가정하였던 것이다.



〈그림 5A.2〉

하지만 실제로 이 가정은 만족되지 않을 수도 있는 것이다. 예를 들어 k = 10이며, $b_0 = b_{10} = 0$이라고 가정하자. 게다가 0이 아닌 모수들이 〈그림5A.2〉에서 나타낸 것과 같은 2차다항식의 선상에 있다고 하여보자. $i = 0$과 $i = 10$의 경우 다항식의 값이 0이라는 가정은 다음을 의미한다. 곧, 다항식을 그림으로 나타내면, 그것은 $i = 0$과 $i = 10$에서 i축과 교차할 것이다. 〈그림5A.2〉는 그렇지 않을 수도 있음을 보여주고 있다. 그림에서 보면 다항식은 $i = 0$과 $i = 1$ 사이 $i = 9$와 $i = 10$사이의 어딘가에서 (점선으로 표시된 것처럼) 축을 통과하고 있는

것이다. 그러므로 (5A.4)와 같이 위의 예처럼 조작한다면 결국 제약을 부과하는 것은 타당하지 못하게 될 것이다. 그 결론은 그에 따른 추정량이 偏倚가 있는 추정량(biased estimators)이라는 것이다. 요약하면, 세밀한 분석결과 종점제약이 타당하다고 볼 수 없다면 그 제약을 부과해서는 안되는 것이다. 이러한 이유로 말미암아 종점모수의 값에 관한 정보를 처리하는 직접적인 방법(예를들면 계수 b_0가 0이라 생각하면 (5A.1)에서 X_t를 제거시키는 것)이 보다 가능성이 있다.

부록B. 한개이상의 회귀모수를 포함한 가설검정

다음의 회귀모형을 생각해보기로 한다.

$$Y_t = a_0 + a_1X_{1t} + \cdots + a_kX_{kt} + u_t, \quad t = 1, \ldots, n, \tag{5B.1}$$

여기에서 설명변수 X_1 , $\cdots X_k$와 교란항은 표준적인 모든 가정을 만족한다고 한다. 이는 교란항이 정규분포라는 것도 포함하는 것이다.

가끔 (5B.1)과 같은 모형에서 경제학자들은 두개 이상의 모수(more than one parameter)와 관련된 가설을 검정하고자 할 때가 있다. 이러한 가정들은 항상 두가지 형태로 나타난다. 그 첫번째는 (5B.1)의 계수에 대한 선형제약(linear restriction)과 관련된다. 그러한 가설에 대한 몇가지 예가 다음과 같다.

$$\begin{aligned} a_1 &= a_2 \\ a_3 &= 2a_5 \\ a_0 + a_1 + a_2 + \cdots + a_k &= 1 \end{aligned} \tag{5B.2}$$

그러한 가설의 두번째 형태는 일련의 설명변수의 유의성(significance)

과 관련된다. 가령 (5B.1)에서 Y_t는 X_1, X_2 또는 X_3에 의존하고 있지 않다는 가설을 검정하기로 가정하자. 그러면 다음과 같은 가설을 검정하는데 관심을 두게 될 것이다.

$$a_1 = a_2 = a_3 = 0 \tag{5B.3}$$

이러한 두 경우 모두 (5B.2)와 (5B.3)에 표시한 가설을 귀무가설로 하고, 대립가설은 단지 귀무가설의 여집합(곧, H_0가 아니다)을 취하게 될 것이다. 예를들면 (5B.3)에서 표시한 귀무가설에 대한 대립가설은 최소한 모수 a_1, a_2, a_3중 하나는 0이 아니다라는 가설인 것이다. 마찬가지로 (5B.2)에 표시한 가설에 대한 대립가설은 다음과 같을 것이다.

$$\begin{gathered} a_1 \lesseqgtr a_2 \\ a_3 \lesseqgtr 2a_5 \\ a_0 + a_1 + \cdots + a_k \lesseqgtr 1 \end{gathered} \tag{5B.4}$$

다행히 이러한 가정을 검정하기 위한 매우 간단한 기법이 있다. 추정절차는 다음과 같이 약술해 볼 수 있다.

제 1 단계 : 회귀모형에 고려할 귀무가설을 부과한다. 가령 가설이 $a_1 = a_2$이면 회귀모형(5B.1)은 다음과 같이 다시 쓰게 될 것이다.

$$Y_t = a_0 + a_1(X_{1t} + X_{2t}) + a_3X_{3t} + \cdots + a_kX_{kt} + u_t \tag{5B.5}$$

또 다른 예로서 귀무가설이 $a_1 = a_2 = a_3 = 0$이면 다음을 얻게 된다.

$$Y_t = a_0 + a_4X_{4t} + \cdots + a_kX_{kt} + u_t \tag{5B.6}$$

마지막으로 세번째 예로서 귀무가설이 $a_1 + 2\,a_2 + 5\,a_3 = 10$이면, $a_1 = 10 - 2\,a_2 - 5\,a_3$이며, 따라서 다음과 같게 된다.

$$Y_t = a_0 + 10X_{1t} + a_2(X_{2t} - 2X_{1t}) + a_3(X_{3t} - 5X_{1t}) + a_4X_{4t} + \cdots + a_kX_{kt} + u_t \qquad (5B.7)$$

이것을 다음과 같이 다시 쓸 수 있다.

$$(Y_t - 10X_{1t}) = a_0 + a_2(X_{2t} - 2X_{1t}) + a_3(X_{3t} - 5X_{1t}) + a_4X_{4t} + \cdots + a_kX_{kt} + u_t \qquad (5B.8)$$

제 2 단계 : 그에 따라 제약이 주어진 형태의 모형을 추정하여, 오차자승합 ESS_R*(error sum of squares)를 결정한다.

제 3 단계 : 회귀모형 (5B.1)의 처음의 (제약이 없는) 형태를 추정하여 오차자승합, ESS_U를 결정한다.

제 4 단계 : 제약의 주어진 모형과 제약이 없는 모형사이의 모수의 수의 차이를 결정한다. d를 이 차이라고 하자. 가령, (5B.2)의 가설 $a_1 = a_2$의 경우 $d = 1$이다. 두번째 예를들면, (5B.3)에서의 가설은 $d = 3$을 의미한다.

제 5 단계 : 다음의 비율을 계산한다.

$$\frac{(ESS_R - ESS_U)/d}{ESS_U/(n - k - 1)} \qquad (5B.9)$$

여기에서 n은 관찰치의 수이며, (k + 1)은 처음의 (제약이 없는) 모형의 모수의 수이다. 지금 논의중인 귀무가설은 (5B.9)에서의 비율의 크기에 의해 채택 또는 기각된다. 예를들면, 고려되고 있는 귀무가설이 잘못되었다 (false)고 가정하자. 그러면 제약이 주어진 회귀모형은 잘못된 가설에 기초하고 있으며, 따라서 잘못 표기하였다고 말할 것이다. 다

---

* $ESS = \Sigma(Y_t - \hat{Y}_t)^2$임을 제 4 장과 제 5 장에서 상기해 볼 것.

른 한편 처음의 회귀모형이 잘못 표기되었다고 결할 것이다. 이는 다음을 의미한다. 곧 ESS_R과 ESS_U는 크기에서 상당히 다른 것이다. 특히 $ESS_R > ESS_U$임을 예상하게 될 것이다.

다른 한편, 고려중인 가설이 사실이(true)라고 가정하자. 그러면, 제약이 주어진 회귀모형은 적절하게 표기되었다고 말하게 될 것이다. 하지만 처음의 또는 제약이 없는 회귀모형은 또한 적절히 표기되었다고 말하게 될 것이다. 비록 이상해 보일지 모르지만 이것이 왜 사실인지를 알게 되는 것은 어려운 일이 아니다. 모형을 완전히 설명할 때 회귀모수들에 대한 가정이란 그것들이 상수라는 것 뿐이다. 분명히 이 가정은 (5B.2) 또는 (5B.3)에서의 가정이 사실인지 아닌지에도 불구하고 만족될 것이다. 가령 $a_1 = a_2$라는 조건은 (5B.1)과 같은 모형에 관계된 어떠한 가정에도 위배되지 않는다. 마찬가지로 0은 상수이기 때문에 $a_2 = a_3 = 0$의 경우라면, 이는 다시 (5B.1)에서의 어떠한 가정에도 위배되지 않을 것이다.

뒷장에서 제약이 주어진 형태의 회귀모형을 고려해볼 때 고려될 귀무가설이 사실이라면 추정량의 특성과 관련한 몇가지 잇점이 있음을 설명할 것이다. 하지만 여기에서는 다음의 것만 주목하기 바란다. 곧, 제약이 주어진 형태의 회귀모형과 제약이 없는 형태의 회귀모형 둘다 타당하다고 말하면 ESS_U와 ESS_R은 단지 "작은(small)" 크기의 차이밖에 없다고 예상하게 될 것이다. 이러한 모든 것은 다음과 같이 요약될 수 있다. 곧,(5B.9)에서 비율의 값이 "크다는(large)"는 의미는 고려되는 귀무가설이 틀렸다는 것이며, 그 값이 "작다(small)"는 의미는 귀무가설이 옳다는 것이다.

이를 공식으로 나타내기 위해서는 다음을 보일 수 있다.* 곧, 고려될 귀

---

* Arthur S.Goldberger, Econometric Theory(New York:Wiley,1964), pp. 173-177을 볼 것.

무가설이 옳다면, (5B.9)에서의 비율은 d와 $(n-k-1)$의 자유도를 가진 F변수($F_{d,n-k-1}$)일 것이다.

제 6 단계 : 이러한 비율의 작은 값이 귀무가설의 채택(acceptance)과 관계되므로 다음과 같다면 가령 5%의 신뢰수준하에서 H_0를 채택하게 될 것이다.

$$\frac{(ESS_R - ESS_U)/d}{ESS_U/(n-k-1)} < F^{0.95}_{d,n-k-1} \qquad (5B.10)$$

여기에서 $F^{0.95}_{d,n-k-1}$은 다음과 같다.

$$Prob(F_{d,n-k-1} < F^{0.95}_{d,n-k-1}) = 0.95 \qquad (5B.11)$$

F분포에 대한 표준표(standard table)로부터 $F^{0.95}_{d,n-k-1}$을 구할 수 있다. 참고하기 위해 이 책 끝부분에 F분포에 대한 그와 같은 표를 통계표 3으로 하여 첨가하였다.

한가지 예로 가설 $a_1 = a_2 = a_3 = 0$를 고려한다고 하자. 게다가 $n = 49$이며 k = 8이라고 가정하자. 통계표 3에서 $F^{0.95}_{3,40} = 2.84$임을 알게 된다. $ESS_R = 35$이며, $ESS_U = 20$이라 결정하였다고 해보자. 그러면 비율은 다음과 같게 될 것이다. 곧,

$$\frac{(35-20)/3}{20/40} = \frac{15/3}{1/2} = 10 > 2.84$$

결과적으로 5%의 신뢰수준하에서 $a_1 = a_2 = a_3 = 0$이라는 가설을 기각하게 될 것이다.

문 제

1. 생산함수 $Q_t = (1/A)\, L_t^{a}\, K_t^{b}\, e^{ut}$ 을 생각하기로 한다. 여기에서 Q_t, L_t, K_t는 t시점에서의 산출량, 노동, 자본이며 u_t는 그와 관련된 오차항이다. $E(u_t) = 0$, $E(u_t^2) = \sigma^2$이며, u_t는 L_t와 K_t와는 무관하다고 가정하자. A, a와 b를 추정하기 위한 방법을 서술하라.

2. 어떤 해의 민간투자액은 이자율과 여당에 의존한다고 가정한다. 곧, 그것은 대통령이 민주당원보다는 공화당원일 때 투자가 보다 높아지게 되는 경우인 것이다. 단지, 두 정당만이 존재한다라고 가정하는 시계열회귀모형의 방정식을 세워보라.

3. 모형 $C_t = a_0 + a_1 F_t Y_t + a_2 Y_t^{1/2} + a_3(1/A_t) + u_t$를 생각한다. 여기에서 C_t는 t번째 가계의 소비지출이고, Y_t는 그 가계의 소득이며, F_t는 그 가계의 가족규모이고, A_t는 그 가계의 유동자산(liquid assets)이다. 이 모형을 선형모형으로 전환시켜 보라.

4. 단순선형소비함수 $C_i = a + b\, Y_i + u_i$를 n명의 개인에 대해 추정하고자 한다고 하자. 만일 절편이 개인의 주거위치에 의해 영향을 받는다고 가정한다면 도시의 소비자와 농촌의 소비자사이에 함수의 이동이 가능하다고 생각할 수 있는가?

5. 어느 기업의 투자지출은 이자율, 이윤율과 예상의 지표로서 판매변화량에 의존한다고 가정하자.

   a. 그와 관련된 회귀모형을 세워보라.

   b. 동일한 기간동안 이 기업의 이윤이 각각 매 시점마다 15%라고 가정하자. 그에 따른 추정의 문제에 대해 논의하여 보라.

6. 어떤 시점 t의 투자율 I_t가 그 시점의 이자율 r_t와 그 시점의 판매액 S_t와 판매율의 시차치 7개에 의존한다고 가정하자. S의 시차치

와 관련된 가중치가 처음에는 증가하여 정점에 도달하였다가 이후 감소한다고 가정하자.

a. 제약이 없는 형태의 모형을 세워보라.

b. 알몬형태의 모형을 세워보라.

c. 알몬형태의 경우 정규방정식을 자세히 써보라.

7. 다음의 다중회귀모형을 생각하기로 한다.

$$Y_t = a + b_0 X_t + \cdots + b_6 X_{t-6} + \varepsilon_t$$

이 모형의 모수를 추정하기 위해 4차다항식을 가진 알몬기법을 이용하기로 한다. 그 결과는 $\hat{\alpha}_0 = 1$, $\hat{\alpha}_1 = 3$, $\hat{\alpha}_2 = 5$, $\hat{\alpha}_3 = 4$, 그리고 $\hat{\alpha}_4 = -10$이라고 가정하자.

a. b_2의 추정치는 얼마가 될 것인가?

b. $b_0 + b_1 + \cdots + b_6 = 1$을 가정한다고 하자.

이 정보를 알몬다항식근사형에서의 α에 대한 제약으로 표시해 보라.

8. 때로는 어떤 회귀모형의 실증결과는 현저히 경험의 영역밖에 있는 사상 (events)을 예측하는데 쓰여서는 안된다고들 한다. 이를 논의하라. (도움말 : 다항식근사형의 기초가 되는 몇가지 가정을 생각해보자)

9. 다음의 코익모형을 보다 간단한 형태로 바꾸어 보자.

$$Y_t = a_0 + a_1 X_t + b_0 Z_t + b_1 Z_{t-1} + \cdots + u_t$$

여기에서

$$b_i = b_0 \lambda^i, \quad i = 1, 2, \ldots$$

10. 다음의 알몬시차모형을 생각한다.

$$Y_t = b + b_0 X_t + b_1 X_{t-1} + \cdots + b_{10} X_{t-10} + u_t$$

여기에서 다음과 같은 2차의 근사형을 가정한다. 곧,

$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2, \quad i = 0, 1, \ldots, 10$$

$b_5 = 3$이라고 가정하자. 그에 따른 α의 제약을 도출하고, 이에 따라 제약이 주어진 형태의 회귀모형을 도출하라.

11. 소비지출 C가 소득 Y_t에 의존하며, 만일 이자율 r_t가 0.05를 초과한다면 r_t에도 의존한다고 가정해 보자. 만일 r_t가 0.05를 초과하지 않으면 단지 Y_t에만 의존하게 되는 것이다. 이 관계를 회귀모형으로 공식화할 수 있을 것이다.

12. 다음의 모형을 다중선형회귀모형으로 전환시키고, 정규방정식을 써보라.

$$\log Y_t = a_0 + a_1 e^{X_{1t}} + a_2 \left(\frac{1}{1 + X_{1t}X_{2t}}\right) + u_t$$

제 6 장 회귀분석상의 제 문제

앞에서는 일련의 변수들 사이의 관계를 추정하기 위한 기법을 개발하였으며, 가설을 검정하고, 어느 한 변수가 또다른 한 변수에 주는 영향을 예측할 수 있도록 정보를 이용하는 방법을 학습하였다. 이러한 기법들은 회귀모형을 설명할 때 열거, 논의되었던 많은 가정들에 의존하고 있다. 하지만 실제로 어느 한가지 이상의 가정이 함수관계의 특성에 의해 위배되거나, 또는 그 대신에 변수에 대한 특정 관찰치집단으로부터 심각한 어려운 문제가 발생하는 일이 확실히 (또는 다른 경우에는 적어도 매우 그럴듯하게) 있다. 계량경제학에서 최근의 많은 시도가 이러한 문제를 다루기 위한 수정된 추정기법을 개발하려 하였다.

이러한 내용이 책의 마지막 두 章의 주제인 것이다. 여기에서는 회귀모형에 대한 가정에 위배되는 사항을 발생시키기거나 적어도 이용상의 난점, 곧 유효성 (effectiveness)의 문제를 일으키는 몇가지 유형의 상황을 검토하여보고, 이에 따라 이러한 문제를 고려할 수 있도록 추정절차를 수정할 수 있는 방법을 논의하여 볼 것이다.

1 . 多重共線性

몇군데에서 이미 다중공선성 (multicollinearity)의 문제를 토론하였다. 정형적으로 이 문제는 하나의 독립변수가 여타 변수의 선형결합일때 발생하는 것으로 그 결과로 독립적인 정규방정식의 수가 부족한 나머지, 모든 계수에 대한 추정량을 도출해낼 수가 없는 것이다. 간략하게 복습하는 의미에서 다음을 보자.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \qquad (6.1)$$

방정식 (6.1) 에서 X_1 과 X_2 의 값은 항상 일치한다고 가정하자. 그러면 다음과 같게 된다.

$$X_{1t} = X_{2t}, \qquad t = 1, 2, \ldots, n$$

이는 X_1 의 매시점마다의 움직임은 바로 X_2 의 동일한 움직임과 정확하게 일치함을 뜻한다. 만일 이것이 사실이라면 Y에 대한 X_1 의 영향을 Y에 대한 X_2 의 영향으로부터 간단하게 분리해낼 수는 없는 것이다. 하지만 Y에 대한 X_1 과 X_2 의 결합된 영향을 추정할 수는 있음을 기억해야 할 것이다. 곧, (6.1) 에서 X_{2t} 대신 X_{1t} 를 대체함으로써 (6.2) 에서 $b_3 = (b_1 + b_2)$ 의 값을 추정할 수 있다.*

$$Y_t = b_0 + (b_1 + b_2)X_{1t} + u_t = b_0 + b_3X_{1t} + u_t \qquad (6.2)$$

만일 X_{1t} 와 X_{2t} 의 다중공선성이 표본을 넘어서는 시점에도 계속 유지된다면, 방정식 (6.2) 의 추정형태는 종속변수 Y_i 의 미래치에 대한 예측에 유용할 것이다.

바로 지금 설명한 경우가 "완전"다중공선성의 하나로 불린다. 곧 하나 (또는 그 이상의) 독립변수는 여타 변수의 정확한 (exact) 선형결합인 것이다. 이 경우는 앞장에서 토론하였던 것으로 假變數 (dummy variables) 가 (삼계절 대신) 사계절 모두에 포함된 경우처럼 회귀방정식

---

* 정형적으로 모수의 수가 독립적인 정규방정식의 수와 일치할 수 있도록 추정하기 위한 모수의 수를 하나로 줄였기 때문에 (6.2) 는 추정이 가능하다.

이 정식화되는 방법으로부터 항상 발생한다. 또다른 한가지 예로서 다음 형태의 소비함수를 가정하기로 한다.

$$C_t = b_0 + b_1 Y_{dt} + b_2 Y_{d(t-1)} + b_3(\Delta Y_{dt}) + u_t \qquad (6.3)$$

여기에서 Y_{dt}는 소비에 대한 현재소득의 영향을 반영하고, $Y_{d(t-1)}$은 과거의 소득수준 또는 습관의 영향을 가리키며, $\triangle Y_{dt} = (Y_{dt} - Y_{d(t-1)})$은 최근의 소득수준변화로부터 발생하는 "기대효과(expectation effect)"를 반영한다. $\triangle Y_{dt} = (Y_{dt} - Y_{d(t-1)})$이 Y_{dt}와 $Y_{d(t-1)}$의 선형결합이므로 (6.3)에서 모든 계수의 추정은 불가능한 것이다. (6.3)을 추정가능하고, C_t의 값을 예측하는데 이용할 수 있는 형태로 만들기 위한 한가지 예로 남겨두기로 한다.

이와 같은 것들은 완전(perfect) 다중공선성의 경우이다. 하지만 이는 모든 단계에서 일어나는 것으로, 전형적으로 연구자들을 괴롭히는 덜 완전한(less-than - perfect) 다중공선성이다. 어느 정도 직관적으로 말하면 독립변수들이 비록 완전하지는 않을지라도 상관관계가 높을 때 문제가 발생한다. 예를 들어 (6.4)에서의 수요함수를 추정하기로 가정하자.

$$Q_t = b_0 + b_1 P_t + b_2 Y_t + u_t \qquad (6.4)$$

어떤 특정상품집단, 가령 수입품의 경우, 여기에서는 수요량(Q)이 국내에서 생산한 재화의 가격수준(P)과 소비자의 소득수준(Y)에 의존한다고 가정한다. 국내의 가격수준이 높을수록 사람들은 보다 값싼 외제상품을 사고자 할 것이며, 따라서 b_1이 양수라고 예상하게 될 것이다. 마찬가지로 소비자의 소득이 높을수록, (수입품을 포함한) 모든 재화와 용역을 사람들이 더욱 많이 구입한다고 예상하게 될 것이며, 따라서 또한

b_2도 양수일 것이다. 자료를 훑어보면, (항상 사실이지만) 급속한 국내의 인플레이션기간동안 수입품은 증가하였으며, 소득이 급속하게 상승하는 기간에도 수입품 또한 증가하였음을 보게 된다. 문제점은 소득이 급속히 상승하는 기간은 일반적으로 높은 인플레이션의 기간이며, 또한 그 반대의 경우도 사실이라는 것이다. 또는 바꿔 말하면 P와 Y사이에 높은 (비록 완전하지 않은) 正의 상관관계가 존재한다는 것이다. 이러한 높은 상관관계의 효과는 Q에 대한 P와 Y의 영향을 분리해내기 어렵게 한다는 것이다. 만일 소득과 국내가격이 증가하는 것과 동시에 수입품이 급속도로 늘어난다면, 수입품의 증가를 초래하는데 대한 인플레이션과 보다 높은 소득의 상대적 효과를 판단하기 곤란한 것이다.

가. "불완전한 (imperfect)" 다중공선성 : 몇가지 결론

다중공선성이라는 심각한 문제가 존재할 때 연구자는 그것을 어떻게 알 것인가? 이미 지적하였듯이 다중공선성은 여러 단계에서 나타나며, 경우에 따라 다루기 곤란한 것으로 또는 그렇지 않은 것으로 증명된다. 하지만 아마도 높은 단계의 "불완전한 (imperfect)" 다중공선성에 의해서만 설명할 수 있는 일련의 방정식의 결과가 존재한다. 곧, 독립변수의 계수에 통계적으로 무의미한 (statistically insignificant) 추정치에 수반되는 큰 결정계수 (R^2)가 그것이다. 이의 의미는 다음과 같다. 독립변수 중 어떤 것 (최소한 하나)이 (높은 R^2에 의해 지적되었듯이) 종속변수에 구조적으로 영향을 미치는 것으로 나타나지만 어느 것인지는 말할 수 없다는 것이다.

보다 정형화하면 문제는 높은 단계의 다중공선성은 계수의 추정량에 대한 큰 분산을 초래한다는 것이다. 생각해야 할 것은 큰 분산이 해당모수에 대한 어떤 주어진 비율 (예를 들어 95%)의 신뢰구간이 상대적

으로 넓음을 의미하다는 것이다. 곧 아마도 0의 값을 포함한 모수의 값의 큰 범위가 구간과 일치하게 될 것이다. 이의 의미는 다음과 같다. 관련된 독립변수가 종속변수에 어떤 중요한 영향을 미친다고 할지라도, 다중공선성문제는 그 변수의 효과를 정확하게 추정하기에 곤란하다는 것이다. 결론적으로 말해서 이러한 추정치에 기초한 정책예측의 신뢰성은 거의 없을 수 있다는 것이다.

불완전한 다중공선성이 추정량에 대한 큰 분산(따라서 큰 표준편차)을 가져온다는 것을 알기 위해 추정량 $\hat{b}_i$의 분산이 다음과 같다는 것을 제 4 장으로부터 생각해 내도록 하자.

$$\text{var}(\hat{b}_i) = \frac{\sigma_u^2}{\sum \hat{v}_{it}^2} \tag{6.5}$$

여기에서 $\hat{v}_{it}$는 모형의 여타 모든 설명변수에 대한 i번째 설명변수 X_{it}의 회귀에서의 殘差(residual)이다. 곧, $\hat{v}_{it} = X_{it} - \hat{X}_{it}$인 것이다. 이제 만일 독립변수사이에 밀접한 선형관계(linear relationship)가 존재한다고 하면, $\hat{v}_{it}$는 작아지게 될 것이다. 왜냐하면 계산된 값$\hat{X}_{it}$이 설명변수의 실제적인 값 X_{it}에 가까울 것이기 때문이다. 결국 $\hat{b}_i$의 분산이 크다는 사실에 의해 (6.5)의 분모는 작아지게 될 것이다.

이 모든 것의 의미를 올바르게 해석하는 것이 중요하다. 유의해야 할 것은 다음과 같다. 곧, 그것의 의미가 계수의 추정량에 偏倚가 있음(biased)을 뜻하지는 않는다는 것이다. 제 4 장의 부록에서 수식에 의해 증명하였듯이 그것들은 여전히 불편추정량이다. 다중공선성이 가져오는 것은 추정량의 不正確性이다. 곧 그것들의 분산이 크며, 따라서 매우 신뢰성이 있는 것은 아니다. 앞에서도 논의하였듯이, 문제는 각각의 독립변수의 경우 여타 독립변수를 별도로 하더라도 그것들이 종속변수에 특별히

미치는 영향을 판단하기 곤란하다는 것이다.

나. 추가 설명

그에 따라 내재적으로는 모형의 모든 설명변수들이 서로 높게 관련되었다고 가정하였다. 꼭 그럴 필요는 없다. 어떤 회귀모형이 설명변수 X_{1t}, X_{2t}와 X_{3t}를 포함하고 있다고 가정하자. 또한 X_{1t}와 X_{2t}가 높게 (그러나 불완전하게) 관련되지만 X_{3t}는 X_{1t}과 X_{2t}에 상대적으로 관련되지 않는다고 가정하자. 그러면 (6.5)의 분산식이 의미하는 바는 X_{1t}와 X_{2t}의 계수와 관련된 추정량, 가령 $\hat{a}_1$과 $\hat{a}_2$의 분산이 클 것이라는 것이다. 하지만 X_{3t}의 계수의 추정량, 곧 $\hat{a}_3$의 분산은 크지 않아도 된다. 직관적으로 말해서 X_{3t}가 X_{1t}와 X_{2t}에 높게 관련되지 않는다면 $\hat{v}_{3t}^2$는 일반적으로 클 것이다. 왜냐하면 $\hat{X}_{3t}$는 X_{3t}의 상대적으로 빈약한 예측자 (Predictor)이기 때문이다. (곧, X_{3t}는 X_{1t}와 X_{2t}에 의해 잘 설명될 수 없는 것이다.) 이제 다중공선성문제를 설명하는데 유용할지 모르는 하나의 정식을 부여하기로 한다. (6.5)의 분모에 나타나는 식 $\Sigma \hat{v}_{it}^2$는 X_{it}를 회귀모형의 여타 모든 설명변수에 대해 회귀함으로써 구한 誤差自乘合(error sum of squares)이다. 이 오차자승합을 ESS_i로 표기하자. $TSS_i = \Sigma X_{it}^2$ 이라 하자. 그러면 $TSS_i = RSS_i + ESS_i$이다. 여기에서 $RSS_i = TSS_i - ESS_i$이다. 분명히 $ESS_i \geq 0$, $TSS_i \geq 0$이다. 또한 $RSS_i \geq 0$임을 보일 수 있다.* 그러므로 $TSS_i \geq RSS_i$이고 $TSS_i \geq ESS_i$인 것이다.

---

* 제 4 장의 부록에 있는 (4A.5)를 보자. 그리고 유의해야 할 것은 k번째 설명변수의 경우 $TSS_k = \Sigma X_{kt}^2$, $ESS_k = \Sigma \hat{v}_{kt}^2$, $RSS = \Sigma \hat{X}_{kt}^2$ 이라는 것이다. 왜냐하면 $(\Sigma \hat{X}_{kt} \hat{v}_{kt}) = 0$이기 때문이다.

$r_i^2 = \mathrm{RSS}_i / \mathrm{TSS}_i$ 라고 하자. 그러면 $0 \leq r_i^2 \leq 1$ 이다. 분명히 X_{it} 가 여타 설명변수와 보다 높게 관련되어있으면, ESS_i 는 보다 작을 것이며 따라서 r_i^2 은 1에 보다 가까워질 것이다. 역으로 여타 설명변수에 대한 X_{it} 의 관계가 보다 약할수록 ESS_i 는 더욱 클 것이고, 따라서 r_i^2 은 0에 보다 가까워질 것이다. 그러므로 r_i^2 은 X_{it} 를 모형의 여타 설명변수와 관련시키는 多重決定係數 (multiple coefficient of determination)와 비슷하다.

이제 분산식 (6.5)의 분모를 $\mathrm{ESS}_i = \mathrm{TSS}_i - \mathrm{RSS}_i \equiv \mathrm{TSS}_i (1 - r_i^2)$ 으로 나타내기로 한다. 이에따라 $\mathrm{Var}(\hat{b}_i)$ 는 다음과 같이 표현할 수 있다.

$$\mathrm{var}(\hat{b}_i) = \frac{\sigma_u^2}{\mathrm{TSS}_i(1 - r_i^2)} \equiv \frac{\sigma_u^2}{\sum X_{it}^2(1 - r_i^2)} \tag{6.6}$$

식 (6.6)은 분명히 어떤 추정량의 분산에 대한 부분적 多重共線性 (partial multicollinearity) 영향을 가리킨다. 부분적 다중공선성이 없다는 것은 $r_i^2 = 0$ 인 경우에 해당한다. r_i^2 이 1에 가깝게 증가함에 따라 문제는 악화되고 결국 보다 큰 분산을 갖게된다. 마지막으로 유의해야 할 것은 r_i^2 이 모든 $i = 1, \cdots, k$ 의 경우에 대해 똑같이 크지 않아도 된다는 것이다.

다. 몇가지 해결책

多重共線性은 해결하기 쉬운 문제는 아니다. 연구자는 가능한 한 관찰치의 수를 늘림으로써 추정량의 정확성 (Precision)을 항상 올리려 할 수 있다.(곧, 분산을 줄이려고 할 수 있다.) 가령 (6.5)에서 $\hat{v}^2$ 이 아무리 작을지라도 분명한 것은 n 이 증가함에 따라 $\mathrm{Var}(\hat{b}_i)$ 이 감소한다는 것이다. 그러나 명백하게도 표본크기를 반드시 증가시킬 수 있는 것은 아

니며, 다중공선성이 아주 심각한 경우에는 표본수가 아주 많이 증가하지 않으면, 그것은 그다지 도움되지 않을 수도 있을 것이다.

한가지 대안적인 접근법은 개개의 계수의 값을 추정키 위해 이용할 수 있는 몇가지 추가적인 정보를 도입하는 것이다. 가령 어떤 특정 상품의 경우 방정식 (6.7)에 표현된 생산함수를 추정하고자 한다고 하자.

$$Q_t = AL_t^{\alpha}K_t^{\beta}e^{u_t} \tag{6.7}$$

여기에서 Q_t는 t시점의 생산량이고, L_t는 노동투입시간이며, K_t는 자본투입량이고, u_t는 교란항이며, 그리고 A, α와 β는 추정하고자 하는 모수이다. 로그변환을 이용함으로써 (6.7)을 다음과 같은 추정가능한 형태로 바꿀 수 있음을 기억해야 할 것이다.

$$Q_t^* = A^* + \alpha L_t^* + \beta K_t^* + u_t \tag{6.8}$$

여기에서 별표는 이것들이 방정식 (6.7)에서의 변수의 로그임을 가리킨다. 설명을 위해 표본에서 L과 K의 相關이 높다는 의미에서 부분적 다중공선성이 존재한다고 가정하자. (특히) 이 경우 L과 K사이의 높은 상관은 생산함수의 탄력성모수, 곧, α와 β의 추정량의 분산이 크다는 결과를 낳는다.

이제 또다른 자료로부터 얻은 정보에 기초하여 이 산업의 특징은 규모에 대한 수확불변(constant returns to scale)이라는 충분한 증거가 있다고 하자. 앞의 장에서의 생산함수에 대한 논의에 비추어볼 때, 이는 ($\alpha + \beta$) = 1 임을 의미한다. 이 정보를 이용하여 $\beta = (1 - \alpha)$를 (6.7)에 대입하면 이로부터 다음을 얻게된다.

$$Q_t = AL_t^{\alpha}K_t^{(1-\alpha)}e^{u_t} \tag{6.9}$$

로그를 취하면, 이제 다음을 얻게 된다.

$$Q_t^* = A^* + \alpha L_t^* + (1 - \alpha)K_t^* + u_t \tag{6.10}$$

여기에서 별표는 역시 원래 변수의 로그를 가리킨다. (6.10)에서 항을 다시 정리하면, 다음과 같다.

$$Q_t^* - K_t^* = A^* + \alpha(L_t^* - K_t^*) + u_t \tag{6.11}$$

또는

$$Y_t^* = A^* + \alpha Z_t^* + u_t$$

여기에서 $Y_t^* = (Q_t^* - K_t^*)$ 이고, $Z_t^* = (L_t^* - K_t^*)$ 이다.

선험적인(a Priori) 정보에 의해 모형을 단지 하나의 독립변수, 곧 Z_t^* 만이 존재하는 것으로 축소할 수 있다. 독자들이 유념해야 할 것은 다음과 같다. 곧, L_t^*과 K_t^*이 매우 강하게 관련된다고 할지라도 (6.11)을 추정하는데 다중공선성으로부터 초래되는 어려움이란 일반적으로 존재하지 않는다는 것이다. 가령 $L_t^* = 4K_t^*$이라고 가정하자. 그러면 추정문제는 없을 것이다. 왜냐하면 (6.11)에서 모형은 다음과 같이 축소될 것이기 때문이다.*

$$Y_t^* = A^* + \alpha(3K_t^*) + u_t \tag{6.12}$$

---

* 이 예가 논증하는 것은 다음과 같다. 곧, 단지 $L_t^* - K_t^*$가 상수와 같다면, 어려움이 존재한다는 것이다. 이는 L_t가 K_t에 정확하게 비례할 때 발생한다. 곧, $L_t = dK_t$일 때이다. 여기에서 d는 상수이다. 이러한 특별한 경우 방정식(6.11)의 $Z_t^* = (L_t^* - K_t^*)$는 상수일 것이며, α에 대한 추정량을 구할 수 없을 것이다. 이러한 점에서 제 2장의 이변량회귀모형에 대한 논의로부터 다시 생각해 보아야 할 것은 다음과 같다. 곧, 회귀계수를 추정할 수 있기 위해서는 독립변수가 적어도 두가지 상이한 값을 취해야만 하는 것이다.

요약하면 그 산업의 특징이 규모에 대한 수확불변이라는 추가적인 정보에 의해 모수A^*와 α의 경우 보다 작은 분산을 가진 추정량을 구할 수 있다는 것이다. 이에 따라 β의 추정량은 단지 다음과 같게 된다.

$$\hat{\beta} = 1 - \hat{\alpha} \tag{6.13}$$

라. 예측에 관한 결론

여러 많은 경우, 보충적인 정보는 이용불가능하며, 따라서 연구자는 보다 신뢰하기 어려운 일련의 모수추정치 때문에 난처하게 된다. 심지어 이러한 경우에도 추정방정식이 예측(forecasting)의 목적을 합리적으로 만족시킬 수도 있다는 사실을 안다면 조금은 위안이 될 수 있는 것이다. 어떤 극단적인 예를 들기 위해 다음과 같은 관계를 생각해 보기로 하자. 곧 그 예란 변수를 정의하는 방법때문에 완전 다중공선성을 나타내는 경우인 것이다.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \tag{6.14}$$

여기에서 $X_{1t} = 3 X_{2t}$ 이다. 앞에서도 지적한 바와같이 X_1 이 X_2 에 대해 완전선형관계(perfect linear relationship)를 포함하고 있기 때문에 b_1 과 b_2 를 추정할 수 없는 것이다. 하지만 예측을 위해서는 b_1 과 b_2 의 값에 관심을 갖는 것이 아니라 그보다는 X_1 과 X_2 가 관련된 Y_t 의 평균치에 관심을 갖는 것이다.

곧, 다음과 같다.

$$\begin{aligned} Y_t^m &= b_0 + b_1 X_{1t} + b_2 X_{2t} \\ &= b_0 + (3b_1 + b_2) X_{2t} \end{aligned} \tag{6.15}$$

앞의 논의로부터 다음과 같은 추정량에 의해 Y_t의 평균치를 추정할 수 있음을 알 수 있다.

$$\hat{Y}_t = \hat{b}_0 + \hat{b}^* X_{2t} \tag{6.16}$$

여기에서 $\hat{b}^*$은 합($3b_1 + b_2$)의 추정량이다. 비록 Y에 대한 X_1과 X_2의 영향을 추정할 수는 없지만, $X_{1t} = 3X_{2t}$의 관계가 계속 유지되는 한, X_{2t}의 어떠한 값에도 대응되는 Y의 값에 관한 예측은 할 수 있는 것이다.* 이에 대한 증명은 이 책의 범위를 벗어나는 것이지만 유사한 결론이 덜 완전한 (less-than perfect)다중공선성의 경우에도 확장된다. 특히 독립변수의 개별 효과의 추정량이 큰 분산을 갖고 있을지라도, 곧 Y_t의 평균Y_t^m으로 주어지는 Y_t에 대한 결합효과의 추정량$\hat{Y}_t$는 작은 분산을 갖고 있을지도 모른다. 예측이 Y_t의 평균치를 추정하는 것을 포함하고 있기 때문에, 그리고 그 평균을 아주 정확하게 추정할 수 있기 때문에 다중공선성은 예측의 목적에 지나치게 곤란한 점은 아닐 수도 있는 것이다.

2. 自己相關의 문제

회귀모형에서 기본적인 가정중의 하나가 어떤 한 시점의 교란항의 값은 여타 다른 시점의 그 값과는 무관하다는 것이다. 따라서 다음과 같게

---

* 예측의 우량성(quality), 또는 정확성(precision)은 다음의 두가지 고려사항에 달려있음이 분명하다. 곧, (1) Y_t의 평균치의 추정량의 정확성(곧, Y_t^m의 추정량으로서의 $\hat{Y}_t$의 분산)과 (2) 교란항 u_t의 분산의 크기, 불행히 다중회귀의 구조에서 예측의 정확성을 나타내는 공식은 여기에서 이용되는 것 이상의 통계학적 개념이 필요하다.

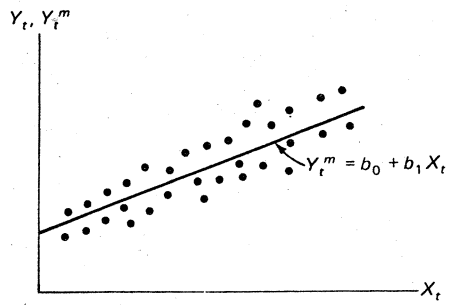
된다.

$$\text{cov}(u_t, u_s) = 0 \quad (\ t \neq s \text{ 일때 }\ )$$

이는 다음을 의미한다. 곧 설명변수의 어떤 값에 대해 Y_t는 그 평균, 곧 Y_t^m과 다른데, 그 크기는 여타 시점의 차이의 크기와는 무관하다는 것이다. 다음과 같은 형태의 회귀모형을 생각하여 보기로 하자.

$$Y_t = b_0 + b_1 X_t + u_t$$

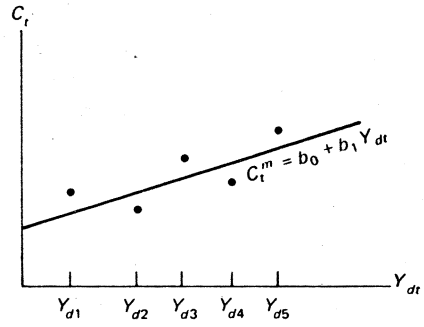
이에 관해 〈그림 6.1〉의 것과 같은 散布圖 (scatter diagram)를 예상해 볼 수 있다. 여기에서 관찰된 점들은 回歸線 (regression line)에 대해 "임의적으로 산포되어 있는 (randomly scattered)" 것이다.



〈 그림 6.1 〉

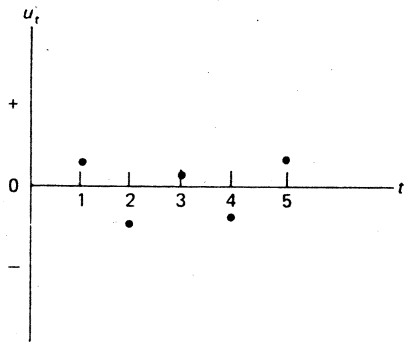
하지만 Cov (u_t, u_s) $\neq 0$ 이라고 가정해보자. 그러면 교란항의 연속적인 값들은 서로 무관한 것이 아니다. 예를 들어 어떤 개인을 고려하되 그의 소비행위를 다음과 같이 나타낼 수 있다고 하자.

$$C_t = b_0 + b_1 Y_{dt} + u_t \qquad (6.17)$$



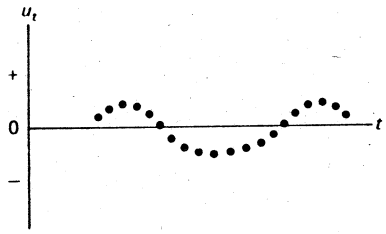
〈그림 6.2〉

하지만 여기에서 u_t 의 값은 이전의 값과는 무관한 것으로 한다. 가령 그가 1시점에 "너무 많이" 지출하였다면(아마도 친구가 갑자기 방문한 결과로), 따라서 $u_t > 0$ 이면, 그는 평소보다 적게 지출함으로써 그시점에서 보상하려 할 것이다. 따라서 $u_2 < 0$이 될 것이다. 다음을 유념해야 한다. 곧, 이것의 의미는 보다 일반적으로 u_t 는 u_{t+1}과 負의 相關(negative correlation)을 갖고 있다는 것이다. 만일 소득수준이 연속적인 시점을 따라 증가한다면, 교란항의 연속적인 값들사이의 그러한 負의 관계는 〈그림6.2〉에서 묘사한 것과 어느 정도 비슷한 산포도를 만들어 낼 것임을 예상해 볼 수 있다. 만일 시간에 걸쳐 교란항의 값을 그림으로 나타낸다면, 〈그림6.3〉에서와 같은 유형을 관찰하게 될 것이다.



〈그림 6.3〉

그 대신 교란항의 값이 시간에 걸쳐 正의 相關 (positive correlation)을 갖는 경우도 존재할 수 있다. 예를 들면 행동에서 완만한 조정의 결과로 u의 陽의 값이 전형적으로 u의 또다른 양의 값을 가져오고, 또 그 逆도 성립하는 것을 보게 된다. 이에 따라 $\mathrm{Cov}(u_t, u_{t+1}) > 0$인 것이다. 그 효과가 임의적으로 正에서 負로 변경되는 외부적인 힘에 의해 교란항의 값이 부분적으로 결정되는 것을 가정하면 (이는 나중에 설명할 것이다), u에 대해 시간에 걸친 값의 유형을 예상해 볼 수 있는데, 그 중에 陽의 값과 陰의 값의 行程 (runs)이 있는 것이다. 예를 들어 만일 외부적인 힘이 u의 양의 값을 만들어 낸다면, 그 이상의 양의 값이 뒤따르는 경향이 있게 된다. 그리고 음의 값이 만들어 진다면, 또 음의 값이 따른다는 것이다. 그러한 하나의 유형이 <그림 6.4>에 나타나 있다.



< 그림 6.4 >

이러한 교란항의 연속적인 값들의 상호종속 (interdependence)의 문제는 自己相關 (autocorrelation)으로 알려져 있다. 몇가지 가정하에서 어떤 주어진 관찰치의 경우에도 $E(u_t) = 0$임이 사실이라는 것을 보일 수 있다. 게다가 교란항은 여전히 독립변수와 상관되어 있지 않음도 보일 수 있으며, 따라서 계수의 추정량은 불편추정량인 것이다. 보게 되겠지만, 자기상관이 초래하는 문제는 추정량의 분산과 관련이 있다. 특히 자기상관이 존재한다면 분산에 대해 도출하는 식은 인정되지 못하는 것이

다. 만일 이 식을 계속적으로 사용하는 경우, 잘못된 t 비율을 만들게 되며, 이는 모형의 계수의 값에 대한 가설검정을 타당하지 않은 것으로 간주할 것이다. 결과적으로 가령, 사실상 그다지 0이나 다름없는 어떤 모수에 대한 추정치를 통계학적으로 유의한 것으로 채택할 수도 있는 것이다.

가. 자기회귀 모형

지금까지의 논의를 수식으로 나타내기 위해 자기회귀과정(autoregressive process)—하나의 교란항이 다른 교란항과 관련되는 양식—이 다음의 형태를 보인다고 가정하자.

$$u_t = \gamma u_{t-1} + \varepsilon_t, \qquad t = \ldots, -2, -1, 0, 1, 2, \ldots \tag{6.18}$$

여기에서 ε_t는 정규분포를 하는 확률변수(random variable)로 평균이 0이고 곧 $E(\varepsilon_t)=0$이고, 여타 어떠한 시점의 값과 무관하며, 따라서 $\mathrm{cov}(\varepsilon_t, \varepsilon_s)=0$이며, 상수의 분산을 갖고 있다. 곧, $E(\varepsilon_t^2)=\sigma_\varepsilon^2$ 이다. 또한 아래의 몇가지 이유로 인해 γ의 절대값은 1보다 작다고 가정하자 곧, $|\gamma| < 1$이다. 간단히 말해서 어떤 시점에서도 교란항의 값은 단순선형회귀모형에 의해 바로 이전의 값과 관련된다고 가정하고 있는 것이다. 유의해야 할 것은 (6.18)이 모든 시점, 곧 과거와 현재의 두 시점의 경우에도 유지된다고 가정하는 것이다. 이 모형에서 교란항, 곧 u의 연속적인 값은 비록 완전하지는 않을지라도 서로 상관관계를 맺고 있다. 특히, γ이 양수이면 u_t의 값은 바로 이전의 값, u_{t-1}과 正의 상관관계에 있을 것이고, 그 반면 γ이 음수이면, 이는 負의 상관인 것이다. 후자의 경우는 어떤 시점, 가령 t시점에 과잉지출(overspend)을 하면 다음의 시점에서 과소지출을 함으로써 보상하려는 어떤 개인에 대한 앞의 예에 해

당된다. 하지만(6.18)로부터 유의해야 할 것은 그 개인이 다음의 시점, 곧(t + 1)에서 실제로 과소지출을 하지 않을지도 모른다는 것이다. 왜냐하면, 또다른 기대치 못한 사정이 그로 하여금 또다시 과잉지출을 하게끔 할 수도 있기 때문이다. (예를 들면 $\varepsilon_{t+1} > 0$인 것이다.)

우선 (6.18)로 나타나는 자기회귀체계하에서 $E(u_t) = 0$ 임을 보이기로 한다. (6.18)에서 유의할 것은 다음과 같다.

$$u_{t-1} = \gamma u_{t-2} + \varepsilon_{t-1} \tag{6.19}$$

따라서 (6.18)에서 (6.19)를 u_{t-1}에 대입하면 다음을 얻게 된다.

$$u_t = \gamma(\gamma u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varepsilon_t + \gamma\varepsilon_{t-1} + \gamma^2 u_{t-2} \tag{6.20}$$

다음으로 u_{t-2}를 치환하고, 이에 따라 u_{t-3}을 치환하는 등의 과정을 계속한다하더라도 (6.18)이 모든 시점의 경우에도 계속 유지된다는 것을 생각하면 다음의 식을 얻게 된다.

$$u_t = \varepsilon_t + \gamma\varepsilon_{t-1} + \gamma^2\varepsilon_{t-2} + \gamma^3\varepsilon_{t-3} + \cdots \tag{6.21}$$

그런데 이는 u의 時差値(lagged value)를 포함하지 않고 있다. 왜냐하면 그 계수는 무한대의 제곱을 한 γ일 것 (따라서 0이 되는 것)이기 때문이다. (6.21)에서 u_t의 기대치를 구하면, 다음과 같다.

$$\begin{aligned} E(u_t) &= E(\varepsilon_t + \gamma\varepsilon_{t-1} + \gamma^2\varepsilon_{t-2} + \gamma^3\varepsilon_{t-3} + \cdots) \\ &= E(\varepsilon_t) + \gamma E(\varepsilon_{t-1}) + \gamma^2 E(\varepsilon_{t-2}) + \gamma^3 E(\varepsilon_{t-3}) + \cdots = 0 \end{aligned} \tag{6.22}$$

왜냐하면 $s = t$, $t-1$, $t-2$, …, 의 경우에 $E(\varepsilon_s) = 0$이기 때문이다. 마찬가지로 항상 해온 것처럼 ε_t가 설명변수, 가령 모든 시점에서 X_t

의 값과 무관하다고 가정하면 (ε_t 가 모든 t 와 s 에 대해 무관하다고 가정하면), u_t 는 X_t 와 무관할 것이며, 따라서 X_t 와는 상관이 없을 것임을 보일 수 있다. 곧, (6.21)을 이용하여 $\mathrm{Cov}(u_t, X_t) = E(u_t X_t) = 0$ 임을 보일 수 있는 것이다.

이하의 언급을 위해 다음에 유의하기로 한다. 곧, u_t 가 ε_t, ε_{t-1}, …의 선형결합이며, 모든 ε 이 서로 독립적이기 때문에 u_t 의 분산은 다음과 같다.*

$$\begin{aligned} \sigma_{u_t}^2 &= \sigma_\varepsilon^2 + \gamma^2\sigma_\varepsilon^2 + (\gamma^2)^2\sigma_\varepsilon^2 + (\gamma^3)^2\sigma_\varepsilon^2 + \cdots \\ &= \sigma_\varepsilon^2[1 + \gamma^2 + (\gamma^2)^2 + (\gamma^2)^3 + \cdots] \\ &= \frac{\sigma_\varepsilon^2}{1 - \gamma^2} = \sigma_u^2 \end{aligned} \tag{6.23}$$

왜냐하면 모든 ε 은 동일한 분산을 갖고 있으며, $|\gamma| < 1$ 이라는 가정 때문이다. (6.23)으로부터 다음을 알고 있다. 곧, u_t 의 분산은 t 를 포함하지 않으며, ε 과 마찬가지로 모든 u_t 는 동일한 분산을 갖고 있다. 곧, $\sigma_{ut}^2 = \sigma_{us}^2 = \sigma_u^2$ 또한 (6.23)으로부터 $|\gamma| < 1$ 이라는 가정의 이유를 알 수 있을 것이다. 이러한 가정없이는 (6.23)에서의 수열은 수렴하지 않게 된다. 곧 u_t 의 분산은 무한한 값을 갖게 될 것이다.

다음으로 교란항의 분산을 검토해 보기로 하자. u_t 를 (6.18)로 치환하고, (6.23)을 이용하면 다음을 얻게 되는 것이다.

---

* 여기에서는 등비수열의 합에 대한 기본적인 가정을 이용하고 있다. 곧 그것은 다음과 같다.

$$s = a(1 + \alpha + \alpha^2 + \alpha^3 + \cdots), \quad |\alpha| < 1$$

$$s = \frac{a}{1 - \alpha}$$

$\alpha \geq 1$ 이면, 수열은 수렴하지 않는다.

$$
\begin{aligned}
E(u_t u_{t-1}) &= E[(\gamma u_{t-1} + \varepsilon_t)u_{t-1}] \\
&= \gamma E(u_{t-1}^2) + E(\varepsilon_t u_{t-1}) \\
&= \gamma E(u_{t-1}^2) + 0 = \gamma \sigma_u^2
\end{aligned} \tag{6.24}
$$

왜냐하면 (6.21)에 의해 u_{t-1}은 단지 ε_{t-1}과 그 시차치에 달려있기 때문이다. 이 의미는 u_{t-1}과 ε_t가 독립적이며, 따라서 다음과 같게 된다는 것이다.

$$
E(\varepsilon_t u_{t-1}) = \mathrm{cov}(\varepsilon_t, u_{t-1}) = 0
$$

그러므로 (6.24)로부터 다음의 사실을 알게 된다. 곧, γ이 0이 아닌 한 방정식(6.18)의 자기회귀모형은 교란항간의 共分散 0이라는 가정에 위배되지 않는다는 것이다.

나. 추정량의 분산에 관한 결론

이제 이러한 사실이 추정량의 분산의 경우에 상이한 수식으로 나타남을 보이기로 한다. 제 2 장에서 서술한 형태의 단순이변수모형이 있다고 가정하자. 곧,

$$
Y_t = b_0 + b_1 X_t + u_t \tag{6.25}
$$

다음의 사실을 생각해야 한다. 곧, (교란항간의 공분산이 0이라는 것을 포함한) 모형의 가정에 따르면 추정량 $\hat{b}_1$의 조건분산(conditional variance)이 다음과 같음을 알게 된다.

$$
\mathrm{var}(\hat{b}_1) = \frac{\sigma_u^2}{\sum (X_t - \bar{X})^2} \tag{6.26}
$$

이 결과를 도출해 내기 위해 방정식(2.71)을 이용하였다. 이는 $\hat{b}_1$을

다음과 같이 표현할 수 있음을 의미한다.

$$\hat{b}_1 = b_1 + \frac{\sum (X_t - \bar{X})u_t}{\sum (X_t - \bar{X})^2} \tag{6.27}$$

그러면 $A = \Sigma (X_t - \bar{X})^2$ 이며 $w_t = (X_t - \bar{X})$라고 하고 (6.27)을 확장된 형태로 다시 쓰면 다음과 같다.

$$\hat{b}_1 = b_1 + \frac{w_1}{A} u_1 + \cdots + \frac{w_n}{A} u_n \tag{6.28}$$

X의 값이 주어지면(따라서 w와 A의 값이 주어지면) $\hat{b}_1$은 단지 교란항의 선형결합인 것이다. 그리하여 상관이 없는 (uncorrelated) 확률변수의 선형합(linear sum)의 분산에 대한 식을 이용하여 (6.28)을 구할 수 있다. 하지만 자기상관이 있다면, 이 식을 이용할 수 없다. 왜냐하면 u_t와 그에 따른 (6.28)의 항이 상관되어 있기 때문이다. 이는 (6.26)은 이미 $\hat{b}_1$의 분산에 대한 정확한 표현이 아님을 의미한다. 결과적으로, 표준식을 이용한 가설검정은 더이상 타당한 것이 아니다.*

다. 추정량의 평균

自己相關은 $\hat{b}_1$에서의 偏倚를 초래하지 않는다는 사실을 (6.28)로부터 쉽게 알아낼 수 있다. 위의 가정하에서 교란항 u_t는 여전히 X_t의 모든 값(따라서 w_t의 모든 값)에 독립적이고 이에 따라 그것들과는 상관이 없으며, 여전히 평균은 0이다. X_t의 주어진 어떠한 값에 대해서도 (6.28)

---

* (6.28)을 이용하면 다음과 같은 사실을 보일 수 있다. 곧, 자기상관이 존재한다면 $\hat{b}_1$의 분산, $E(\hat{b}_1 - b_1)^2$은 (6.28)의 교적항(cross-product terms)의 기대치를 포함한다는 것이다. 하지만 자기상관이 존재하지 않는 경우, 모든 교차항은 0이며, 따라서 그것은 없어지게 되며, 이에 따라 $\hat{b}_1$의 분산은 (6.26)의 식으로 나타날 것이다.

의 기대치를 구하면 다음과 같다.

$$E(\hat{b}_1) = b_1 + \frac{w_1}{A} E(u_1) + \cdots + \frac{w_n}{A} E(u_n) = b_1 \tag{6.29}$$

마찬가지로 식 $\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$를 이용하면, 다음과 같은 사실을 보일 수 있다.

$$\begin{aligned} E(\hat{b}_0) &= b_0 + b_1\bar{X} + E(\bar{u}) - \bar{X}E(\hat{b}_1) \\ &= b_0 \end{aligned} \tag{6.30}$$

자기상관이라는 문제의 일반적인 성격은 이제 분명하다. 그것이 존재하는 경우, 연속적인 관찰치에 대해 교란항의 값에서 체계적인 변동(systematic variation)이 있는 것이다. 이러한 변동의 유형은 편의가 있는 모수추정량을 가져오지는 않는다. 하지만 분산식은 더 이상 유지될 수 없으며, 따라서 그 이상의 결과없이는 가설검정과 신뢰구간추정이 불가능하다. 분명히 그 절차는 무언가를 요구하는 것이다. 더군다나 직관에 의해 다음과 같이 말할 수 있을 것이다. 곧 표준적인 추정절차는 분명히 자기상관을 설명할 수 없으므로 그것은 가장 신뢰할만한 모수추정량을 구할 수 없을지도 모른다는 것이다. 곧, 교란항중에서 어떤 유형의 변동이 존재하는 경우, 이러한 추가적인 정보를 계산에 포함시키게 되면 보다 나은 추정과 예측이 가능하다는 것이다. 이러한 것을 어떻게 할 수 있는지를 다음에서 알아보기로 한다.

라. 일반화된 추정기법

모형이 다음과 같이 이루어져 있다고 가정하자. 곧,

$$Y_t = b_0 + b_1 X_t + u_t \tag{6.31}$$

그리고

$$u_t = \gamma u_{t-1} + \varepsilon_t \tag{6.32}$$

여기에서 $|\gamma| < 1$이며, ε_t는 위에서 설정한 모든 가정을 만족한다. 문제는 방정식(6.31)에서의 모수추정량을 개선하기 위해 (6.32)에서 제공되는 정보를 어떻게 이용할 것인가 하는 것이다.

우선 γ의 값을 알고 있다고 가정하자. 만일 (6.31)의 시차형태(lagged form)를 구하여 γ을 곱하면, 다음을 얻게 된다.

$$\gamma Y_{t-1} = \gamma b_0 + \gamma b_1 X_{t-1} + \gamma u_{t-1} \tag{6.33}$$

(6.31)에서 (6.33)을 빼면, 다음과 같게 된다.

$$Y_t - \gamma Y_{t-1} = (b_0 - \gamma b_0) + (b_1 X_t - \gamma b_1 X_{t-1}) + (u_t - \gamma u_{t-1}) \tag{6.34}$$

(6.32)에서 각 항을 다시 정리함으로써 다음을 얻게 된다.

$$\varepsilon_t = (u_t - \gamma u_{t-1})$$

이를 (6.34)의 마지막항에 대입하면 다음을 얻게 된다.

$$Y_t - \gamma Y_{t-1} = (b_0 - \gamma b_0) + (b_1 X_t - \gamma b_1 X_{t-1}) + \varepsilon_t \tag{6.35}$$

이제 (6.35)를 다음과 같이 다시 쓸 수 있다.

$$Y_t' = B + b_1 X_t' + \varepsilon_t \tag{6.36}$$

여기에서

$$Y_t' = Y_t - \gamma Y_{t-1}$$
$$B = b_0 - \gamma b_0$$
$$X_t' = X_t - \gamma X_{t-1}$$

이다.

그런데 유의할 것은 (6.31)을 (6.36)으로 바꿀 때 하나의 관찰치를 잃게 된다는 것이다. 왜냐하면 (6.34)에서 시차가 주어지고, 뺄셈을 하기 때문인 것이다.

방정식(6.36)은 회귀모형의 표준형에 속한다. 특히, ε_t는 (u_t와는 달리) 교란항의 특성에 관한 모든 가정을 만족하고 있다. 그러므로 앞에서 설정한 추정절차를 단지 따를 수 있는 것이다. $\sum_{t=2}^{n} \hat{\varepsilon}_{t=0}$ 과 $\sum_{t=2}^{n} (X_t' \hat{\varepsilon}_t) = 0$의 조건을 부과하여 두개의 정규방정식을 도출하게 된다. 그러면 그 해로부터 (6.36)의 두 모수에 대한 추정량, $\hat{B}$와 $\hat{b}_1$을 얻게 되는 것이다. 그러면 이로부터 b_0의 추정량을 다음과 같이 구할 수 있게 된다.

$$\hat{b}_0 = \frac{\hat{B}}{1 - \gamma} \tag{6.37}$$

또한 다음의 사실도 보일 수 있다.

$$\text{var}(\hat{b}_0) = \left(\frac{1}{1 - \gamma}\right)^2 \text{var}(\hat{B}) \tag{6.38}$$

왜냐하면 $\hat{b}_0$는 $\hat{B}$에 완전하게, 그리고 선형적으로 관련되어 있기 때문이다. ε_t는 표준적인 모든 가정을 만족하기 때문에 $\hat{B}$과 $\hat{b}_1$의 분산은 다음과 같이 표준적인 식으로 주어질 것이다.

$$\text{var}(\hat{B}) = \frac{\sigma_\varepsilon^2 \sum_{t=2}^{n} (X_t')^2}{n' \sum_{t=2}^{n} (X_t' - \bar{X}')^2}$$

$$\text{var}(\hat{b}_1) = \frac{\sigma_\varepsilon^2}{\sum_{t=2}^{n} (X_t' - \bar{X}')^2} \tag{6.39}$$

여기에서 $n' = n - 1$ 이다. 왜냐하면 X'_t을 얻기 위해 시차를 주고, 뺄셈을 하는 과정에서 한개의 관찰치를 상실하기 때문이다.

이제 (6.36)에 의해 기본적인 모수를 추정한다면, 최소한 원칙적으로는 가설을 검정할 수 있으며, 타당한 신뢰구간을 설정할 수 있다. 또한 지적해야 할 것은 (6.36)으로부터 얻은 추정량이 "유효하다(efficient)"는 것이다. 직관적으로 말하면,* 이 의미는 다음과 같다. 표본크기가 크다면 (6.36)으로부터 구한 추정량의 분산은 B와 b_1의 여타 불편추정량의 분산보다 작거나 같다는 것이다. 표본크기에도 불구하고 추정량이 최소분산을 가진다고 말할 수 없는 이유는 기법이 한개의 관찰치를 상실하기 때문인 것이다. 본질적으로 이러한 한개의 관찰치의 중요성은 표본크기가 증가함에 따라 무시할 수 있는 것이다.**

이제 교란항 자체간의 관계에 대한 이용가능한 정보를 추정절차에 통합하기 위해 바람직한 특성을 지닌 기법을 알게 되었다. 바로 자기상관이 있는 교란항을 가진 회귀모형을(교란항간의 공분산이 0임을 포함하는) 기본적인 회귀모형의 모든 가정을 만족하는 회귀모형으로 변형시켜서 단지 표준적인 기법을 적용시키는 것이다. 이러한 절차를 수행하는데 따르는 어

---

* 그 이하의 사항은 유효추정량(efficient estimator)의 공식적인 정의는 아니다. 그 정의는 이 책의 범위를 벗어나며, 앞으로 나타날 결과를 이해하는 데 불필요한 것이다.

** 유효성(efficiency)은 항상 소위 평균평방오차(mean square error, M.S.E.)에 의해 정의된다. 곧, $\hat{\alpha}$를 α의 추정량이라 하자. 그러면 $\hat{\alpha}$의 M.S.E.는 그것의 분산 $\sigma^2_{\hat{\alpha}}$의 합과 그것의 편의의 제곱 $[E(\hat{\alpha}) - \alpha]^2$과 동일하다. 위의 경우 추정량은 불편추정량이며, 따라서 M.S.E.는 분산과 동일하다. 여하튼 약간의 세밀한 차이를 무시하면, $\hat{\alpha}$이 α의 유효추정량일 때, (표본이 크다면) $\hat{\alpha}$의 M.S.E.는 α의 여타 일치추정량의 M.S.E보다 작거나 동일함을 예상할 수 있다.

려움이란 일반적으로 γ을 모른다는 것이다. 따라서 γ의 값을 먼저 추정해야만 한다.*

원래의 방정식(6.31)로부터 殘差(residuals)를 살펴봄으로써 그 추정은 가능하게 된다. $E(u_t)=0$이며, $\mathrm{Cov}(u_t, X_t)=E(u_t X_t)=0$임을 생각하면, $\Sigma_{t=1}^{n} \hat{u}_t = 0$이며, $\Sigma_{t=1}^{n} (\hat{u}_t X_t)=0$이라는 조건을 부과함으로써 (6.31)에서 모수의 불편추정량을 도출할 수 있다. 이로부터 다음과 같은 통상적인 정규방정식을 얻게 된다.

$$\begin{aligned} \sum Y_t &= n\hat{b}_0 + \hat{b}_1 \sum X_t \\ \sum (X_t Y_t) &= \hat{b}_0 \sum X_t + \hat{b}_1 \sum X_t^2 \end{aligned} \tag{6.40}$$

불편추정량 $\hat{b}_0$와 $\hat{b}_1$을 얻기 위해 이 방정식을 풀 수 있다. 그러면 교란항의 값에 대한 추정량 $\hat{u}_t$을 구하기 위해 $\hat{b}_0$와 $\hat{b}_1$을 이용할 수 있는 것이다.

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - (\hat{b}_0 - \hat{b}_1 X_t) \tag{6.41}$$

γ를 추정하기 위해서는 $\hat{u}$의 값을 (6.32)에 의해 "제시된" 관계에 그대로 대입한다. 곧,

$$\hat{u}_t = \gamma \hat{u}_{t-1} + \varepsilon_t \tag{6.42}$$

---

* 가끔 이는 포함된 모든 변수가 제1차 계차의 형태(first-difference form)속에 들어있는 이론상의 회귀방정식에서 보인다. 곧 이변수의 경우 $(Y_t - Y_{t-1})$이 $(X_t - X_{t-1})$에 대해 회귀되는 것이다. (6.36)으로부터 알 수 있는 것은 그러한 방정식이 $\gamma = 1$인 자기회귀모형의 특별한 경우라는 것이다. (6.23)에서 유의할 것은 그러한 모형의 경우 u_t의 분산이 무한이라는 것이다. 더군다나 이 절차는 아주 제한적이다. 왜냐하면 그 결과는 $\gamma = 1$을 조건으로 하고 있기 때문이다.

(6.42)를 하나의 회귀모형으로 생각하기로 한다. ε_t가 u_{t-1}과는 무관하므로 ε_t가 $\hat{u}_{t-1}$과 상관이 없도록 하여보자.* 이러한 가정을 이용하여 표준적인 기법에 의해 (6.42)에서 γ을 추정할 수 있을 것이다. 특히 (약간의 복습으로서) (6.42)를 다음과 같이 다시 쓰기로 한다.

$$\hat{u}_t = \hat{\gamma}\hat{u}_{t-1} + \hat{\varepsilon}_t \tag{6.43}$$

여기에서 $\hat{\varepsilon}_t = (\hat{u}_t - \hat{\gamma}\ \hat{u}_{t-1})$은 교란항의 추정량이고, $\hat{\gamma}$는 γ의 추정량이다. $\mathrm{Cov}(\varepsilon_t,\ \hat{u}_{t-1}) = 0$이라는 가정이 의미하는 것은 (한개의 관찰치가 상실된다는 것을 생각해볼 때) $\Sigma_{t=2}^{n}(\hat{\varepsilon}_t\ \hat{u}_{t-1}) = 0$이라는 조건이 부과된다는 것이다. 이러한 내용을 알기 위해 (6.43)의 항에 $\hat{u}_{t-1}$을 곱하여 표본 전체에 대한 합을 구한 다음, 조건을 부과하여 다음을 얻기로 한다.

$$\sum_{t=2}^{n}(\hat{u}_t\hat{u}_{t-1}) = \hat{\gamma}\sum_{t=2}^{n}(\hat{u}_{t-1})^2 + \sum_{t=2}^{n}(\hat{\varepsilon}_t\hat{u}_{t-1}) = \hat{\gamma}\sum_{t=2}^{n}(\hat{u}_{t-1})^2 \tag{6.44}$$

(6.44)로부터 γ의 추정량은 다음과 같게 된다.

$$\hat{\gamma} = \frac{\sum_{t=2}^{n}(\hat{u}_t\hat{u}_{t-1})}{\sum_{t=2}^{n}(\hat{u}_{t-1})^2} \tag{6.45}$$

---

* 어느 정도 보다 정형화하면, ε_t와 $\hat{u}_{t-1}$간의 의존도(dependence)는 표본크기가 제한없이 증가함에 따라 0으로 감소한다. 곧, 표본크기가 작으면 ε_t와 $\hat{u}_{t-1}$은 상관관계에 있을 것이다. 왜냐하면 $\hat{u}_{t-1}$은 $\hat{b}_0$와 $\hat{b}_1$에 의존하며, 다시 그것들은 ε_t를 포함한 모든 ε에 의존하기 때문이다. 하지만 표본의 크기가 증가함에 따라 $\hat{b}_0$와 $\hat{b}_1$이 일치추정량이기 때문에 그것들은 확률상으로는 b_0와 b_1에 수렴한다. 그러므로 극한에서 $\hat{u}_{t-1}$은 확률이 0으로 되어, u_{t-1}로부터 발산하게 된다. 방정식 (6.42)는 단지 표본이 무한한 크기이면, 확률이 1과 같을 때, 유지된다고 가정할 수도 있을 것이다. 간단히 말하면, (6.42)를 근사(또는 대표본)방정식으로 보아야 한다는 것이다.

이제 회귀모형에서 계수추정량을 얻기 위해서는 γ 대신 $\hat{\gamma}$를 이용함으로써 앞에서 서술하였던 절차를 사용할 수 있는 것이다. 이변량의 경우에 대해서는 이제 다음을 얻게 된다.

$$\hat{b}_1 = \frac{\sum_{t=2}^{n} (X_t^* - \bar{X}^*)Y_t^*}{\sum_{t=2}^{n} (X_t^* - \bar{X}^*)^2} \tag{6.46}$$

그리고

$$\hat{b}_0 = \frac{\hat{B}}{1 - \hat{\gamma}} \tag{6.47}$$

여기에서

$$X_t^* = X_t - \hat{\gamma}X_{t-1}$$
$$Y_t^* = Y_t - \hat{\gamma}Y_{t-1}$$

그리고

$$\hat{B} = \bar{Y}^* - \hat{b}_1\bar{X}^*$$

이다.

마찬가지로 아래에서의 논의에 따르면, 분산식으로서 다음을 얻게 된다.

$$\text{var}(\hat{b}_0) = \frac{1}{(1 - \hat{\gamma})^2}\text{var}(\hat{B})$$
$$\text{var}(\hat{b}_1) = \frac{\sigma_\varepsilon^2}{\sum_{t=2}^{n} (X_t^* - \bar{X}^*)^2} \tag{6.48}$$

여기에서

$$\text{var}(\hat{B}) = \frac{\sigma_\varepsilon^2 \sum_{t=2}^{n} (X_t^*)^2}{n' \sum_{t=2}^{n} (X_t^* - \bar{X}^*)^2} \qquad (6.49)$$

이다.

여기에서도 $n' = n - 1$ 이다. 간단히 말하면, 일단 $\hat{\gamma}$ 를 이용함으로써 방정식을 변형하는 경우, $\hat{\gamma}$ 가 γ 인 것처럼 $\hat{\gamma}$ 을 취급하게 될 것이며, 따라서 표준식을 모두 이용하게 될 것이다. 이는 위의 분산식에서 σ_ε^2 의 추정을 포함하게 된다. 특히, (6.36)으로부터 분명한 것은 다음과 같다. 곧, σ_{ε^2} 의 추정량은 다음과 같다는 것이다.

$$\hat{\sigma}_\varepsilon^2 = \sum_{t=2}^{N} \frac{(Y_t^* - \hat{B} - \hat{b}_1 X_t^*)^2}{(n' - 2)} \qquad (6.49\,\text{A})$$

이제 가설검정과 신뢰구간에 대해 논의하기로 하자. (6.45)로부터 ^는 추정된 교란항에 달려 있다는 사실을 알고 있다. 게다가 (6.46)으로부터 $\hat{b}_1$ 는 非線型的인 방식으로 $\hat{\gamma}$에 의존한다는 사실도 알고 있다. 이에 따라 무엇보다도 $\hat{b}_1$는 비선형적인 방식으로 추정된 교란항에 달려 있는 것이다. 더구나 추정된 교란항은 교란항의 실제치에 부분적으로 달려 있으므로〔(6.41)과 (6.25)를 볼 것〕, 이에 따라 $\hat{b}_1$는 비선형적인 방식으로 실제 교란항에 달려 있는 것이다. 이 때문에 $\hat{b}_1$는 정규분포하지 않으며, 비율 $(\hat{b}_1 - b_1) / \hat{\sigma}_{\hat{b}_1}$ 는 $(n - 2)$의 자유도를 가진 "t"변수가 아니다. 여기에서 $\hat{\sigma}_{\hat{b}_1}$ 는 $\hat{b}_1$의 추정표준편차이다. 유사한 결론이 $\hat{B}$와 $\hat{b}_0$에도 적용된다. 왜냐하면 그것들은 비선형적인 방식으로 지나칠 정도로 γ에 의존하고 있기 때문이다.

다행히, 비율 $(\hat{b}_0 - b_0) / \hat{\sigma}_{\hat{b}_0}$, $(\hat{b}_1 - b_1) / \hat{\sigma}_{\hat{b}_1}$, 과 $(\hat{B} - B) / \hat{\sigma}_B$이 標準

. . . . .
化正規變數(standard normal variables)임을 가정함으로써 가설은 어떤 "근사한(approximate)" 방법으로 검정할 수 있으며, 또는 "근사한" 신뢰구간을 설정할 수 있다. 만일 표본크기가 무한하다면, 진정 이러한 비율들은 표준화정규변수일 것이다. 표본크기가 한정된 전형적인 경우, 正規性의 假定(normality assumption)은 하나의 近似化(approximation)로 간주될 것임에 틀림없다. 그러므로 이러한 정규성의 가정에 기초한 가설검정, 또는 신뢰구간은 근사한 것으로 간주됨에 틀림없다. 유사한 맥락에서 유의해야할 것은 다음과 같다. 위의 분산식은 단지 무한한 표본의 경우에 엄밀히 유지된다는 점에서 $\hat{\gamma}$가 이용되는 경우에 대해 또한 근사한 것이라는 점이다.

위의 논의에서 따르는 사실은 b_1에 대한 근사한 95퍼센트의 신뢰구간은 $(\hat{b}_1 \pm 1.96\sigma_{\hat{b}_1})$이라는 것이다. 만일 5퍼센트의 유의수준하에서 귀무가설 $H_0: b_1 = 0$과 대립가설 $H_1: b_1 \neq 0$에 대해 검정하고자 한다면, $|\hat{b}_1/\hat{\sigma}_{\hat{b}_1}| < 1.96$일때 귀무가설 H_0는 채택될 것이고, 그 반대일 때 기각될 것이다. 그러므로 다시 소위 계속적으로 t비율이라고 부르는 것, 곧 $\hat{b}_1 / \sigma_{\hat{b}_1}$에 관심을 가지게 된다. 단지 그 차이란 정확한 臨界値(critical value)가 t분포보다는 정규분포에 관련된다는 것이다. 이러한 이유로 인해 연구자들은 독자들의 편의를 위하여 회귀결과의 비율을 흔히 계산하고, 그것을 보고하는 것이다.

게다가 지적할 것은 $\hat{\gamma}$에 기초한 추정량은 이미 불편추정량이 아니지만 一致性(consistency)이라는 바람직한 특성을 가지고 있다는 것이다.* 더

---

* 위의 분석에 포함된 논의를 수식으로 표현한 것에 대해서는 다음을 볼 것. Arthur S.Goldberger, Econometric Theory (New York : Wiley, 1964), chap.6.

구나 이러한 추정량들은 또한 有效推定量이며, 따라서 적어도 대표본인 경우 모수의 보다 우수한 일치추정량을 이용할 수 없는 것이다.

마. 다중회귀모형의 확장

위의 기법을 다중회귀의 경우로 확장시키는 것은 간단하다. 가령, 지금 k개의 설명변수가 있다는 것을 제외하고는 위의 모든 가정이 유지되고 있다고 가정하자, 그러면 모형은 다음과 같게 될 것이다.

$$\begin{aligned} Y_t &= b_0 + b_1 X_{1t} + \cdots + b_k X_{kt} + u_t, \\ u_t &= \gamma u_{t-1} + \varepsilon_t \end{aligned} \tag{6.50}$$

우선 표준적인 기법에 의해 회귀모수 $b_0, b_1, \cdots, b_k$를 추정한 다음, 교란항을 추정하면, $\hat{u}_t = Y_t - \hat{Y}_t$이 된다. 다음으로 (6.45)에 의해 γ를 추정하고, 이에 따라 종속변수를 $Y_t^* = Y_t - \hat{\gamma} Y_{t-1}$로 변형시키고, 독립변수를 $X_{it}^* = X_{it} - \hat{\gamma} X_{i(t-1)}$로 변형시킨다. 그러면 이제 다음과 같은 회귀모형을 고려하게 될 것이다.

$$Y_t^* = B + b_1 X_{1t}^* + \cdots + b_k X_{kt}^* + \varepsilon_t \tag{6.51}$$

그리고 표준적인 기법에 의해 추정량의 분산뿐만 아니라 $B, b_1, \cdots, b_k$를 추정할 것이다. b_0에 대한 결과는 바로 위에서 말한 것과 같이 B에 대한 결과로부터 도출될 것이다. 모수추정량들은 편의를 갖고 있을지라도 일치추정량이며 유효추정량일 것이다. 결국, 비율 $(\hat{b}_i - b_i) / \hat{\sigma}_{\hat{b}_i}$이 표준화정규변수라고 가정함으로써 가설을 검정하거나 신뢰구간을 설정하게 될 것이다. 위의 경우에서처럼, 그와 같은 가설검정, 신뢰구간, 또는 심지어 모수추정량에 대한 분산식은 단지 표본이 무한한 경우에서만 확실히 타당할

것이며, 따라서 유한한 크기의 표본의 경우에는 결과를 "근사한 것"으로 해석해야만 할 것이다.

한가지 덧붙인다면, 지금까지 논의한 기법이 자기상관을 수정하기 위한 유일한 기법은 아니라는 것을 지적해야 할 것이다. 널리 쓰이는 두가지 다른 기법이 소위 코크란-오컷 (Cochrane-Orcutt)방법과 힐드레드-루 (Hildreth-Lu)방법이다.* 하지만 지금까지 대상으로 하였던 모형의 경우 이러한 기법들은 대표본에서의 특성이 앞에 서술한 추정량의 특성과 동일한(identical) 추정량을 산출한다는 점에서, (곧 그 추정량들이 일치추정량이자 유효추정량이라는 점에서) 위의 기법들과 다름없는 것이다.

바. 자기상관에 대한 더빈-왓슨검정

회귀모형의 교란항이 자기상관을 내포하고 있을 때 그 관계는 위의 모형(6.32)에 의해 주어진 형태를 취한다고 가정해보자. 현시점에서 자기상관문제를 처리하기 위한 방법은 있지만, 아직 우선 자기상관을 포함한 오차항이 있는지의 여부를 검사하기 위한 수단을 설정하지는 않았다. 그 대신 위의 추정절차는 (6.32)에서 $\gamma \neq 0$이라는 가정에 기초하고 있다. 따라서 분명히 그것은 이러한 가설을 추정할 수 있기 위해서는 바람직한 것이다.

한가지 간단한 추정절차는 이러하다. 곧 귀무가설로서 $\gamma = 0$을 취한 다음 어떤 특정한 유의수준하에서 $\gamma \neq 0$에 의해 그 가설이 기각될 수 있는지를 알아보는 것이다. 제 3 장의 접근법과 유사하게 γ의 경우 추정량에 대한 신뢰구간을 설정하게 될 것이다. 만일 $\hat{\gamma}$의 신뢰구간이 0을 포함한

---

* 이러한 방법과 여타 방법에 관한 논의에 대해서는 다음을 볼 것.
S. Goldfeld and R. Quandt, Non-Linear Methods in Econometrics (Amsterdam : North Holland, 1972), pp.183-186.

다면, $\gamma = 0$이라는 귀무가설을 채택할 것이다. 하지만 만일 ($\hat{\gamma}$이 충분히 음수이거나 양수이기 때문에) 그렇지 않다면, 가설 $\gamma = 0$을 기각하게 될 것이다. 이 후자의 경우, 교란항에 자기상관이 포함되었다는 가설을 채택할 것이다. 이에 따라 이 절에서 개발한 수정된 추정절차를 이용하게 될 것이다.

다행히 자기상관에 관한 이러한 일반적인 종류의 검정은 더빈(J. Durbin)과 왓슨(G. Watson)이 개발한 것이다.* 이 검정은 항상 더빈-왓슨 d 통계량(Durbin-Watson d statistic)이라 부르는 것을 이용하는 데, 이는 추정된 교란항의 연속적인(successive) 값들의 차이의 제곱의 합에 기초하고 있다. 곧,

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \tag{6.52}$$

이다.

어느 정도 직관적으로 말하면 다음의 사실을 알 수 있다. 곧, 만일 正의 자기상관(positive autocorrelation)이 있다면, 교란항의 연속값들은 비정상적으로 서로간에 밀접해지려는 경향이 있을 것이다. 시점 t의 교란항의 양의 값을 대부분 ($t+1$)시점의 양의 값이 따르게 될 것이다. 이는 (6.52)의 분자에 있는 항이 상대적으로 작다는 것을 의미한다. 그

---

* J. Durbin and G.S. Watson, "Testing for Serial Correlation in Least-Squares Regression," Parts I and II, Biometrika 37 (1950), pp.409-428과 38(1951), pp.159-178. 또한 검정에 관한 논의는 다음의 책들을 볼 것. Arthur S. Goldberger, Econometric Theory (New York : Wiley, 1964), pp.243-244; J.Johnston, Econometric Methods, 2nd ed.(New York : McGraw-Hill, 1972), pp.249-254.

러므로 正의 자기상관은 d의 작은 값을 가져온다고 예상하게 될 것이다. 반대로 負의 自己相關은 u_t의 연속값 사이에 커다란 차이를 낳는 경향이 있다. 이러한 형태의 자기상관에 대한 신호는 항상 d의 값이 크다는 것이다.

원래의 회귀모형으로부터 $s \neq t$의 경우 $E(u_s u_t) = 0$이라는 가정이 옳다고 하자. 그러면 자기상관은 존재하지 않게 된다. 이러한 경우 또한 추정된 잔차 $\hat{u}_t$와 $\hat{u}_{t-1}$사이의 共分散(covariance)은 0에 가까울 것으로 예상하게 된다. 이것이 사실일 때 (6.52)의 d통계량에 대한 수식의 분자를 전개함으로써 다음의 사실을 알 수 있다. 곧, n이 크면 d는 2에 "가까운" 값을 가지게 된다는 것이다.*

$$
\begin{aligned}
d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} &\doteq \frac{2\sum_{t=2}^{n} \hat{u}_t^2 - 2\sum_{t=2}^{n} (\hat{u}_t \hat{u}_{t-1})}{\sum_{t=1}^{n} \hat{u}_t^2} \\
&= \frac{\left[2\sum_{t=2}^{n} \hat{u}_t^2/(n-1)\right] - \left[2\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}/(n-1)\right]}{\left[\sum_{t=1}^{n} \hat{u}_t^2/(n-1)\right]} \\
&\doteq 2
\end{aligned}
\tag{6.53}
$$

왜냐하면 $\Sigma_{t=2}^{n} (\hat{u}_t \hat{u}_{t-1}) / (n-1) \fallingdotseq 0$, $\Sigma_{t=2}^{n} \hat{u}_t^2 \fallingdotseq \Sigma_{t=1}^{n} \hat{u}_t^2$, 그리고 $\Sigma_{t=2}^{n} \hat{u}_t^2 \fallingdotseq \Sigma_{t=2}^{n} \hat{u}_{t-1}^2$이기 때문이다.

실제 어느 정도 일반화된 것으로, (6.53)로부터 다음과 같은 사실이 유도된다. 곧, n이 크면 그리고 앞에서의 自己回歸模型, $u_t = \gamma u_{t-1} + \varepsilon_t$를 가정하면 다음과 같다.

---

* 만일 표본크기 n이 무한대이면, d는 1과 동일한 확률에서 2의 값을 취하게 될 것이다.

$$d \doteq \frac{2\,\mathrm{var}(u_t) - 2\,\mathrm{cov}(u_t, u_{t-1})}{\mathrm{var}(u_t)}$$
$$= \frac{2\sigma_u^2 - 2\gamma\sigma_u^2}{\sigma_u^2} = 2(1-\gamma) \qquad (6.54)$$

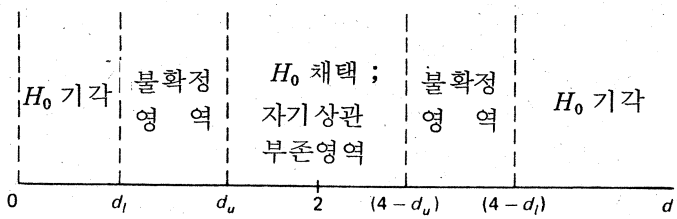
왜냐하면 $\hat{u}_t$ 은 u_t의 一致추정량이며, (6.24)로부터 $\mathrm{Cov}(u_t, u_{t-1}) = \gamma\,\sigma_u^2$ 이기 때문이다. 간단히 하면 다음과 같은 사실을 알게 된다. 곧,

$\gamma = 0$ 은 $d \doteq 2$ 를 의미한다.
$\gamma = 1$ 은 $d \doteq 0$ 를 의미한다. (6.55)
$\gamma = -1$ 은 $d \doteq 4$ 를 의미한다.

이러한 모든 것의 의미는 다음과 같다. 곧, 자기상관이 없는 귀무가설, $H_0 : \gamma = 0$과 자기상관이 있는 대립가설, $H_1 : \gamma \neq 0$을 검정하고자 한다면, d가 충분히 2에 가까운 값을 가질 때 H_0를 채택하고, 그렇지 않을 때 H_1을 채택하게 될 것이다. 그 대신 0 또는 4에 가까운 d의 값은 $H_1 : \gamma \neq 0$를 채택하도록 할 것이다.

불행히도 d통계량의 몇가지 통계적 특성때문에 문제는 이보다 훨씬 더 복잡한 것이다. 특히 자기상관이 0이라는 귀무가설의 채택과 기각에 해당하는 d의 영역이 d의 가능한 모든 값에 완전히 해당되는 것이 아니다. 따라서 H_0를 기각할 수도 채택할 수도 없는 값의 범위가 존재하게 된다. 특히, $H_0 : \gamma = 0$과 $H_1 : \gamma \neq 0$에 대한 더빈-왓슨 양측검정의 경우 〈그림 6.5〉에 나타낸 것과 같은 d의 값에 대한 다섯개의 영역이 있게 된다. 만일 d가 d_l보다 작거나 $(4 - d_l)$보다 크면 귀무가설을 기각하고 대립가설을 채택하게 되는데, 이는 자기상관이 있음을 의미한다. 그와는 달리 d가 2근방의 값이거나, d_u와 $(4 - d_u)$사이의 값이면, 자기상관이 없다는 귀무가설을 채택하게 될 것이다. 하지만 만일 d의 값이 d_l과 d_u

사이에 있거나 $(4-d_u)$와 $(4-d_l)$사이에 있으면 더빈-왓슨검정은 결론에 도달하지 못하게 된다. 이와 같이 d의 값의 경우 어떤 특정 유의수준하에서 오차항간에 자기상관이 있는지 또는 없는지의 여부를 결론내릴 수 없다는 것이다. 곧, 앞의 검정들과는 달리 더빈-왓슨검정은 몇가지 통계적 난점때문에 불확실한 영역을 포함하고 있는 것이다.



〈그림 6.5〉

단측검정절차는 바로 위의 결과로부터 나온다. 가령, 가설 $H_0: \gamma = 0$과 $H_1: \gamma > 0$에 관심을 두기로 하자. 그러면 d가 0에서부터 충분하게 "멀리" 있는 경우 H_0를 채택할 것이다. 〈그림 6.5〉에 의해 보다 정형화하면, $d > d_u$이면 H_0를 채택할 것이고, $d < d_l$이면 H_0를 기각할 것이며, $d_l < d < d_u$이면 결론없는 검정이 될 것이다. 마찬가지로 대립가설이 $\gamma < 0$인 경우, $d < 4-d_u$이면 H_0를 채택할 것이고, $d > 4-d_l$이면 H_0를 기각할 것이며, $4-d_u < d < 4-d_l$이면 결론없는 검정이 될 것이다.

d_u와 d_l의 값에 대한 표는 책말미에 통계표 4로 만들어 놓았다. 대상으로 하는 문제에 대한 d_u와 d_l의 특정치를 알기 위해서는 단측검정 또는 양측검정인 경우의 유의수준과, 표본크기 n, 그리고 회귀방정식에서의 독립변수의 수 k′를 알아야 한다.* 통계표 4로 돌아가서 예를 들어 양

* k′은 상수항을 포함하지 않는다. 가령, k′은 기울기모수 (slope parameters)의 수로 정의할 수 있다.

측검정의 경우 5%의 유의수준($\alpha = 0.05$)하에서 관찰치가 50개이고 곧 $n = 50$이고, 방정식의 독립변수가 3개이면, 곧 $k' = 3$이면, $d_l = 1.34$ 이며 $d_u = 1.59$임을 알게 된다. 이러한 수치를 구할 때 유의해야 할 것은 표에 있는 d_u와 d_l의 값은 2½퍼센트의 유의수준(곧, 각 꼬리부분에 대해 2½의 유의수준)에 해당한다는 것이다. 이런 경우, 〈그림 6.5〉의 다섯개 영역은 다음과 같게 된다.

a. $(0, d_l) = (0, 1.34)$
b. $(d_l, d_u) = (1.34, 1.59)$
c. $(d_u, 4 - d_u) = (1.59, 2.41)$
d. $(4 - d_u, 4 - d_l) = (2.41, 2.66)$
e. $(4 - d_l, 4) = (2.66, 4.00)$

따라서 방정식(6.53)을 이용하여 변수의 관찰치로부터 d의 실제치를 계산하고, 교란항사이에 자기상관이 있는지를 알아보기 위해 그 값이 다섯개의 영역중 어디에 있는지를 결정할 수 있게 될 것이다.

사. 응용

자기상관에 대한 처리를 복습하기 위해서는 실제 계산을 통한 설명이 효과적일 것이다. 그러면 다음과 같은 형태의 단순소비함수를 추정하기로 하자.

$$C_t = b_0 + b_1 Y_{dt} + u_t \tag{6.56}$$

여기에서 (앞에서와 마찬가지로) C_t는 소비지출이며, Y_{dt}는 가처분소득이다. 1951 ~ 1969년간의 미국의 총소비지출과 가처분소득에 대한 일련의 연도별 관찰치를 이용할 수 있다. 이는 〈표 6.1〉에 나와 있다. 만일 〈표 6.1〉의 자료를 이용하여 Y_d에 대해 C를 회귀한다면, 다음을 얻게

될 것이다.

$$\hat{C}_t = \underset{(1.5)}{3.29} + \underset{(162.0)}{0.906 Y_{dt}} \qquad n = 19, \quad R^2 = 0.999 \tag{6.57}$$

여기에서 t 비율의 값은 계수추정치 하단에 나와있다. 방정식은 분명히 (거의 1이나 다름없는 R^2이 나타내는 바와 같이) 소비에서의 변동량을 대부분 설명하고 있다. 게다가 해당 t 비율의 아주 큰 값이 증명하고 있듯이, $\hat{b}_1$의 분산은 극히 작은 값이다.

이제 자기상관의 증거가 있는지를 알기 위해 교란항의 추정치를 검토해 보기로 하자. 회귀분석에 대한 대부분의 컴퓨터프로그램은 더빈-왓슨 d 통계량의 값을 계산한다. 따라서 이 문제는 첫눈에 그 답을 알 수 있다. 하지만 이 기법을 더욱 잘 알기 위해 〈표 6.2〉는 실제의 계산결과를 보여주고 있다. 추정된 교란항을 계산하기 위해서는 우선 추정된 방정식(6.57)을 이용하여 매년도 소비에 대한 예측치를 계산한 다음 이것을 실제 소비액에서 제함으로써 4번째 열에 나타나 있는 교란항들이 추정치를 구하면 된다. 만일 단지 이 열만을 훑어본다면, 음수를 가진 일련의 추정된 교란항 다음에 양수를 가진 교란항들이 따른다는 것에 주목해야 할 것이다. 이는 바로 독자들을 의심케 할 것이다. 왜냐하면 그것은 교란항 사이에 正의 상관관계가 있음을 나타내는 것이며, 이에 따라 d 통계량의 경우 상당히 낮은 값을 예상하게 될 것이기 때문이다. 실제로 d 통계량의 값은 2 이하이다. 만일 더빈-왓슨 d 통계량에 대한 통계표 4로 되돌아 갈 경우 양측검정의 경우* $n = 19$, k′

---

* 가설은 결과를 검토하기 전에 정식화 하여야 한다. 곧 이 때문에 양측검정을 이용하고 있는 것이다. 이 의미는 다음과 같다. 곧, 실제로 잔차를 검토하기 전에 자기상관이 있다면 그것은 正의 자기상관일 것이라고 예상할 아무런 이유가 없다는 것이다.

$=1$, $\alpha=0.05$이면 下限(lower bound)인 d_l의 값은 1.06임을 알게 된다. d통계량은 이 하한보다 아래에 있게 되며, 자기상관이 없다는 귀무가설을 기각하고 교란항에 자기상관이 있다는 대립가설, 곧 $\gamma \neq 0$을 채택하게 되는 것이다.

자기상관을 수정하기 위해 앞에서 설명한 절차를 이용하려면, 우선 교란항사이의 관계를 추정하여야 한다. 이러한 관계가 다음의 형태라고 가정해 보자.

$$u_t = \gamma u_{t-1} + \varepsilon_t$$

여기에서 ε_t는 앞의 모든 가정을 만족하고 있다. 〈표 6.2〉에서 교란항에 대한 추정치를 구하면 (6.45)의 식을 이용하여 γ의 값을 추정할 수 있다.

$$\hat{\gamma} = \frac{\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{n} \hat{u}_{t-1}^2} = 0.48 \qquad (6.58)$$

γ의 추정치, 0.48을 이용하면 다음을 계산하게 된다. 곧,

$$C_t^* = C_t - \hat{\gamma} C_{t-1} = C_t - 0.48 C_{t-1}$$
$$Y_{dt}^* = Y_{dt} - \hat{\gamma} Y_{d(t-1)} = Y_{dt} - 0.48 Y_{d(t-1)}$$

Y_d^*에 대해 C^*을 회귀하면 다음의 사실을 알게 된다.

$$\hat{C}_t^* = \underset{(1.0)}{2.12} + \underset{(98.9)}{0.905} Y_{dt}^*, \quad n = 18, \quad R^2 = 0.998 \qquad (6.59)$$

마지막으로 상수항과 그에 관한 분산의 추정치는 다음과 같다.

$$\hat{b}_0 = \frac{\hat{B}}{1 - \hat{\gamma}} = \frac{2.12}{(1 - 0.48)} = 4.08$$

$$\widehat{\text{var}}(\hat{b}_0) = \frac{1}{(1 - \hat{\gamma})^2} \widehat{\text{var}}(\hat{B}) = \frac{1}{(1 - 0.48)^2} (4.2) = 15.5$$

따라서 자기상관을 수정한 추정방정식은 다음과 같다.*

$$\hat{C}_t = \underset{(1.0)}{4.08} + \underset{(98.9)}{0.905 Y_{dt}} \qquad (6.60)$$

이러한 경우 비록 MPC, b_1의 예측치가 실제로 통상적인 회귀방정식 (6.57)에서와 동일할지라도 그에 해당되는 t 비율은 자기상관을 수정하였을 때 (비록 여전히 극도로 클지라도) 상당히 보다 작다는 것을 유의해야 한다. 여타의 경우, 이는 모수에 대해 그 값이 0이라는 귀무가설을 기각하거나 채택하는 것 사이에 차이를 만들 수도 있다.

---

* 상수항의 경우 t 비율은 $\hat{b}_0$의 표준편차 (분산의 제곱근)를 b_0로 나눔으로써 계산하게 된다.

〈표 6.1* 〉

년도	소비지출	가처분소득
1951	206.3	226.6
1952	216.7	238.3
1953	230.0	252.6
1954	236.5	257.4
1955	254.4	275.3
1956	266.7	293.2
1957	281.4	308.5
1958	290.1	318.8
1959	311.2	337.3
1960	325.2	350.0
1961	335.2	364.4
1962	355.1	385.5
1963	375.0	404.6
1964	401.2	438.1
1965	432.8	473.2
1966	466.3	511.9
1967	492.1	546.3
1968	536.2	591.0
1969	579.6	634.2

* 단위는 경상 1억달러

출전 : Economic Report of the President (Washington, D.C.: U.S. Government Printing Office, Jan.1972), p.212.

〈표 6.2〉 더빈-왓슨 d 통계량의 계산*

연 도	실제의 소비액 (C_t)	예측된 소비액 ($\hat{C}_t$)	$\hat{u}_t = C_t - \hat{C}_t$	$\hat{u}_t^2$	$\hat{u}_{t-1}$	$\hat{u}_t - \hat{u}_{t-1}$	$(\hat{u}_t - \hat{u}_{t-1})^2$
1951	206.3	208.6	-2.3	5.2			
1952	216.7	219.2	-2.5	6.2	-2.3	-0.2	0.0
1953	230.0	232.1	-2.1	4.6	-2.5	0.4	0.1
1954	236.5	236.5	0	0	-2.1	2.1	4.6
1955	254.4	252.7	1.7	2.9	0	1.7	2.8
1956	266.7	268.9	-2.2	5.0	1.7	-3.9	15.3
1957	281.4	282.8	-1.4	1.9	-2.2	0.8	0.7
1958	290.1	292.1	-2.0	4.1	-1.4	-0.6	0.4
1959	311.2	308.9	2.3	5.4	-2.0	4.3	18.8
1960	325.2	320.4	4.8	23.1	2.3	2.5	6.2
1961	335.2	333.4	1.8	3.1	4.8	-3.0	9.3
1962	355.1	352.4	2.7	7.4	1.8	1.0	0.9
1963	375.0	369.9	5.1	26.5	2.7	2.4	5.8
1964	401.2	400.2	1.0	1.0	5.1	-4.2	17.2
1965	432.8	432.0	0.8	0.6	1.0	-0.2	0.0
1966	466.3	467.1	-0.8	0.6	0.8	-1.6	2.4
1967	492.1	498.2	-6.1	36.7	-0.8	-5.4	28.8
1968	536.2	538.7	-2.5	6.4	-6.1	3.6	13.0
1969	579.6	577.9	1.7	3.0	-2.5	4.3	18.2
				$\Sigma\hat{u}_t^2 = 143.7$			$\Sigma(\hat{u}_t - \hat{u}_{t-1})^2 = 144.5$

$$d = \frac{\Sigma(\hat{u}_t - \hat{u}_{t-1})^2}{\Sigma\hat{u}_t^2} = \frac{144.5}{143.7} = 1.01$$

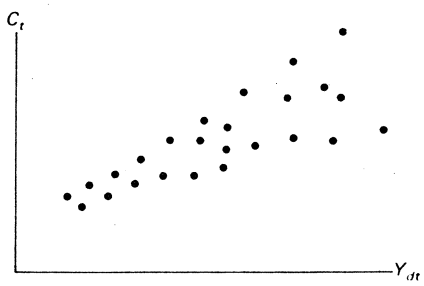
* 수치는 반올림에 의해 정확한 값이 아닐수도 있다.

3 . 異分散性

이 절에서는 교란항에 관한 한가지 가정이 위배됨으로부터 발생하는 또 다른 문제를 살펴보기로 한다. 기본적인 회귀모형에서 다음과 같이 전제하였음을 기억해야 할 것이다.

$$\text{var}(u_t) = E(u_t^2) = \sigma_u^2$$

곧, 모든 교란항은 동일한 분산 σ_u^2 을 갖고 있음을 가정하였다. 이와 같이 일정한 분산이란 조건은 均等分散(homoscedasticity)이라 한다. 하지만 모든 교란항이 동일한 분산을 가지지 않는 경우도 있을 수 있는 것이다. 이와같이 일정하지 않은 분산이란 조건은 말그대로 異分散性(heteroscedasticity)이라 한다. 한가지 예로서 상이한 가처분소득을 가진 가계들의 소비지출수준을 연구할 때, 소비에서의 분산이 소득수준과 함께 증가한다는 것을 보게 된다. 가령 보다 많은 소득을 가진 가계들은 단지 소비에 관한한 보다 신축적인 것이다. 그러한 조건이 〈그림 6.6 〉에 설명되고 있는데 여기에서 가상적인 점들의 집합의 변동량이 보다 높은 소득수준에서 증가하고 있음을 보게 된다. 이와같은 경우에 소비함수의 교란항은 이분산성을 가진다고 가정하게 될 것이다.



〈 그림 6.6 〉

가. 공식적인 모형

보다 정형화하면 다음과 같은 형태의 소비함수가 있다고 가정하자.

$$C_t = b_0 + b_1 Y_{dt} + b_2 A_t + u_t \tag{6.61}$$

여기에서

C_t = t 번째 가구의 소비지출

Y_{dt} = t 번째 가구의 가처분소득

A_t = t 번째 가구의 유동자산

u_t = 교란항

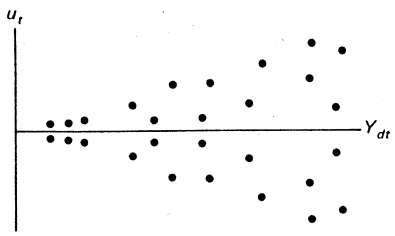
여기에서 이제 설명변수의 일련의 어떠한 값에 대해서도 u_t 는 자기상관이 없는 정규분포변수이며, 그 분산이 t 번째 가구의 소득에 비례한다고 가정한다. 곧, $\text{Var}(u_t) = Y_{dt}\,\sigma_u^2$ 이다. 그러므로 소득수준이 높으면 높을수록, 소비의 분산은 더욱 큰 것으로 관찰될 것이다.

앞의 모형에서는 교란항이 모든 설명변수에 독립적인 것으로 가정하였다. 이제는 이러한 가정은 세울 수 없다. 왜냐하면 이제는 교란항의 분산이 하나의 설명변수, 곧 Y_{dt} 의 값에 의존한다고 가정하였기 때문이다. 그러므로 교란항은 그 변수와는 독립적이지 못하다. 추가적인 가정을 세우지 않는다면, 표준적인 추정절차를 이행할 수 없는 것이다. 왜냐하면, 교란항과 설명변수 Y_{dt} 사이의 공분산의 값이 0이 아니기 때문이다. 앞으로 보게 되듯이 이분산성을 설명하고, 교란항과 각 설명변수사이의 공분산이 0이라고 가정할 수 있게 해주는 추가의 가정은 바로 다음과 같다. 곧, 어떠한 관찰치의 경우에도 설명변수 Y_d 와 A의 값이 어떠하든지 교란항의 평균이 0이라는 것이다. 보다 정형화하면, Y_{ds} 와 A_s 의 어떠한 값에 대해서도 모든 t 와 s 의 경우 $E(u_t) = 0$ 인 것이다. 방정식 (6.61)과 관련하여 이의 의미는 C_t 의 평균 C_t^m 은 여전히 $C_t^m = b_0 + b_1 Y_{dt} + b_2 A_t$ 라는 것

이다.

이 가정의 또다른 의미는 다음과 같다. 곧, 교란항은 각각의 설명변수와 상관이 없다는 것이다. $\mathrm{Cov}(u_t, Y_{dt}) = \mathrm{Cov}(u_t, A_t) = 0$. 가령, u_t가 Y_{dt}와 상관이 있다면, 그 값은 Y_{dt}가 증가함에 따라 증가하거나 또는 감소하는 것으로 예상하게 될 것이다. 하지만 어떠한 Y_{dt}의 값에 대해서는 u_t의 평균이 0이라는 가정은 위의 사실을 의미하지 않는다. 곧, Y_{dt}가 증가함에 따라 u_t의 기대치는 불변이라는 것이다. 즉, 0이다. 그 결과 u_t와 Y_{dt}는 상관이 없다는 것이다.

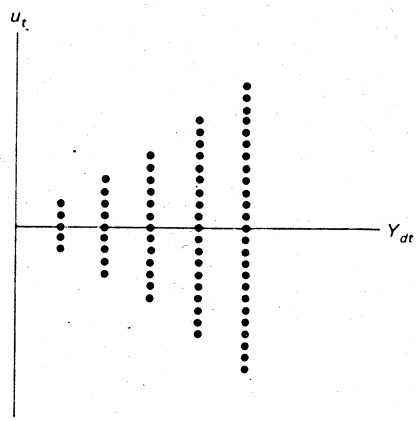
이러한 결론은 u_t의 분산이 Y_{dt}와 함께 증가한다는 가정에 비추어볼 때 혼란스러울 수도 있다. 하지만 〈그림 6.7〉을 잠깐 보면 그 혼란은 사라지게 될 것이다. 가상적인 점들의 집합에서 볼때, 분명히 u_t의 분산은 Y_{dt}에 따라 증가하고 있다. 그러나 또한 u_t는 Y_{dt}와 상관이 없다는 것도 분명하다. 왜냐하면 Y_{dt}의 어떠한 값에 대해서도 u_t의 평균치는 0이기 때문이다.



〈그림 6.7〉

위에서 유의해야 할 것은 u_t와 Y_{dt}(그리고 u_t와 A_t)가 상관이 없다는 것이다. 왜냐하면 u_t의 평균이 Y_{dt}의 어떠한 값에 대해서도(또한 A_t의 경우에도) 0이기 때문이다. 만일 단지 "Y_{dt}의 어떠한 값에 대해서도"라는 조건을 언급하지 않고서 u_t의 평균이 0이라는 가정을 세웠더라면 이 결론에 도달할 수 없었을 것이다. 한가지 예로 X_1은 그 평균이

0인, 곧 $E(X_1)=0$인 변수라고 하자. 그리고 $X_2=2X_1$이라고 하자. 그러면 X_2의 평균도 0일 것이다. 곧, $E(X_2)=2E(X_1)=0$. 하지만 X_2의 평균은 X_1의 어떠한 값에 대해서도 0인 것은 아니다. 예를 들어 $X_1=3$이면 X_2의 평균은 6인 것이다. 이러한 경우 X_1과 X_2는 완전히 상관관계에 있는 것이다.



〈그림 6.8〉

추정에 대한 문제로 돌아가기 전에 모든 독자들에게 명확하지 않을 수도 있는 마지막 한가지 논점을 알아보기로 한다. 〈그림 6.8〉(〈그림 6.7〉을 정교화한것)에서 Y_{dt}의 어떠한 값에(수직으로) 관련된 점들은 0의 평균치를 가지는 것으로 보인다. 이는 u의 평균이 Y_d의 어떠한 값에 대해서도 0이라는 가정을 반영하고 있다. 이제 유의해야 할 것은 그림에서의 모든 점들의 평균치가 0인 것처럼 보인다는 것이다. 이는 u의 평균(때로는 소위 "전체평균")이 0이라는 조건에 해당한다. 이제는 Y_d의 어떠한 값에도 관련되는 u의 평균이 0이라는 가정은 u의 전체평균이 0임을 의미한다는 것이 분명하다. 하지만 그 역은 사실이 아니다. 가령, u의 평균이 Y_d의 어떤 값에 대해서는 양수일 수도 있으며, 다른 값에 대해서는 음수일 수도 있지만, u는 여전히 0의 전체평균을 가질 수도 있는 것이다.

나. 추정량에 대한 결론

이분산성을 가진 교란항을 포함하고 있는 방정식에 대해 통상적인 추정절차를 이용하면 어떠한 일이 일어날 것인가? 어느 정도 직관적으로 말하면, 교란항의 평균이 0이고, u_t는 각 설명변수와 상관이 없기 때문에〔따라서 방정식(6.62)에서 $E(u_t Y_{dt})=0$이고 $E(u_t A_t)=0$이므로〕 모수추정량은 일치추정량이자 불편추정량일 것이다.* 요점은 $E(u_t)=0$, $E(u_t Y_{dt})=0$, $E(u_t A_t)=0$이라는 조건이 여전히 $\sum \hat{u}_t=0$, $\sum(\hat{u}_t Y_{dt})=0$, $\sum(\hat{u}_t A_t)=0$이라는 것을 의미한다는 것이다. 곧, 기본적으로 정규방정식과 다를 바 없는 것이다. 하지만, 자기상관의 경우처럼 모수추정량의 분산의 경우 그 표현이 상이하게 된다. 곧, 일단 다시 한번 이러한 분산을 추정하기 위해 통상적인 식을 이용한다면 그에 따른 가설검정과 신뢰구간은 모호하게 될 것이다.

이는 명확한 사실이다. 왜냐하면, 기본적인 회귀모형은 일정한 분산을 가정하고 있으며, 추정절차는 이러한 분산의 추정량을 산출하는 것이기 때문이다. 하지만 교란항의 분산이 이분산성인 경우 그와 동일하지는 않다. 곧, 그 자체가 하나의 변수인 것이다. 이는 다음을 의미한다. 사실상 표준적인 추정량은 교란항의 상이한 분산들의 일종의 평균을 나타내는 것이다. 그리하여 이 추정량은 실제로 의미가 거의 없으며, 따라서 예를 들면 방정식의 모수에 대한 타당한 신뢰구간(또는 t비율)을 설정할 수 없다. 자기상관의 경우처럼 보다 신뢰할만한 계수의 추정량(곧, 보다 작은 분산을

---

* 관심이 있는 독자들은 제4장 부록에 있는 논의를 복습함으로써 표준추정량이 불편추정량임을 알 수 있다. 이때, 유의해야 할 것은 도출에 필요한 가정은 설명변수들의 어떠한 값에 대해서도 교란항의 평균이 0이라는 가정이라는 것이다. 표준적인 모형에서 이 조건은 독립성가정에서 따르는 사실이다. 곧 이분산성의 모형에서는 분명히 이 가정을 세우게 된다.

가진 추정량)을 구하는 것은 가능하며, 그 분산들의 추정량도 산출해낼 수 있다. 이는 교란항의 진실된 특성들에 관한 정보를 추정절차에 포함시킴으로써 가능하다.

다. 추정절차

이 문제를 보다 자세히 검토하기 위해 방정식(6.62)에서의 소비관계로 돌아가기로 하자. 여기에서 $\mathrm{Var}(u_t)=Y_{dt}\,\sigma_u^2$ 로 한다. 이제 보일 수 있는 것은 다음과 같다. 곧, (6.61)을 $\sqrt{Y_{dt}}$ 로 나누면, 균등분산을 가진 교란항을 포함하고 있는 방정식을 얻게 된다. 이러한 계산절차를 통해 다음을 구하게 된다.

$$\frac{C_t}{\sqrt{Y_{dt}}} = b_0\left(\frac{1}{\sqrt{Y_{dt}}}\right) + b_1\sqrt{Y_{dt}} + b_2\left(\frac{A_t}{\sqrt{Y_{dt}}}\right) + u_t^* \qquad (6.62)$$

여기에서

$$u_t^* = \frac{u_t}{\sqrt{Y_{dt}}}$$

이다.

Y_{dt} 의 어떠한 값에 대해서도 u_t 의 평균치가 0이므로 다음과 같아진다.*

$$E(u_t^*) = E\left(\frac{u_t}{\sqrt{Y_{dt}}}\right) = \left(\frac{1}{\sqrt{Y_{dt}}}\right)E(u_t) = 0 \qquad (6.63)$$

이제 u_t^*의 분산으로 돌아가서 동일한 가정하에 다음을 얻게 된다.

$$\begin{aligned}\mathrm{var}(u_t^*) &= E(u_t^*)^2 = E\left(\frac{u_t^2}{Y_{dt}}\right) = \frac{1}{Y_{dt}}E(u_t^2)\\ &= \frac{1}{Y_{dt}}(Y_{dt})\sigma_u^2 = \sigma_u^2\end{aligned} \qquad (6.64)$$

---

* (6.63)의 의미는 분명하다. 만일, Y_{dt}가 900으로 주어지면 $E(u_t^*)=E(u_t)/30=0$이다. 왜냐하면 Y_{dt}의 어떠한 값에 대해서도 u_t의 평균은 0이기 때문이다.

변형된 모형 (6.62)는 교란항 u_t^*가 0의 평균과 일정한 분산을 갖고 있는 것임을 알게 된다.

분석을 진행하여 보면, u_t^*가 (6.62)의 독립변수와 상관이 없다는 사실을 어렵지 않게 알 수 있다. 가령 (6.62)의 설명변수의 주어진 어떤 값에 대해서도, 또는 마찬가지로 Y_{ds}와 A_s의 주어진 어떤 값에 대해서도 $E(u_t^*) = (1/\sqrt{Y_{dt}})\,E(u_t) = 0$이다.

위의 결론으로부터 u_t^*는 (6.61)의 설명변수와 상관이 없으며, 따라서 다음과 같은 사실이 도출된다. 곧,

(1) $$E\left[u_t^*\left(\frac{1}{\sqrt{Y_{dt}}}\right)\right] = 0$$

(2) $$E(u_t^*\sqrt{Y_{dt}}) = 0$$

(3) $$E\left[u_t^*\left(\frac{A_t}{\sqrt{Y_{dt}}}\right)\right] = 0$$

간단하게 말하면, 각 관찰치에 대해 단지 C_t, Y_{dt}와 A_t를 $\sqrt{Y_{dt}}$로 나누면 이러한 새로운 일련의 "수정된" 관찰치에 대한 해당회귀모형은 (6.61)일 것이며 이 모형은 모든 기본적인 가정을 만족할 것이다. 이에 따라 이러한 경우 회귀모수와 추정량의 분산의 불편추정량을 얻기 위해 단지 표준적인 추정절차를 이용할 수 있다.* 특히, 가정(1)—(3)을 이용하면 정규방

---

* 앞의 모형과는 달리 (6.65)의 정규방정식과 관련된 방정식 (6.62)은 상수항을 가지고 있지 않다. 그 결과로 추정량의 분산에 대한 공식에 약간의 변화가 있게 된다. 특히 이 공식의 $\hat{v}_{it}$항은 이제 여타의 설명변수에 대한 i번째 설명변수의, 상수항이 없는 회귀로부터의 잔차로 정의될 것이다. 예를 들면, (6.65)에서의 $\hat{b}_2$의 분산은 $\sigma_u^2 / \sum \hat{v}_{2t}^2$일 것이며 $\hat{v}_{2t}$은 다음과 같은 회귀로부터의 잔차이다.

$$\frac{A_t}{\sqrt{Y_{dt}}} = \gamma_1\left(\frac{1}{\sqrt{Y_{dt}}}\right) + \gamma_2(\sqrt{Y_{dt}}) + v_{2t}$$

정식은 다음과 같게 된다.*

$$\sum\left(\frac{C_t}{Y_{dt}}\right) = \hat{b}_0 \sum\left(\frac{1}{Y_{dt}}\right) + \hat{b}_1 n + \hat{b}_2 \sum\left(\frac{A_t}{Y_{dt}}\right)$$

$$\sum C_t = n\hat{b}_0 + \hat{b}_1 \sum Y_{dt} + \hat{b}_2 \sum A_t$$

$$\sum\left(\frac{C_t A_t}{Y_{dt}}\right) = \hat{b}_0 \sum\left(\frac{A_t}{Y_{dt}}\right) + \hat{b}_1 \sum A_t + \hat{b}_2 \sum\left(\frac{A_t^2}{Y_{dt}}\right) \qquad (6.65)$$

또한 방금 고려하였던 예는 이분산성의 문제에 몇가지 추가적인 시각을 제공한다. 모수추정량을 구할 때 정규적인 추정절차는 각 관찰치에 동일하게 가중치를 부여한다. 하지만 지금의 설명은 이분산성의 경우 관찰치에 대해 상이한 가중치를 이용해야 한다는 것을 말한다. 보다 분명하게 하기 위해 앞의 예를 들면 각 관찰치에 $(1/\sqrt{Y_{dt}})$만큼의 가중치를 부여해야 하는 것이다. 이의 의미는 다음과 같다. 보다 큰 분산에 해당하는 관찰치들은 보다 작은 분산에 해당하는 관찰치들보다 작은 가중치를 부여해야 한다는 것이다. 직관적으로 말하면 이는 어떤 의미를 지니고 있다. 작은 분산에 해당되는 관찰치는 진정한 회귀선에 보다 "가까운" 것으로 보인다는

---

* 위의 정규방정식은 가정(1)—(3)을 부과함으로써 얻게 된다는 사실을 강조해야 할 것이다. 이러한 경우 가정 $E(u_t^*)=0$은 사용되지 않았다. 곧 비록 추정할 모수가 단지 b_0, b_1, b_2의 세개일지라도 교란항에 관한 4개의 가정이 있다. 이처럼 가정이 너무 많다고 하는 "외견상의" 문제는 항상 이분산성의 문제를 해결하는 가운데 발생한다. 일반적으로 그 해결책은 바로 위에서 행한대로 하는 것이다. 곧, 단지 변형된 모형에서 설명변수의 분산과 관련된 가정만을 이용하는 것이다. 이는 변형된 방정식에서의 교란항의 평균이 0이라는 가정을 포기해야 한다는 것을 뜻한다. 비록 이에 대한 증명이 이 책의 범위를 벗어날지라도, 이 절차를 따르는 경우 그에 따른 b_0, b_1과 b_2의 추정량이 0의 평균의 조건과 위의 세가지 가정중 (어느)두가지를 부과하는 경우보다도 작은 분산을 가지게 된다는 사실을 보일 수가 있다.

것이다.* 곧, 추정절차시에 어떤 의미에서 평균적으로 보다 멀리 있는 점들보다는 진정한 회귀선에 보다 가까이 있다고 믿을만한 이유가 있는 점들에 대해 더욱 주목할 필요가 있다는 것이다. 회귀선의 위치를 추정할 때 보다 작은 분산에 해당하는 관찰치들은 단지 그 선으로부터의 예상편차가 보다 큰 것들보다 훨씬 더 가치가 있는 것이다.

라. 이분산성 : 추가적인 접근법

이분산성이 있을 때를 어떻게 알 것이며, 보다 일반적으로 그에 대처하기 위해 추정절차를 어떻게 변경시킬 것인가? 이것들은 답하기 쉬운 문제가 아니다. (그래서 흔히 불행히도 무시되어 버린다). 한가지 상식적인 접근법은 우선 연구중인 관계를 검토하여 오차항이 균등분산을 갖고 있지 않다고 믿을만한 이유가 있는지의 여부를 알아보는 것이다. 이분산성은 흔히 모형 자체의 정식화에 의해 나타나게 된다. 예를 들어, 달러로 측정된 이윤수준(이를 π라 하자)이 가령 그 기업의 자산(A)의 가치로 표시되는 기업의 규모에 달려있다는 가설을 고려해보자. 그러면 다음과 같은 관계를 전제로 할 것이다.

$$\pi_t = b_0 + b_1 A_t + u_t \tag{6.66}$$

여기에서 n개의 기업의 π와 A에 대한 관찰치가 있다고 한다. 이러한 경우에 교란항이 동일한 분산을 가진다고 믿기는 어려울 것이다. 확실히 달러이윤(dollar profits)의 분산은 지방의 사탕가게와 철물점보다는 제너럴모터스사와 스탠더드오일사와 같은 기업들중에서 보다 클 것이다. 왜냐하면 이윤의 절대크기에서 상당한 차이가 존재하기 때문이다. 여기서 강

* 진정한 회귀선이란 종속변수와 독립변수를 관련시키는 평균(mean) 방정식을 일컫는 말이다.

조할 점은 단지 다음과 같다. 곧 A_t의 값과 u_t의 분산사이에 正의 관계가 있는 것으로 예상할 강력한 이유가 있다는 것이다. 덧붙여 말하면, 방정식(6.66)에 여타 독립변수들이 있지만 이분산성의 문제는 정상적으로 하나의 독립변수와 교란항의 분산사이의 관계로 집중되는 것이다. 여하튼 이와 같은 경우 이분산성의 존재가능성은 매우 크다.

기본적으로 위의 것과 같은 모형과 관련된 이분산성문제를 해결하기 위해 채택할 수 있는 접근법은 두가지가 있다. 첫째, 이분산성의 존재가능성을 제거할 수 있는 방식으로 관계를 다시 정식화시킬 수 있다. 곧, 이분산성을 조잡하게 정식화한 모형의 결과로 간주할 수 있는 것이다. 보다 우수한 모형을 정립함으로써 문제를 효과적으로 해결하게 될 것이다. 앞의 예에 의하면, π와 A의 관계보다는 이윤율〔곧, $\pi^* = (\pi/A)$〕과 기업규모사이의 관계를 검토하는 것이 보다 의미가 있을런지도 모른다. 그리하여 다음과 같은 방정식을 추정하게 될 것이다.

$$\pi_t^* = b_0 + b_1 A_t + u_t \tag{6.67}$$

그리고 이윤율이 보다 큰 기업들 또는 보다 작은 기업들중에서 훨씬 더 실질적으로 변동할 것이라고 예상할 수 있는 특정한 이유없이 어느 정도 보다 큰 신뢰속에서 교란항 사이에 분산이 일정하다고 가정할 수 있는 것이다.

둘째, 이분산성의 유형을 판단하여 이 정보를 추정절차에 포함시키는 시도를 행할 수 있다. 가령, 앞에서 검토하였던 소비함수(6.61)에서 이러한 형태를 알고 있다고 가정하였다. 곧, u_t의 분산은 소득수준에 비례하였던 것이다. 회귀모형의 항들을 소득의 제곱근으로 나눔으로써 문제를 해결하였던 것이다. 하지만 많은 경우에 어떤 특정한 유형을 가정할 근거가 없을 수도 있는 것이다. 예를 들면, 이윤-자산모형(6.66)에서 교란항의 분

산은 자산수준 A_t, A_t^2 또는 A_t의 어떤 다른 함수에 비례하는가? 선험적으로 이 문제에 대한 답을 모를 수도 있는 것이다.

이분산성의 유형을 판단하는 것은 어려운 문제이다. 몇가지 제안된 해결책은 이 책의 범위를 벗어나는 것이다.* 하지만 다행히도 간단하면서도 직관적으로 끌리는 기법이 있다. 그것에 의해 몇가지 가정하에서 이분산성의 존재를 검정하는 동시에 그 유형을 추정할 수 있는 것이다. 게다가, 이러한 기법은 앞의 장들에서 개발한 소재에 의거하고 있다. 그것을 통해 이해하기가 상대적으로 쉬워지며 또한 유용하게 복습할 수 있는 것이다.

다음과 같은 관계를 추정하고자 한다고 하자.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \tag{6.68}$$

그리고 u_t의 분산이 가령 X_{2t}의 값과 구조적으로 관련될지도 모른다고 생각하기로 한다. 곧 다음과 같이 생각하게 될 것이다.

$$\sigma_{u_t}^2 = f(X_{2t}) \tag{6.69}$$

함수 $f(X_{2t})$의 특정한 형태를 알고 있다면, 단지 방정식 (6.68)을 $\sqrt{f(X_{2t})}$로 나눔으로써 이분산성의 문제를 (앞에서와 마찬가지로) 해결할 수 있는 것이다. 왜냐하면 그에 따른 교란항 $u_t^* = (u_t / \sqrt{f(X_{2t})})$의 분산은 상수, 곧 1일 것이기 때문이다.** 이에 따라 새로운 방정식

---

* 보다 수준높은 해결책에 대해서는 다음을 볼 것. J.Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), pp.214-221.

** X_{2t}의 어떤 값에 대해 다음과 같은 사실에 유의해야 한다. 곧,

$$E(u_t^{*2}) = E\left[\frac{u_t^2}{f(X_{2t})}\right] = \frac{1}{f(X_{2t})} E(u_t^2) = \frac{f(X_{2t})}{f(X_{2t})} = 1$$

은 균등분산의 가정을 만족할 것이며, 따라서 통상적인 경우와 마찬가지로 진행할 수 있는 것이다.

그런데 당면하는 문제는 함수 $f(X_{2t})$를 모른다는 것이다. 절차는 우선 $f(X_{2t})$의 근사형을 만들어 추정한 다음, (6.68)을 $f(X_{2t})$의 추정치의 제곱근으로 나누는 것이다. 그러면 수정된 회귀방정식을 이용하여, 일련의 정규방정식을 도출할 수 있게 되고, 이에 따라 모수추정량을 구할 수 있게 된다.

우선 유의해야 할 것은 (6.69)에서 이분산성의 가정이 어떤 X_{2t}의 값에 대해 다음과 같음을 의미한다는 것이다.

$$E(u_t^2) = f(X_{2t}) \tag{6.70}$$

이제 다음과 같다고 하자.

$$\varepsilon_t = u_t^2 - f(X_{2t}) \tag{6.71}$$

그러면 X_{2t}의 어떤 값에 대해 다음과 같은 사실에 주목해야 할 것이다.

$$\begin{aligned} E(\varepsilon_t) &= E(u_t^2) - f(X_{2t}) \\ &= f(X_{2t}) - f(X_{2t}) = 0 \end{aligned}$$

ε_t의 평균은 0인 것이다.

이제 (6.71)을 u_t^2에 대해 풀면, 다음을 얻게 된다.

$$u_t^2 = f(X_{2t}) + \varepsilon_t \tag{6.72}$$

(6.72)에 대한 해석은 간단하다. 변수 u_t^2은 그 평균 $f(X_{2t})$와 평균으로부터의 편차를 반영하는 변수 ε_t와의 합으로 나타난다. (6.72)가 상

당히 회귀모형과 흡사함을 유의해야 할 것이다.

우선은 u_t에 대한 관찰치가 존재한다고 가정하자. 더욱이 앞에서 관계에 대한 지식에 견주어볼 때, u_t가 (6.70)에서처럼 이분산성을 갖고 있으면, 함수 $f(X_{2t})$는 k차다항식에 의해 "정확하게" 근사화할 수 있을지도 모른다는 것을 가정하자.* 이러한 가정하에서 (6.73)을 표준적인 형태를 취하는 회귀모형으로 변환시킬 수 있다. 곧,

$$u_t^2 = a_0 + a_1 X_{2t} + \cdots + a_k X_{2t}^k + \varepsilon_t \tag{6.73}$$

이미 X_{2t}의 어떠한 값에 대해서도 $E(\varepsilon_t)=0$임을 알고 있다. 또한 위에서 이러한 0의 평균이란 조건이 ε_t가 X_{2t}와 상관이 없음을 의미한다는 것을 알고 있다. X_{2t}의 어떤 값은 (6.73)에 나타나는 X_{2t}의 모든 各乘의 값을 수반하므로 이에 따라 ε_t는 또한 이러한 X_{2t}의 모든 各乘과도 상관이 없게 된다.** 이의 의미는 다음과 같다. 곧, (6.73)을 정규방정식을 만들어내는 조건을 만족하는 교란항 ε_t를 포함하는 회귀모형으로 간주할 수 있다는 것이다.

다음과 같이 두고, 표준적인 기법을 이용함으로써 (6.73)을 추정하여 보기로 하자.

$$\sum \hat{\varepsilon}_t = 0, \sum (\hat{\varepsilon}_t X_{2t}) = 0, \ldots, \sum (\hat{\varepsilon}_t X_{2t}^k) = 0$$

그러면 몇가지 가정을 추가하면 그에 따른 추정량 $\hat{a}_0, \cdots \hat{a}_k$은 일치추정량

---

* 전형적으로 k는 $k \leq 3$으로 취하게 된다. 수식의 차원에서 다항식의 "근사화"가 완전할 때만 이 절에서의 결과들은 인정된다. 실제로 이는 사실인 듯이 보이지는 않으므로 이러한 모든 결과들을 근사한 것으로 간주해야만 할 것이다.

** 예를 들어 ε_t의 평균이 $X_{2t}=3$일때 0이면 또한 그것은 $X_{2t}^2=9$일때도 0인 것이다.

임을 보일 수 있다. 그러므로 $f(X_{2t})$의 일치추정량은 다음과 같다.

$$\widehat{f(X_{2t})} = \hat{a}_0 + \hat{a}_1 X_{2t} + \cdots + \hat{a}_k X_{2t}^k \tag{6.74}$$

이제 나머지 절차는 분명한 것이다. 특히, 원래의 모형(6.68)의 양변을 $\hat{f}_t = [\hat{f}(X_{2t})]^{1/2}$로 나누게 되면, 다음의 정규방정식으로부터 b_0, b_1과 b_2의 추정량을 얻게 될 것이다.

$$\Sigma\left(\frac{Y_t}{\hat{f}_t}\right) = \hat{b}_0 \Sigma\left(\frac{1}{\hat{f}_t}\right) + \hat{b}_1 \Sigma\left(\frac{X_{1t}}{\hat{f}_t}\right) + \hat{b}_2 \Sigma\left(\frac{X_{2t}}{\hat{f}_t}\right)$$

$$\Sigma\left(\frac{Y_t X_{1t}}{f_t}\right) = \hat{b}_0 \Sigma\left(\frac{X_{1t}}{f_t}\right) + \hat{b}_1 \Sigma\left(\frac{X_{1t}^2}{f_t}\right) + \hat{b}_2 \Sigma\left(\frac{X_{1t}X_{2t}}{f_t}\right)$$

$$\Sigma\left(\frac{Y_t X_{2t}}{f_t}\right) = \hat{b}_0 \Sigma\left(\frac{X_{2t}}{f_t}\right) + \hat{b}_1 \Sigma\left(\frac{X_{1t}X_{2t}}{f_t}\right) + \hat{b}_2 \Sigma\left(\frac{X_{2t}^2}{f_t}\right)$$

$\hat{f}_t$이 f_t의 일치추정량이므로 지금까지의 가정하에서 그에 따른 추정량 $\hat{b}_0$, $\hat{b}_1$과 $\hat{b}_2$이 일치추정량이자 유효추정량임을 보일 수 있다. 덧붙여 말하면 표본크기가 무한하면 통상의 분산식이 인정된다. 그러면 $\hat{\sigma}^2_{\hat{b}_i}$ ($i = 0, 1, 2$)을 통상의 공식에 의해 주어지는 $\hat{b}_i$의 분산의 추정량이라고 하자. 그러면 $(\hat{b}_i - b_i)/\hat{\sigma}_{\hat{b}_i}$이 표준화정규변수라고 가정함으로써 가설검정과 신뢰구간설정이 가능하게 된다. 또한 그러한 결과는 단지 표본크기가 무한한 경우에만 확실히 타당할 것이다. 따라서 실제로 결과는 근사한 것으로 생각해야 할 것이다. 자기상관이 있는 모형의 경우와 유사하게 추정량 $\hat{b}_i$이 $\hat{f}_i$에 의존함으로써 교란항에서 비선형이기 때문에 문제가 복잡한 것이다.

위의 절차에서 명백한 문제점은 다음과 같다. 실제로는 u_t^2에 대한 관찰치가 존재하지 않는다는 것이다. 이러한 절차를 수행하기 전에 먼저 u_t^2의 값을 추정하여야만 한다. 하지만 이러한 일은 쉽게 진전될 수 있다. 왜냐하면 모수의 일치추정량을 구할 수 있으며, 따라서 표준적인 기법에 의

해 불균등분산이 존재하는 최초의 모형(6.68)의 교란항의 일치추정량을 구할 수 있기 때문이다. 단지 통상적인 방식으로 (6.68)의 모수를 추정하여 $\hat{u}_t = Y - \hat{b}_0 - \hat{b}_1 X_{1t} - \hat{b}_2 X_{2t}$ 를 교란항의 추정량으로 간주하게 된다. 따라서 $\hat{u}_t^2$ 으로 대체한 u_t^2 을 이용하여 위의 절차를 수행하게 되면 위의 모든 결과들은 계속적으로 인정된다는 사실을 보일 수 있는 것이다.

마. 이분산성의 검정

이제 이분산성을 가진 교란항이 존재할 때 추정절차를 수정하기 위한 기법을 알아보기로 한다. 이 절의 결론을 도출하기 위해 회귀모형이 사실상 이분산성의 경우인지의 여부를 판단하는 논의로 돌아가기로 하자. 앞에서는 (기업의 자산에 대한 이윤의 가설적 회귀에서) 모형의 공식화 자체가 이분산성의 존재가능성을 나타내는지를 알기 위해 어떻게 회귀방정식을 검토할 수 있는지를 논의하였다. 하지만 교란항이 이분산성을 가지고 있는지를 판단하기 위한 보다 정형적인 절차를 얻는 것이 바람직할 것이다. 여기에서는 그러한 검정을 알아보기로 한다. 검정할 가설은 다음과 같다. 곧, 처음의 모형(6.68)에서의 교란항이 (6.69)에서 규정한 것처럼 이분산성을 가지고 있다는 것이다. (6.69)의 공식적인 표현은 "교란항 u_t 가 이분산성을 갖고 있으면, 그 방식은 (6.69)로 주어진 것이다"라는 것임을 이해하여야 한다.

이제 검정을 하기 위해 $f(X_{2t})$ 의 다항식근사화(polynomial approximation)를 이용하기로 한다. 특히 (6.69)에서의 규정에 대한 검정은 $H_0 : a_1 = a_2 = \cdots = a_k = 0$ 이라는 (6.73)에서의 결합가설(joint hypotheses)을 검정함으로써 수행할 수 있다. 만일 H_0 가 채택되면 u_t 의 분산은 X_{2t} 에 의존하고 있지 않다고 결론내릴 수 있으며, 따라서 u_t 가 균등분산을 갖는 것으로 간주하게 될 것이다. 하지만 만일 H_0 가 기각되면 u_t 는 이분산성을 갖고 있는 것으로 결론지을 것이고, 따라서 그 수

정을 위한 위의 추정기법을 검토하는 방향으로 진행시켜야 할 것이다.*

제 5 장 부록B에서 설명한 절차를 변형시킴으로써 위의 가설H_0에 대한 대표본검정을 설정할 수 있다. 특히 다시한번 $\hat{u}_t$를 표준적인 절차를 모형 (6.8)에 적용함으로써 얻게 되는 i번째 추정교란항이라 하자. ESS_u를 (6.73)의 (k+1)개의 변수, 곧 상수항, $X_{2t}, \cdots, X_{2t}{}^{k}$에 대해 $\hat{u}_t^2$을 회귀함으로써 얻게 되는 오차자승합(error sum of squares) 이라고 하고, $\hat{\sigma}_\varepsilon^2$을 통상적인 방식으로 얻은 교란항분산의 추정치, 곧 $ESS_u/(N-k-1)$이라고 하자. 여기에서 N은 표본크기이다. 마지막으로 ESS_R을 단지 상수항에 대해 $\hat{u}_t{}^2$을 회귀하여 얻은 오차자승합이라고 하자. 이러한 경우에 $ESS_R=\sum_{t=1}^{N}(\hat{u}_t{}^2-A)^2$이다. 여기에서 $A=\sum_{t=1}^{N}\hat{u}_t{}^2/N$이다. 그러면 $N=\infty$이면, $(ESS_R-ESS_u)/\hat{\sigma}_\varepsilon^2$은 k개의 자유도를 가진 카이제곱변수(chi-square variable)임을 보일 수 있다. 제 5 장 부록B에서의 논의와 유사하게 H_0가 틀린 것이면, 따라서 교란항이 (6.73)의 성질상 이분산성을 가지고 있으면, ESS_R은 ESS_u에 비하여 큰 영향이 있다는 것을 논증할 수 있다. 그러므로 카이제곱통계량$(ESS_R-ESS_u)/\sigma_\varepsilon^2$의 "큰"값은 H_0를 기각시키게 할 것이다. k개의 자유도를 가진 카이제곱변수를 $\chi^2{}_k$으로 표기하기로 하자. 그러면 검정에 대해 5퍼센트의 유의수준을 가정한다면, 통계량의 "큰"값은 $\chi^2{}_{k,0.05}$를 초과하는 값으로 정의될 것이다. 여기에서 $\text{Prob}(\chi^2{}_k \leq \chi^2{}_{k,0.05})=0.95$이다. 이때, $\chi^2{}_{k,0.05}$의 값은 카이제곱변수에 대한 작성표에서 구할 수 있다.

물론 실제로는 표본크기가 무한하지 않을 것이며, 따라서 검정결과는 단지 근사한 것일 뿐이다. 관심사항으로 지적할 것은 $N=\infty$이면, 이러한 카

---

* 공식적으로 말하면, 그것의 수정을 시작하기 전에 새로운 표본을 얻어야만 한다. 하지만 많은 경우에 그것은 불가능하며, 따라서 처음의 표본으로 계속 수행하게 된다.

이제곱검정은 u_t^2을 $\hat{u}_t^2$으로 교체한 뒤에 방정식(6.73)에 대해 제5장 부록B에 개관한 절차를 따름으로써 만들 수 있는 F검정과 동일한(equivalent)것임을 보일 수 있다.

바. 이분산성 : 총귀결

미국의 밀수요를 추정하는 가운데 일어나는 실제적인 경우를 이용하여 이분산성의 처리에 대한 결론을 내려보기로 한다.*

이 경우에는 이분산성의 근원이 앞에서 검토하였던 것과는 약간 다른 것이었다. 특히 연구의 진행은 다음과 같은 표준적인 류의 수요함수에서 시작하였던 것이다.

$$Q_t = b_0 + b_1 P_t^w + b_2 P_t^g + b_3 Y_t + b_4 D_t + b_5 S_{1t} + b_6 S_{2t} + b_7 S_{4t} + u_t \qquad (6.75)$$

이 방정식에서 t시점의 밀수요(Q_t)는 밀의 경상가격(P_t^W), 여타 곡물의 가격(P_t^g), 1인당 소득수준(Y_t)과 4개의 가변수에 의존하고 있다. 가변수 D_t는 연구대상기간중 일정기간동안에 국내식량가공업자가 구입해야 하는 일종의 시장판매허가증의 비용을 설명한다. 곧, D_t는 시장판매허가증이 유효한 기간동안은 1의 값을 취하며, 나머지 기간동안은 0의 값을 취하게 된다. 나머지 가변수 S_1, S_2와 S_4는 각각 연차순의 제1사분기, 제2사분기와 제4사분기동안의 계절적 가변수(seasonal dummies)이다.

이 경우에 문제의 근원은 바로 자료의 속성이다. 미국농업성은 밀과 여타 곡물의 수량과 가격에 대한 기간별 수치를 제공하고 있다. 이러한 변수들에 대한 자료는 1964년도의 제3사분기 이래로 분기별로 이용할 수

* 이 연구에 필자중의 한사람이 참가하였다. 다음을 볼 것. David F. Bradford and Harry H.Kelejian, A Quarterly Demand Model for wheat (unpublished manuscript,1976).

있다. 하지만, 그 시점 이전에는 밀의 수요변수와 관련한 수치는 단지 6개월간의 형태로만 이용가능하다.

예를 들면, 1964년도의 제3사분기 이전의 수요변수에 대해 이용가능한 관찰치는 1964년도의 전반기(처음의 두 분기)에 해당한다. 이 시점보다 이전의 이용가능한 관찰치는 1963년도 후반기(마지막 두 분기)에 해당한다. 그러므로 단일한 수요함수를 추정하기 위해 1964년도의 제3사분기이전과 이후의 자료를 이용하고자 한다면 분명히 문제가 있는 것이다. 단일한 회귀방정식을 추정하기 위해서는 6개월별 자료와 분기별 자료를 어떻게 함께 이용할 수 있을까?

보다 일반적인 공식을 이용하여, 종속변수 Y_t의 분기별 수치를 설명하는 회귀모형이 다음과 같다고 가정하자.

$$Y_t = a_0 + a_1 X_{1t} + \cdots + a_n X_{nt} + u_t \tag{6.76}$$

여기에서 시간하첨자 t는 연차순의 분기를 일컫는 것이고, 교란항 u_t는 표준적인 모든 가정을 만족하고 있다. 특히, u_t는 설명변수의 시차치, 현재치, 미래치 모두와 무관하고, 자기상관이 없으며, 평균이 0, 곧 $E(u_t) = 0$인 동시에 분산이 일정하다고, 곧 $E(u_t^2) = \sigma_u^2$이라고 가정한다.

종속변수에 대한 분기별 관찰치는 단지 시점 $t = T, T+1, T+2, \cdots, T+N$동안만 이용가능하다고 가정한다. T이전의 시점동안에는 단지 중복되지 않는 6개월별 관찰치, 곧 $(Y_{T-1} + Y_{T-2})$, $(Y_{T-3} + Y_{T-4})$, …, $(Y_{T-\mathscr{S}} + Y_{T-\mathscr{S}_1})$만이 있다. 여기에서 $\mathscr{S}$는 이용가능한 6개월별 관찰치의 수를 가리키는 홀수인 정수이다. 이에 대해서는 아래에서 논의할 것이다. 위의 예에서 $(Y_{T-1} + Y_{T-2})$는 1964년의 전반기에 해당하며, $(Y_{T-3} + Y_{T-4})$는 1963년의 후반기에 해당하는 것등이다. 모형(6.76)로 돌아가서 분기별 관찰치는 관계되는 기간동안에는 (위의 예에서는 1964년도

의 제3사분기 이전과 이후동안에)각 설명변수에 대해 이용가능한 것으로 가정한다.

만일 (6.76)이 분기별 변수Y_t를 설명하는 회귀모형이라면 그에 따라 6개월별 변수($Y_{T-j}+Y_{T-j-1}$)에 대한 회귀모형은 다음과 같을 것이다.

$$(Y_{T-j}+Y_{T-j-1}) = 2a_0 + a_1(X_{1,T-j}+X_{1,T-j-1}) + \cdots + a_n(X_{n,T-j}+X_{n,T-j-1}) + (u_{T-j}+u_{T-j-1}) \qquad (6.77)$$

$$j = 1, 3, 5, \ldots, \mathscr{S}$$

여기에서 $\mathscr{S}$는 홀수인 정수로 그 값은 (아래에서 논의하는 것과 같이) 이용가능한 6개월별 관찰치의 수를 결정한다. (6.77)에서의 모형은 종속변수와 관련된 분기, 곧 $T-j$와 $T-j-1$기간에 대해 (6.76)의 우변을 단지 합산함으로써 얻게 된다. 지금까지의 가정하에서 6개월별 모형(6.77)에서의 교란항($u_{T-j}+u_{T-j-1}$)은 평균이 0이며, 자기상관이 없다. 왜냐하면, 분기별 자료들은 중복되지 않으며 설명변수의 시차치, 현재치와 미래치 모두에 대해 독립적이기 때문이다. 그리고 또한 일정한 분산을 가지고 있다. 곧,

$$E(u_{T-j}+u_{T-j-1})^2 = 2\sigma_u^2 \qquad (6.78)$$

그러므로 6개월별 관찰치와 관련되는 (6.77)의 모형은 표준적인 모든 가정을 만족하고 있다. 하지만 또한 (6.76)의 분기별 모형도 표준적인 모든 가정을 만족하는 한편, 동일한 미지의 모수를 포함하고 있다. 이제 이 두 모형을 표준적인 모든 가정을 마찬가지로 만족하면서 모수추정에 분기별 자료와 6개월별 자료 둘다 사용할 수 있는 하나의 모형으로 결합할 수 있음을 논증하기로 한다.

우선 종속변수에 대한 이용가능한 관찰치를 연차순으로 다음과 같이 배

열할 수 있음에 유의해야 할 것이다. 곧,

$$(Y_{T-\mathscr{S}}+Y_{T-\mathscr{S}-1}),(Y_{T-\mathscr{S}+2}+Y_{T-\mathscr{S}+1}),\ldots,(Y_{T-5}+Y_{T-6}),$$
$$(Y_{T-3}+Y_{T-4}),(Y_{T-1}+Y_{T-2}),Y_T,Y_{T+1},\ldots,Y_{T+N}. \qquad (6.79)$$

유의할 점은 $\mathscr{S}=0$이면 세개의 6개월별 관찰치, 곧 $(Y_{T-5}+Y_{T-6})$, $(Y_{T-3}+Y_{T-4})$와 $(Y_{T-1}+Y_{T-2})$가 있게 된다는 것이다. 또한 $(5+1)/2=3$임을 유의해야 할 것이다. 또다른 한가지 예로서 $\mathscr{S}=3$이면 두개의 6개월별 관찰치가 있게 되는 것이 분명하다. 또한 $(3+1)/2=2$임에 유의해야 할 것이다. 이러한 설명을 통해서 볼 때, 분명히 일반적으로 어떠한 홀수인 정수$\mathscr{S}$에 대해 6개월별 관찰치의 수는 $(\mathscr{S}+1)/2$이다. 그러므로 (6.79)에 나열된 관찰치의 총수는 $\{[(\mathscr{S}+1)/2]+[N+1]\}$이다.

이제 (6.79)에 있는 $\{[(\mathscr{S}+1)/2]+[N+1]\}$개의 관찰치를 y_t ($t=1, 2, \cdots, \{[(\mathscr{S}+1)/2]+[N+1]\}$)로 표기하기로 하자. 곧, y_1은 (6.79)에서 최초의 관찰치, 곧 $(Y_{T-\mathscr{S}}+Y_{T-\mathscr{S}-1})$을, y_2는 다음의 관찰치를 가리킨다. 이와 비슷하게 (6.77)에서는 각 설명변수에 대해 $[(\mathscr{S}+1)/2]$개의 관찰치가 있음을 알고 있다.* 덧붙여 말하면, 종속변수에 대한 분기별 자료가 존재하는 각 기간 동안에는, 곧 시점 $T, T+1, \cdots, T+N$ 동안에는 각 설명변수에 대한 분기별 관찰치가 이용가능하다고 가정하였다. X_{jt} ($t=1, \cdots, \{[(\mathscr{S}+1)/2]+[N+1]\}$)을 (6.79)에서와 꼭 같은 방식으로 연차순에 따라 배열된 설명변수 X_{jt}에 대한 6개월별 관찰치와 분기별 관찰치, 곧 $\{[(\mathscr{S}+1)/2$

---

* 설명변수에 대한 분기별 관찰치가 이용가능하다고 가정하였기 때문에 이러한 변수들의 6개월별 형태에 대한 관찰치를 쉽게 구성할 수 있게 된다.

〕+〔$N+1$〕}개의 관찰치라고 표기하자. 그러면 (6.76)과 (6.77)에서의 결론은 다음을 의미한다. 곧, y_t는 다음과 같이 $x_{1t}, x_{2t}, \cdots, x_{nt}$와 관련되어 있다는 것이다.

$$y_t = a_0 x_{0t} + a_1 x_{1t} + \cdots + a_n x_{nt} + v_t$$
$$t = 1, 2, \ldots, \left(\frac{\mathscr{S}+1}{2}\right) + (N+1) \tag{6.80}$$

여기에서 $t \leq$〔($\mathscr{S}+1$)／2〕일 때 t가 6개월별 관찰치에 해당되면 $x_{0t}=2$이며, 그렇지 않으면 $x_{0t}=1$이다. 그리고 v_t는 이분산성이 존재한다는 가정을 제외하고는 표준적인 모든 가정을 만족하는 교란항이다. 특히, $t \leq$〔($\mathscr{S}+1$)／2〕이면 $E(v_t^2)=2\sigma_u^2$이며, 그렇지 않으면 $E(v_t^2)=\sigma_u^2$이다.

(6.80)의 모형은 표준적인 모든 가정을 만족하는 것으로 전환할 수 있다. 특히 $t \leq$〔($\mathscr{S}+1$)／2〕이면 $d_t=\sqrt{2}$로 하고, 그렇지 않으면 $d_t=1$로 하자. 그러면 앞 절에서 개괄한 절차에 따라서 (6.80)의 양변을 d_t로 나눔으로써 이분산성문제를 제거할 수 있게 된다. 그에 따른 모형은 다음과 같다.

$$\left(\frac{y_t}{d_t}\right) = a_0\left(\frac{x_{0t}}{d_t}\right) + a_1\left(\frac{x_{1t}}{d_t}\right) + \cdots + a_n\left(\frac{x_{nt}}{d_t}\right) + w_t \tag{6.81}$$
$$t = 1, 2, \ldots, \left(\frac{\mathscr{S}+1}{2}\right) + (N+1)$$

여기에서 $w_t = v_t / d_t$이다. 분명히 $t=1, 2, \cdots,$〔($\mathscr{S}+1$)／2〕+〔$N+1$〕}의 모든 경우에 대해 $E(w_t^2)=\sigma_u^2$이다.

(6.81)의 모형은 표준적인 모든 가정을 만족하며, 종속변수에 대한 이용가능한 6개월별 관찰치및 분기별 관찰치와 모두 관련되어 있다. 이 모형은 표준적인 절차에 의해 추정할 수 있는 것이다. 특히 정규방정식은

다음의 조건에 의해 결정될 것이다.

$$\sum \hat{w}_t\left(\frac{x_{0t}}{d_t}\right) = 0, \ldots, \sum \hat{w}_t\left(\frac{x_{nt}}{d_t}\right) = 0 \tag{6.82}$$

여기에서 각각의 합계는 $t = 1, \cdots, \{[(\varphi + 1)/2] + [N+1]\}$로부터 나온 것이고, $\hat{w}_t = (y_t/d_t) - \hat{a}_o(x_{ot}/d_t) - \cdots - \hat{a}_n(x_{nt}/d_t)$이며, $\hat{a}_0, \cdots, \hat{a}_n$은 $a_o, \cdots, a_n$의 추정량이다.

유의해야 할 것은 일단 추정량 $\hat{a}_o, \cdots, \hat{a}_n$을 구하게 되면 y_t의 값은 다음과 같은 모형에 의해 설명할 수 있다는 점이다. 곧,

$$\left(\frac{\hat{y}_t}{d_t}\right) = \hat{a}_0\left(\frac{x_{0t}}{d_t}\right) + \cdots + \hat{a}_n\left(\frac{x_{nt}}{d_t}\right) \tag{6.83}$$

이다.

또는 다음과 같이 d_t를 소거하면,

$$\hat{y}_t = \hat{a}_0 x_{0t} + \cdots + \hat{a}_n x_{nt} \tag{6.84}$$

이 된다.

이는 분기별 수치가 다음과 같이 설명될 것임을 뜻한다.

$$\hat{Y}_t = \hat{a}_0 + \hat{a}_1 X_{1t} + \cdots + \hat{a}_n X_{nt} \qquad t = T, T+1, \ldots, T+N \tag{6.85}$$

6개월별 수치는 다음과 같이 설명될 것이다.

$$(Y_{T-j} + Y_{T-j-1}) = 2\hat{a}_0 + \hat{a}_1(X_{1,T-j} + X_{1,T-j-1}) + \cdots + \hat{a}_n(X_{n,T-j} + X_{n,T-j-1}) \qquad j = 1, 3, 5, \ldots, \mathscr{S}$$

4. 변수선정의 문제

지금까지 회귀모형의 변수들은 어느 정도 주어진 것이며, 문제란 단지 모형의 추정, 가설검정, 자기상관의 수정등인 것으로 가정하였다. 하지만

실제로 모형에 포함시킬 변수를 선정하여야만 한다. 실험자는 전형적으로 종속변수의 결정과 관련된 이론을 참고하여, 그 이론을 가장 잘 설명하는 독립변수를 지정하려고 한다. 이렇게 하는 중에 두 종류의 오류가 발생할 수 있다. 첫째, 모형에서 중요한 독립변수를 포함시키지 않을 수도 있는 것이다. 곧, 종속변수를 결정하는 중요한 요인을 단지 간과하게 될지도 모른다. 둘째, 어떤 특정한 요인이 실제로는 그렇지 않지만 종속변수를 결정하는데 중요한 것으로 가정할 수도 있다. 그렇게 할 경우에 모형에는 불필요한 변수가 포함되는 결과가 나올 것이다. 이 절에서는 이러한 유형의 오류에서 나오는 결과들을 생각하기로 한다.

가. 생략된 변수

우선 가설에 따른 관계 (hypothesized relationship) 로부터 하나의 설명변수를 빠뜨린 경우를 생각해 보기로 하자. 예를 들어 실제의 (하지만 알지 못하는) 관계가 다음과 같다고 하자.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \tag{6.87}$$

여기에서 u_t 는 표준적인 모든 가정을 만족하는 교란항이다. 하지만 X_2 를 간과한 채, 그대신에 다음의 방정식을 고려하게 된다. 곧,

$$Y_t = b_0 + b_1 X_{1t} + r_t \tag{6.88}$$

여기에서 r_t 를 교란항으로 간주한다. 이제 표준적인 기법에 의해 (6.88) 을 추정하면, 그에 따른 b_0 와 b_1 의 추정량은 일반적으로 편의를 가진 동시에 일치성이 없는 추정량임을 보이게 될 것이다. 그 이유는 다음과 같다. 곧 모형에 X_2 를 포함시키지 못함으로써 일반적으로 교란항, 이 경우에는 r_t 에 관한 주요가정을 위반하게 될 것이기 때문이다.

보다 명확하게 하면 (6.87) 을 (6.88) 과 비교해볼 때 지금 고려

되고 있는 모형에서의 교란항이 부분적으로 X_{2t}에 의존하고 있음을 알게 된다.

$$r_t = b_2 X_{2t} + u_t \tag{6.89}$$

기대치를 구하면 다음을 알게 된다.

$$E(r_t) = E(b_2 X_{2t}) + E(u_t) = b_2 \mu_2 + 0 = b_2 \mu_2 \tag{6.90}$$

여기에서 μ_2는 X_{2t}의 평균이다. 분명히 교란항 r_t의 기대치는 Y_t가 X_{2t}에 의존하고 있지 않음을 의미하는 $b_2 = 0$일 때를 제외하면 일반적으로 0이 아닐 것이다.* 여기에서 알 수 있는 것은 Y가 X_2의 함수이지만 회귀방정식에 X_2가 포함되지 않으면 전형적으로 방정식의 교란항이 0이 아닌 평균을 가지게 된다는 것이다. 덧붙여 말하면, X_{2t}가 X_{1t}와 상관이 있을 때 r_t는 X_{1t}와 상관이 있음이 분명하다. 그 결과로 Cov (r_t, X_{1t})는 일반적으로 0이 아닐 것이다. 적어도 직관적으로 보면 그에 따라 그러한 조건하에서 b_0와 b_1의 추정량은 편의가 있는 동시에 일치성이 없는 추정량일 것이다. 가령 $E(r_t) \neq 0$이지만 $\sum \hat{r}_t = 0$으로 놓음으로써 첫번째 정규방정식을 얻게 되면, 절차는 단지 본질적으로 일치하지 않는다.

보다 직관적인 차원에서 이러한 편의(bias)의 특성을 고려해 보는것도 유용할 것이다. 어떤 소비함수를 추정하기로 하고, 그 실제의 관계가 다음과 같다고 가정하자.

$$C_t = b_0 + b_1 Y_{dt} + b_2 A_t + u_t \tag{6.91}$$

---

* $E(r_t) = 0$인 한가지 또다른 경우는 $\mu_2 = 0$인 우연적이고 비정상적인 경우인 것이다.

여기에서

C_t = t 시점의 소비지출

Y_{dt} = t 시점의 가처분소득

A_t = t 시점의 유동자산스톡

u_t = 교란항

이다.

A_t는 C_t에 대해 正의 영향을 미칠 것으로 예상하게 된다. 따라서 $b_2 > 0$이다. 또한 A_t와 Y_t는 正의 相關을 갖고 있다고 가정하자. 곧,가처분소득이 증가함에 따라 전형적으로 유동자산의 가치도 상승할 것이다(그 역도 사실일 것이다). 하지만 실제로 추정하는 방정식에서 A_t가 빠져있다고 가정하자. 곧,

$$C_t = b_0 + b_1 Y_{dt} + r_t \tag{6.92}$$

(위에서와 마찬가지로)이러한 경우에 추정하기로 선정한 방정식에서의 교란항의 기대치는 일반적으로 0이 아닐 것이다. 곧,

$$E(r_t) = E(b_2 A_t) + E(u_t) = b_2\mu_A + 0 = b_2\mu_A \gtrless 0 \tag{6.93}$$

이다.

여기에서 μ_A는 A_t의 평균이다. 마찬가지로 r_t는 Y_t와 상관이 있는 것으로 예상하게 된다. 그러므로 b_0와 b_1(여기에서 후자는 MPC이다.)의 추정량은 편의가 있는 것으로 계산될 것이다. 더구나 A_t가 C_t에 대해 正의 영향을 미치므로 (곧 $b_2 > 0$이므로), 또한 A_t와 Y_{dt}의 正의 상관때문에, 전형적으로 MPC의 과대추정치(overestimate)를 얻게 될것이다. 곧, $E(\hat{b}_1) > b_1$일 것이다. 직관적으로 말하면, 이 이유는 방정식(6.92)에서의 Y_{dt}가 그 자신과 A_t에 대한 代理變數(proxy variable)로서 기능하기 때문이다. 곧, 어떤 의미에서는 그로 인하여 C_t에 대해 Y_{dt}와 A_t가 正의 영향을 미치는 것이다. 요점은 이러하다. 곧, Y_{dt}

가 증가함에 따라 전형적으로 A_t도 증가하지만 추정된 방정식에서 A_t가 없음으로 해서 C_t에 대한 A_t의 正의 영향은 Y_{dt}에 속한다는 것이다. 그 결과로 C_t에 대한 Y_{dt}의 측정된 영향은 과장된다. 그 대신에 A_t가 C_t와 負의 相關關係에 있지만, 여전히 Y_t와는 正의 關係에 있을 경우 (A_t가 없을 때) b_1의 추정량은 전형적으로 하향편의를 가질 것이다. 곧, C_t에 대한 A_t의 負의 영향은 어느 정도 $\hat{b}_1$에 반영될 것이다. 마지막으로 지적해야 할 것은 Y_{dt}와 A_t가 負의 상관관계에 있으면 이런 경우의 편의의 방향은 일반적으로 반대라는 것이다. 중요한 변수가 방정식으로부터 생략되면 가장 중요한 점은 편의의 방향을 판단하는 것이라기 보다는 추정량이 편의를 갖고 있으며 일치성이 없다는 것을 깨닫는 것이다. 사실상 독립변수가 많은 대부분의 회귀모형에서도 항상 편의의 방향을 판단할 수 있는 것은 아니다.

나. 너무 많은 변수

다음으로 설명변수가 너무 많은 경우를 생각해 보기로 하자. 예를 들어 실제의 관계가 다음과 같다고 가정한다.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \tag{6.94}$$

여기에서 교란항은 설명변수들의 모든 값에 독립적이며, 여타의 표준적인 모든 가정을 만족하고 있다고 하자. 이 방정식이 말하고 있는 것은 다음과 같다. 곧, 모형에 X_{3t}를 포함시킬 필요가 없다는 것이다. 왜냐하면 Y_t가 그것에 의존하지 않기 때문이다. 곧,

$$b_3 = 0$$

하지만 이는 모형을 다음과 같이 쓰지 못하게 하는 것이 아님에 유의해야 할 것이다. 곧,

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + b_3 X_{3t} + u_t \qquad (6.95)$$

여기에서 $b_3 = 0$이다. 제5장 부록B에서의 논의로부터 다시 생각해 내어야 할 것은 다음과 같다. 곧, 0이 상수이기 때문에 그 값이 0인 회귀계수를 포함시키는 것은 기본적으로 어떤 잘못도 없다.

사실상 b_3이 0임을 모른다고 가정하자. 그 결과로 (6.95)를 모형으로 간주하여, 보통의 방식으로 b_0, b_1, b_2와 b_3의 추정량을 산출하기로 한다. 그런데 고려해야 할 문제는 표준적인 추정기법이 계속적으로 회귀계수와 추정량의 분산에 대한 불편추정량을 제공해 줄 것인가 하는 것이다. 다행히도 (6.95)에서의 교란항에 관한 기본적인 가정중 어느 것도 위배되지 않으면 그 대답은 긍정이다. 예를 들어 u_t가 X_{3t}와 무관하다고 가정하자. 그러면 (6.95)에서 교란항이 0의 평균을 가지고 모든 설명변수와는 무관하며, 따라서 그것과 상관이 없으며 다른 모든 가정을 만족한다고 하자. 그에 따라 b_0, b_1, b_2와 b_3의 추정량은 불편추정량일 것이다.* 어느 정도 직관적으로 말하면, Y_t에 대한 X_{3t}의 영향이 0임을 모른다면 자료는 그러한 점에서 b_3의 추정량이 불편추정량이라 할 것이다. 곧, $E(\hat{b}_3) = b_3 = 0$이다. 그 대신 5퍼센트의 유의수준하에서 귀무가설 $b_3 = 0$을 검정한다면, 그것을 채택할 가능성은 95퍼센트가 될 것이다.

다. 몇가지 추가설명

위의 결과들은 다음을 의미할 수도 있다. 곧, 추정할 방정식에서 아주 모호하지만 중요한 설명변수 전체를 포함시키면 아무것도 상실하지 않는다

---

* 이것은 책에서 고려되는 통례적인 경우이다. 만일 X_{3t}가 단지 교란항과 상관이 없으면 몇가지 추가적인 가정하에서 추정량이 일치추정량임을 보일 수 있다. 마지막으로 X_{3t}가 교란항과 상관이 있으면 추정량은 편의를 가지는 동시에 일치추정량이 아님은 분명하다.

는 것이다. 하지만 이는 그렇지 않다. 우선 유의해야 할 것은 이러하다. 곧, 설명변수가 교란항과 무관할지라도, 따라서 위반되는 가정이 전혀 없다 할지라도 어떤 특정한 모수가 0인지의 여부에 대한 귀무가설을 기각할 가능성이 계속 존재한다는 것이다. 다른 말로 하면 제1종의 과오(Type 1 error)가 발생할 수도 있다는 것이다.

둘째, 그리고 아마도 보다 중요하겠지만, 표본크기가 주어졌을 때 모수추정량의 분산은 일반적으로 설명변수의 수에 따라 증가한다. 곧, $b_3 = 0$임을 알고 있으며, 따라서 (6.94)로부터 가령, b_2의 추정량을 구하면 일반적으로 그 추정량은 (6.95)로부터 얻은 b_2의 추정량보다 더욱 작은 분산을 갖게 된다. 어떤 의미에서 (6.95)를 이용하면, 이용할 자료가 더욱 많이 필요하게 된다. 왜냐하면 3개의 모수가 아닌 4개의 모수를 추정해야 하기 때문이다. 그 대신에 (6.95)와는 달리 모형(6.94)는 $b_3 = 0$이라는 정보를 포함시키며, 추정량의 분산은 이러한 추가적인 정보를 반영하고 있는 것이다. 그러므로 독립변수의 수가 증가함에 따른 비용이 존재하게 된다. 보다 정확도가 떨어지는 계수추정량의 결과는 보다 큰 분산인 것이다. 결국 신뢰구간의 범위는 보다 넓어질 것이며, 사실상 종속변수에 구조적인 영향을 미치는 변수를 통계학적으로 무의미한 것으로 기각시킬지도 모르는 것이다.

이에 따라 딜레머에 빠지게 된다. 무언가를 배제시키면 편의가 있는 결과를 얻게 되지만, 너무 많이 포함시키면 추정량의 분산은 증가하는 것이다. 이 의미는 독립변수의 선정에 어느 정도 사려분별이 필요하다는 것이다. 예를 들면 (연구자가 도심에서 소매가 판매세율에 미치는 영향을 판단하려고 하는 제4장의 실례에서와 마찬가지로) 단지 이변수사이의 관계에만 관심을 기울일지라도 종속변수의 값을 결정하는 여타 변수들을 모형에 포함시키는 것이 매우 중요하다. 일반적으로 여타 설명변수들을 포함시

키지 않으면 계수의 추정량에 편의가 있게 된다. 그 실례로 돌아가서 도시에서 소매를 결정하는 여타 변수들을 방정식에 포함시키지 않으면, 도심의 판매량에 대한 보다 높은 세율의 추정효과는 편의를 가지게 될 것이다. 다른 한편, 자료가 있는 모든 변수를 설명변수로 포함시킬 아무런 이유도 없다. 선험적인(a priori)논거에서 종속변수에 영향을 줄 것으로 믿는 변수들을 선정해야 하는 것이다. 경제학자들이 흔히 사용하는 한가지 접근법은 다음과 같다. 곧, 중요한 것으로 생각하는 변수를 "사용" 하여 0과 다를 바 없는 것으로 판정되면 방정식에서 제외하는 것이다. 가령 다음과 같은 방정식을 추정하기로 하자.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_t \tag{6.96}$$

여기에서 X_2가 Y에 영향을 주고 있다고 믿을 이유가 모호하기는하지만 몇가지 있다고 하자. 그러면 $b_2=0$이라는 귀무가설을 기각시킬 수 없음을 알고 있다. b_1의 추정량의 분산을 감소시키는 절차는 방정식에서 X_2를 제거하여 다음과 같이 수정된 방정식을 추정하는 것이 직관적으로 호소력이 있다.

$$Y_t = b_0 + b_1 X_{1t} + u_t \tag{6.97}$$

이처럼 널리 이용되는 기법은 외관상 중요하지 않은 변수를 제거시킴으로써 어떤 실용적인 가치를 지니는 반면, 그것을 정상적으로 이용하는 방식이 결코 타당한 것만은 아님을 알아야 한다. 특히, 연구자가 방정식(6.96)으로부터 X_2를 뺀 뒤에는 일련의 새로운 자료를 이용하여 방정식(6.97)의 b_0와 b_1을 추정하여야 하는 것이다. 만일 (6.97)을 추정하기 위해 처음의 자료를 "다시 이용하는"경우에는 순환성을 띤 요소가 포함되어 그에 따른 결과는 편의를 가진 것으로 나타날 수가 있는 것이다. 이러한 순환성을 띤 요소는 모형(6.97)이 자료에 대해 수행된 검

정에 기초하여 정식화되었기 때문에 일어나는 것이다. 경제학자들은 전형적으로 일련의 새로운 자료를 모두 이용할 수 있는 것은 아니며, 따라서 이 때문에 (불행히도) 이 문제를 무시하게 된다.

문 제

1. 다음의 모형을 생각하기로 한다.

$$Y_t = a + bX_t + u_t$$

관찰치는 다음과 같다.

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	2	2	2	1	3	5	6	6	10	10	10	12	15	10	11

제 1 종의 과오 $\alpha = 0.05$ 를 이용하여 자기상관을 검정하라. 표준적인 조건을 가정하라.

2. 다음과 같은 선형생산모형을 생각하기로 한다.

$$Q_t = a + bL_t + cK_t + u_t$$

여기에서 t 시점의 전체경제의 총생산량 Q_t 는 총노동투입량과 총자본투입량, 곧, L_t 와 K_t 와 관련되어 있다.

a. 위의 모형에서 u_t 는 공분산성을 갖고 있다고 말할 수 있는가? 논평해보라.

b. 자본을 무시하기로 하자. 그러면 모형 $Q_t = a + bL_t + u_t$ 가 추정될 것이다. b 의 추정량이 상향편의를 가질 것으로 추측하게 되는 이유를 설명하라.

3. t 시점의 각 개인 i 에 대한 소비함수가 다음과 같다고 하자.

$$C_{it} = a + b_1 Y_{it} + b_2 Y_{it}^2 + u_{it} \qquad i = 1, \ldots, N \tag{1}$$

C_t 와 Y_t 를 t 시점에서의 평균소비지출과 평균소득이라고 하자。곧,

$$C_t = \sum_{i=1}^{N} \frac{C_{it}}{N}, \qquad Y_t = \sum_{i=1}^{N} \frac{Y_{it}}{N} \tag{2}$$

그러면 개인의 소비함수 (1)에 의해 다음과 같은 거시적 시계열회귀모형 (macro time-series regression model)을 생각해 볼 수 있다.

$$C_t = a + b_1 Y_t + b_2 Y_t^2 + u_t, \qquad t = 1, \ldots, T \tag{3}$$

a。이 거시적 시계열모형이 表記錯誤 (specification error)를 포함하고 있음을 증명하라。이 착오는 미시적 관계를 합계함으로써 발생하기 때문에 總計偏倚 (aggregation bias)라 한다.

b。$b_2 = 0$ 임을 가정하면 총계편의는 존재하지 않게 된다. $T = 2$ 년도에 $N = 3$ 의 개인에 대한 관찰치가 있다고 하자. 방정식(1)에 해당하는 관찰치행렬을 기초형태로 쓰라。

c。위의 Y_{it}^2 과 같이 전제한 횡단면 (cross-sectional)회귀모형이 비선형성 (nonlinearity)을 포함한다면 일반적으로 왜 총계편의가 존재하는가?

4。다음의 화폐수요방정식을 생각하자。

$$M_{dt} = b_0 + b_1 i_t + b_2 i_{(t-1)} + b_3(\Delta i_t) + u_t$$

여기에서 M_{dt} 는 화폐수요이고, i_t 는 이자율이며, $\triangle i_t = i_t - i_{t-1}$ 이다. i_{t-1} 을 습관 또는 관성 (inertia)의 영향을 반영하며, $\triangle i_t$ 는 이자율의 최근의 변화로부터 발생하는 "기대효과 (expectation effect)"를 반영하고 있다고 하자。

a. 추가적인 정보없이는 b_1, b_2와 b_3의 추정이 불가능함을 보여라.

b. M_{dt}의 값을 예측하기 위해 방정식을 추정과 검정이 가능한 형태로 변형하라.

5. 다음의 생산함수를 가정하자.

$$Q_t = AL_{1t}^{\alpha_1}L_{2t}^{\alpha_2}K_t^{\alpha_3}e^{u_t}$$

여기에서 L_1=생산부문노동자의 수, L_2=비생산부문노동자의 수, K=자본스톡, u는 교란항이며 아래첨자 t는 시점을 말한다. 기업은 항상 10,000명의 노동자를 고용하고 있다고 하자. 그러면 $L_{1t}+L_{2t}=$ 10,000이다. α_1, α_2와 α_3를 추정할 수 있는가? 이를 설명하라.

6. 다음의 모형을 생각하기로 한다.

$$Y_t = a + bX_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

다음과 같은 통상의 공식을 써서 $\hat{b}$을 계산하기로 하자. 곧,

$$\hat{b} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X - \bar{X})^2}$$

자기상관때문에 다음과 같은 $\hat{b}$의 분산에 대한 통상의 공식이 더 이상 유지될 수 없음을 보여라.

$$\sigma_{\hat{b}}^2 = \frac{\sigma_u^2}{\sum (X_t - \bar{X})^2}$$

7. 소비함수가 다음의 형태라고 하자.

$$C_t = b_0 + b_1Y_t + b_2A_t + u_t$$

여기에서 Y_t와 A_t의 어떠한 값에 대해서도 $E(u_t)=0$이며 $\mathrm{Var}(u_t) = Y_t^2\sigma_u^2$이라고 가정한다. 교란항이 균등분산을 가지는 형태로 위의 방정식을 변환하라. 그리하여 정규방정식을 도출하라.

8. 다음의 모형을 생각하기로 한다.

$$I_t = a_0 + a_1 \Delta Y_t + a_2 r_t + \varepsilon_t$$
$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t$$

여기에서 I=투자, $\triangle Y$=소득의 변화분, 그리고 r_t=이자율이다. u_t가 모든 설명변수와는 무관하고, 자기상관을 갖고 있지 않으며, 0의 평균과 일정한 분산을 갖고 있다고 하자. ρ_1과 ρ_2을 모를 때 자기상관을 설명하는 a_0, a_1과 a_2의 추정치를 얻기 위한 절차를 약술하라.

제 7 장 방정식 체계

지금까지는 단일한 방정식의 추정을 연구하였을 뿐, 그것이 하나의 부분을 이루는 보다 큰 경제모형은 다루지 않았다. 예를 들면, 어떤 특정한 상품의 수요방정식은 전형적으로 그 상품시장에서의 균형가격과 수량을 결정하는 하나의 방정식체계 (system of equations)중에서 하나인 것이다. 곧 시장에 대한 경제모형은 일반적으로 수요방정식, 공급방정식, 그리고 균형도달과정 (equilibrating process)을 나타내는 세번째 방정식 (예를들면, 수요량이 공급량과 일치하기에 이르는 방정식)을 포함하고 있다. 이 장에서는 추정하고자 하는 방정식이 보다 복잡한 모형의 여타 방정식들과 서로 관련되어 있을 때 일어나게 될 문제점들을 명확하게 고려하여 볼 것이다. 특히 어떤 여건하에서는 정규적인 추정절차로는 더 이상 불편추정량 (심지어 일치추정량)을 얻을 수 없다는 사실을 알게 될 것이다. 그러한 경우에 추정기법을 수정해야만 할 것이다.

1. 연립방정식 偏倚

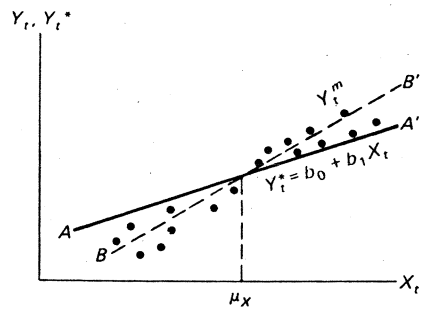
제 2 장에서 회귀모형에 대해 처음 서술할 때, 다음과 같은 회귀방정식의 특성에 관해 두가지 중요한 가정을 세웠음을 기억해야 할 것이다.

$$Y_t = b_0 + b_1 X_t + u_t$$

교란항의 평균이 0 이며, 곧 $E(u_t) = 0$ 이며, 교란항은 설명변수와 무관하고, 따라서 교란항과 독립변수사이의 공분산은 0 이고, 곧 $E(X_t u_t) = 0$ 이라고 가정하였다. 이러한 가정에 의거해 볼 때, 추정절차에서 다음과 같이 정하기로 한다. 곧,

$$\sum \hat{u}_t = 0 \quad \text{이며} \quad \sum (X_t \hat{u}_t) = 0$$

이로부터 두개의 정규방정식을 얻게 되었다. 이를 풀어서 모수추정량 $\hat{b_0}$와 $\hat{b_1}$을 구하였던 것이다. 따라서 이러한 계수의 추정량이 不偏推定量임을 보일 수 있었던 것이다.



〈그림 7.1〉

또한 이러한 추정량에 편의가 없는 것은 가정의 타당성에 기인한 것임을 보였다. 되돌아 가서, X_t와 u_t의 공분산이 0이 아니라고 가정하자. 곧, cov (X_t , u_t) $= E (X_t u_t) \neq 0$인 것이다. 예를 들어 공분산이 陽이라고 하자. 이는 u_t의 평균보다 큰 값(곧, u_t의 평균이 0이기 때문에 양의 값)이 X_t의 평균보다 큰 값과 관련되며, 또한 그 역도 사실임을 의미한다. 다시 이는 X_t가 평균치 μ_x보다 클 때 Y_t의 평균치는 $Y_t^* = (b_0 + b_1 X_t)$ 보다 클 것이며, X_t가 μ_x보다 작을 때 Y_t^*보다 작을 것임을 의미한다.〈그림 7.1〉에 있는 가설적인 散布圖(scatter diagram)에 의해 Y_t^*를 포함하는 관계가 AA'선으로 표시된다고 하자. 그러므로 u_t의 陰의 값이 항상 X_t의 보다 작은 값을 수반하는 반면 X_t가 클 때 u_t의 陽의 값이 전형적으로 일어나므로 관찰된 점들의 산포정도는 BB'와 같은 어떤 직선 가까이에 있어야 한다. 만일 $E (X_t u_t) = 0$이라고 잘못 가정하여 그에 따라 조건도 ($X_t \hat{u}_t$) $= 0$임을 부과하면 결국 관계 BB'를 추정하게 될 것이다. 산포도의 모든 점의 "중간"에 있는 것으로 믿는 Y_t와 X_t사이의 관계를 추정할 것이다. 이는 계수 b_0와 b_1의 추정

량에 편의가 있는 것으로 귀결될 것이다. 곧, 이러한 경우에 b_0의 값(AA'의 수직절편)을 과소추정(under estimate)하게 될 것이며, b_1 (AA'의 기울기)을 과대추정(over estimate)하는 경향이 있다. 표본크기가 커짐에 따라 이 편의가 주목할 정도로 줄어들지 않는다는 것을 강조할 것이다. 추정량은 편의를 갖고 있으며, 또한 일치추정량이 아닌 것이다.

이제 간단한 소비함수를 다시 생각해 보기로 한다. 곧, *

$$C_t = b_0 + b_1 Y_t + u_t \tag{7.1}$$

(7.1)에서 교란항 u_t는 정규분포를 하고, 평균이 0이며, 곧 $E(u_t) = 0$이며, 분산이 일정하다. 곧, $E(u_t^2) = \sigma_u^2$이다. 그리고 자기상관이 존재하지 않는다. 이제 토론하게 될 이유때문에 u_t는 Y_t에 독립적이라거나 상관이 없다라고 가정하지는 않는다.

거시경제학 이론에서 알고 있는 것은 (적어도) 이 모형에 하나의 추가적인 방정식이 있다는 것이다. 이는 총수요($C_t + I_t$)가 총공급(Y_t)과 일치하도록 산출량과 소득수준이 조정된다는 것을 일컫는 균형식인 것이다. 곧,

$$Y_t = C_t + I_t \tag{7.2}$$

여기에서 I_t는 개인과 기업에 의한 투자지출수준이다.** I_t가 외생변수 (exogeneous variable)라고 가정하자. 이는 어떠한 t시점에서도 그 값이 모

---

* 이 절에서 개발하고 있는 단순모형에서 총소득과 가처분소득을 구별하지 않을 것이다. 대부분의 교과서에서의 관례에 따라 소득을 Y로 표기하며 단지 소비를 소득의 함수로 간주할 것이다.

** 방정식(7.2)는 분명히 정부와 순 해외수요를 무시한 고도로 단순화한 균형식이다. 따라서 수요에 대한 이런 요소들을 쉽게 모형에 포함시킬 수 있는 것이다. 하지만 이것은 여기에서 고려중인 문제에는 불필요하다.

형 외부의 요소에 의해 결정됨을 의미한다. 가령, 투자지출은 습관이란 힘에 의해 결정되거나 (항상 어느 한 시점의 투자는 그 이전 시점의 투자보다 2퍼센트 높다), 사회학적인 조건들에 의해 결정되기도 한다. 여하튼 이 모형의 목적상 이러한 "외부적인 힘들"이 어떠한 것이든지 그것들은 (7.1)에서의 교란항 u_t와 관련이 없다고 가정한다. 그러면 I_t와 u_t 사이의 공분산은 0인 것이다. 곧,

$$E(I_t u_t) = \text{cov}(I_t, u_t) = 0$$

이제 방정식(7.2)에서의 Y_t와 C_t사이의 관계때문에 (7.1)에서의 소득Y_t와 교란항u_t사이의 공분산이 0이 아님을 증명하여 보기로 한다. 이는 다음에 의해 쉽게 볼 수 있다. 곧, (7.1)에서의 C_t의 식을 (7.2)에 대입하면 다음을 얻게 된다.

$$Y_t = b_0 + b_1 Y_t + u_t + I_t \tag{7.3}$$

Y_t에 대해 (7.3)을 풀면 다음을 구하게 된다.

$$Y_t = \frac{b_0}{1 - b_1} + \frac{I_t}{1 - b_1} + \frac{u_t}{1 - b_1} \tag{7.4}$$

다음으로 (7.4)에 u_t를 곱하여 기대치를 구하면 다음과 같다.*

---

* 또한 u_t의 평균이 0이므로 u_t와 Y_t의 공분산은 $E(u_t Y_t)$ 임을 기억해야할 것이다. 이는 다음과 같기 때문이다.
$E[(u_t - 0)(Y_t - u_Y)] = E(u_t Y_t) - E(u_t u_Y) = E(u_t Y_t) - u_Y E(u_t) = E(u_t Y_t)$
여기에서 u_Y는 Y의 평균이다.

$$\begin{aligned}\mathrm{cov}(Y_t, u_t) = E(Y_t u_t) &= E\left(\frac{b_0 u_t}{1-b_1} + \frac{I_t u_t}{1-b_1} + \frac{{u_t}^2}{1-b_1}\right) \\ &= \frac{b_0}{1-b_1} E(u_t) + \frac{1}{1-b_1} E(I_t u_t) + \frac{1}{1-b_1} E({u_t}^2)\end{aligned} \tag{7.5}$$

u_t 의 평균치가 0이고, u_t 와 I_t 사이의 공분산이 0이라고 가정하였기 때문에 (7.5)의 마지막 식에서 처음의 두항은 0이지만, 마지막 항은 0이 아니다. 명확히 하면 다음과 같다.

$$E(Y_t u_t) = \frac{1}{1-b_1} E({u_t}^2) = \frac{{\sigma_u}^2}{1-b_1} \neq 0 \tag{7.6}$$

추가된 방정식(7.2) 때문에 (7.1)에서 독립변수 Y_t와 교란항 u_t 사이의 공분산이 0이라고 더 이상 가정할 수 없음을 알게 되었다. 방정식(7.6)은 회귀모형의 기본적인 가정중의 하나가 타당하지 않다는 것을 말한다. 곧, 앞에서의 몇가지 이유로 인하여 표준적인 추정절차를 적용시키면 b_0와 b_1의 추정량은 편의가 있는 동시에 일치성이 없는 추정량인 것이다.

이러한 편의의 근원을 보다 직관적으로 논의하기 위해 정형화하면 유용할 것이다. 방정식(7.1)에서 t 시점의 소비지출C_t가 동일한 시점의 소득Y_t에 의존한다고 전제하였다. 하지만 Y_t 또한 C_t에 의존하고 있음을 (7.2)로부터 알고 있다. 두 방향에서 "인과관계 (causation)"가 작용하고 있는 것이다. 곧, 두 변수는 상호종속되어 있는 것이다. 표준적인 기법을 써서 (7.1)을 추정할 때 C_t와 Y_t 사이의 통계적인 관계의 유형을 판단하게 된다. 만일 분명히 인과관계가 Y_t에서 C_t로의 방향이라면, 추정치가 C_t에 대한 Y_t의 영향을 가리키는 것으로 해석해 볼 수 있다. 하지만 〔(7.2)에 의해 드러난 것처럼〕 상호종속의 경우, C_t와 Y_t 사이의 이러한 측정된 통계적 관계는 한 변수의 다른 변수에 대한 각각의

영향의 혼합(mix)을 반영하게 된다. 이제는 더 이상 추정치가 한 변수 여기에서는 Y_t 의 다른 변수, C_t에 대한 영향을 분명히 나타내는 측도로서 해석할 수 없는 것이다. 필요한 것은 두가지 효과를 해명할 수 있게 해주는 수정된 기법인 것이다. 곧 Y_t에 대한 C_t 의 영향으로부터 C_t 에 대한 Y_t 의 영향을 분리해내는 수단이 필요하다.

이것이 바로 聯立方程式偏倚(simultaneous-equation bias)의 문제인 것이다. 그것은 일반적으로 하나의 독립변수의 값 자체가 종속변수의 함수일 때 발생한다. 다음과 같은 다중회귀모형을 생각해 보기로 한다.

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + u_t \tag{7.7}$$

여기에서 X_{1t} 는 교란항과 무관하다는 점에서 통상적인 조건들을 만족하는 설명변수이다. 만일 X_{2t} 가 Y_t에 의존하는 것이 사실이라면 두번째 방정식을 고려해야 할 것이다. 아마 그 형태는 다음과 같을 것이다. 곧,

$$X_{2t} = c_0 + c_1Y_t + r_t \tag{7.8}$$

여기에서 r_t 는 또한 X_{1t} 와 무관하다고 가정한 교란항이다. 일반적으로 다음과 같은 사실을 (앞에서 처럼) 보이는 것은 과제로 남기기로 한다.

$$E(X_{2t}u_t) \neq 0$$

이러한 일련의 방정식들은 회귀모형의 주요가정중의 한가지에 위배됨을 뜻하며, (7.7)을 추정하기 위해 통상적인 절차를 이용하려는 시도는 방정식의 모든 모수의 추정량이 편의가 있는 반면, 일치성이 없는 결과를 낳을 것이다.

그러므로 이제 추정량의 一致性(consistency)에 대해서만 논의하기로 한다.

이렇게 하는 것은 계량경제학자들이 일반적으로 일치추정량을 구함으로써 "체계문제(the systems problem)를 해결하기" 때문이다. 왜냐하면 일반적으로 보다 큰 방정식체계의 부분에 속하는 방정식의 모수에 대해서는 불편추정량을 구할 수 없기 때문이다.

2. 이단계 최소자승법 : 단순한 경우

최근에 계량경제학자들은 연립방정식체계와 관련한 추정문제를 처리하는 여러가지 상이한 기법들을 개발하였다.* 여기에서는 그중 한가지를 관찰하여 볼 것이다. 곧 그것은 이단계 최소자승법(two-stage least squares, TSLS)이다. 이 특이한 기법은 몇가지 매력적인 특성을 가지고 있다. 첫째, 직관적으로 그것은 호소력이 있다. 곧 그것은 이해하기가 상대적으로 쉬우며, 이미 이 책에서 개발하였던 절차에 상당히 기초하고 있는 것이다. 둘째, 통상적인 조건하에서 고려중인 방정식을 추정할 수 있다면 (몇가지 다른 기법들과는 달리) TSLS는 항상 효력이 있다. 곧, 그것은 추정가능한 모수의 일치추정량을 구할 것이다.** 셋째, TSLS는 소위 "制限情報(limited-information)" 절차라 한다. 이는 다음을 의미한다. 곧, 여타 방정식에 대해 단지 "모호하게" 알고 있으면서도 TSLS기법을 써서 연립방정식체계 속에 포함되어 있는 어떤 방정식을 추정할 수도 있다는 것이

---

* 연립방정식체계의 추정에 대한 보다 진전된 처리방식은 다음을 볼 것 J.Johnston, Econometric Methods, 2nd ed. (New York : McGraw-Hill, 1972), chaps. 12 and 13, 그리고 Arthur S.Goldberger Econometric Theory (New York : Wiley, 1964), chap.7.

** 간단히 알아 보겠지만 단지 그중에서 몇가지 방정식 또는 어떤 모수들은 추정불가능하다. 실제로 다른 내용, 다중공선성중의 한가지에서 이미 추정할 수 없는 모수들이 존재하는 방정식을 보았던 것이다.

다. 여러가지 많은 여타 기법들은 훨씬 더 상세한 정보가 있어야만 적용될 수 있다. 그리고 넷째, TSLS는 상대적으로 연산을 적게 요구하는 편이다.

가. 설명 : 일치추정량

우선 TSLS에 대해 직관적으로나마 소개하기로 한다. 문제는 "이중인과관계 (two-way causation)"의 존재가 교란항과 하나의 (또는 그 이상의) 설명변수사이에 0의 공분산을 야기시키는 것임을 생각해야 할 것이다. 만일 표준적인 추정절차 (OLSQ)를 이용하고자 한다면,* 방정식이 회귀모형의 가정을 만족할 수 있도록 이러한 0의 공분산을 어느 정도 제거시켜야만 하는 것이다. 이것이 바로 TSLS기법이다. 그것은 2단계추정절차이다. 제1단계에서 독립변수 (들)로부터 교란항과 상관이 있는 부분을 제거한다. 곧, 이는 의심스러운 독립변수들에 대한 일련의 수정된 값을 만들어 내는 것을 포함한다. 이러한 "수정된 (revised)" 값들은 더 이상 교란항과 상관이 없으며, 따라서 단지 제2단계는 표준적인 기법을 이용하여 모수를 추정하는 것이다.

이러한 방법을 알아 보기 위해 국민소득에 대한 단순이변수방정식모형으로 돌아가기로 하자. 그러면 다음과 같다.

---

* 代變數 (instrumental – variable)推定은 추정된 회귀선으로부터의 관찰점의 편차의 자승합을 최소화하는 것과 같다는 점에서 "최소자승법"의 특성을 지니고 있다는 것을 제2장에서 기억해야 할 것이다. 이 때문에 표준적인 결과를 통상최소자승 (ordinary least – squares, OLSQ)결과라 하며, 그 결과가 일치하는 한에서 표준적인 절차를 OLSQ로 약식표기 한다. 이 표기는 표준적인 절차, OLSQ를 TSLS 와 구분할 수 있는 간편한 방식인 것이다.

$$C_t = b_0 + b_1 Y_t + u_t \tag{7.1}$$

$$Y_t = C_t + I_t \tag{7.2}$$

앞에서와 마찬가지로 (7.1)에서의 C_t의 식을 (7.2)에 대입하여 Y_t에 대해 풀면 다음을 얻게 된다.

$$Y_t = \frac{b_0}{1 - b_1} + \frac{I_t}{1 - b_1} + \frac{u_t}{1 - b_1} \tag{7.4}$$

표기를 간단히 하기 위해 (7.4)를 다음과 같이 쓰기로 한다.

$$Y_t = a_0 + a_1 I_t + q_t \tag{7.9}$$

여기에서

$$a_0 = \frac{b_0}{1 - b_1}$$

$$a_1 = \frac{1}{1 - b_1}$$

$$q_t = \frac{u_t}{1 - b_1}$$

이제 Y_t를 단지 I_t와 교란항 q_t의 함수라고 한다. 이때 유의해야 할 것은 (u_t와는 달리) q_t의 평균이 0이라는 것이다. 당분간 a_0와 a_1의 값을 알고 있다고 가정한다. 이러한 가정하에서 I_t에 대해 어떤 주어진 값과 관련하여 Y_t의 평균, 곧 Y_t^m에 대한 값을 도출할 수 있을 것이다. 곧,

$$Y_t^m = a_0 + a_1 I_t \tag{7.10}$$

또한 (7.9)로부터 다음을 알 수 있을 것이다.

$$Y_t = Y_t^m + q_t \tag{7.11}$$

$q_t = u_t / (1 - b_1)$임에 유의하면 (7.12)를 다음과 같이 쓸 수 있다.

$$C_t = b_0 + b_1 Y_t^m + \frac{u_t}{1 - b_1} = b_0 + b_1 Y_t^m + q_t \tag{7.13}$$

이제 (7.1)과는 달리 방정식(7.13)은 회귀모형의 기본가정들을 만족하고 있음을 보이게 되는 것이다. 첫째, 교란항의 평균이 0임을 보일 수 있다.

$$E(q_t) = E\left(\frac{u_t}{1 - b_1}\right) = \frac{1}{1 - b_1} E(u_t) = 0$$

그리고 둘째, 독립변수가 더 이상 교란항과 상관이 없음을 알게 된다. 만일 (7.10)에 q_t를 곱하여 기대치를 구하면 다음을 얻게 된다.

$$E(Y_t^m q_t) = E(a_0 q_t + a_1 I_t q_t) = \left(\frac{a_0}{1 - b_1}\right) E(u_t) + \left(\frac{a_1}{1 - b_1}\right) E(I_t u_t) = 0 \tag{7.14}$$

왜냐하면 u_t의 기대치가 0이며, 가정에 의해 I_t가 u_t와는 무관하기 때문이다.* 그러므로 정규적인 추정절차를 이용하여 b_0와 b_1의 일치추정량을 얻을 수 있다. $\Sigma \hat{q}_t = 0$과 $\Sigma (Y_t^m \hat{q}_t) = 0$의 조건을 부과하면 다음과 같은 정규방정식을 얻게 될 것이다.

$$\sum C_t = n\hat{b}_0 + \hat{b}_1 \sum Y_t^m$$
$$\sum (C_t Y_t^m) = \hat{b}_0 \sum Y_t^m + \hat{b}_1 \sum (Y_t^m)^2$$

이를 풀면 $\hat{b}_0$와 $\hat{b}_1$을 얻을 수 있다. 추가로 유의해야 할 것은 다음과 같다. 이러한 방정식들은 바로 원래의 방정식(7.1)에서 단지 Y_t를 Y_t^m

---

* 또한 분명한 것은 다음과 같다. 곧, (u_t가 자기상관을 가지지 않으므로) 교란항 q_t는 자기상관을 가지지 않을 것이며, 분산이 일정할 것이다. 곧, $[\sigma_u^2 / (1 - b_1)^2]$. 덧붙이면, u_t가 정규분포하므로 q_t도 정규분포할 것이다.

으로 대체하여 통례대로 정규방정식을 도출하면 구할 수 있는 것이다.

돌이켜 보면 체계의 문제를 처리하기 위해 우선 투자 I_t의 각각의 값과 관련된 소득의 평균치, Y_t^m을 결정하였다. 두번째 단계는 단지 Y_t를 이러한 새로운 변수로 치환하여 통상적인 방법으로 모수를 추정하는 것이다. 물론 기본방정식(7.1)에서 Y_t를 Y_t^m으로 치환할 때, 유의해야 할 점은 그에 따른 방정식(7.13)의 교란항 q_t가 원래의 교란항과는 같지 않다는 점이다. q_t가 u_t와는 기본적인 특성에서 동일하다는 것을 보였으므로 목적상 단지 표기에서의 이러한 변화는 중요한 것이 아닌 것이다.

TSLS절차는 원래의 교란항 u_t와 상관이 있는 요소인 q_t를 독립변수 Y_t로부터 제거시키는 것으로 사실상 이루어진다. $Y_t = Y_t^m + q_t$이며, 앞에서 보았듯이 Y_t^m은 q_t와 (따라서 u_t와) 상관이 없다는 것을 기억해야 할 것이다. Y_t대신에 Y_t^m을 이용할 때 Y_t로부터 효과적으로 q_t를 제거시켰던 것이다. 곧, $Y_t^m = Y_t - q_t$. 어떤 의미에서는 설명변수(Y_t)가 종속변수(C_t)의 교란항에 미치는 영향을 나타내는 Y_t의 부분을 제거시킨 것이다. 곧 이러한 방식으로 TSLS를 이용하여 앞에서 서로 얽혀있던 효과들을 분리해 낼 수 있다.

원칙으로는 이러한 절차가 나무랄데 없는 반면, 문제점은 실제로는 직접 그것을 수행할 수 없다는 것이다. 왜냐하면 Y_t^m의 값을 모르기 때문이다. 방정식(7.10)에서 다음과 같은 사실을 기억해내야 할 것이다.

$$Y_t^m = a_0 + a_1 I_t \tag{7.10}$$

그리고 a_0와 a_1의 값은 알고 있다고 가정하였다. 이는 일반적으로 사실이 아니지만 그 값들을 추정할 수는 있다. 원래의 방정식으로부터 도출해낸 하나의 방정식, 곧 (7.9)를 되돌아보면, 그것이 기본적인 회귀모형의

모든 가정을 만족하고 있음을 증명할 수 있다. 곧,

$$Y_t = a_0 + a_1 I_t + q_t \tag{7.9}$$

$\Sigma \hat{q}_t = 0$ 과 $\Sigma (\hat{q}_t I_t) = 0$ 으로 놓음으로써 만드는 정규방정식으로부터 a_0 와 a_1 의 추정량, 가령 $\hat{a}_0$ 과 $\hat{a}_1$ 을 구할 수 있다. 그러므로 $\hat{a}_0$ 과 $\hat{a}_1$ 을 이용하여 다음과 같은 $Y_t{}^m$ 의 추정량을 구할 수 있는 것이다.

$$\hat{Y}_t^m = \hat{a}_0 + \hat{a}_1 I_t \tag{7.15}$$

앞장에서의 표기를 생각해보면, $\hat{Y}_t^m$ 은 회귀방정식 (7.9) 와 관련된 Y_t 의 계산치에 지나지 않음을 알게 된다. 곧, $\hat{Y}_t{}^m \equiv \hat{Y}_t$.

제 2 단계는 ($Y_t{}^m$ 대신에) $\hat{Y}_t$ 을 이용하여 소비방정식을 추정하는 것이다. 곧, 회귀모형을 다음과 같이 취하게 될 것이다.

$$C_t = b_0 + b_1 \hat{Y}_t + u_t^* \tag{7.16}$$

여기에서 $u_t^* = u_t + b_1 \hat{q}_t$ 이다. 왜냐하면 $Y_t = \hat{Y}_t + q_t$ 이기 때문이다. 통상적인 절차를 따르면 단지 $\Sigma \hat{u}_t^* = 0$ 과 $\Sigma (\hat{a}_t^* \hat{Y}_t) = 0$ 이라는 조건을 부과함으로써 다음과 같은 정규방정식을 얻을 수 있다.

$$\Sigma C_t = n\hat{b}_0 + \hat{b}_1 \Sigma \hat{Y}_t$$
$$\Sigma (C_t \hat{Y}_t) = \hat{b}_0 \Sigma \hat{Y}_t + \hat{b}_1 \Sigma \hat{Y}_t^2$$

이에 따라 $\hat{b}_0$ 와 $\hat{b}_1$ 을 구하게 될 것이다.*

---

* 기술적인 면이지만 (7 . 9) 에서 $\Sigma \hat{q}_t = 0$ 과 $\Sigma (\hat{q}_t I_t) = 0$ 이라는 조건을 부과함으로써 a_0 과 a_1 을 추정한다는 것을 유의해야 한다.
$\hat{Y}_t = \hat{a}_0 + \hat{a}_1 I_t$ 이므로 $\Sigma (\hat{q}_t \hat{Y}_t) = 0$ 이 되는 것이다. 이제 (7 . 16) 에서 $u_t^* = u_t + b_1 \hat{q}_t$ 이므로 다음과 같게 된다. 곧,
$\Sigma u_t^* = \Sigma u_t$ 이며, $\Sigma (u_t^* \hat{Y}_t) = \Sigma (u_t \hat{Y}_t)$.
이러한 조건들을 (7 . 16) 을 추정하기 위해서는 다음과 같이 놓아야 함을 의미한다. 곧, (다음 페이지 계속)

그러므로 일반적인 조건하에서 $\hat{b}_0$과 $\hat{b}_1$이 일치추정량임을 보일 수 있다.

요약하면 지금까지의 사례에서는 TSLS추정절차는 먼저 외생변수 (exogenous variable), I_t에 대해 의심스러운 독립변수Y_t를 회귀한 다음, 이 추정방정식을 이용하여 새로운 독립변수$\hat{Y}_t$을 만들어 내는 것으로 이루어진다. 제 2 단계는 원래의 방정식에서 Y_t를 $\hat{Y}_t$으로 대체하여 통례대로 방정식을 추정하는 것이다. 이러한 절차에서 유의해야 할 것은 제 2 단계에서의 교란항u_t^*는 원래의 교란항u_t를 약간 수정한 것이라는 것이다.

나. 추가 결론

기법을 만들기 전에 유의해야 할 점은 일단 b_0와 b_1의 일정추정량을 얻게 되면 다음과 같은 식에 의해 (7.1)에서의 원래의 교란항의 일치추정량을 구할 수 있다는 것이다.

$$\hat{u}_t = C_t - (\hat{b}_0 + \hat{b}_1 Y_t)$$

이를 이용하면서 다음과 같은 통상의 수식에 의해 u_t의 분산의 일치추정량을 얻게 되는 것이다. 곧,

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_t^2}{n-2} \tag{7.17}$$

TSLS기법이 지닌 또다른 한가지 매력은 다음과 같다. 곧,일단 Y_t를 $\hat{Y}_t$으로 바꾸면 해석이 약간 달라질 뿐 $\hat{b}_0$과 $\hat{b}_1$의 분산식은 계속 유지되는 것이다. (7.16)을 참고로 하여 $\hat{Y}_t$의 표본평균이 $\overline{\hat{Y}} = \overline{Y}$임을 생각하

(앞 페이지 계속)

$\Sigma \hat{u}_t^* = \Sigma \hat{u}_t = 0$ ($\because E(u_t) = 0$)

$\Sigma(\hat{u}_t^* Y_t) = \Sigma(\hat{u}_t \hat{Y}_t) = 0$ ($\because E(u_t Y_t^m) = 0$).

바꿔말하면 u_t^*의 요소인 $\hat{q}_t$은 (7.16)의 추정에 아무런 역할을 하지 않는다. 아래에서 볼 분산식에서 이러한 결과를 보게 될 것이다.

면 다음과 같은 통상의 분산식은 표본크기가 무한할 때 유지되는 것임을 보일 수 있다.

$$\sigma_{b_0}^2 = \sigma_u^2 \left[\frac{\sum \hat{Y}_t^2}{n \sum (\hat{Y}_t - \bar{Y})^2} \right] \tag{7.18}$$

$$\sigma_{b_1}^2 = \sigma_u^2 \left[\frac{1}{\sum (\hat{Y}_t - \bar{Y})^2} \right] \tag{7.19}$$

그런데 표본크기가 무한한 것을 결코 얻을 수 없으므로 이 식들을 근사한 식으로 해석하여야만 할 것이다.

기민한 독자라면 이 분산식들에서의 미묘한 차이에 주목하게 될 것이다. 명확하게 하면 제 2 단계에서 추정절차는 교란항이 (u_t 가 아니라) u_t^* 인 방정식 (7.16)을 포함하고 있기 때문에 (7.18)과 (7.19) 에서의 $\hat{b}_0$ 과 $\hat{b}_1$ 의 분산식은 u_t^* 의 분산, 곧 σ_u^2 대신 가령 σ_u^{2*} 을 포함하여야 할 것이다. 하지만 이는 그렇지 못하다. 왜냐하면 u_t^*, 곧 $u_t^* = u_t + b_1 \hat{q}_t$ 의 성분인 $\hat{q}_t$ 이 $\hat{b}_0$ 과 $\hat{b}_1$ 의 식에서 생략되어 버리기 때문이다. 따라서 이 두 추정량의 값은 u_t^* 이 아니라 u_t 에 의존한다는 사실을 보일 수 있다.

이를 알아보기 위해 추정량 $\hat{b}_1$ 이 다음과 같음에 유의해야 할 것이다.

$$\hat{b}_1 = \frac{\sum (\hat{Y}_t - \bar{Y}) C_t}{\sum (\hat{Y}_t - \bar{Y})^2} \tag{7.20}$$

C_t 에 대해 (7.16) 을 (7.20) 에 대입하여 정리하면 다음을 얻게 된다.

$$\hat{b}_1 = b_1 + \frac{\sum (\hat{Y}_t - \bar{Y}) u_t^*}{\sum (\hat{Y}_t - \bar{Y})^2} \tag{7.21}$$

$\Sigma \hat{q}_t = 0$ 이고 $\Sigma (\hat{q}_t \hat{Y}_t) = 0$ 이므로 다음과 같게 될 것이다.

$$\Sigma (\hat{Y}_t u_t^*) = \Sigma (\hat{Y}_t u_t) + b_1 \Sigma (\hat{Y}_t \hat{q}_t) = \Sigma (\hat{Y}_t u_t)$$
$$\Sigma (\bar{Y} u_t^*) = \bar{Y} \Sigma u_t + \bar{Y} b_1 \Sigma \hat{q}_t = \bar{Y} \Sigma u_t$$

이제 $\hat{b}_1$ 을 다음과 같이 나타낼 수 있을 것이다. 곧,

$$\hat{b}_1 = b_1 + \frac{\Sigma (\hat{Y}_t - \bar{Y}) u_t}{\Sigma (\hat{Y}_t - \bar{Y})^2} \tag{7.22}$$

따라서 적어도 직관적으로 보면 $\hat{b}_1$ 의 대표본분산 (large-sample variance) 은 u_t^* 이 아니라 u_t 의 분산을 포함하게 될 것이다. $\hat{b}_0$ 도 그러한 이유로 해서 u_t 에 의해 표현될 수 있음을 보이는 것은 독자들에게 맡기기로 한다.

비록 σ_u^2 을 항상 알 수 없다할지라도 단지 σ_u^2 을 그것의 일치추정량으로 대체함으로써 $\hat{b}_0$ 과 $\hat{b}_1$ 의 분산의 일치추정량을 얻을 수 있는 것이다. 곧,

$$\hat{\sigma}_{\hat{b}_0}^2 = \hat{\sigma}_u^2 \left[\frac{\Sigma \hat{Y}_t}{n \Sigma (\hat{Y}_t - \bar{Y})^2} \right] \tag{7.23}$$

$$\hat{\sigma}_{\hat{b}_1}^2 = \hat{\sigma}_u^2 \left[\frac{1}{\Sigma (\hat{Y}_t - \bar{Y})^2} \right] \tag{7.24}$$

마지막으로 비율 $(\hat{b}_i - b_i) / \hat{\sigma}_{\hat{b}_i}$ ($i = 0, 1$) 이 표준화정규변수임을 가정함으로써 가설검정, 또는 신뢰구간설정이 가능하다. 또한 실제로 표본크기는 한정되어 있으며, 따라서 그 결과들은 근사치로 간주해야만 할 것이다. 이 경우에 이처럼 복잡해지는 것은 $\hat{b}_0$ 과 $\hat{b}_1$ 이 $\hat{Y}_t$ 에 의존함으로 해서 교란항에 대해 선형적이지 못하기 때문이다. 〔예로서는 (7.22) 를 볼 것. 〕

3. 방정식체계 : 보다 일반화된 논의*

가. 모형에 대한 설명

이제 연립방정식모형에 대한 논의를 일반화하기로 한다. 다음과 같은 다중회귀 방정식체계를 생각해 보자.

$$Y_{1t} = b_0 + b_1 Y_{2t} + b_2 Y_{3t} + a_1 X_{1t} + \cdots + a_k X_{kt} + u_t \tag{7.25}$$

$$Y_{2t} = c_0 + c_1 Y_{3t} + c_2 Y_{1t} + d_1 Z_{1t} + \cdots + d_r Z_{rt} + \varepsilon_t \tag{7.26}$$

$$Y_{3t} = g_0 + g_1 Y_{2t} + g_2 Y_{1t} + h_1 W_{1t} + \cdots + h_s W_{st} + e_t \tag{7.27}$$

여기에서 u_t, ε_t 와 e_t 는 교란항이다. 이러한 교란항 각각의 평균이 0 이라고 가정해보자. 곧,

$$E(u_t) = 0, \quad E(\varepsilon_t) = 0, \quad E(e_t) = 0$$

그리고 분산은 일정하고, 자기상관이 없으며 정규분포를 하고 있다고 가정한다.

$$E(u_t^2) = \sigma_u^2, \quad E(\varepsilon_t^2) = \sigma_\varepsilon^2, \quad E(e_t^2) = \sigma_e^2$$

게다가 이 교란항들은 식 (7.25) - (7.27) 에서 나타나는 모든 X, Z 및 W 와는 상관이 없다고 가정한다. 표기는 그렇지 않을 경우를 나타내

---

* 설명에서는 세가지 방정식모형에 의한 논의를 서술하고 있다. 하지만 그 결과는 모든 규모의 모형에 마찬가지로 적용되는 것이다.

지만 이러한 변수들중에서 한가지 또는 그 이상이 한가지 이상의 방정식에도 나타날 수 있는 것이다. 달리 말하면 X_t, Z_t 와 W_t 들은 공통의 요소일 수도 있는 것이다.

또한 교란항이 Y_{1t} , Y_{2t}, Y_{3t} 와 상관이 없음을 가정할 수 없다는 것은 분명하다. 일반적으로 각각의 Y_t 는 세가지 모든 교란항과 상관이 있을것이다. 가령 방정식 (7.25)는 바로 다음을 뜻한다. 곧, Y_t 는 u_t 에 의존하며, 따라서 전형적으로 이 두 변수는 상관관계에 있다는 것이다. 게다가 방정식 (7.26)은 Y_{2t} 가 Y_{1t} 에 의존하며, 따라서 u_t 에도 의존함을 가리키고 있다. 일반적으로 Y_{2t} 또한 u_t 와 상관관계에 있음을 알게 된다. 마찬가지로 (7.27)은 Y_{3t} 가 일반적으로 u_t 와 상관관계에 있음을 뜻한다. 이 주장은 분명히 여타 두 교란항, ε_t , e_t 와 Y_{1t} , Y_{2t} 와 Y_{3t} 의 분산이 0 이 아님을 증명하기 위해 확장시킬 수 있다. 요약하면 두 변수가 피드백 (feedback) 관계에 있으면, 일반적으로 각 변수는 다른 변수에 대한 방정식의 교란항과 상관관계에 있게 되는 것이다.

이 세가지 방정식의 모형에서 Y_{1t}, Y_{2t}와 Y_{3t} 는 內生 (endogeneous) 變數이다. t 시점에서의 그것들의 값은 방정식 (7.25) – (7.27)에서 설명한 모형에 의해 결정된다. 이것들은 그 값이 모형에 의해 설명되는 변수이다. 이 체계에서 여타 변수들 (X, Z 와 W)은 事前決定變數 (predetermined variables)라 한다. 그것들은 교란항과는 상관이 없으며, t 시점에서의 값이 t 시점에서의 모형에 의해 결정되지 않는다. 간단하게나마 그것들이 어떻게 결정되는지를 검토해 볼 것이다. 한편, 교란항의 값과 더불어 사전결정변수의 값들은 t 시점에 모형의 내생변수의 값들을

결정하는 것은 분명하다. 가령, 일반적으로 (7.25), (7.26)과 (7.27) 과 같은 일련의 선형방정식들은 내생변수 Y_{1t} , Y_{2t}와 Y_{3t}를 얻기 위해 사전결정변수와 교란항에 의해 풀 수 있는 것이다. 곧, (7.25)- (7.27)의 방정식들을 풀어서 Y_t 각각을 다음과 같은 형태로 나타낼 수 있게 된다.

$$Y_{1t} = l_0 + \sum_{i=1}^{k} l_i X_{it} + \sum_{i=k+1}^{k+r} l_i Z_{it} + \sum_{i=k+r+1}^{k+r+s} l_i W_{it} + \alpha_1 u_t + \alpha_2 \varepsilon_t + \alpha_3 e_t \qquad (7.28)$$

$$Y_{2t} = m_0 + \sum_{i=1}^{k} m_i X_{it} + \sum_{i=k+1}^{k+r} m_i Z_{it} + \sum_{i=k+r+1}^{k+r+s} m_i W_{it} + \gamma_1 u_t + \gamma_2 \varepsilon_t + \gamma_3 e_t \qquad (7.29)$$

$$Y_{3t} = d_0 + \sum_{i=1}^{k} d_i X_{it} + \sum_{i=k+1}^{k+r} d_i Z_{it} + \sum_{i=k+r+1}^{k+r+s} d_i W_{it} + \beta_1 u_t + \beta_2 \varepsilon_t + \beta_3 e_t \qquad (7.30)$$

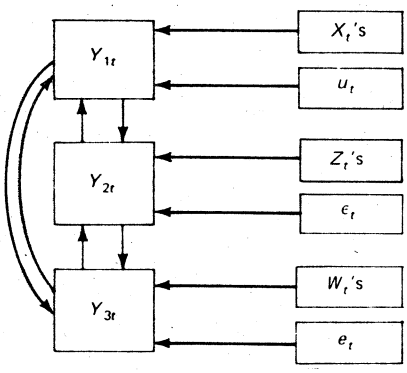
여기에서 사전결정변수와 교란항의 값이 완전히 내생변수의 값을 결정한다는 사실을 알 수 있다. 이는 사전결정변수와 교란항이 공통된 어떤 것을 갖고 있음을 의미하는 것이다. 곧 어느 것도 t 시점의 모형에 의해 결정되지 않으며, 둘다 내생변수의 값의 결정에 필요하다는 것이다. 물론 그것은 사실이다. 하지만 중요한 차이가 있다. 곧, 교란항과는 달리 매 t 시점의 사전결정변수의 값은 알고 있거나, 또는 관찰하게 된다고 가정한다. 달리 말하면, 어떤 특정한 교란항의 값에 대한 관찰치가 있으면 그 교란항은 사전결정변수인 것으로 간주할 수 있는 것이다. 그 대신 교란항은 평균이 0이며, 관찰된 사전결정변수 모두와는 상관이 없는 관찰불가

능한 사전결정변수로 간주해볼 수 있을 것이다.

요약하면, 〈그림 7.2 〉에 (7.25) – (7.27)에서의 일련의 관계들의 인과구조에 대한 도식적 표현을 나타내고 있다. (화살표가 가리키고 있듯이) 유의해야 할 것은 사전결정변수와 교란항은 직접 내생변수들의 값에 영향을 주고 있지만 다시 그것들에 의해 영향받지는 않는다는 것이다. 그와는 달리 내생변수사이의 피드백관계 (또는 상호의존) 가 존재한다. 예를 들면 Y_{1t}는 Y_{2t}와 Y_{3t}에 의존할 뿐만 아니라 Y_{2t}와 Y_{3t}의 값에 영향을 주고 있는 것이다.

나. 事前決定變數의 속성

사전결정변수 자체는 두가지 형태를 띤다. 첫째, 外生變數 (exogeneous variables)가 있다. 단순소비모형 (7.1)과 (7.2)에 대한 논의에서 지적하였듯이, 그 값들이 모형에 대해 외적인 힘들에 의해 결정되는 변수들이 그러한 것이다. 어느정도 보다 정형화하면, 외생변수의 값은 어떠한 방식으로든 내생변수, 또는 교란항과는 관련이 없는 변수들에 의존하는 것으로 가정하는 것이다. 어떤 의미에서 이러한 변수들은 분석범위



〈그림 7.2〉

를 벗어나는 것이다.* 그것들을 설명하려는 시도없이 그것들의 값은 단지 주어진 것으로 한다. 소비함수에서 외생적인 것으로 간주될 수도 있는 변수의 한가지 예가 가족규모일 것이다. 이는 사회체계에 의해 경제모형에 "공급되는 (fed in)"것으로 간주하는 변수이다. 곧 이것은 소비와 같은 경제변수들을 결정하는데 중요하지만, (적어도 단기에서는) 특정한 모형에서 경제변수들에 의해 영향받지 않을 수 있는 것이다. 외생변수의 또다른 한가지 예는 밀생산을 설명하는 방정식에서 연간 강우량일 것이다. 첫번째 경우 (가족규모) 에 외생변수를 설명하기 위해 분석의 범위를 확장시키는 경우를 생각할 수 있다. 가령 가족규모는 고용기회 등에 의해 설명될 수 있을 것이다. 만일 그러하다면 이 변수는 더 이상 외생적이지 않다. 두번째의 경우에는 아마도 강우량을 분석의 확장가능성 정도를 벗어나는 것으로 간주할 수 있다.

사전결정변수의 제 2 유형은 시차가 주어진 내생변수 (lagged endogeneous variable) 이다. 예를 들면 소비모형이 다음과 같다고 하자.

$$C_t = b_0 + b_1 Y_t + b_2 Y_{t-1} + u_t \tag{7.31}$$

$$Y_t = C_t + I_t \tag{7.32}$$

통상의 가정하에서 Y_t 는 u_t 와 상관이 있음을 알 수 있다. 왜냐하면 Y_t 에 대해 방정식을 풀면 (앞에서도 지적하였듯이) Y_t 가 u_t 에 의존한다는 것을 알기 때문이다. 곧,

---

* "분석범위 (the scope of analysis)"를 정의하는 것은 간단한 일이 아니다. 문제는 보다 복잡한 분석은 "보다 광범한 범위 (wider scope)"를 필요로 하지만, 이는 전형적으로 모형의 복잡화를 가중시킨다는 것이다.

$$Y_t = \left(\frac{b_0}{1-b_1}\right) + \left(\frac{b_2}{1-b_1}\right) Y_{t-1} + \left(\frac{1}{1-b_1}\right) I_t + \left(\frac{1}{1-b_1}\right) u_t \qquad (7.33)$$

하지만 이것은 Y_{t-1}의 경우에는 사실이 아니다. 곧 Y_{t-1}의 값은 前期에서 결정되며, 그 값은 C_t, u_t, 또는 t 시점에서의 여타 변수의 값에 의존할 수 없는 것이다. 예를 들어 (7.33)에서 t를 ($t-1$)로 대체하면 Y_{t-1}은 u_t가 아니라 u_{t-1}에 의존한다는 사실을 알게 되는 것이다. 곧,

$$Y_{t-1} = \left(\frac{b_0}{1-b_1}\right) + \left(\frac{b_2}{1-b_1}\right) Y_{t-2} + \left(\frac{1}{1-b_1}\right) I_{t-1} + \left(\frac{1}{1-b_1}\right) u_{t-1}$$

이는 다음을 뜻한다. 곧, 교란항 u_t의 값이 각각의 매 시점마다 독립적으로 결정되면 (따라서 u_t가 자기상관이 없으면), u_t와 Y_{t-1}은 상관이 없다는 것이다.

요약한다면, Y_{t-1}의 값이 t 시점동안 모형에 의해 결정되지 않으며, t 시점의 교란항과 상관이 없기 때문에 Y_{t-1}을 사전결정변수로 분류할 수 있는 것이다. 몇가지 가정하에서 단지 시차가 주어진 모든 내생변수들을 사전결정변수로 취급할 수 있다.* 사전결정변수와 교란항사이의 공분산이

---

* 이에 대한 예외가 교란항이 자기상관을 갖고 있는 경우이다. 곧 이 경우에 시차가 주어진 내생변수는 모형에서 t시점의 교란항과 상관이 있을 수 있다. 관심있는 독자를 위해 체계문제와 자기상관이 있는 교란항을 포함하고 있는 모형을 추정하는 방법을 이 장 부록에서 논의하고 있다. 경제학자들이 동일한 모형에서 한가지 이상의 추정 문제를 어떻게 처리하는지를 알아보는데는 이것이 도움될 것임을 알게 될 것이다.

0 이라는 기본가정은 체계내의 외생변수와 시차가 주어진 내생변수의 두 경우 모두에도 유지된다.

다. 構造方程式과 縮小型 方程式

추정기법을 일반화하는 문제를 고려하기 전에 마지막으로 한가지 규정할 필요가 있다. 내생변수들의 행태와 상호관계를 설명하는 (7.25), (7.26), (7.27)과 같은 기본방정식들은 構造方程式(structural equations)이라 한다. 이것들은 경제이론에 의해 고안된 방정식이다. 앞에서는 간단한 소득결정모형에 의해 다음과 같은 두개의 구조방정식체계를 전제하였다. 곧,

$$C_t = b_0 + b_1 Y_t + u_t \tag{7.1}$$
$$Y_t = C_t + I_t \tag{7.2}$$

이 방정식들은 소비지출이 소득(7.1)에 의존하며, 총산출물은 총수요(7.2)와 동일한 균형수준으로 이동한다는 이론적 전제들을 담고 있다. 다시 한번 말한다면 구조방정식은 기본적인 경제모형에 대한 정형적인 서술인 것이다.

만일 (7.28) – (7.30) 에서처럼 내생변수를 얻기 위해 구조방정식을 사전결정변수와 교란항에 의해 풀면 그에 따른 방정식을 縮小型方程式(reduced-form equations)이라 한다. 이것들은 각각의 내생변수가 인과적으로 사전결정변수와 교란항에 의해 어떻게 결정되는지를 설명하는 방정식이다.

한가지 예로서 두 내생변수에 대해 구조방정식(7.1)과 (7.2)를 풀면 다음을 얻게 된다.

$$C_t = \frac{b_0}{1 - b_1} + \frac{b_1 I_t}{1 - b_1} + \frac{u_t}{1 - b_1} \tag{7.34}$$

$$Y_t = \frac{b_0}{1 - b_1} + \frac{I_t}{1 - b_1} + \frac{u}{1 - b_1} \tag{7.35}$$

(7.34)와 (7.35)를 다음과 같이 다시 쓸 수 있다. 곧,

$$C_t = a_0 + a_1 I_t + q_t \tag{7.34A}$$

$$Y_t = d_0 + d_1 I_t + q_t \tag{7.35A}$$

여기에서 $a_0 = d_0 = b_0/(1 - b_1)$, $a_1 = b_1/(1 - b_1)$, $d_1 = 1/(1 - b_1)$, $q_t = u_t/(1 - b_1)$이다. 방정식 (7.34A)와 (7.35 A), 또는 (7.34)와 (7.35)는 각각 C_t와 Y_t에 대한 축소형방정식이다. 유의해야 할 점은 다음과 같다. 곧, 이러한 축소형방정식들은 C_t와 Y_t가 완전히 I_t와 u_t에 의해 결정되는 것임을 의미한다는 것이다.

또다른 한가지 예로서 임금-물가모형을 고려하기로 한다.

$$\dot{W}_t = a_1 + b_1 \dot{P}_t + c_1 R_{t-1} + \varepsilon_{1t} \tag{7.36}$$

$$\dot{P}_t = a_2 + b_2 \dot{W}_t + b_3 \dot{T}_{t-1} + \varepsilon_{2t} \tag{7.37}$$

여기에서 $\dot{W}_t$와 $\dot{P}_t$은 각각 화폐임금과 소비자물가의 백분율변화이고, R_{t-1}은 前期의 실업률이며, $\dot{T}_{(t-1)}$은 전기의 원료 가격의 백분율변화이고, ε_{1t}와 ε_{2t}는 교란항이다. 방정식 (7.36)과 (7.37)은 모형의 구조방정식이다. 그것들은 임금변화율이 가격변화율과 (전기의 실업률로 나타낸) 노동시장조건에 의존하는 반면, 다시 가격변화율은 임금변화률과 전기동안의

원료가격의 변화율로 나타낸 여타 비용의 변화율에 의존한다는 가정을 표현하고 있는 것이다. 내생변수는 $\dot{W}_t$과 $\dot{P}_t$인 반면, 사전결정변수는 $R_{(t-1)}$과 $\dot{T}_{(t-1)}$이다. 이 경우에 두 사전결정변수 모두 시차가 주어진 외생변수인 것이다.* 시차가 주어진 내생변수는 없다. 마지막으로 $\dot{W}_t$와 $\dot{P}_t$에 대해 (7.36)과 (7.37)을 풀면, 다음과 같은 축소형방정식을 얻게 된다.

$$\dot{W}_t = d_0 + d_1\dot{T}_{t-1} + d_2R_{t-1} + d_3\varepsilon_{1t} + d_4\varepsilon_{2t} \tag{7.38}$$

$$\dot{P}_t = e_0 + e_1\dot{T}_{t-1} + e_2R_{t-1} + e_3\varepsilon_{1t} + e_4\varepsilon_{2t} \tag{7.39}$$

여기에서 d_0, d_1, d_2, d_3, d_4와 e_0, e_1, e_2, e_3, e_4는 다음과 같다.

$$d_0 = \frac{a_1 + b_1a_2}{1 - b_1b_2}, \quad d_1 = \frac{b_1b_3}{1 - b_1b_2}, \quad d_2 = \frac{c_1}{1 - b_1b_2},$$

$$d_3 = \frac{1}{1 - b_1b_2}, \quad d_4 = \frac{b_1}{1 - b_1b_2}$$

---

* (7.36)과 (7.37)에서의 좁은 "분석범위"를 따를 수도 있다. 곧, 賃金-物價上昇現象(wage-price spiral)에 대한 완전한 설명은 실업율을 설명하는 방정식을 포함시켜야만 한다. (곧, 실업률은 임금-물가상승모형에 외생적인 것으로 간주되어서는 안된다.) 하지만 위에서 언급하였듯이 그러한 보다 복잡한 설명은 복잡성만 증가시킬 것이다. 또한 분석범위와 관련된 한계선을 긋는 위치는 바로 연구자의 판단에 달려 있는 것이다.

$$e_0 = \frac{a_2 + b_2 a_1}{1 - b_1 b_2}, \quad e_1 = \frac{b_3}{1 - b_1 b_2}, \quad e_2 = \frac{b_2 c_1}{1 - b_1 b_2},$$

$$e_3 = \frac{b_2}{1 - b_1 b_2}, \quad e_4 = \frac{1}{1 - b_1 b_2}$$

유의할 점은 축소형방정식이 사전결정변수와 교란항에 대해 선형(linear)이라는 것이다.

4. 이단계 최소자승법 : 일반화

가. 개 관

이제 추정문제로 돌아가기로 하자. 일단 다시 한번 (7.25)부터 (7.27)까지의 일련의 방정식을 생각하고, 이 모형에서 방정식(7.25)의 계수를 추정하기로 하자.

$$Y_{1t} = b_0 + b_1 Y_{2t} + b_2 Y_{3t} + a_1 X_{1t} + \cdots + a_k X_{kt} + u_t \tag{7.25}$$

$$Y_{2t} = c_0 + c_1 Y_{3t} + c_2 Y_{1t} + d_1 Z_{1t} + \cdots + d_r Z_{rt} + \varepsilon_t \tag{7.26}$$

$$Y_{3t} = g_0 + g_1 Y_{2t} + g_2 Y_{1t} + h_1 W_{1t} + \cdots + h_s W_{st} + e_t \tag{7.27}$$

TSLS하에서의 추정절차는 우선 세 방정식모형의 모든 사전결정변수에 대해 방정식에서 나타나는 내생변수, Y_{2t}와 Y_{3t}(곧, 교란항과 상관이 있는 설명변수)를 회귀하는 것이다. 명확히 말하면, 통상의 절차를 써서 다음의 방정식을 추정하게 된다. 곧,

$$Y_{2t} = m_0 + m_1 X_{1t} + \cdots + m_k X_{kt} + m_{(k+1)} Z_{1t} + \cdots + m_{(k+r)} Z_{rt} + m_{(k+r+1)} W_{1t} + \cdots + m_{(k+r+s)} W_{st} + \theta_{1t} \tag{7.40}$$

여기에서 θ_{1t} 는 교란항이다.* 마찬가지로 동일한 독립변수집합에 대해 Y_{3t} 를 회귀할 것이다. 그리하여 추정방정식을 이용하여 계산치, $\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 을 결정하게 될 것이다. 제 2 단계는 방정식 (7.25) 에서 Y_{2t} 와 Y_{3t} 각각에 대해 $\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 으로 대체하여 교란항에 별표를 한 다음 통례대로 제 2 단계의 정규방정식을 도출하여 그것들을 풀어 추정량, $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ 와 $\hat{a}_1$ $\hat{a}_2$, ⋯, $\hat{a}_k$ 을 구하는 것이다. 또한 $\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 은 각각 방정식의 교란항 u_t 와는 상관이 없는 Y_{2t} 와 Y_{3t} 의 부분의 추정량임이 판명된다.

나. 수식화

이를 보다 더 분명하게 알기 위해 유의할 것은 다음과 같다. 세 방정식 모형이 선형이기 때문에 가령 Y_{2t} 에 대한 解 (또는 축소형방정식)가 X_t, Z_t , W_t 와 u_t, ε_t , e_t 에 대해 선형이라는 것이다. 표기를 절약하기 위해 교란항을 제외하고는 축소형방정식 (7.29 를 볼 것) 의 모든 항의 합을 Y_{2t}^m 으로 지정하기로 한다. 곧, Y_{2t}^m 은 X_t, Z, W 의 선형결합인 것이다.** 곧,

$$Y_{2t}^m = m_0 + m_1 X_{1t} + \cdots + m_k X_{kt} + m_{(k+1)} Z_{1t} + \cdots + m_{(k+r)} Z_{rt} + m_{(k+r+1)} W_{1t} + \cdots + m_{(k+r+s)} W_{st} \qquad (7.41)$$

이제 Y_{2t} 에 대한 축소형방정식을 다음과 같이 보다 간단하게 나타낼 수 있게 된다.

$$Y_{2t} = Y_{2t}^m + \gamma_1 u_t + \gamma_2 \varepsilon_t + \gamma_3 e_t \qquad (7.42)$$

---

* 앞으로 식을 통해 보일 것이지만 교란항 θ_{1t} 는 방정식 (7.29) 에서 처럼 세 교란항 ($\gamma_1 u_t + \gamma_2 \varepsilon_t + \gamma_3 e_t$) 의 가중합이다.

** 예를 들어 모형 (7.36) 과 (7.37) 에 대해서는 다음과 같다.

$$\dot{W}_t^m = d_0 + d_1 \dot{T}_{(t-1)} + d_2 R_{(t-1)}$$

여기에서 d_i 는 (7.38) 에서 정의되고 있는 것이다.

여기에서 γ_1, γ_2와 γ_3는 상수이다.

다음으로 유의해야 할 점은 이러하다.

$$E(\gamma_1 u_t + \gamma_2 \varepsilon_t + \gamma_3 e_t) = \gamma_1 E(u_t) + \gamma_2 E(\varepsilon_t) + \gamma_3 E(e_t) = 0$$

왜냐하면 모든 교란항의 평균이 0이기 때문이다. 이 세 항을 결합하여 하나의 결합교란항(composite disturbance term)으로 만들어보자. 곧,

$$\theta_{1t} = \gamma_1 u_t + \gamma_2 \varepsilon_t + \gamma_3 e_t \tag{7.43}$$

그러면 〔 7.42)를 다음의 형태로 쓸 수 있다.

$$Y_{2t} = Y_{2t}^m + \theta_{1t} \tag{7.44}$$

여기에서 E (θ_{1t})= 0 이다. 마찬가지로 다음을 보일 수 있는 것이다.

$$Y_{3t} = Y_{3t}^m + \theta_{2t} \tag{7.45}$$

여기에서 E (θ_{2t}) = 0 이며, Y_{3t}^m은 X_t, Z_t와 W_t의 선형결합이다. 〔 (7.30)을 볼 것.〕 마지막으로 바로 단순한 설명변수의 경우에서처럼 (7.44)와 (7.45)를 이용하여 첫번째 구조방정식 (7.25)을 다음과 같이 나타낼 수 있는 것이다. 곧,

$$Y_{1t} = b_0 + b_1 Y_{2t}^m + b_2 Y_{3t}^m + a_1 X_{1t} + \cdots + a_k X_{kt} + q_{1t} \tag{7.46}$$

여기에서 $q_{1t} = u_t + b_1 \theta_{1t} + b_2 \theta_{2t}$ 이며,

따라서 $E(q_{1t}) = 0$ 이다.

이제 단순한 경우와 유사하다는 것은 분명하다. 곧, Y_{2t}^m과 Y_{3t}^m에 대한 관찰치가 있다면 단지 통례대로 (7.46)을 추정할 수 있는 것이다.

이 이유는 다음과 같다. 곧, Y_{2t}^m과 Y_{3t}^m이 단지 X_t, Z_t와 W_t에 의존하며, 이 사전결정변수들이 교란항과는 상관이 없기 때문에 그 결과 Y_{2t}^m과 Y_{3t}^m 또한 교란항과는 상관이 없음이 분명한 것이다. 단순한 경우에서처럼 Y_{2t}^m과 Y_{3t}^m을 모르며 따라서 "통례대로 진행하기" 전에 먼저 ($\hat{Y}_{2t}$과 $\hat{Y}_{3t}$에 의해) 그것들을 추정해야만 하는 것이다.

지적할 것은 다음과 같다. 곧, 앞의 이 변수의 경우에서처럼 TSLS에 의해 만들어진 추정량들은 일반적으로 편의를 갖고 있지만 일치추정량이라는 것이다. 또한 지적해야 할 것은 다음과 같다. 회귀모수의 일치추정량을 이용하여 다음과 같은 수식으로부터 가령 (7.25)에서 교란항의 일치추정량을 만들어 낼 수도 있는 것이다.

$$\hat{u}_t = Y_{1t} - (\hat{b}_0 + \hat{b}_1 Y_{2t} + \hat{b}_2 Y_{3t} + \hat{a}_1 X_{1t} + \cdots + \hat{a}_k X_{kt}) \tag{7.47}$$

그러면 다음과 같이 그 교란항의 분산의 일치추정량을 얻게 된다.*

$$\hat{\sigma}_u^2 = \sum_{t=1}^{n} \frac{\hat{u}_t^2}{n - k - 3} \tag{7.48}$$

또한 증명없이 다음을 언급하기로 한다. 곧, 일단 내생설명변수를 계산된 부분으로 대체하면 또한 바로 설명변수가 하나인 경우에서처럼 대표본의 결과로서 모수추정량에 대한 모든 분산식을 승인하게 된다. 가령 방정식 (7.25)를 참고로 하면 TSLS 추정량 $\hat{b}_1$의 대표본분산은 다음과 같게 될 것이다.

$$\sigma_{\hat{b}_1}^2 = \sigma_u^2 \left(\frac{1}{\sum \hat{v}_{1t}^2} \right) \tag{7.49}$$

---

* (7.48)에서 $(n-k-3)$으로 나눈다. 왜냐하면 해당 회귀모형 (7.46)에서 모수는 $(k+3)$개이기 때문이다.

여기에서 $\hat{v}_{1t}$은 $\hat{Y}_{3t}, X_{1t}, \cdots, X_{kt}$에 대한 $\hat{Y}_{2t}$에서의 잔차이다. 또한 유의해야 할 점은 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$는 제 2 단계에서의 설명변수이며, 그 계수들은 b_1과 b_2라는 것이다. 마찬가지로 (7.49)에서의 分散式은 제 2 단계의 회귀와 관련된 교란항, 곧 $u_t^* = u_t + b_1\hat{\theta}_{1t} + b_2\hat{\theta}_{2t}$이 아니라 u_t의 분산을 포함하고 있다는 사실도 알게 된다. 마지막으로 일반적으로 σ_u^2을 모르기 때문에 $\hat{b}_1$의 분산의 일치추정량은 다음과 같을 것이다.

$$\hat{\sigma}_{b_1}^2 = \hat{\sigma}_u^2 \left(\frac{1}{\sum \hat{v}_{1t}^2} \right) \qquad (7.50)$$

유사한 식이 여타 모수추정량의 경우에도 유지될 것이다.

이러한 분산식들을 이용하여 근사한 것으로서 또한 정규분포를 사용함으로써 신뢰구간설정과 가설검정을 행할 수 있을 것이다. 실제로 그러한 결과들은 근사한 것으로 간주되어야 할 것이다. 왜냐하면 표본크기가 유한하기 때문이다.

다. 없어진 변수를 갖는 TSLS

추가되는 실례에 대해서는 명확화가 필요하다. (7.25)에서 TSLS 하의 제 1 단계의 절차는 모형의 모든 사전결정변수에 대해 내생변수 각각을 회귀하는 것이었다. 그런데 그것은 자료가 이러한 모든 변수들에 대해 이용할 수 있는 상황이 아닌 경우일 수도 있다. 방정식 (7.26)과 (7.27) 각각으로부터 Z_{1t}와 W_{1t}에 대해 관찰치가 없다고 가정하자. 이로 인해 일반적으로 (7.25)를 추정하기 위해 TSLS를 이용하는 것이 불가능 하여진다. 이러한 경우의 제 1 단계절차는 모든 이용가능한 사전결정변수 (곧 현재의 실례에서는 Z_{1t}와 W_{1t} 이외의 변수)에 대해 Y_{2t}와 Y_{3t}를 회귀하는 것이 된다. 그리하여 이러한 제 1 단계의 방정식을 이용하여 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$을 얻을 수 있을 것이다. 곧 Y_{2t}와 Y_{3t}을 $\hat{Y}_{1t}$와 $\hat{Y}_{3t}$로 각각 대

체하면 앞에서와 마찬가지로 제2단계의 추정을 시작할 수 있으며, 앞에서 서술한 모든 수식과 결과들이 여전히 유지될 것이다. 그러므로 비교적 덜 완전한 자료로 TSLS기법을 이용할 수 있는 것이다.

이것은 왜 그러한지를 좀 더 공식적으로 논의해 보기로 한다. 우선 Y_{2t}와 관련된 내용을 알아보기로 하자. Z_{1t}와 W_{1t}에 대한 자료를 이용할 수 없으므로 제1단계의 회귀는 Z_{1t}와 W_{1t}가 없다는 것만 다를뿐 (7.40)과 유사하게 될 것이다. 곧,

$$Y_{2t} = \pi_0 + \pi_1 X_{1t} + \cdots + \pi_k X_{kt} + \pi_{(k+1)} Z_{2t} + \cdots + \pi_{(k+r-1)} Z_{rt} + \pi_{(k+r)} W_{2t} + \cdots + \pi_{(k+r+s-2)} W_{st} + \theta_{3t} \quad (7.51)$$

다음과 같이 놓음으로써 얻은 이 모형의 모수추정량으로부터 $\hat{Y}_{2t}$을 구하게 되는 것이다. 곧,

$$\sum \hat{\theta}_{3t} = 0, \quad \sum (\hat{\theta}_{3t} X_{jt}) = 0, \quad \sum (\hat{\theta}_{3t} Z_{it}) = 0, \quad \sum (\hat{\theta}_{3t} W_{mt}) = 0 \quad (7.52)$$

(j = 1, ···, k ; i = 2, ···, r ; m = 2, ···, S)

보다 명확하게 하면 다음과 같다.

$$\hat{Y}_{2t} = \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \cdots + \hat{\pi}_k X_{kt} + \hat{\pi}_{(k+1)} Z_{2t} + \cdots + \hat{\pi}_{(k+r-1)} Z_{rt} + \hat{\pi}_{(k+r)} W_{2t} + \cdots + \hat{\pi}_{(k+r+s-2)} W_{rt} \quad (7.53)$$

통례대로 Y_{2t}를 다음과 같이 표현할 수 있다.

$$Y_{2t} = \hat{Y}_{2t} + \hat{\theta}_{3t} \quad (7.54)$$

여기에서 $\hat{\theta}_{3t}$는 ($Y_{2t} - \hat{Y}_{2t}$)으로 정의된다. 또한 $\hat{Y}_{2t}$는 단지 X_{1t}, ···, X_{kt}, Z_{2t}, ···, Z_{rt}와 W_{2t}, ···, W_{st}에 의존하기 때문에 그에 따라 (7.52)로 부터 다음과 같게 될 것이다.

$$\sum (\hat{Y}_{2t}\hat{\theta}_{3t}) = 0 \qquad (7.55)$$

$\hat{Y}_{3t}$ 를 얻기 위해 동일한 집합의 사전결정변수에 대해 Y_{3t} 를 회귀함으로써 Y_{3t} 에 대한 유사한 수식을 얻을 수 있다. 다음에 주목해야 할 것이다. 곧,

$$Y_{3t} = \hat{Y}_{3t} + \hat{\theta}_{4t} \qquad (7.56)$$

여기에서 $\hat{\theta}_{4t} = Y_{3t} - \hat{Y}_{3t}$ 이다. 이에 따라 다음과 같게 된다.

$$\sum (\hat{Y}_{3t}\hat{\theta}_{4t}) = 0 \qquad (7.57)$$

마지막으로 $\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 은 둘다 $X_{1t}, \cdots X_{kt}$, Z_{2t}, $\cdots$, Z_{rt} 와 W_{2t}, $\cdots$, W_{st} 의 선형결합이기 때문에 (7.52)와 $\hat{\theta}_{4t}$ 에 대한 관련조건으로부터 다음의 사실을 얻게 된다.

$$\sum (\hat{Y}_{2t}\hat{\theta}_{4t}) = 0 = \sum (\hat{Y}_{3t}\hat{\theta}_{3t}) \qquad (7.58)$$

이제 방정식 (7.25)로 돌아가기로 하자. Y_{2t} 와 Y_{3t} 에 (7.54)와 (7.56)을 대입하면 다음을 얻게 된다.

$$Y_{1t} = b_0 + b_1\hat{Y}_{2t} + b_2\hat{Y}_{3t} + a_1X_{1t} + \cdots + a_kX_{kt} + u_t' \qquad (7.59)$$

여기에서 $u_t' = b_1\hat{\theta}_{3t} + b_2\hat{\theta}_{4t} + u_t$ 이다.

(7.52)와 $\hat{\theta}_{4t}$ 에 대한 관련조건으로부터 다음과 같게 된다.

$$\sum u_t' = \sum u_t, \qquad \sum (u_t'\hat{Y}_{2t}) = \sum (u_t\hat{Y}_{2t}),$$
$$\sum (u_t'\hat{Y}_{3t}) = \sum (u_t\hat{Y}_{3t}), \qquad \sum (X_{jt}u_t') = \sum (X_{jt}u_t) \qquad (7.60)$$

$(j = 1, \cdots, k)$

곧, (7.60)에서의 조건들은 다음을 의미한다. 추정을 위해 u_t와 동일한 방식으로 u_t'을 처리할 수 있다는 것이다.

가령, 이변수의 경우에서처럼 (7.60)에서의 조건들은 다음을 의미한다. (7.59)를 추정하기 위해 다음과 같이 놓게 된다는 것이다.

$$\sum \hat{u}_t' = \sum \hat{u}_t = 0 \qquad (\because E(u_t) = 0)$$

$$\left.\begin{aligned} \sum (\hat{u}_t' \hat{Y}_{2t}) &= \sum (\hat{u}_t \hat{Y}_{2t}) = 0 \\ \sum (\hat{u}_t' \hat{Y}_{3t}) &= \sum (\hat{u}_t \hat{Y}_{3t}) = 0 \end{aligned}\right\}$$

왜냐하면 u_t는 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$이 의존하는 모든 사전결정변수와 상관이 없기 때문이다.

$$\sum (\hat{u}_t' X_{jt}) = \sum (\hat{u}_t X_{jt}) = 0 \qquad (\because E(u_t X_{jt}) = 0 \qquad (7.61)$$

$$j = 1, 2, \ldots, k)$$

만일 통상의 절차로서 (7.61)의 조건들을 부과함으로써 (7.59)의 모수를 추정한다면 그에 따른 추정량들은 일치추정량임을 보일 수 있다.

요약하면, (7.59)를 (7.25)와 비교하면 또한 단지 Y_{2t}와 Y_{3t}를 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$으로 치환하여 교란항의 기호를 바꾼 다음 통례의 절차를 따르면 된다는 것이다.

이는 TSLS 절차의 매우 매력적인 속성이다. 관심을 가지는 방정식을 추정하기 위해 전방정식체계를 추정할 필요는 없는 것이다. 곧 모형에서의 여타 방정식에 대한 자료를 전부 규정하여 그것을 완전히 얻을 필요는 없다는 것이다. 필요한 것이란 내생독립변수(교란항과 상관이 있는 변수)들을 회귀하는 방정식체계로부터의 사전결정변수의 "적당한 집합(adequate set)"인 것이다. 사전결정변수의 "적당한 집합"이 의미하는것은 아래에서 논의할 것이다. 이 단계에서 지적할 것은 사전결정변수의 적당한 집합이 추정될 방정식에 나타나는 모든 사전결정변수를 반드시 포함하여야만 한다는 것이다. 가령 위의 (7.25)의 추정에서 Z_{1t}와 W_{1t}가

$\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 의 결정에 이용할 수 없다고 가정하였다. 그것들의 생략은 용인될 수 있다. 왜냐하면 Z_{1t} 도 W_{1t} 도 구조방정식 (7.25)에서 나타나지 않기 때문이다. 하지만 만일 X_{1t}가 $\hat{Y}_{2t}$ 와 $\hat{Y}_{3t}$ 을 만드는데 이용되지 않는다면, 위의 절차는 일치추정량을 이끌어 내지 못하게 될 것이다. 이를 알기 위해 유의해야 할 것은 만일 X_{1t}가 그렇게 이용되지 않으면 (7.52)로부터 $\Sigma (\hat{\theta}_{3t} X_{1t}) = 0$ 을 도출할 수 없다는 것이다. 왜냐하면 Y_{2t}는 X_{1t}에 대해 회귀하지 않았기 때문이다. 사실상 이 합계는 일반적으로 0이 아닐 것이다. 이러한 경우에 (7.60)에서 X_{1t}에 대한 관련방정식은 전형적으로 유지되지 않는다. 곧,

$$\sum (u_t' X_{1t}) \neq \sum (u_t X_{1t})$$

분명히 (7.61)에서 관련된 합을 0과 같다고 두는 것은 당연하지 않다.

마지막으로 지적할 점은 다음과 같다. 곧, TSLS 절차의 제 1 단계에서 체계의 모든 사전결정변수를 이용할 필요는 없지만 그렇게 하는데는 한가지 장점이 있다는 것이다. 일반적으로 TSLS절차의 제 1 단계에서 사용하는 모형의 사전결정변수가 많으면 많을수록 제 2 단계에서 만드는 계수추정량의 대표본분산이 보다 작아진다. 비록 그에 대한 증명은 이 책의 범위를 벗어나는 것이지만 그 결과는 그다지 놀라운 것은 아니다. 만일 추가적인 사전결정변수의 형태로 보다 많은 "정보"를 제 1 단계에서 이용하면 그에 따라서 제 2 단계의 추정량은 개선되기 마련이다. 이제 "識別의 問題 (identification problem)"로 알려진 것을 고려하여봄으로써 사전결정변수의 적당한 집합이 무엇으로 이루어져 있는지에 대한 문제로 돌아가기로 한다.

5. 識別의 문제 *

TSLS를 소개하는 가운데 그것의 한가지 장점은 방정식을 추정할 수 있으면 계수의 일치추정량을 제공해 준다는 의미에서 항상 TSLS는 적용되는 것임을 기억할 수 있을 것이다. 하지만 몇가지 (또는 어쩌면 모든) 모수의 값을 추정할 수 없는 방정식체계의 상황이 발생할 수도 있는 것이다.

가. 보기 1

다음과 같은 어느 특정상품의 시장에 대한 단순모형을 생각해 보자.

$$Q_t^d = a_0 + a_1 P_t + u_t \qquad (7.62)$$

$$Q_t^s = b_0 + b_1 P_t + \varepsilon_t \qquad (7.63)$$

$$Q_t^d = Q_t^s = Q_t \qquad (7.64)$$

모형은 세개의 방정식으로 이루어진다. 곧 수요함수(7.62), 공급함수(7.63), 그리고 시장균형식(7.64)가 그것들이다. 여기에서 Q_t^d와 Q_t^s는 각각 t시점의 수요량과 공급량이고, P_t는 t시점의 상품가격이며, u_t와 ε_t는 교란항으로 평균이 0이고, 분산이 일정하며 자기상관이 없는 동시에 정규분포를 하고 있다. 방정식(7.64)에 의해 Q_t^d와 Q_t^s가 t시점동안 실제로 교역이 이루어진 양, Q_t와 일치하는 것으로 규정할 수 있는 것이다. 그 의미는 각 시점동안 수요량이 공급량과 일치하도록 가격이 조정된다는 것이다.

---

* 이 주제에 대한 고전적인 저작은 다음과 같다. Franklin M.Fisher, The Identification Problem in Econometrics, (New York:McGraw-Hill.1966). 이하의 논의는 부분적으로 이 책을 따르고 있다.

이제 수요함수의 모수를 추정하는데 관심을 가져보기로 하자. 우선 P_t가 u_t와 상관이 있음을 지적하기로 한다. 그러면 표준적인 추정절차는 일치성이 없는 계수의 추정량을 제공하여 줄 것이다. 이를 알아보기 위해 방정식 (7.62)와 (7.63)의 우변의 식이 서로 같다고 규정하면 다음과 같게 된다. 곧,

$$a_0 + a_1 P_t + u_t = b_0 + b_1 P_t + \varepsilon_t \qquad (7.65)$$

P_t에 대해 (7.65)를 풀면 P_t에 대한 축소형방정식을 얻게 된다.

$$P_t = \frac{(b_0 - a_0)}{(a_1 - b_1)} + \frac{\varepsilon_t}{(a_1 - b_1)} - \frac{u_t}{(a_1 - b_1)} \qquad (7.66)$$

(7.66)에 u_t를 곱하여 기대치를 구하면 다음과 같다.

$$\begin{aligned} E(P_t u_t) &= E\left[\frac{(b_0 - a_0)u_t}{(a_1 - b_1)} + \frac{\varepsilon_t u_t}{(a_1 - b_1)} - \frac{u_t^2}{(a_1 - b_1)}\right] \\ &= \frac{(b_0 - a_0)}{(a_1 - b_1)} E(u_t) + \frac{E(\varepsilon_t u_t)}{(a_1 - b_1)} - \frac{E(u_t^2)}{(a_1 - b_1)} \qquad (7.67) \\ &= 0 + \frac{\mathrm{cov}(\varepsilon_t, u_t)}{(a_1 - b_1)} - \frac{\sigma_u^2}{(a_1 - b_1)} \end{aligned}$$

cov (ε_t, u_t)와 σ_u^2이 같다고 예상할 아무런 이유가 없기 때문에 일반적으로 다음을 가정할 수 있다.

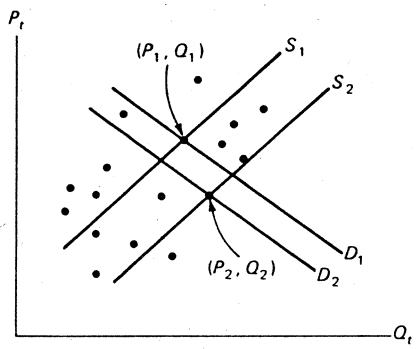
$$E(P_t u_t) = \mathrm{cov}(P_t, u_t) \neq 0 \qquad (7.68)$$

(7.68)은 일반적으로 0이 아니기 때문에 설명변수 P_t는 u_t와 상관이 있게 되고 따라서 체계문제가 존재한다. 이 문제를 해결하기 위해 이단계최소자승법(TSLS)에 의해 (7.62)를 추정하기로 하자. 제 1 단계는 모형의 모든 사전결정변수에 대해 내생독립변수(여기에서는 P_t)를 회귀하는 것으로 이루어짐을 기억할 것이다. 하지만 세 방정식모형을 잠깐

보면 체계에 사전결정변수가 하나도 없음을 알게 될 것이다.*

P_t에 대한 축소형방정식, 곧 (7.66)의 우변에는 하나의 사전결정변수도 없는 것이다. 따라서 TSLS 추정기법을 이용할 수 없다.

이러한 경우에 수요함수는 (공급함수도 마찬가지로) 식별되지 않으며 (unidentified), 이는 모수를 추정할 수 없음을 말한다. 유의할 점은 識別의 問題(identification problem)가 자료의 문제(data problem)가 아니라는 것이다. 곧 가격과 수량에 관한 관찰치를 아무리 많이 얻을지라도 여전히 수요방정식이나 공급방정식의 계수를 추정할 수 없는 것이다. 식별의 문제는 하나의 規定(specification)이다. 곧, 추정을 불가능하게 하는 것은 모형의 구조와 이용가능한 정보의 속성인 것이다.

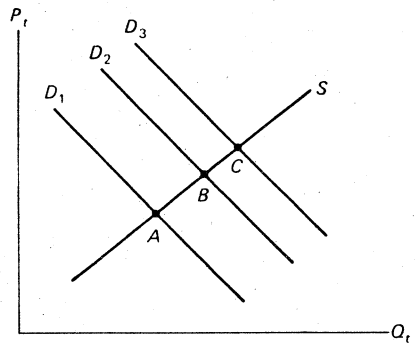


〈그림 7.3〉

---

* 실제로 다음에서 논의되듯이 상수항은 사전결정변수로 간주될 수도 있다. 만일, 그렇다고 한다면, P_t에 대한 축소형방정식(7.66)은 하나의 사전결정변수를 포함할 것이다. 하지만 앞으로 보게 되겠지만 수요함수(7.62)를 추정하는 것은 여전히 불가능하다.

이 결론을 지지하는 직관적인 서술을 알아보는 것도 도움이 될 것이다. 〈그림 7.3〉에 상이한 시점의 가격과 구매량에 대한 관찰치를 나타내는 散布圖가 있다. 이 점들은 수요곡선과 공급곡선을 추정하는데 필요한 정보들을 나타내고 있다. 당면한 문제는 산포도의 각 점이 수요곡선과 공급곡선 둘다에 의해 결정되는 것이다. 곧, ε_t와 u_t가 변화하기 때문에 수요와 공급스케줄은 시점에 따라 이동하며, 그에 따라 시장의 가격과 수량도 시점에 따라 변동한다. 이는 〈그림 7.3〉에서 P_1과 Q_1, P_2와 Q_2를 결정하는 시점 1과 시점 2에서의 수요곡선과 공급곡선의 두 집합에 의해 묘사되고 있다. 관찰할 것은 〈그림 7.3〉에서와 같이 흩어진 점들이다. 하지만 단지 수요의 변화로부터 결과한, 그리고 알고 있는 점이란 하나도 없는 것이다. 만일 있다고 한다면 공급곡선의 계수를 추정하기 위해 이러한 점들을 이용할 수 있는 것이다. 예를 들어 〈그림 7.4〉에서 점A, B, C가 동일한 공급곡선을 따라 이동하는 세가지 상이한 수요곡선에 의해 만들어지는 것임을 안다면 직관적으로 보아 공급곡선의 기울기와 절편모수를 추정하기 위해 이 세점을 이용할 수 있음이 분명한 것이다. 그러나 〈그림 7.3〉에서 단지 수요의 변화로부터 결과한 그러한 점들을 구분해낼 방도가 없기 때문에 수요곡선이나 공급곡선을 "식별할" 수 있는 방법이 〈그림 7.3〉의 점들로부터는 없다는 것을 알게 된다.



〈그림 7.4〉

나. 보기 2

다음으로 공급-수요모형의 변형을 생각해 보기로 한다.

$$Q_t^d = a_0 + a_1 P_t + a_2 P_{t-1} + u_t \qquad (7.69)$$

$$Q_t^s = b_0 + b_1 P_t + b_2 P_{t-1} + \varepsilon_t \qquad (7.70)$$

$$Q_t^d = Q_t^s = Q_t \qquad (7.71)$$

세 방정식모형, (7.69)-(7.71)은 시차가격변수 P_{t-1}이 수요방정식과 공급방정식에서 함께 나타난다는 예외를 제외하고는 앞의 모형과 일치한다.* 교란항이 자기상관이 없다고 가정하였기 때문에 P_{t-1}은 u_t 또는 ε_t와 상관이 없을 것이고, 따라서 사전결정변수로 취급될 것이다. 하지만 현재 가격변수 P_t는 또한 u_t와 ε_t와 상관이 있으며, 따라서 수요방정식과 공급방정식을 추정하기 위해서는 수정된 기법을 이용해야 할 것이다.

수요방정식의 모수를 추정하기 위해 TSLS를 이용하기로 하여보자. 우선 모형의 모든 사전결정변수, 이 경우에는 P_{t-1}에 대해 P_t를 회귀할 것이다. 그리고 나서 P_t의 계산치를 만들 것이다. 그것은 가령 다음과 같다.

$$\hat{P}_t = \hat{c}_0 + \hat{c}_1 P_{t-1} \qquad (7.72)$$

그러면 (7.69)에서 P_t를 $\hat{P}_t$로 치환하여 제2단계에서 다음의 방정식을 추정하려 할 것이다.

$$Q_t^d = a_0 + a_1 \hat{P}_t + a_2 P_{t-1} + u_t^* \qquad (7.73)$$

---

* P_{t-1}의 존재는 구매자와 판매자가 어느 정도 과거의 습관 또는 정보에 기초하여 행동하는 것을 가리키거나 또는 P_t와 P_{t-1} 모두 기대치에 영향을 주는 의사결정에 예상미래가격(expected future prices)이 영향을 미친다는 것을 가리킨다.

여기에서 물론 $Q_t{}^d = Q_t$ 이다. 하지만 (7.72)에 의거해 볼 때, 또한 추정절차가 무너지게 된다. 왜냐하면 $\hat{P}_t$ 의 값은 P_{t-1} 의 값과 완전하게, 그리고 선형적으로 관련이 있을 것이기 때문이다. 이는 완전다중공선성의 경우로 이 때문에 a_o, a_1 과 a_2 의 각각의 추정량을 구할 수 없을 것이다. 여전히 공급방정식과 수요방정식은 식별되지 않는다.

다. 보기 3

공급-수요모형의 세번째 형태를 알아보기로 하자.

$$Q_t^d = a_0 + a_1 P_t + a_2 Y_t + u_t \qquad (7.74)$$

$$Q_t^s = b_0 + b_1 P_t + \varepsilon_t \qquad (7.75)$$

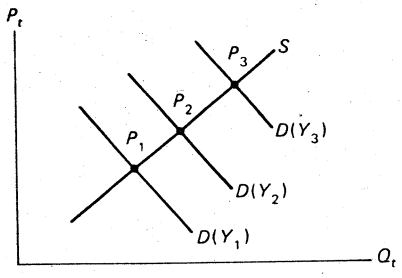
$$Q_t^d = Q_t^s = Q_t \qquad (7.76)$$

이 모형은 추가변수인 소득수준(Y_t)이 수요방정식에서 나타난다는 사실을 제외하면 원래의 모형과 유사하다. 이 예의 목적에 따라 Y_t 는 u_t 와 ε_t 모두와 무관하다고 가정한다. 이 경우에 수요함수나 공급함수를 추정할 수 있는지를 알아보기로 하자. 우선 수요방정식(7.74)을 고려하기로 한다. 상수항을 별도로 하면 체계내에는 사전결정변수가 한개 있다. 곧, 그것은 Y_t 이다. 그러므로 우선 $\hat{P}_t$ 를 가령 $\hat{P}_t = \hat{h}_o + \hat{h}_1 Y_t$ 로 얻기 위해 Y_t 에 대해 P_t 를 회귀한다. 수요방정식(7.74)에서 P_t 를 $\hat{P}_t$ 로 치환하면 앞의 경우에서와 마찬가지로 a_o, a_1 과 a_2 에 대한 추정량을 구할 수 없음을 알게 된다. 왜냐하면 완전다중공선성이 존재하기 때문이다. 하지만 공급방정식(7.75)의 계수들은 추정할 수 있다. 곧, (7.75)에

서 P_t 를 $\hat{P}_t$ 로 대체하면 TSLS 절차를 이용할 수 있게 된다. 왜냐하면 완전다중공선성의 문제가 없기 때문이다. $\hat{P}_t$ 에 대해 Q_t 를 회귀함으로써 b_0 와 b_1 의 일치추정량을 얻을 수 있는 것이다. 이런 경우에 공급방정식의 계수는 추정할 수 있지만, 수요방정식의 계수는 추정할 수 없다.

이러한 결과에 대한 직관적인 정당화 또는 명확화는 도형으로도 할 수 있다. 〈그림 7.3〉과 비슷한 P_t 와 Q_t 의 산포도가 또한 있다고 하자. 하지만 이제 단지 수요만의 변화로부터 결과한 점들의 분리와 관련되는 추가정보가 있는 것이다. 특히 교란항 u_t 와 ε_t 가 변화할 때 수요곡선과 공급곡선이 이동한다는 사실을 방정식에 대한 규정으로부터 알고 있다. 하지만 이러한 교란항의 효과뿐만 아니라 Y_t 가 변화하면 공급곡선이 아니라 수요곡선이 이동하게 될 것이다. 산포도에 의하면 이는 다음을 의미한다. 곧, 교란항이 0에서 불변인 채로 있으면 공급곡선을 긋는 Y_t 의 상이한 값들과 관련된 일련의 점들을 관찰하게 된다는 것이다. 그러한 상황은 〈그림 7.5〉에 묘사되고 있다. 이 경우에 교란항이 반드시 0은 아닐지라도 그 대신에 시점에 따라 상이한 값을 취하게 된다는 사실을 어느 정도 설명할 수 있다면 공급곡선을 추정할 수 있게 된다. 어느 정도 직관적으로 말하면 이러한 교란항의 변동은 교란항이 소득수준 Y_t 와는 무관하며, 따라서 그것과 상관이 없다는 가정에 의해 설명되는 것이다. 이 조건과 교란항의 평균이 0이라는 가정에 의해 추정기법을 이용하여 다소간 교란항의 영향을 평균할 수 있게 된다. 그 대신에 〈그림 7.5〉에서의 곡선의 각각의 소득수준과 관련된 "평균(mean)" 곡선으로 간주할 수 있을 것이다. 곧, Y_t 의 주어진 어떠한 값에 대해서도 $E(u_t)=E(\varepsilon_t)=0$ 이며, 따라서 〈그림 7.5〉의 곡선의 위치에 대한 교란항의 영향은 0인 것이다. 하지만 유의해야 할 점은 다음과 같다. 곧, 수요곡선을 추정할 수 있음을 뜻하는 논거는 없다는 것이다. 왜냐하면 공

급곡선에서의 변수로서 (수요함수에서는 나타나지 않는 것에 의해) 공급곡선의 이동이 수요곡선을 따라 발생하도록 하는 것은 하나도 없기 때문이다. 이러한 세번째의 경우에 공급방정식은 식별되지만 수요방정식은 식별되지 않는 것이다.



〈그림 7.5〉

라. 보다 일반화된 표현

이제 훨씬 더 일반적인 예를 몇가지 생각해 보기로 한다. 어떤 연립방정식체계의 첫번째 구조방정식이 다음과 같다고 하여 보자.

$$Y_{1t} = b_0 + b_1 X_{1t} + b_2 Y_{2t} + b_3 Y_{3t} + \varepsilon_t \qquad (7.77)$$

여기에서 Y_{1t}, Y_{2t}와 Y_{3t}는 내생변수이고, X_{1t}는 사전결정변수이며, 교란항 ε_t는 통상적인 모든 가정을 만족한다. 또한 이 방정식이 그 일부분이 되는 전체체계가 단지 하나의 추가사전결정변수, 가령 X_{2t}를 포함하고 있다고 가정하자. 그러면 (7.77)을 추정하기 위해 TSLS절차를 이용할 것이다. 왜냐하면 Y_{2t}와 Y_{3t}는 ε_t와 상관이 있는 것으로 기대될 것이다. X_{1t}와 X_{2t}에 대해 Y_{2t}와 Y_{3t}를 회귀하면 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$을 얻게

될 것이다.

$$\hat{Y}_{2t} = \hat{\gamma}_0 + \hat{\gamma}_1 X_{1t} + \hat{\gamma}_2 X_{2t} \qquad (7.78)$$

$$\hat{Y}_{3t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} \qquad (7.79)$$

그러면 (7.78)과 (7.79)를 (7.77)에 대입하면, 제 2 단계회귀모형은 다음과 같을 것이다.

$$Y_{1t} = b_0 + b_1 X_{1t} + b_2 \hat{Y}_{2t} + b_3 \hat{Y}_{3t} + \varepsilon_t^* \qquad (7.80)$$

여기에서 ε_t^*는 새로운 교란항이다. 그러면 통례대로 (7.80)을 추정하게 될 것이다. 하지만 또한 절차는 설명변수 사이에 완전다중공선성때문에 무너지게 된다. 이를 알아보기 위해 방정식(7.78)에 $\hat{\alpha}_2$를, 방정식(7.79)에 $\hat{\gamma}_2$를 곱하고, 전자로부터 후자를 제하여 X_{2t}항을 소거하면 다음을 얻게 된다.

$$\hat{Y}_{2t}\hat{\alpha}_2 - \hat{Y}_{3t}\hat{\gamma}_2 = (\hat{\alpha}_2\hat{\gamma}_0 - \hat{\alpha}_0\hat{\gamma}_2) + (\hat{\gamma}_1\hat{\alpha}_2 - \hat{\alpha}_1\hat{\gamma}_2)X_{1t} \qquad (7.81)$$

방정식(7.81)은 $\hat{Y}_{2t}$, $\hat{Y}_{3t}$과 X_{1t}의 값들사이에 완전선형관계가 존재함을 가리킨다. 이에 따라 하나의 정규방정식은 여타의 것들과 선형적으로 독립이 아니며, 따라서 (7.80)의 모수를 추정할 수 없는 것이다.

(7.80)과 관련된 정규방정식들을 고려하여 봄으로써 이 결과를 이해할 수 있다. $\Sigma \hat{\varepsilon}_t^* = 0$과 $\Sigma(\hat{\varepsilon}_t^* X_{1t}) = 0$으로 놓으면 처음의 두 정규방정식을 얻게 된다. $\Sigma(\hat{\varepsilon}_t^* \hat{Y}_{2t}) = 0$으로 놓음으로써 얻게 되는 세번째 정규방정식은 처음의 두 방정식과는 무관하다. 왜냐하면 $\hat{Y}_{2t}$이 X_{2t}에 의존하기 때문이다. 〔(7.78)을 볼 것〕 그런데 만일 그렇지 않다고 한다면, 세번째 방정식은 첫번째를 $\hat{\gamma}_0$배한 것과 두번째를 $\hat{\gamma}_1$배한 것을 합한 것과 동일할 것이다. 어떤 의미에서 X_{2t}는 처음의 두 정규방정식

으로부터 세번째 정규방정식을 "분리한다". 하지만 $\Sigma(\hat{Y}_{3t}\,\hat{\varepsilon}_t^*)=0$으로 놓음으로써 얻는 네번째 정규방정식에 이르면 어려움은 깊어진다. 실제로 이미 처음의 두 정규방정식으로부터 세번째 정규방정식을 분리해내기 위해 X_{2t}를 사용해 버린 것이다. 다시 처음의 세 정규방정식으로부터 네번째 정규방정식을 분리해내기 위해 그것을 사용할 수 있을까? * 어느 정도 직관적으로 말하면, 각각의 선형독립정규방정식은 새로운 정보, 곧 새로운 변수를 요구한다. 하지만 남아있는 것이 하나도 없기 때문에 (7.80) 또는 (7.77)의 모수들은 독립적인 네번째 정규방정식의 누락으로 인하여 식별되지 않는다.**

(7.77)을 포함하는 방정식체계가 두개의 추가사전결정변수, 가령 X_{2t}와 X_{3t}를 포함한다고 가정하자. 그러면 제1단계에서 (7.78)과 (7.79)는 하나의 추가독립변수, 곧 X_{3t}를 포함할 것이다. 앞에서와 마찬가지로 이러한 제1단계의 방정식으로부터 X_{2t}를 소거하면 $\hat{Y}_{2t}$, $\hat{Y}_{3t}$, X_{1t}과 X_{3t}과 관련된 선형방정식을 얻게 될 것이다. 이러한 X_{3t}항 때문에 $\hat{Y}_{3t}$은 X_{1t}와 $\hat{Y}_{2t}$의 완전결합(perfect combination)이 아닐 것이다.

---

* (7.81)에 비추어 네번째 정규방정식은 첫번째의 $(\hat{\alpha}_0\hat{r}_2-\hat{\alpha}_2\hat{r}_0)/\hat{r}_2$배, 두번째의 $(\hat{\alpha}_1\hat{r}_2-\hat{r}_1\hat{\alpha}_2)/\hat{r}_2$배, 세번째의 $\hat{\alpha}_2/\hat{r}_2$배를 모두 합한 것과 일치할 것임을 당연히 보일 수 있어야 한다.

** 이러한 경우의 보다 간단한 실례는 다음의 모형이다.

$$Y_{1t}=b_0+b_1Y_{2t}+b_3Y_{3t}+u_t,$$

여기에서 Y_{2t}와 Y_{3t}는 내생변수이며, 방정식체계의 나머지부분에는 단 하나의 사건결정변수(가령 Z_t)만이 존재한다. 이러한 경우에 제1단계의 회귀와 계산은 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$을 산출할 것이다. 하지만 $\hat{Y}_{2t}$과 $\hat{Y}_{3t}$의 값은 Z_t의 값의 정확한 선형결합이며, 서로간에 완전상관관계에 있을 것이다. 설명변수사이의 완전다중공선성때문에 추정절차는 분명히 실패하게 될 것이다.

이러한 경우에 TSLS절차의 제 2 단계에서 완전다중공선성이 존재하지 않을 것이며, 이제 방정식 (7.77)은 식별될 것이다. 두개 이상의 추가사전결정변수가 방정식 (7.77)이 그 일부분이 되는 체계에 나타나면 이러한 결론들은 분명히 입증된다. 곧 〔방정식 (7.77)에는 나타나지 않는〕 체계의 추가사전결정변수의 수가 내생설명변수의 수인 2 보다 크거나 같으면 고려중인 방정식이 식별되는 것으로 보인다.

마지막으로 한가지 지적할 점은 다음과 같다. $\Sigma \varepsilon_t^* = 0$으로 놓음으로써 얻는 첫번째 정규방정식은 상수항과 관련이 있는 것으로 생각해볼 수 있다.*

그 대신에 상수항을 하나의 사전결정변수로 생각해 볼 수 있다. 그와 같은 경우 아래의 예가 지적하고 있듯이 추정하고 있는 방정식은 상수항을 가지고 있지 않지만 체계의 여타 방정식중의 하나 또는 그 이상이 상수항을 가지면, 상수항은 추정할 방정식에는 나타나지 않는 체계의 하나의 사전결정변수로 간주될 것이다.

예를 들어 두 방정식 모형을 생각해 보기로 한다.

$$Y_{1t} = b_1 Y_{2t} + b_2 X_t + \varepsilon_{1t} \qquad (7.82)$$

$$Y_{2t} = a_1 + a_2 Y_{1t} + \varepsilon_{2t} \qquad (7.83)$$

---

* 가령 방정식 (7.80)은 다음과 같이 표현할 수 있다.

$$Y_{1t} = b_0 X_{0t} + b_1 X_{1t} + b_2 \hat{Y}_{2t} + b_3 \hat{Y}_{3t} + \varepsilon_t^*$$

여기에서 X_{0t}는 매시점마다 1인 것으로 정의된다. 곧, 모든 t에 대해 $X_{0t} \equiv 1$이다. 그러므로 $\Sigma \hat{\varepsilon}_t^* = 0$이라는 조건은 $\Sigma (\hat{\varepsilon}_t^* X_{0t}) = 0$으로 생각해 볼 수 있는 것이다. (다음 페이지 계속)

여기에서 X_t는 사전결정변수이며, ε_{1t}와 ε_{2t}는 기본가정을 만족하는 교란항이다. 우선 보기에 독자들은 (7.82)에 TSLS절차를 적용시킬 수 없는 것이 아닌가하고 생각할지도 모른다. 왜냐하면 체계내에 사전결정변수는 X_t로 하나밖에 없으며, 또 그 변수는 추정하고자 하는 방정식에서 나타나기 때문이다. 하지만 상수항이 (7.82)에서 배제되어 버리기 때문에 TSLS를 이용하여 (7.82)를 추정할 수 있음을 알고 있다. 왜냐하면 어떤 의미에서 배제된 상수항은 추가로 필요한 "사전결정"변수를 제공하는 것이기 때문이다. 가령 TSLS절차의 제 1 단계로부터 다음을 얻게 될 것이다.

$$\hat{Y}_{2t} = \hat{c}_1 + \hat{c}_2 X_t \qquad (7.84)$$

조건 (7.84)는 다음과 같이 놓음으로써 얻은 세가지 정규방정식중에서 두가지만이 독립적임을 의미한다.

$$N_1: \sum \hat{\varepsilon}^*_{1t} = 0, \quad N_2: \sum (\hat{\varepsilon}^*_{1t} \hat{Y}_{2t}) = 0, \quad N_3: \sum (\hat{\varepsilon}^*_{1t} X_t) = 0 \qquad (7.85)$$

그러나 (7.82)가 상수항을 포함하지 않기 때문에 단지 두개의 정규방정식만이 필요한 것이다. (7.85)에서 N_2와 N_3와 관련된 방정식만을 취하게 될 것이다. 왜냐하면 이러한 조건들은 방정식 (7.82)에서의 설명변수와 부합하기 때문이다.

마. 일반적인 설명 *

이러한 것을 바탕으로 하여, 이제 TSLS를 써서 일치추정량을 얻는 일반적인 규칙을 설명하기로 하자. 곧, 회귀방정식의 모수의 일치추정량을

---

* 여기에서 전개하고 있는 일반적인 설명은 분석의 맥락내에서 타당한 것이다. 예를 들면 연구자가 (다음 페이지 계속)

얻기 위해 TSLS를 이용하려면 추정할 방정식에서 설명변수로 나타나는 내생변수의 수가 모형 전체에서 나타나면서 그 방정식에서는 배제되는 사전결정변수의 수를 초과할 수 없다는 것이다. 그 대신에 일반적으로 어떤 모형의 어느 특정방정식은 $K_2 \geq K_1$ 이면 식별된다(그리고 그 추정량은 일치성을 지닐 수 있다.). 여기에서 K_2 는 주어진 방정식으로부터 배제된 모형의 사전결정변수의 수와 같으며, K_1 은 그 방정식에서 설명변수로서 나타나는 내생변수의 수이다.**

앞의 예들을 돌이켜보면 다음의 사실을 알게 된다. 곧, 단순소득결정모형에서 설명변수(Y_t)로서 하나의 내생변수가 존재하고, 모형의 일부분이지만 소비방정식에서는 나타나지 않는 하나의 사전결정변수(I_t)가 존재하기 때문에 소비방정식을 추정할 수 있었다는 것이다. 처음의 두 공급-수요모형에서는 공급방정식이나 수요방정식을 추정할 수 없었다. 이러한 두 경우 모두 하나의 내생설명변수는 있지만 배제된 사전결정변수는 하나도 없었던 것이다. 마지막으로 세번째 예에서는 공급방정식이 식별되었다. 왜냐하면 공급함수의 하나의 내생설명변수(P_t)가 단지 수요방정식에서만 나타나는 사전결정변수(Y_t)와 짝지어지기 때문이다. 하지만 수요함수로부터 배제된 사전결정변수는 없으며, 따라서 그 방정식을 추정할 수 없음을 알았다.

(앞페이지 계속) 어떤 분산들의 값, 특정 모수사이의 관계 등을 알고 있는 경우는 고려하지 않았던 것이다. 하지만 여기에서 서술한 분석은 실제로 대다수의 경우에 만나게 되는 것이라고는 말할 수 없을 것이다.

** 이것들은 실제로 방정식을 식별하기 위한 필요조건이지만 충분조건은 아니기 때문에 "일반적으로"라는 문구를 사용해야만 한다. 〔다음을 볼 것. Arthur S.Goldberger, Econometric Theory(New York:Wiley, 1964), pp.306-329〕. 충분조건에 관한 논의는 이 책의 범위를 벗어난다. 하지만 실제적인 관점에서 볼 때, 항상 실제로 고려되는 조건은 여기에서 논의한 것들이다.

이러한 규칙의 이론적 근거는 앞의 몇가지 예로부터 분명하다. 방정식으로부터 배제된 사전결정변수의 수가 불충분한 경우에 제 2 단계에서의 완전다중공선성 때문에 TSLS 추정절차는 붕괴될 것이다. 결론적으로 이것은 그 자체로는 TSLS의 특유한 결점은 아니라는 것을 강조하고자 한다. 곧 다른 추정기법들도 식별되지 않는 방정식의 일치추정량들을 또한 제공할 수 없다. 변수사이의 상호의존을 해결할 수 없는 것은 모형 자체의 구조와 이용가능한 정보의 속성때문이다.

6. TSLS 추정 : 두가지 실례

TSLS 추정절차에 보다 익숙해지기 위해 두가지 실례를 생각하여봄으로써 이 장의 결론을 맺기로 한다. 어떤 특정상품에 대한 가설적인 수요곡선에 대한 추정을 포함하는 첫번째 실례는 단지 기법을 철저히 수행할 수 있도록 해주는 것이다. 두번째 실례는 지방재정에 관한 최근의 잡지논문으로부터의 실증연구를 따온 것이다.

가. 수요와 공급모형

먼저 다음과 같은 농산물(가령 아티초크)시장의 단순모형을 고려하여 보기로 한다.

$$Q_t^d = b_0 + b_1 P_t + b_2 Y_t + u_t \tag{7.86}$$
$$Q_t^s = a_0 + a_1 P_{t-1} + a_2 W_t + e_t \tag{7.87}$$
$$Q_t^d = Q_t^s = Q_t \tag{7.88}$$

방정식 (7.86), 곧 수요함수는 t 시점의 수요량은 가격 (P_t)과 소득 (Y_t)에 의존한다는 것을 가리킨다. 공급방정식 (7.87)은 아티초크의 생산이 (파종결정을 내려야 하는) 전기의 가격 P_{t-1}과 기후 W_t에 의존함을 말한다.

기후조건의 측도로서는 인치(inches) 단위의 강우량을 사용할 것이다. 마지막으로 표준적인 시장청산방정식(7.88)은 t시점의 가격이 수요량과 공급량이 일치할 수 있도록 조정됨을 의미한다.

Y_t와 W_t가 체계밖의 요인들에 의해 결정된다고 가정하자. 곧, 그것들은 외생변수이다. 더구나 교란항 u_t와 e_t는 평균이 0이고 분산이 일정하며 정규분포를 하고, 자기상관이 없으며, 사전결정변수에 독립적이라고 가정하자. 이제 아티초크의 수요함수의 계수를 추정하기로 한다. 우선 유의해야 할 것은 하나의 설명변수, 곧 P_t가 모형의 교란항과 상관이 있을 것으로 예상하는 내생변수라는 것이다. OLSQ의 이용은 일치성이 없는 계수의 추정량을 낳기 때문에 TSLS를 이용하여 (7.86)을 추정하기로 한다. 두번째로 지적할 것은 수요방정식이 식별된다는 것이다.

곧, 그것은 하나의 내생설명변수를 포함하지만 체계에서는 나타나나 수요방정식에서는 나타나지 않는 두개의 사전결정변수(P_{t-1}과 W_t)가 존재한다는 것이다. 이러한 변수들에 대한 자료가 있다고 가정하면 TSLS를 이용하여 이러한 수요함수를 추정할 수 있는 것이다.

모형의 변수들에 대한 일련의 가설적인 자료는 〈표7.1〉에 나타나 있다. 유의할 점은 단지 관찰치가 10개로 "대표본"의 결과를 제공하는 추정절차의 이용을 정당화시키기에는 충분하지 못하다는 것이다. 이 실례의 목적은 단지 어떤 실제의 수를 이용하여 TSLS 추정절차의 단계를 완전히 수행하는 것임을 강조할 것이다.

첫번째 작업은 체계의 모든 사전결정변수에 대해 수요함수의 내생독립변수(곧, P_t)를 회귀하는 것이다. 상수항을 제외하면 그러한 사전결정변수는 세개 있다. 곧, 두개의 외생변수(Y_t와 W_t)와 한개의 시차내생변수(P_{t-1})가 그것이다. 추정회귀방정식은 다음과 같다.

$$\hat{P}_t = -8.60 + 3.75Y_t - 0.22W_t + 0.42P_{t-1} \qquad (7.89)$$

이제 (7.89)를 이용하여 P_t에 대한 일련의 "수정된" 값을 계산하게 된다. 곧, 각 시점 $t=1, \cdots, 10$에 대해 Y_t, W_t, P_{t-1}의 값을 (7.89)에 대입하여 $\hat{P}_t$에 대한 값을 계산하는 것이다. (관찰치 P_t와 함께) 가격변수에 대한 이러한 일련의 새로운 "정화된" 값들은 〈표7.2〉에 나타나고 있다.* TSLS의 제2단계는 수요방정식에서 P_t를 $\hat{P}_t$으로 치환한 다음 단지 통례대로 새로운 방정식을 추정하는 것이다. Y_t와 $\hat{P}_t$에 대해 Q_t를 회귀하면 다음과 같다.

$$\hat{Q}_t = \underset{(2.9)}{-39.9} - \underset{(3.7)}{1.3\hat{P}_t} + \underset{(4.1)}{9.5Y_t} \qquad (7.90)$$

여기에서 괄호안의 숫자는 해당하는 t비율의 절대치이다. 그러므로 $\hat{b}_0 = -39.9$, $\hat{b}_1 = -1.3$, 그리고 $\hat{b}_2 = 9.5$인 것이다.

만일 아티초크의 수요를 추정하기 위해 단지 通常 最小自乘法(OLSQ)을 이용하였다면, (곧, $\hat{P}_t$ 대신에 P_t를 이용하였다면) 다음의 결과를 얻었을 것이다.

$$\hat{Q}_t = \underset{(1.9)}{-25.1} - \underset{(2.1)}{0.7P_t} + \underset{(2.8)}{6.2Y_t} \qquad (7.91)$$

---

* $\hat{P}_t$에 대한 관찰치가 단지 9개임을 유념해야 한다. 곧, (7.89)에서 시차변수 P_{t-1}을 이용하는 과정에서 하나의 관찰치를 "상실하는" 것이다.

〈표 7.1〉 아티초크시장에 대한 가설적 자료[a]

시점 (t)	수량 (Q)	단위가격 (P)	소득 (Y)	강우량 (W)
1	11	20	8.1	42
2	16	18	8.4	58
3	11	22	8.5	35
4	14	21	8.5	46
5	13	27	8.8	41
6	17	26	9.0	56
7	14	25	8.9	48
8	15	27	9.4	50
9	12	30	9.5	39
10	18	28	9.9	52

a 이러한 변수들에 대한 단위는 다음과 같다.

Q = 아티초크의 톤수

P = 아티초크의 단위가격 (센트)

Y = 평균 1 년가계소득 (1 천달러)

W = 연도별 강우량 (인치)

(7.90)과 (7.91)의 추정계수들 사이의 차이에 유의해야 할 것이다. TSLS 방정식은 OLSQ 방정식 보다 아티초크의 수요가 가격의 변화와 소득의 변화에 훨씬 더 민감하다는 것을 가리킨다. 표본의 관찰치의 수가 작다는 사실에 비추어 볼 때, 이와 같은 경우에 TSLS 추정치가 보다 신뢰할만한 것이라고 아주 믿을 수는 없었던 것이다. 하지만 이러한 결과

〈표 7.2 〉

지 점	P_t	$\hat{P}_t$
1	20.0	-
2	18.0	18.5
3	22.0	23.1
4	21.0	22.4
5	27.0	24.2
6	26.0	24.2
7	25.0	25.1
8	27.0	26.1
9	30.0	29.8
10	28.0	29.7

들이 변수들에 대한 관찰치의 대표본에(그리고 물론 가설적이라기 보다는 실제적인 자료에) 기초하고 있다면, OLSQ추정량의 不一致性 때문에 TSLS 기법에 기초한 결과들을 선호할 충분한 이유가 있는 것이다.

나. 지방재정 모형

몇가지 가설적인 수치를 이용하여 TSLS기법을 완전히 수행하여 보았으므로 다음으로 TSLS기법을 이용하는 실증연구로 돌아갈 것이다. 경제학자들에게 뿐만 아니라 지방재정공무원들에게 진정한 관심거리는 지방재정정책이 개인의 의사결정에 미치는 영향이다. 예를 들면 높은 지방재산세가 잠재적인 새로운 거주민과 기업의 진입을 억제할 것인가? (실제의, 그리고 잠재적) 주민들은 지방학교의 자질에 얼마만큼 관심을 가질 것인가? 그들의 주거지결정에 영향을 미치기에 충분한가? 이것들은 대답하기에 쉽

지 않은 문제들이지만 몇가지 명백한 이유로 인하여 지방공무원들의 걱정거리인 것이다.

몇년전, 찰스 티보(Charles Tiebout)는 이러한 몇가지 논점을 다루는 이론적인 지방재정모형을 진전시켰다.* 티보는 많은 수의 지방정부로 구성된 하나의 체계를 전제하였는데 그것들은 상이한 공공서비스라는 산출물을 제공하며, 자동차소비자들은 이러한 서비스에 대한 자신의 선호에 따라 주거할 지역을 선정하는 것이었다. 예를 들면 교육에 대해 높게 수요하는 사람이면 아마도 상위의 학교가 있는 지역에 함께 모일 것이다. 티보모형의 매력은 다음과 같다. 곧, 그것은 사람들이 지방서비스에 대한 수요 곧 그들의 주거지결정을 표현하고 만족하는 메카니즘을 제공하는 것이다. (그것이 때로는 "발로 투표한다"로 표현되듯이)

하지만 실증의 차원에서 사람들이(적어도 몇몇 사람들이) 이러한 방식으로 행동하는지의 여부에 대한 문제가 남아있다. 티보모형에서 생각하는 완전한 이동성에 대한 장애는 많이 있다. 곧 이동비용, 직장의 위치, 통근비용 등이 그것이다. 그럼에도 불구하고 대도시권내의 구조와 이동성을 볼 때 티보유형의 행태는 완전히 받아들이기 어려운 것은 아니다. 곧, 중심도시에서 일하는 개인들은 흔히 주거지로 도시근교의 지역을 폭넓게 선정하며, 가령 지방공립학교의 자질이 주거지역의 선정에 실질적으로 중요할 수도 있다.

그러나 그러한 행태의 존재에 대한 가설검정을 어떻게 할 것인가? 한가지 접근법은 다음과 같다. 곧, 티보유형의 행태가 중요하다면 관찰하기로 예상하는 것이 무엇인지를 자문하는 것이다. 만일 실제로 학교자질과

---

* Charles Tiebout, "A Pure Theory of Local Expenditures," Journal of Political Economy, 64 (Oct. 1956), pp. 416-424

세금부담과 같은 재정변수들이 개인의 주거지결정에 중요한 변수라면, 그것들은 재산가치에 어느정도 영향을 미치는 것으로 예상할 수도 있는 것이다. 예를 들면, 상위학교가 있는 지역은 다른 조건이 일정하다면 살기에 상대적으로 바람직한 장소일 것이다. 곧, 사람들은 그러한 지역에 주거를 정하려고 하는 가운데 다른 지역에서 그 수준 이상의 재산가치는 억제하게 될 것이다. 마찬가지로 다른 조건이 일정하다면, 높은 세율은 지방의 재산가치를 감소시키려는 경향이 있다. 요약하면 티보의 세계에서는 재산가치에서의 지역간 차이에 의해 지역사이의 재정격차가 드러나게 됨을 예상하게 된다. 이는 상이한 지역에 대한 재정변수와 재산가치 사이의 관계를 검토할 것임을 의미한다. 다음과 같은 일반적인 형태의 관계를 전제할 수 있을 것이다.

$$V_t = b_0 + b_1 X_{1t} + \cdots + b_k X_{kt} + b_{(k+1)} Z_{1t} + \cdots + b_{(k+l)} Z_{lt} + b_{(k+l+1)} T_t + u_t \qquad (7.92)$$

여기에서

V_t: t 번째 지역의 재산가치

X_{1t}, …, X_{kt} : 지방재산가치에 영향을 주는 비재정변수

Z_{1t}, …, Z_{lt} : l 개의 상이한 공공서비스의 산출물수준

T_t : 지방재산세율

u_t : 교란항

그러면 어떤 지역표본으로부터의 자료를 이용하여 이 방정식을 추정할 수 있으며, 공공서비스, 조세변수와 지방재산가치 사이의 어떤 체계적인 관계가 발견되는지를 알아볼 수 있다.

이제 그러한 연구를 서술해 보기로 한다.* 이 연구는 (보다 큰 뉴욕 대도시권내의 모든 지역인) 북동뉴저지의 53개 지방도시의 표본을 이용하여 (7.92)와 유사한 방정식을 추정하였다. 미국인구와 주거에 관한 국세조사는 미국도시의 인구와 주거의 특성과 관계되는 충분한 양의 정보를 제공하고 있다. 뉴욕시 자료집으로부터 지방지출과 세율에 대한 예산정보를 이 자료에 보충하면 (7.92)의 변수의 측정치를 모을 수 있다. 지방재산 가치의 지표로서 이 연구는 다음을 이용하였다.**

V_t = t번째 도시에서의 소유주택의 중위수.

주어진 지역의 주거단위가치는 분명히 많은 비재정변수에 의존한다. 곧, 지방도시의 중심도시에의 근접성, 주택단위의 물리적 특성과 여러가지 무형의 "환경적" 고려사항들이 그것이다. 이것들은 (7.92)에서 X_{it}인 것이다. 연구는 이러한 영향들을 측정하기 위해 설명변수로서 마일로 측정한 뉴욕시로부터의 지방도시의 거리(M_t), 지방도시의 소유주택당 방수의 중위수(R_t), 주택공급스톡의 연수의 측도로서 1950년 이래 세워진 지방도시의 주택의 비율(N_t), 그리고 지역의 무형의 특성에 대한 대리변수로서 지방도시의 중위가계소득(Y_t)을 이용하였다. 여기에서의 가정은 보다 높은 소득을 받는 가계가 보다 "매력적인" 지역에서의 거주지를 선택하는 경향이 있다는 것이다. 곧, 그에 따라 중위가계소득은 지역거주의 무형의 특성치를 나타낸다. 재정변수의 경우에 연구는 지방재산에 대한 실효세율 또는 실질세율(T_i) (과세의 다양화를 위해 수정한 명목세율)을 포함하며, 지방서비스의 측도로서 지방공립학교의 학생당 지출(E_t)을 포함하고 있

---

* W.E.Oates, "The Effects of Property Taxes and the Tiebout Hypothesis," Journal of Political Economy, 77(Nov-Dec.1969), pp.957-971.

** 이러한 모든 자료는 1960년에 대한 것이다.

는 것이다.*

제 1 단계는 통상최소자승법을 이용하여 이러한 독립변수집합에 대해 V_t를 단지 회귀하는 것이었다. 이는 다음과 같은 추정방정식을 도출하였던 것이다.**

$$\hat{V} = \underset{(2.4)}{-21} - \underset{(4.1)}{3.6 \log T} + \underset{(2.1)}{3.2 \log E} - \underset{(4.8)}{1.4 \log M}$$

$$+ \underset{(4.1)}{1.7R} + \underset{(3.9)}{0.05N} + \underset{(8.9)}{1.5Y} + \underset{(3.6)}{0.3P}, \qquad R^2 = 0.93 \qquad (7.93)$$

모든 설명변수가 예상한 대로의 부호를 가진 추정계수를 갖고 있음을 알 수 있고, t비율로부터 관계가 없다라는 어떠한 귀무가설을 고려하였을지라도(예를 들어 $H_0 : b_i = 0$; $H_1 : b_i \neq 0$), 5퍼센트의 유의수준하에서 그것을 기각할 것임을 알 수 있다. 그 결과들은 재정변수들이 거주지 결정에 상당한 영향을 미친다는 개인행태모형을 지지하는 것으로 보인다. 곧 (다른 조건이 불변일 때) 세율이 높으면 높을수록 전형적인 주거단위의 가치는 낮아지게 된다. 그리고 역으로 학생당 지출이 크면 클수록 그 가치는 높아진다. 사람들은 낮은 세율과 상위학교의 조건을 갖춘 지역에서 사는데 많은 비용을 지불할 용의가 있는 것으로 보인다.

회귀모형과 추정절차를 약간 생각해보더라도 몇가지 심각한 문제점을 제기해 보게 된다. 도시의 물리적이며 어떤 환경적인 특성을 외생변수로 간주하고자 하는 반면에 두개의 재정변수는 분명히 내생변수인 것이다. 예를 들면 세율은 공공예산의 규모에 의존하며, 그리고 물론 재산가치에 부

---

* 실제의 회귀방정식에서 쓰인 이러한 변수들과 특정한 변환에 대한 보다 자세한 서술(과 정당화)에 대해서는 논문을 볼 것.

** 변수 p는 3,000 달러 미만의 소득을 가진 지방도시의 가계의 비율이다. 이는 소득변수 Y_t의 결함을 수정하기 위해 이용하고 있다. 논문의 pp. 962-963을 볼 것.

분적으로 달려있는 세원에도 의존하는 것이다. 마찬가지로 학교에 대한 지출수준은 소득과 인구의 여타 특성과 함께 세율(그리고 그에 따라 재산가치)에 의해 결정될 것이다.* 사실상 (7.93)에서 세율과 거주단위의 가치사이의 관찰된 負의 관계는 단지 보다 가치있는 재산이 있을 때 주어진 양의 수입은 보다 낮은 세율에 따라 증가할 수 있다는 사실을 반영하는 것이라고 분명히 주장할 수 있을 것이다. 곧, V는 T를 결정하고 있지만, 그 역은 사실이 아닌 것이다. 결론은 이러하다.

곧, 다음과 같은 일반적인 형태의 체계에서 두개의 추가방정식의 결과를 고려하여야만 한다는 것이다.

$$T_t = f(V_t, E_t, \ldots) \tag{7.94}$$

$$E_t = g(T_t, Y_t, \ldots) \tag{7.95}$$

방정식 (7.94)는 세율이 공공지출과, 세원의 규모를 결정하는 (V_t와 같은) 여타 변수들의 함수임을 가리키는 반면, (7.95)는 세율, 소득과 학교재정이 t번째 지역주민들의 속성을 반영하는 여타 변수들의 함수임을 말한다. 요약하면, 재정변수 〔이 경우에는 ($\log T_t$)와 ($\log E_t$)〕가 모형에서 내생변수이다. 이는 그것들이 교란항과 상관이 있으며 그 결과로 (7.93)의 추정계수들은 연립방정식 편의에 빠지기 쉽다는 것을 의미한다. 따라서 OLSQ는 회귀모형을 추정하기 위한 적당한 기법이 아니다.

---

* 세율은 공공지출에 영향을 미친다. 왜냐하면 실제로 그것은 공공서비스의 "가격"을 나타내기 때문이다. 예를 들면, 상대적으로 거대한 상공업의 세원을 가진 지역에 사는 거주민들은 상대적으로 낮은 "가격"으로 교육단위를 "구입"할 수 있다. 왜냐하면 학생당 지출의 대부분이 지방기업들이 지불한 조세로부터 나오기 때문이다. 그러한 지역의 거주민들은 거대한 상공업의 세원이 없는 경우보다 더 규모가 큰 학교예산을 선택할 것으로 예상된다.

이 때문에 연구는 처음으로 되돌아가서 TSLS를 이용하여 방정식을 다시 추정하게 된다. 이는 다음을 의미한다.

곧, 체계의 사전결정변수에 대해 두개의 내생변수, $E_t' = (\log E_t)$와 $T_t' = (\log T_t)$를 회귀하여 정화된 변수, $\hat{E}_t' = (\log \widehat{E}_t)$와 $T_t'^{\wedge}=(\log \widehat{T}_t')$를 만든 다음 $\log E_t$와 $\log T_t$ 대신에 $\hat{E}_t' = (\log \widehat{E}_t)$와 $\hat{T}_t' = (\log \widehat{T}_t)$를 이용하여 원래의 방정식을 추정하여야 한다는 것이다.* 이 점에서 연구는 TSLS의 한가지 매력적인 특성을 이용하고 있다.

앞의 논의에서 다시 생각해 보아야 할 것은 TSLS가 모형의 여타 방정식에 대한 완전한 규정을 요구하지는 않는다는 것이다. 곧, (7.94)와 (7.95)의 모든 독립변수들을 완전히 규정하고 그에 대한 자료를 얻을 필요는 없다는 것이다. 그들중 몇가지에 대한 정보가 필요할 뿐이다. 보다 명확하게 말하면 절차를 진행시키기 위해서는 (7.93)에서 나타나는 것 이외에 두개의 사전결정변수가 적어도 있어야 한다는 것이다. 연구에서의 절차는 방정식 (7.94)와 (7.95)에 설명변수로서 참여하는 많은 추가사전결정변수들을 유리시켜 버리는 것이었다. 곧, 이것들은 세율과 학교지출에 영향을 미치는 변수들, 다시 말해서 1인당 상공업재산가치, 인구의 교육수준, 학교에 등록한 인구의 비율등과 같은 변수들인 것이다.**

---

* 독자들은 "정화된(Purged)" 변수가 $\log(\hat{T}_t)$와 $\log(\hat{E}_t)$이 아님을 유념해야 한다. 곧, 우선 $\hat{T}_t$과 $\hat{E}_t$을 얻기 위해 사전결정변수에 대해 T_t와 E_t를 회귀한 다음 로그를 취하는 것이 아니다. 그 대신에 변형된 변수로서 $\log T_t$와 $\log E_t$, 곧 $\hat{T}_t$과 $\hat{E}_t$을 취한 다음 $\hat{T}_t'$과 $\hat{E}_t'$을 얻기 위해 사전결정변수에 대해 $\log T_t$와 $\log E_t$를 회귀한다. 이러한 미묘한 점을 유념해야 할 것이다. 왜냐하면 제8장에서 보게 되겠지만 $\log(\hat{T}_t)$와 $\log(\hat{E}_t)$의 이용은 일치성이 없는 추정량을 결과하기 때문이다.

** 실제로 $\log \widehat{E}_t$과 $\log \widehat{T}_t$을 만드는데 쓰이는 추가사전결정변수는 7개였다. 그 완전한 일람표는 인용된 논문의 965페이지에 있다.

연구는 그러한 사전결정변수에 대해 $\log E_t$ 와 $\log T_t$ 을 회귀하여 $\log \widehat{E}_t$ 과 $\log \widehat{T}_t$ 을 계산한 후에 제 2 단계의 회귀방정식을 추정하여 다음의 결과를 얻고 있다.

$$\hat{V} = \underset{(2.3)}{-29} - \underset{(3.1)}{3.6 \log T} + \underset{(2.1)}{4.9 \log E} - \underset{(4.0)}{1.3 \log M} + \underset{(3.6)}{1.6R}$$

$$+ \underset{(3.9)}{0.06N} + \underset{(7.7)}{1.5Y} + \underset{(3.1)}{0.3P} \qquad (7.96)$$

이 경우에는 앞의 예와는 달리 TSLS 방정식의 추정계수가 일반적으로 OLSQ 추정치에 매우 가깝다는 것이 흥미롭다. 그것들 모두는 동일한 부호, 거의 동일한 값과 거의 동일한 t 비율을 가진다. 유일하게 현저한 차이란 〔비록 OLSQ 추정치, 3.2 가 (7.96) 의 정보를 이용하여 설정한 95 퍼센트 신뢰수준하의 모수에 대한 신뢰구간내에 있지만〕 학교지출 변수에 대한 추정계수가 OLSQ 방정식에서 보다 TSLS 방정식에서 훨씬 크다는 것이다. 그 결과들은 다음을 의미한다. 곧, 이러한 특수한 연구에서 연립성문제에 의해 도입된 불일치성은 분명히 지나치게 심각하지는 않다는 것이다. 다른 경우에 OLSQ 추정치와 TSLS 추정치는 아주 다를 수도 있다. 여하튼 여기에서는 개인이 근교도시의 선택에 지방재정변수를 고려하는 모형에 부합되는 증거를 그 결과가 제공하고 있음을 알게 된다.

부록. 연립방정식모형에서 자기상관을 갖는 교란항*

제 6 장과 제 7 장의 내용에서는 기본적인 회귀모형에 대한 가정이 만족되지 않을 때 발생하는 추정상의 몇가지 문제점을 검토하였다. 그 절차는 이러한 문제를 각각 차례로 고려하고, 그것들을 처리할 수 있도록 추정절차를 수정하는 다소 만족할만한 방법을 개발하는 것이었다. 하지만 어떤 경우에는 이러한 추정문제 몇가지가 동일한 회귀모형에 나란히 나타나기도 하고, 그리하여 어떻게 일을 진행시켜야할지 반드시 분명한 것은 아니었다.

이 부록에서는 바로 그러한 경우, 곧 체계문제와 자기상관을 가진 교란항을 포함하고 있는 모형을 고려하여 보고 그러한 모형을 어떻게 추정할 것인지를 연구하여 보기로 한다. 이는 계량경제학자들이 두가지 (또는 그 이상의) 추정문제를 동시에 처리하는 방법에 대한 통찰력을 제공해 줄 것이다.

다음과 같은 연립방정식체계의 방정식들에 관심을 가지기로 하자.

$$Y_{1t} = b_0 + b_1 X_{1t} + b_2 Y_{2t} + b_3 Y_{3(t-1)} + u_t \qquad (7A.1)$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad -1 < \rho < 1 \qquad (7A.2)$$

여기에서 X_{1t} 는 외생변수, Y_{2t} 는 현재 내생변수, $Y_{3(t-1)}$ 은 시차내생

---

* 다양한 각도에서 이 논의는 다음의 논문들을 참고로 하고 있다.
Ray C.Fair,"The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, 38 (May 1970), pp.507-516 ;
J.Phillip Cooper, "Assymptotic Covariance Matrix of Procedures for Linear Regression in the Prsence of First Order Serially Correlated Disturbances," Econometrica,40 (March 1972), pp.305-310.

변수이며 u_t는 교란항이다. 앞에서의 예와는 대조적으로 이제 방정식 (7A.2)에서 묘사한 대로 교란항은 자기상관을 가진다고 가정하자. 여기에서 ε_t는 정규분포를 하는 교란항으로 체계내의 모든 외생변수(와 모든 외생변수의 시차변수)에 독립적이며, 따라서 상관이 없다. 또한 ε_t는 자기상관이 없고 〔$E(\varepsilon_s\ \varepsilon_t)=0$, $S_k \neq t$〕, 평균이 0이며 〔$E(\varepsilon_t)=0$〕 분산이 일정하다. 〔$E(\varepsilon_t{}^2)=\sigma^2$〕.

만일 ρ가 방정식(7A.2)에서 0이면 단지 $Y_{3(t-1)}$을 사전결정변수로 취급하며, 따라서 본문에서 서술한 대로 진행하면 될 것이다. 하지만 여기에서 $\rho \neq 0$이라고 가정하였기 때문에 몇가지 난점이 있다. 자기상관 때문에 $Y_{3(t-1)}$은 u_t와 상관이 있을 것으로 예상하게 될 것이다. 이를 알아보기 위해 유의할 점은 다음과 같다. 곧, Y_{3t}가 내생변수이므로 일반적으로 $Y_{3(t-1)}$은 $Y_{1(t-1)}$에 의존하며, 다시 $Y_{1(t-1)}$은 $u_{(t-1)}$에 의존한다는 것이다. 그 결과 $Y_{3(t-1)}$은 $u_{(t-1)}$에 의존하게 될 것이고, 따라서 (7A.2)에 비추어볼 때 u_t와 상관이 있게 될 것이다. 이는 $Y_{3(t-1)}$을 하나의 사전결정변수로 취급할 수 없음을 의미한다. 이의 일반화는 분명하다. 곧, 교란항이 자기상관이 있으면 일반적으로 시차내생변수는 교란항과 자기상관이 있다.

계속적으로 외생변수를 사전결정변수로 간주할 수도 있음에 유의해야 할 것이다. 왜냐하면 그것들은 교란항과 상관이 없기 때문이다. 방정식 (7A.2)에서 u_t에 반복적으로 시차를 주고 그것들을 대입함으로써 이를 알 수 있다. 곧, 다음과 같을 것이다.

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots \qquad \text{(7A.3)}$$

u_t는 궁극적으로 ε_t의 현재치와 시차치에 의존하며, 가정에 의해 ε_t의 현재치와 시차치는 외생변수와는 무관하므로 그 결과 u_t는 외생변수와 상

관이 없음에 틀림없다. 그러므로 방정식 (7A.1)에서 X_{1t}가 u_t와는 상관이 없는 것으로 추측해 볼 수 있다.

이제 자기상관을 설명하기 위해 TSLS기법을 수정해야만 할 것이다. 앞장의 연구로부터 추측해 볼 수 있듯이 우선 ρ의 추정량을 얻고 방정식 (7A.1)을 변환하여 그것에서 자기상관이 있는 교란항을 제거한 다음 TSLS 추정절차를 이용하여 진행할 것이다.

불행하게도 ρ의 추정량을 도출하는 것은 자기상관이 단절된 것으로 간주하였던 제5장에서와 같은 단순한 과정은 아니다. 만일 방정식(7A.1)이 자기상관을 갖고 있다면 제5장에서 서술한 것처럼 그 절차는 OLSQ에 의해 b_0, …, b_3의 일치추정량, $\hat{b}_0$, …, $\hat{b}_3$을 구하고, 그것을 이용하여 $\hat{u}_t = Y_{1t} - (\hat{b}_0 + \hat{b}_1 X_{1t} + \hat{b}_2 Y_{2t} + \hat{b}_3 Y_{3(t-1)})$에 의해 u_t를 추정하는 것이다. $\hat{u}_t$으로부터 방정식(6.45)의 ρ를 계산하게 될 것이다. 그러나 (7A.1) 또한 체계문제를 갖고 있기 때문에 단순히 OLSQ를 (7A.1)에 적용할 수 없다. 왜냐하면 그에 따른 b_0, …, b_3의 추정량은 일치성을 갖지 않을 것이고 따라서 그것들은 일치성이 없는 ρ의 추정량을 얻게 될 것이기 때문이다. 그러므로 TSLS기법을 이용하여 (7A.1)을 추정해야만 하는 것이다.

이러한 TSLS절차를 수행하기 위해 $Y_{3(t-1)}$이 u_t와 상관이 있으므로 그것을 마치 현재 내생변수, 가령 $Z_t = Y_{3(t-1)}$인 것처럼 취급하기로 가정해 보자. 표기의 간단화를 위해 X_t가 시점 t에 체계내의 (X_{1t}를 포함한) 모든 외생변수를 나타내는 것으로 하자. 그러면 TSLS절차의 제1단계는 Z_t와 Y_{2t}의 "수정된" 또는 "정화된" 값, 곧 $\hat{Z}_t$와 $\hat{Y}_{2t}$을 얻는 것이 될 것이다.

이때 전형적으로 체계내의 모든 외생변수, X_t*를 Y_{2t}에 대해 회귀함

으로써 $\hat{Y}_{2t}$을 얻게 될 것이다.* 마찬가지로 시점을 일치시키려는 가운데 외생변수의 모든 시차치, X_{t-1}에 대해 Z_t를 회귀함으로써 $\hat{Z}_t = \hat{Y}_{3(t-1)}$을 만들게 될 것이다. 하지만 TSLS절차의 기술적 조건을 보장하기 위해 X_t와 X_{t-1}로 표기된 모든 변수들(예를 들어 다른 것들중에서 X_{1t}와 $X_{1(t-1)}$)에 대해 Y_{2t}와 Z_t를 회귀함으로써 $\hat{Y}_{2t}$와 $\hat{Z}_t$를 만들게 될 것이다.** 일단 $\hat{Y}_{2t}$과 $\hat{Z}_t$을 만들었으면, 기법의 나머지는 명백하다. 곧, 우선 Y_{2t}와 $Y_{3(t-1)}$을 각각 $\hat{Y}_{2t}$와 $\hat{Y}_{3(t-1)}$으로 대체함으로써 (7A.1)의 모수를 추정하고, 교란항에 "별표"를 붙인 다음, 표준적인 기법에 의해 제2단계의 정규방정식을 도출할 것이다.

만일 단지 b_0, b_1, b_2와 b_3의 일치추정량을 얻는데 관심을 둔다면, 단지 $\hat{b}_0$, …, $\hat{b}_3$에 대한 정규방정식을 풀 수 있으며, 그것으로 끝마칠 수 있다. 하지만, 항상 신뢰구간설정과 가설검정이 가능하도록 추정량에 대한 분산과, 또 그와 관련된 t비율을 결정하고자 할 것이다. 불행히도 여기에서 추정절차를 그만두게 되면 이러한 분산과 t비율을 얻을 수 없다. 왜냐하면 자기상관문제는 여전히 남아있으며, 모든 분산식을 무효화시키기 때문이다. 추가적인 수정을 행해야만 하는 것이다.

TSLS의 이용은 (7A.1)의 모수의 일치추정량, $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$와 $\hat{b}_3$을

---

* 자기상관 때문에 시차내생변수를 현재내생변수로 취급해야함을 기억해야 한다. 그러므로 외생변수는 체계내의 유일한 사전결정변수인 것이다.

** 독자는 만일 $\hat{Y}_{2t}$이 단지 X_t에 대해 Y_{2t}를 회귀함으로써 얻게 되며, $\hat{Z}_t$는 단지 X_{t-1}에 대해 Z_t를 회귀함으로써 얻게 된다면, 일반적으로 (7.61)에서 서술한 조건이 유지되지 않을 것임을 보일 수 있어야만 한다. 곧, $Y_{2t} = \hat{Y}_{2t} + \hat{\theta}_{1t}$과 $Z_t = \hat{Z}_t + \hat{\theta}_{2t}$으로 놓으면, $\Sigma(\hat{\theta}_{1t}\,\hat{Z}_t) \gtrless 0$과 $\Sigma(\hat{\theta}_{2t}\,\hat{Y}_{2t}) \gtrless 0$이다. (도움말 : $\hat{\theta}_{1t}$은 X_{1t}와 같이 어떠한 현재외생변수의 경우에도 $\Sigma(\hat{\theta}_{1t}\,X_{1t}) = 0$인 것과 같을 것이다. 하지만 $\hat{Z}_t$은 $X_{1(t-1)}$에 의존하게 될 것이다.)

만들기 때문에 다음과 같은 교란항의 일치추정량을 얻을 수 있다.

$$\hat{u}_t = Y_{1t} - (\hat{b}_0 + \hat{b}_1 X_{1t} + \hat{b}_2 Y_{2t} + \hat{b}_3 Y_{3(t-1)}) \tag{7A.4}$$

제 5 장의 절차를 이용하여 이제 다음과 같은 ρ의 일치추정량을 얻게 될 것이다.

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{u}_{t-1}\hat{u}_t}{\sum_{t=2}^{n} \hat{u}_{t-1}^2} \tag{7A.5}$$

이는 방정식 (7A.2)에 따른 ρ의 추정량임을 기억할 것이다. 제 5 장에서 설명하였듯이 이제 방정식 (7A.1)에 시차를 주고, 시차방정식에 $\hat{\rho}$을 곱한 다음 그에 따른 방정식을 (7A.1)에서 빼면 다음을 얻게 된다.*

$$Y_{1t}^* = B + b_1 X_{1t}^* + b_2 Y_{2t}^* + b_3 Y_{3(t-1)}^* + \varepsilon_t \tag{7A.6}$$

여기에서 $B = (b_0 - \hat{\rho}\, b_0)$이고,

$$Y_{1t}^* = Y_{1t} - \hat{\rho} Y_{1(t-1)}, \qquad X_{1t}^* = X_{1t} - \hat{\rho} X_{1(t-1)}$$
$$Y_{2t}^* = Y_{2t} - \hat{\rho} Y_{2t(t-1)}, \qquad Y_{3(t-1)}^* = Y_{3(t-1)} - \hat{\rho} Y_{3(t-2)}$$

효과적으로 자기상관문제를 제거하였으므로 이제 한가지 예외를 제외하고는 앞에서 서술한 것과 유사하게 진행시킬 수 있을 것이다. 위에서 Y_{2t}와 $Y_{3(t-1)}$의 수정치를 만들기 위해 모든 외생변수 X_t와 그것의 시차변수 X_{t-1}을 이용하였다. 이제 (7A.6)에서의 내생변수 Y_2^*t와 $Y_3{}^*{}_{(t-1)}$

---

* 또한 방정식 (7A.6)은 단지 표본크기가 무한하면 확률의 면에서 아주 옳다. 물론 이에 대한 이유는 $\hat{\rho}$이 단지 ρ의 일치추정량이라는 것이다.

에 관심을 갖게 된다.* 이러한 변수들은 각각 Y_{2t}^* 과 $Y_{3(t-1)}^*$ 과 그것들의 첫번째 시차(first lags)의 선형결합인 것이다. 그러므로 변수들의 시점을 같도록 하는 가운데 X_t 와 X_{t-1} 의 해당선형결합, 곧 $X_t^* = (X_t - \hat{\rho} X_{t-1})$ 과 $X_{t-1}^* = (X_{t-1} - \hat{\rho} X_{t-2})$에 대해 Y_{2t}^* 와 가령 $Z_t^* = Y_{3(t-1)}^*$을 회귀함으로써 $\hat{Y}_{2t}^*$ 과 $Y_{3(t-1)}^*$을 만들게 된다.

한가지 예로서 X_t^* 으로 표기한 하나의 변수는 $X_{1t}^* = (X_{1t} - \hat{\rho} X_{1(t-1)})$ 일 것이다. 일단 $\hat{Y}_{2t}^*$ 와 $\hat{Y}_{3(t-1)}^*$를 계산하였다면 Y_{2t}^* 와 $Y_{3(t-1)}^*$ 대신에 그것들을 (7A.6)에 대입하여 교란항 ε_t에 별표를 붙인 다음 제2단계의 정규방정식을 얻게 될 것이다. 일반적인 조건하에서 그에 따른 추정량은 일치추정량이며, 표준식에 의해 주어진 대표본분산을 가지게 됨을 보일 수 있으며, 따라서 표준적인 방식으로 신뢰구간설정과 가설검정이 가능하게 된다. 유의할 점은 b_0의 추정량이 $\hat{b}_0 = \hat{B} / (1 - \hat{\rho})$이라는 것이다. 마찬가지로 $\hat{b}_0$의 대규모분산은 다음과 같을 것이다. 곧,

$$\text{var}(\hat{b}_0) = \frac{1}{(1 - \hat{\rho})^2} \text{var}(\hat{B})$$

이제 결과를 요약하고 일반화하기로 한다. 만일 연립방정식체계내의 어떤 주어진 방정식이 자기상관이 있는 교란항을 가지게 되면, 그 방정식의 시차내생설명변수는 내생변수로서 취급되어야만 한다. 우선 계산치를 만들기 위해 모든 외생변수와 그것들의 시차치에 대해 현재내생설명변수와 시차내생설명변수를 회귀함으로써 그 방정식을 추정하게 될 것이다. 그러면 현재내생설명변수와 시차내생설명변수를 이러한 "정화된"값으로 바꾸고 통례대로 회귀방정식의 계수의 일치추정량을 얻게 된다. 다음으로 이러한 TSLS추정량을 이용하여 교란항의 일치추정량을 만들고 다시 이것을 이

---

* (7A.6)에서의 진정한 모수 대신에 $\hat{\rho}$을 이용하는데 포함되는 몇가지 문제 때문에 여전히 $Y_{3(t-1)}^*$ 을 내생변수로 간주해야만 할 것이다.

용하여 자기상관체계 (autocorrelation scheme)에서 ρ의 추정량 $\hat{\rho}$을 얻게된다. 그에 따라 $\hat{\rho}$을 이용하여 그것을 변환함으로써 원래의 방정식에서 자기상관을 제거할 수 있다. 이제 여기서 지나칠 정도로 시차내생설명변수를 (사전결정변수가 아니라) 외생변수로 취급하여야함을 기억하면서 다시 TSLS 절차를 이용하게 된다. 변환된 현재외생설명변수와 시차외생설명변수에 대해 현재내생변수와 시차내생변수의 "변환"치를 회귀하고, 이러한 설명변수들을 그 수정된 또는 정화된 것들로 대체하여, TSLS 절차의 제 2 단계를 수행하게 될 것이다.

문 제

1. 다음의 모형을 생각하기로 한다.

$$C_t = b_0 + b_1 C_{t-1} + b_2 Y_t + \varepsilon_{1t} \tag{1}$$

$$Y_t = I_t + C_t \tag{2}$$

$$I_t = a_0 + a_1 Y_t + a_2 Y_{t-1} + a_3 r_t + \varepsilon_{2t} \tag{3}$$

여기에서 C, I, Y와 r은 각각 소비자지출, 투자, 소득과 이자율이다. ε_1과 ε_2는 자기상관이 없으며 r_t와 무관하다고 가정한다.

a. 모형의 내생변수와 외생변수를 나열하라.

b. 방정식(1)을 어떻게 추정할 것인가?

c. 방정식(2)를 어떻게 추정할 것인가?

2. 임금-물가행태모형으로서 다음을 취하기로 한다.

$$\dot{W}_t = a_0 + a_1 (UN)_t + a_2 \dot{P}_t + \varepsilon_{1t}$$

$$\dot{P}_t = b_0 + b_1 \dot{M}_t + b_2 (UN)_t + b_3 \dot{W}_t + \varepsilon_{2t}$$

여기에서

$\dot{W}$=임금의 백분율변화

UN=실업률

$\dot{P}$=물가의 백분율변화

$\dot{M}_t$ =화폐공급의 백분율변화

ε_1 , ε_2 =교란항

ε_{1t}와 ε_{2t}가 평균이 0이고 분산은 일정하며 자기상관을 가지지 않으며, $(UN)_t$와 $\dot{M}_t$와 무관하다고 가정한다.

a. 위의 방정식들은 식별될 것인가? 이를 설명하라.

b. 식별된 방정식의 추정절차를 약술하라.

3. 다음의 모형을 생각하기로 한다.

$$L_t = a_0 + a_1 W_t + a_2 S_t + u_{1t} \tag{1}$$

$$W_t = b_0 + b_1 L_t + b_2 P_t + u_{2t} \tag{2}$$

여기에서

L=고용된 노동량

W=임금율

S=판매액

P=노동생산성의 측정치

a. L_t와 W_t에 대한 축소형방정식을 구하라.

b. 방정식(1)을 추정하기 위한 기법을 약술하라.

4. 어느 개인의 신발수요가 다음과 같이 묘사된다고 가정하자.

$$D_{it} = a_0 + a_1 P_t + a_2 D_{i(t-1)} + u_{it} \tag{1}$$

여기에서 D_{it}는 시점 t의 i번째 개인의 신발수요이며, P_t는 그가 직

면한 가격이다. 다음과 같이 가정하기로 한다.

$$u_{it} = \rho u_{i(t-1)} + \varepsilon_{it}, \quad -1 < \rho < 1$$

여기에서 ε_{it}는 평균이 0이고, 분산은 일정하며, 자기상관이 없으며, P_t와 그것의 모든 시차치와는 무관하다.

a. 시차종속변수 $D_{i(t-1)}$가 교란항과는 상관이 없음을 직관적으로 논하라.

b. 방정식(1)이 방정식체계의 일부분이 아니라고 가정하자. 그럼에도 불구하고 TSLS에 의해 추정할 수 있음을 증명하라.

5. 다음과 같은 다중회귀모형을 생각하기로 한다.

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + u_{1t} \quad (1)$$

$$X_{2t} = c_0 + c_1 Y_t + u_{2t} \quad (2)$$

통상의 가정하에서 $E(X_{2t} \; u_{1t}) \neq 0$임을 보여라.

6. 다음의 임금-물가모형을 생각하기로 한다.

$$\dot{W}_t = a_0 + a_1 \dot{P}_t + a_2 (UN_t) + \varepsilon_{1t} \quad (1)$$

$$\dot{P}_t = b_0 + b_1 \dot{W}_t + \varepsilon_{2t} \quad (2)$$

여기에서

$\dot{W}$ = 화폐임금의 백분율변화

$\dot{P}$ = 물가의 백분율변화

UN = 실업률

a. 방정식(1)을 추정하려고 한다면 TSLS를 적용할 수 없음을 보여라.

b. 방정식(2)를 추정하려고 하는 경우에도 TSLS 절차가 붕괴되는가? 설명하라.

7. 연립방정식체계의 일부분인 다음과 같은 구조방정식을 가정한다.

$$Y_{1t} = b_0 + b_1 X_{1t} + b_2 Y_{2t} + b_3 Y_{3t} + u_{1t}$$

여기에서 Y_{1t}, Y_{2t}와 Y_{3t}는 내생변수이고 X_{1t}는 사전결정변수이다. 이 방정식이 그 일부분이 되는 완전한 체계가 10개의 추가사전결정변수를 포함한다고 가정한다. 하지만 그중 하나의 변수, 가령 X_2에 대해서만 관찰치가 있다고 가정한다.

a. 방정식은 식별될 것인가? 그 이유는?

b. TSLS에 의해 이 방정식을 추정할 수 있는가? 설명하라.

8. 다음과 같은 두개의 방정식모형을 생각하기로 한다.

$$Y_{1t} = a_1 + b_1 X_t^2 + c_1 Y_{2t} + \varepsilon_{1t} \quad (1)$$
$$Y_{2t} = a_2 + b_2 X_t + c_2 Y_{1t} + \varepsilon_{2t} \quad (2)$$

여기에서 X_t는 사전결정변수이며, ε_1와 ε_2는 표준적인 가정을 만족한다.

a. 두 방정식은 식별되는가? 그 이유는?

b. 축소형방정식을 도출하라.

c. 위의 모형의 첫번째 방정식을 추정하기 위한 절차를 약술하라.

9. 사적 투자지출이 다음과 같다고 가정하자.

$$I_{it} = a + b_1 r_{it} + b_2 S_{i(t-1)} + u_{it}, \quad i = 1, \ldots, N,$$
$$r_{it} = r_t + b_3 I_{it} + \varepsilon_{it}$$

여기에서

I_{it} = t시점의 i번째 기업의 투자지출

r_{it} = i번째 기업이 투자기금에 대해 지불해야 하는 이자율

$S_{i(t-1)}$ = $t-1$ 시점의 i 번째 기업의 판매액

r_t = 투자기금에 대한 경제전체의 평균이자율

투자지출수준이 이자율에 영향을 미치게끔 이러한 기업의 수, N이 크다고 가정한다. u_{it} 와 ε_{it} 와 관련한 표준적인 조건들을 가정한다. 또한 단지 횡단면자료만이 있다고 가정한다.

a. 방정식들이 식별되는지의 여부를 논하라.

b. I_{it}에 대한 축소형방정식을 구하라.

제 8 장 비선형인 연립방정식 모형

제 7 장에서 논의한 모든 연립방정식 모형은 內生變數로 이루어진 線型이었다. 그러나, 경제학자들이 실제로 고려하고 있는 많은(대부분은 아니라도) 모형은 내생변수로 이루어진 선형이 아니다. 예를 들어, 많은 經濟模型은 賃金(W), 物價(P)를 내생변수로서 포함하고 있다. 이들 모형은 또한 전형적으로 實質賃金率(W/P)로 勞動에 대한 需要를 설명한다. 분명히 실질임금 변수(W/P)는 모형으로 설명되어야 한다. 즉, W와 P가 내생적이면 W/P는 내생적이다. 또한, W/P가 내생변수들의 非線型函數인 것도 확실하다.

내생변수에서의 非線型性에 관한 다른 예는 허다하다. 예컨데, 總收入 R은 보통 P가 가격이고 Q가 판매된 단위수량일 때, 곱(PQ)로 정의한다. 다시, P와 Q가 내생이라면, 총수입도 섞여 있는(아마도 이윤의 한 성분으로서) 모형은 비선형으로 간주하여야 한다. 비슷한 경우로서 임금률 W와 단위 시간당 구입한 노동 단위 L의 곱으로 정의된 賃金基金이 들어가 있는 모형을 생각할 수 있다. 마찬가지로, 대부분의 巨視模型이 고려하는 거의 모든 경제변수들의 實質値(디플레이트된)와 經常値뿐만 아니라 一般物價水準(즉, GNP 디플레이터)의 定量을 설명하고 있음에 유의하자. 한 예로서, 그러한 모형은 GNP를 不變額뿐만 아니라 經常額으로도 설명한다. 따라서 우리의 논의는 더 나아가, 내생변수들의 경상치와 디플레이트한 가격치를 모두 포함하고 있는 모형은 만일 가격 디플레이터(price deflator)가 내생적이라면, 내생변수로 이루어진 非線型으로 간주하여야만 한다는 것을 시사한다.

더 비선형성의 예를 보면, 내생적 생산요소에서 비선형인 생산함수, 내생적으로 결정된 실업률의 역수로 정식화된 필립스곡선 그리고 마

지막으로 경제학자들이 해명하려는 많은 변수의 정의 자체를 들 수 있다. 예를 들어, 실업률은 失業勞動者의 勞動力에 대한 비율로 정의된다. 만일 巨視模型이 실업노동자의 수 뿐만 아니라 노동력의 크기를 설명하려 한다면, 그 모형은 반드시 비선형으로 간주하여야 한다.

이 章에서는 이상과 같은 모형의 분석을 논의할 것이다. 특히 識別問題와 二段階最小自乘의 응용에 관한 분석을 내생변수들의 비선형을 갖고 있지만 그 母數에서는 선형인 모형으로 확대한다.* 이 장의 부록에서는 推定에 관한 우리의 결과를 모수에서도 비선형성을 갖는 모형으로 확대한다. 비선형성은 분석에 좀 더 복잡함을 초래한다는 점을 독자들에게 지적하여 둔다. 결과적으로 이 장의 일부는 각별한 주의를 요구할 것이다. 그러나 그 분석은 직접 그리고 오로지 앞 장들에서 다룬 내용에만 의거한다. 사실 여기서의 분석은 약간 놀랄 정도로 앞의 내용을 기초로 얼마나 더 많은 지식을 얻을 수 있는지를 알게 하는 것이기도 하다.

1. 분 석 틀

맨 먼저 분석의 관심 대상인 비선형모형의 유형을 좀 더 공식적으로

---

* 선형과 비선형 계량경제학 모형 둘다에 대한 고전적인 참고문헌은 F. Fisher, The Identification Problem in Econometrics (New York: McGraw-Hill, 1966)이다. 우리의 식별문제에 관한 논의는 H.H.Kelejian, "Identification of Nonlinear systems: An Interpretation of Fisher," Princeton University, Econometric Research Program, Research Paper No.22(Revised), 1970에 대부분 근거하고 있다. 二段階最小自乘의 절차는 다음을 따른다. H. H. Kelejian, " Two Stage Least Squares and Econometric Models Linear in the Parameters but Nonlinear in the Endogenous Variables," Journal of the American Statistical Association, June 1971, vol.66, pp.373-374.

정의하여야 한다. 그렇게 함으로써 가장 일반적인 형식의 모형을 설명하지는 못하지만 대신에 가장 전형적인 모형을 밝히게 될 것이다. 실제로 고려되는 대부분의 모형은 우리의 분석틀에 맞추어져야 한다. 두 방정식으로 이루어진 모형으로 출발하여 그 결과를 일반화 시키기로 한다.

가. 2개 방정식의 실례

다음의 예로 든 모형을 보자.

$$Y_{1t} = a_0 + a_1 Y_{2t} + a_2[Y_{1t}Y_{2t}/X_{1t}] + a_3 X_{2t} + \varepsilon_{1t} \qquad (8.1)$$

$$Y_{2t} = b_0 + b_1 Y_{1t} + b_2[(Y_{1t} - 2Y_{2t})^2 e^{X_{3t}}] + b_3 X_{4t} + \varepsilon_{2t} \qquad (8.2)$$

여기서 X_{1t}, X_{2t}, X_{3t}와 X_{4t}는 외생변수, ε_{1t}와 ε_{2t}는 교란항 그리고 Y_{1t}와 Y_{2t}는 모형이 설명하려고 하는 변수(즉, 내생변수)이다. ε_{1t}와 ε_{2t}가 모든 t와 s에 걸쳐서 외생변수 X_{1s}, X_{2s}, X_{3s}와 X_{4s}에 독립적이며, $i = 1, 2$에 대하여 $E(\varepsilon_{it}) = 0$ 그리고 $E(\varepsilon_{it}^2) = \sigma_i^2$ 이라고 가정하자. 또한 $t \neq s$에 대하여 ε_{1t}와 ε_{2t}는 ε_{1s}와 ε_{2s}에 독립적이라고 하여 어느 교란항이나 自己相關이 없다고 가정한다.

(8.1)과 (8.2)로 표현된 모형에 관하여 주의해야 할 두가지 주요 특징이 있다. 첫째로 모형은 회귀 모수 $a_0, \cdots, a_3, b_0, \cdots, b_3$에서 線型이라는 점이다. 둘째로, 모형은 모수 a_2와 b_2에 대응하는 괄호 안에 있는 변수들 때문에 내생변수에서 비선형이라는 것이다. 괄호가 쳐진 변수들의 값은 변수 Y_{1t}, Y_{2t}, X_{1t}, X_{3t}의 값으로 결정될 수 있다는 점을 명심하여야 한다. 달리 표현하면, 괄호 안의 변수는 이들 네 변수의 알려진 함수로 볼 수 있는 것이다. 알려진 함수란 그 형식을 알고 있고 미지의 母數를 포함하지 않는 함수를 말한다. 예를 들어 b_2에 대응하여 괄호가 쳐진 변수가 형식이 $[(Y_{1t} - \gamma_1 Y_{2t})^2 e^{\gamma_2 X_{3t}}]$이고 여기서 γ_1과 γ_2가 이 값

을 모르는 모수라면, 알려진 함수가 아니다. 만일 그렇다면, 그 모형은 모수에서 선형이지가 않다.

이제 모형의 이론(즉, 한계소비성향은 陰이 될 수 없다)과 (8.1), (8.2)의 母數들의 값으로 이루어진 집합이 일치하게 주어졌다면, 그 모형은 외생변수 $X_{1t}, \cdots, X_{4t}$와 교란한 ε_{1t}, ε_{2t}로써 내생변수 Y_{1t}와 Y_{2t}를 풀 수 있게 된다. 가능한 모수의 값들의 집합으로서 만일 $a_2 = b_2 = 0$을 예로 들면 간단하게 명시적인 解를 얻을 수 있을 것이다. 또다른 모수의 값들의 집합으로서는 숫자로 된 解를 얻을 수 있을 뿐이다. 예를 들어, 외생변수와 교란항의 수치로 된 값의 집합에 대응하는 숫자로 나타난 Y_{1t}와 Y_{2t}값을 추론할 수 있을 뿐이다. 그러나 어느 경우에서나 모형으로 지적되듯이 내생변수 Y_{1t}와 Y_{2t}의 값은 외생변수와 교란항의 값들에 종속할 것이다. 그러므로 이러한 현상을 내생변수에 대한 모형의 解가 외생변수와 교란항에 종속한다거나 그것의 함수라고 말한다.

Y_{1t}와 Y_{2t}에 대한 (8.1)과 (8.2)의 解는 부분적으로 교란항에 종속하기 때문에, 이들 내생변수와 교란항이 독립적이거나 심지어 상관이 없다고는 일반적으로 가정할 수 없다.* 이러한 결론은 제7장의 논의에서 명백하였다. 이제 a_2에 대응하는 괄호 안의 변수를 보자. Y_{1t}와 Y_{2t}의

---

* 더 많이 알고 있는 독자들을 위하여, 다음의 사실을 밝혀 둔다. 즉, 일반적으로 비선형 방정식의 집합은 한가지 解보다 많은 解를 가진다. 우리의 논의는 만약 모형의 방정식이 한 解보다 많은 解를 정의하면, 예를 들어 가격은 陰이 될 수 없다는 등의 모형의 변수에 대한 확실한(말하지 않더라도) 제한을 두어 한 解를 제외한 나머지 解들을 배제시킨다. 예로써, 두 방정식 $x^2 + y^2 = 20$, $|x| = 2|y|$의 집합은 4개의 가능한 解, 즉 $(x = 4, y = 2)$, $(x = -4, y = 2)$, $(x = 4, y = -2)$와 $(x = -4, y = -2)$을 가지고 있다. 그러나 x와 y 둘다 陽이어야 한다고 하면, 모형의 유일한 해는 $x = 4$, $y = 2$이다.

값이 부분적으로 교란항에 종속하기 때문에, 괄호 안의 변수 그 자체가 부분적으로 교란항에 종속하게 된다. 그 결과로서 일반적으로 그 변수는 교란항과 상관을 갖는다. 마찬가지로 b_2에 대응하는 괄호 안의 변수도 보통 교란항과 상관을 갖게 될 것으로 결론내릴 수 있다. 좀 더 일반화하면, 하나 또는 그 이상의 내생변수로 이루어진 함수는 일반적으로 교란항과 상관을 가질 것으로 결론을 내리게 한다.

나. 분명히 해두어야 할 사항

논의를 진행하기 앞서 독자들이 모두 알고 있지는 못할 것으로 보이는 몇몇 사항에 대하여 분명히 밝혀 두기로 한다. M개의 내생변수 $Y_{1t}, \cdots, Y_{Mt}$와 G개의 외생변수 $X_{1t}, \cdots, X_{Gt}$를 갖는 모형을 가정하자. 또한 그 모형은 M개의 교란항 $\varepsilon_{1t}, \cdots, \varepsilon_{Mt}$를 갖는다고 한다. 앞서의 예에서처럼 이 모형은 외생변수와 교란항으로 내생변수에 대해서 풀 수 있다고 가정하자. 그 解는 다음과 같다.

$$Y_{1t} = F_1(X_{1t}, \ldots, X_{Gt}, \varepsilon_{1t}, \ldots, \varepsilon_{Mt})$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$Y_{Mt} = F_M(X_{1t}, \ldots, X_{Gt}, \varepsilon_{1t}, \ldots, \varepsilon_{Mt})$$

일반적으로 만약 모형이 비선형 모형이면, 내생변수의 교란항과 외생변수에 대한 종속을 의미하는 위의 함수는 선형이 되지 않는다. 모형이 또한 $(Y_{1t}^2 + Y_{3t}Y_{5t})$ 또는 더 일반적인 $K(Y_{1t}, \cdots, Y_{Mt})$를 변수로서 갖고 있다고 가정하자. 그러면, 이들 변수에 대한 모형의 解가 갖는 값은 각각 $F_{1t}^2 + F_{3t}F_{5t}$와 $K(F_{1t}, \cdots, F_{Mt})$이며, 여기서 F_{it}는 $F_i(X_{1t}, \cdots, X_{Gt}, \varepsilon_{1t}, \cdots, \varepsilon_{Mt})$를 단순화시킨 것이다. 비선형 모형에서는 분명히 내생변수의 함수가

교란항과 외생변수에 선형으로 종속하지 않는다. 어쨌든 내생변수는 교란항과 전형적으로 상관을 갖는다. 그 이유는 내생변수의 값이 부분적으로 교란항의 값으로 결정되기 때문이다.

추정을 달성하기 위하여 만일 변수가 교란항과 상관이 있다면 그 변수를 내생으로 정의할 것이다. 그러므로 하나 또는 그 이상의 내생변수의 함수로 구성된 변수들을 내생변수들의 함수들 또는 줄여서 내생변수들이라고 한다. 내생변수들의 함수가 또한 외생변수를 포함하는지의 여부는 별로 쓸모가 없다는 것을 명심하라. 구성된 변수는 하나 또는 그 이상의 내생변수에 대한 종속성 때문에 일반적으로 오직 교란항과 상관이 있다. 덧붙여서 (8.1)과 (8.2)의 두개의 방정식 모형은 외생변수와 내생변수로 4개의 내생변수 전부를 결정한다는 사실에 유의하자. 예를 들어 만일 Y_{1t}와 Y_{2t}가 외생변수와 교란항으로 표현할 수 있다면, 구성된 변수인 $[Y_{1t}Y_{2t}/X_{1t}]$도 그렇게 표현할 수 있다.

다. 또 다른 實例

(8.1)과 (8.2)에 대조적인 것으로서, 다음의 2개 방정식 모형을 보자.

$$\log(Y_{1t}) = a_0 + a_1 Y_{2t}^3 + a_2 X_{1t} + \varepsilon_{1t} \qquad (8.3)$$

$$Y_{2t}^3 = b_0 + b_1[\log(Y_{1t})] + b_2 X_{2t} + \varepsilon_{2t} \qquad (8.4)$$

여기서 Y_{1t}와 Y_{2t}는 내생변수이고 X_{1t}와 X_{2t}는 외생변수 그리고 ε_{1t}와 ε_{2t}는 교란항이다. 교란항은 앞의 모형에서 언급하였던 모든 바람직한 속성을 다 갖고 있다고 가정한다.

얼핏 보면, 이 모형은 비선형인 것 같다. 그러나 추정을 하는 데에는, 이 모형은 線型이다. 이 점을 보기 위하여 Z_{1t}와 Z_{2t}를 다음과 같이

정의 한다.

$$Z_{1t} = \log(Y_{1t}),\ Z_{2t} = Y_{2t}^{3} \qquad (8.5)$$

그러면, (8.3)과 (8.4)의 모형은 변수 Z로 아래처럼 나타낼 수가 있다.

$$Z_{1t} = a_0 + a_1 Z_{2t} + a_2 X_{1t} + \varepsilon_{1t} \qquad (8.6)$$

$$Z_{2t} = b_0 + b_1 Z_{1t} + b_2 X_{2t} + \varepsilon_{2t} \qquad (8.7)$$

위의 형식에서 모형은 母數에서와 두 외생변수 Z_{1t}, Z_{2t}에서 선형으로 보인다. 그러므로 모형의 모수는 제 7 장에서 설명한 절차로 추정할 수 있다. 이제 (8.1)과 (8.2)의 2개 방정식 모형이 4개의 내생변수를 갖고 있었음을 상기하자. (8.1)과 (8.2)의 내생변수의 수가 방정식의 수를 초과하기 때문에, 그 모형을 2개 내생변수로 된 선형모형으로 표현할 수가 없다. 예를 들어 Z_{3t}와 Z_{4t}를 다음과 같이 가정한다.

$$Z_{3t} = \left[\frac{Y_{1t}Y_{2t}}{X_{1t}}\right]; \qquad Z_{4t} = (Y_{1t} - 2Y_{2t})^2 e^{X_{3t}} \qquad (8.8)$$

따라서, 만일 (8.1)과 (8.2)의 모형이 두 내생변수 Z_{3t}와 Z_{4t}로써 2개 방정식 모형으로 표현되는 것이라 하면, 변수 Y_{1t}와 Y_{2t}도 Z_{3t}와 와 Z_{4t}로 나타내야 한다. Y_{1t}와 Y_{2t}는 (8.8)을 Y_{1t}와 Y_{2t}에 대해서 풀면 Z_{3t}와 Z_{4t} (그리고 X_{1t}와 X_{3t})로 표현할 수 있다. 그러나, 그렇게 하여도 Y_{1t}와 Y_{2t}는 Z_{3t}와 Z_{4t}에서 비선형으로 판명될 것이다. 이를 보이기 위해서 그 관계를 다음과 같이 나타내어서

$$\begin{aligned} Y_{1t} &= g_1(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) \\ Y_{2t} &= g_2(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) \end{aligned} \qquad (8.9)$$

(8.9)의 함수가 Z_{3t}와 Z_{4t}에 대하여 선형이 아님에 유의하자. 그러면, (8.1)과 (8.2)의 모형은 Z_{3t}와 Z_{4t}로써 다음과 같이 나타낼 수 있다.

$$g_1(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) = a_0 + a_1 g_2(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) + a_2 Z_{3t} + a_3 X_{2t} + \varepsilon_{1t} \qquad (8.10)$$

$$g_2(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) = b_0 + b_1 g_1(Z_{3t}, Z_{4t}, X_{1t}, X_{3t}) + b_2 Z_{4t} + b_3 X_{4t} + \varepsilon_{2t} \qquad (8.11)$$

분명히 (8.10)과 (8.11)의 모형은 내생변수 Z_{3t}와 Z_{4t}에 대하여 선형 모형이 아니다.

라. 일반화

이상의 결과는 외양과는 다르게 방정식 수와 동일한 수의 내생변수를 갖는 모형은 추정하는 데서 선형모형으로 볼 수 있다는 사실을 시사한다. 만일 내생변수의 수가 방정식의 수를 초과하면, 일반적으로 그 모형은 선형모형으로 변형되지 않는다. 이 점을 더 공식적으로 보기 위하여, K개의 방정식을 갖고 있으며 외생변수와 교란항으로 내생변수의 모든 값을 유일하게 정할 수 있는 한 聯立方程式 模型을 고려하여 보자.* 이 모형은 모수에 대해서 선형이라고 한다. 더구나 모형에서 나타나는 한개 또는 그 이상의 내생변수에 종속하는 변수의 숫자가 K^*라고 가정한다. 마지막으로, K^*개의 변수중의 어느 하나도 다른 변수의 선형결합으로 나타낼 수

---

* 변수가 만족시켜야 하는 다양한 제약조건 때문에 하나 이외의 모든 解가 배제된다고 하면, 그 비선형모형은 그 내생변수들의 모든 값을 유일하게 결정한다. 이는 이미 앞서 언급한 바 있다.

없다고(즉, 이들 변수는 多重共線性이 없다고) 가정하자.* 이제 만일 K^* $> K$라면 그 모형은 비선형이다. 만약 $K^* = K$이면 추정을 하는 데에서 선형모형으로 표현할 수 있다. $K^* < K$인 경우는 고려하지 않는다. 왜냐하면, 이는 過大決定된 體係(overdetermined system ; 즉 변수보다 방정식이 많은 체계)에 상응하는 것이기 때문이다. 만일 방정식들의 일부가 본질적으로 그외 방정식과 같은 것이어서 남아도는 것이 아니라면, 그 모형은 일반적으로 내부적인 일관성이 없다.**

2. 식별문제

가. 실례

다음의 2개 방정식 모형을 보자.

$$Y_{1t} = a_0 + a_1 g(Y_{2t}) + a_2 X_t + \varepsilon_{1t} \qquad (8.12)$$

$$Y_{2t} = b_0 + b_1 Y_{1t} + \varepsilon_{2t} \qquad (8.13)$$

여기서 $g(Y_{2t})$는 Y_{2t}의 비선형함수로 알려져 있고, X_t는 모든 t와 s에 대하여 교란항 ε_{1t}와 ε_{2t}에 독립적인 것으로 가정된 외생변수이다. 교란항은 모든 바람직한 속성을 만족시킨다고 가정한다. 즉, 교란항은 0의 평균과 일정 불변하는 분산을 가지며 自己相關이 없다.

제 7 장에서의 논의는 (8.12)의 모수가 식별되지 않는다고 시사하였다. 그 이유는 (8.12)가 방정식의 우변에 한 개의 내생변수를 포함하고 있

---

* 이 가정은 餘分(redundancy)를 제거하기 위한 가정이다. 예를 들어, 이 가정이 없다면, Y_{1t}, Y_{2t}, $2Y_1$과 $3Y_{2t}$를 포함하고 있는 모형은 4개의 내생변수를 가진 것으로 취급될 것이다.

** 간단한 예로서, 2개 방정식 체계가 한 변수를 갖고 있다고 하자. $3 + 2X = 5$, $X + 10 = 15$의 체계는 일관성이 없다. 왜냐하면, 첫번째 식에서는 $X = 1$인 반면에 두번째 식은 $X = 5$를 의미한다.

지만 事前決定變數들을 배제하지 않았기 때문이다. 그러므로, 제 7 장의 주장에 따르면, (8.12)를 TSLS로 추정하려 할 경우, 그 노력은 이단계 절차에서의 완벽한 多重共線性으로 인하여 수포로 돌아 간다. 그러나, (8.12)와 (8.13)의 모형은 선형모형이 아니어서, 식별에 관한 제 7 장의 결과가 적용되지 않는다. 특히, 어느 정도 상당히 일반화된 가정 아래에서는 식 (8.12)가 식별된다는 것을 보게 될 것이다.

이 점을 알기 위하여, (8.12)와 (8.13)의 2개 방정식 모형은 외생변수 X_t와 교란항 ε_{1t}, ε_{2t}로서 내생변수 Y_{1t}와 Y_{2t}의 값을 유일하게 결정한다고 가정하자. 표기를 간단히 하기 위하여, 다음과 같다고 한다.

$$Z_t = g(Y_{2t}) \qquad (8.14)$$

따라서 만약 모형이 X_t와 ε_{1t}, ε_{2t}로써 Y_t값을 결정한다면, 그 모형은 또한 이들 변수로 Z_t의 값을 결정한다. 이러한 종속을 다음과 같이 표시한다.

$$Z_t = h(X_t, \varepsilon_{1t}, \varepsilon_{2t}) \qquad (8.15)$$

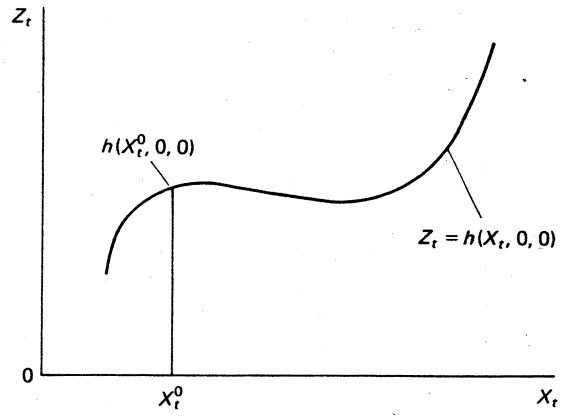
그리고 모형이 비선형이기 때문에, (8.15)의 函數 h는 일반적으로 비선형이 될 것임에 유의하자.

만일 ε_{1t}와 ε_{2t}가 (항상) 0과 같다면, Z_t는 완전히 X_t에 의하여 결정될 것이다. 그러므로, 만일 Z_t와 X_t에 관한 관찰치를 그리면, (8.15)에서 $\varepsilon_{1t} = \varepsilon_{2t} = 0$으로 놓은 다음과 같은 식의 자취를 따르는 곡선이 생긴다.

$$Z_t = h(X_t, 0, 0) \qquad (8.16)$$

한 예로써 그러한 곡선을 〈그림 8.1 〉에 그려 보았다. 이 곡선은 일부

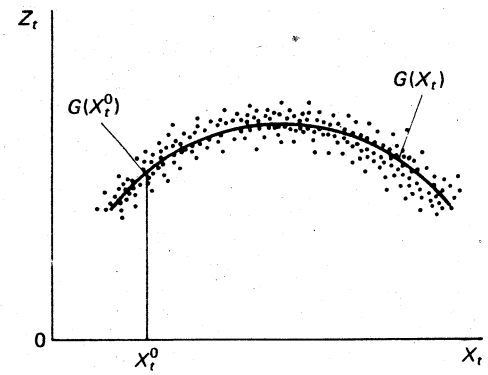
러 비선형 형상으로 그린 것이다. 왜냐하면 (8.12)와 (8.13)의 비선형모형은 변수 $Z_t = g(Y_{2t})$의 외생변수 X_t에 대한 종속이 일반적으로 선형이 아닐 것임을 의미하기 때문이다.



〈그림 8.1 〉

이제 보다 전형적인 사례로서 ε_{1t}와 ε_{2t}가 항상 0과 같지 않을 때를 고려하자. 이 경우에 (8.15)는 Z_t가 완전히 X_t에 의하여 결정되지 않을 것임을 의미한다. 그러나 다시 (8.15)를 참조하면, Z_t가 X_t에 독립적이지는 않을 것이다. 그 이유는 X_t가 Z_t를 결정하는 요소중의 하나이기 때문이다. 설명을 위해서 Z_t와 X_t에 관한 無限한 관찰치가 있다고 가정하자. 그 결과 이상의 논의는 만일 관찰치를 그린다면, 그 흩어져 있는 점들이 X_t에 대한 Z_t의 부분적인 종속을 반영하는 곡선을 보일 것임을 지적하는 것이다. 그러한 散布圖는 〈그림 8.2 〉에 그려져 있다.

〈그림 8.2 〉를 볼 때, 맨 먼저 그림의 모든 점들이 지시된 곡선에 있지 않음에 유의하자. 그 이유는 X_t가 Z_t를 결정하는 요소중의 하나에 불과하기 때문이다. 둘째로, 산포도로 곡선을 그려내는 방법이 많이 있다는 점이



〈그림 8.2〉

다. 〈그림 8.2〉에 나타나 있는 곡선은 주어진 X_t 값에 대응하는 Z_t의 평균값을 나타낸다. 예를 들어, $X_t = X_t^0$에 대응하는 Z_t의 평균값은 곡선의 높이 $G(X_t^0)$로 나타난다. 세째로, 〈그림 8.2〉의 곡선은 의도적으로 〈그림 8.1〉의 곡선과 다르게 그린 것이다. 그렇게 그린 이유는 〈그림 8.1〉에서 정의한 것과 같은 곡선은 〈그림 8.2〉에서의 "평균적 관계"의 곡선과 일반적으로 다를 것이기 때문이다. 비록 이 말은 직관과 정반대되는 것으로 보이지만, 쉽게 설명될 수 있다. 예를 들어, 주어진 X_t 값, 즉 X_t^0에 대응하는 Z_t의 평균값은 (8.15)로부터 다음과 같다.

$$E(Z_t) = E[h(X_t^0, \varepsilon_{1t}, \varepsilon_{2t})] \qquad (8.17)$$

만일 (8.17)의 함수 h가 ε_{1t}와 ε_{2t}에서 비선형이라면, 제1장 부록 B의 결과 특히 (1B.12)에 의거할 때 다음과 같다.

$$E[h(X_t^0, \varepsilon_{1t}, \varepsilon_{2t})] \neq h(X_t^0, E(\varepsilon)_{1t}, E(\varepsilon)_{2t}) = h(X_t^0, 0, 0) \qquad (8.18)$$

그러므로 $X_t = X_t^0$에 대응하는 〈그림 8.1〉의 곡선이 갖는 값, $h(X_t^0, 0, 0)$은 다음과 같은 〈그림 8.2〉의 곡선이 대응하는 값과 동일하지

않다.

$$G(X_t^0) = E[h(X_t^0, \varepsilon_{1t}, \varepsilon_{2t})] \qquad (8.19)$$

이미 말한 바와 같이 〈그림 8.2〉의 곡선은 X_t의 어떤 값에 대응하는 Z_t의 평균값을 나타낸다. 다음과 같이 정의하자.

$$V_t = Z_t - G(X_t) \qquad (8.20)$$

그러면, 어떤 주어진 X_t 값에 대응하는 Z_t의 값은 0이다. 예를 들어, X_t가 X_t^0일 때, V_t의 평균값은,

$$E[V_t] = E[Z_t] - G(X_t^0) = G(X_t^0) - G(X_t^0) = 0 \qquad (8.21)$$

제 6 장의 3절로부터, V_t의 평균이 X_t의 어떠한 값에 대해서도 0이기 때문에 V_t의 전체에 걸친 평균도 0이고 V_t는 X_t와 상관이 없음을 상기하게 된다.

(8.20)의 항들을 재정리하면, 다음과 같다.

$$Z_t = G(X_t) + V_t \qquad (8.22)$$

(8.22)의 관계는 제 5 장 3절의 多項回歸模型에서 도출하였던 관계와 매우 유사하다. 특히, (8.22)로부터 Z_t의 t번째 값이 독립변수 X_t의 t번째 값과 비선형 방식으로 관련되어 있고 또한 V_t항에도 관련되어 있다. V_t항은 어떤 주어진 독립변수의 값에 대응하여 0의 평균값을 갖는 교란항으로 볼 수 있다. 설명을 위하여 제 5 장 3절에서 했던 것처럼 다음과 같이 가정하자.

$$G(X_t) \doteq b_0 + b_1 X_t + b_2 X_t^2 + \cdots + b_k X_t^k \qquad (8.23)$$

그러면 (8.22)로부터,

$$Z_t \doteqdot b_0 + b_1 X_t + b_2 X_t^2 + \cdots + b_k X_t^k + V_t \qquad (8.24)$$

표시를 간단히 하기 위하여 이제 (8.23)과 (8.24)가 等號를 갖는다고 가정한다. 이 가정은 단지 설명만을 위한 것임을 강조하여 둔다. 아래의 결과가 이 가정에 좌우되지 않는다는 사실이 명백하게 될 것이다.

이제 모형 (8.12)와 (8.13)의 변수에 대한 n개 관찰치를 갖고 있다고 가정하자. 그러면, Z_t는 Y_t의 함수, 즉 $Z_t = g(Y_t)$로 알려져 있기 때문에, Z_t에 관한 n개 관찰치를 얻을 수 있다. 그 모형에 따라서 Z_t 값을 계산하면, 아래와 같다.

$$\hat{Z}_t = \hat{b}_0 + \hat{b}_1 X_t + \hat{b}_2 X_t^2 + \cdots + \hat{b}_k X_t^k \qquad (8.25)$$

여기서 $\hat{b}_0, \cdots, \hat{b}_k$는 통상적인 방식으로 얻은 것이다.

이제 (8.12)에 관한 추정문제를 검토하자. 우변의 내생변수는 $g(Y_t)$인데, 이는 Z_t 변수이다. (8.12)에 TSLS 절차를 적용하면, Z_t를 (8.15)에서 주어진 계산치, 즉 $\hat{Z}_t$로 대체할 수 있다. 이렇게 하면, 두번째 단계의 특성이 完全多重共線性(perfect multicollinearity)에 있지는 않을 것이다. $\hat{Z}_t$의 X_t에 대한 종속은 형식상 비선형이기 때문에, $\hat{Z}_t$는 X_t의 완전한 선형 함수가 되지는 않는다. 이는 곧 (8.12)의 모수들의 一致추정량을 구할 수 있으며, 따라서 그 방정식은 식별된다는 점을 시사한다.

나. 세밀한 고찰

이상의 분석에서는 (8.22)의 평균 함수 $G(X_t)$를 X_t의 다항식으로 나타낼 수 있다고 가정하였다. 이 節에서는 이러한 가정이 식별에 불필

요하고 따라서 방정식 (8.12)의 일치성을 가져오는 推定에도 필요하지 않다는 것을 보인다. 또한, (8.12)와 (8.13)의 모형이 $g(Y_{2t}) = Y_{2t}$ 에서 선형이라면, 앞에서 설명한 절차와 비슷한 식으로는 방정식(8.12)의 일치추정량을 구할 수 없다는 점을 보여줄 것이다. $g(Y_{2t}) = Y_{2t}$라고 더 명시적으로 가정한다. 또한 Y_{2t}는 X_t의 멱(power)인 X_t, X_t^2, …, X_t^k에 대하여 회귀하고, Y_t의 계산치는 아래와 같이 얻어진다고 한다.

$$\hat{Y}_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + \cdots + \hat{\alpha}_k X_t^k \qquad (8.26)$$

그러면 Y_{2t}를 $\hat{Y}_{2t}$로 대신하여 (8.12)에 TSLS절차를 적용할 경우, 그 추정 결과는 一致推定量이 아니다.

먼저 뒤의 명제를 설명하고 나서 앞의 명제를 증명하는 것이 편리하다. 우선, $g(Y_t) = Y_{2t}$이면, Y_t에 대한 선형모형 (8.12)와 (8.13) (즉, 축약형 방정식 ; reduced form equation)의 解는 다음과 같이 나타날 것이다.

$$Y_{2t} = \pi_0 + \pi_1 X_t + \psi_t \qquad (8.27)$$

여기서 어떤 주어진 X_t 값에 대하여 변수 Ψ_t의 평균은 0, $E(\Psi_t) = 0$이다.* 따라서, 선형인 경우, 어떤 주어진 X_t 값에 대응하는 Y_{2t}의 평균값은 X_t에서 선형, 즉 $(\pi_0 + \pi_1 X_t)$이다. 따라서 X_t, X_t^2, …, X_t^k에 대하여 Y_{2t}를 관련시키는 회귀모형은 다음과 같이 나타낼 수 있다.

$$Y_{2t} = \pi_0 + \pi_1 X_t + \pi_2 X_t^2 + \cdots + \pi_k X_t^k + \psi_t \qquad (8.28)$$

---

* 제 7 장 4절에서 ψ는 바로 교란항 ε_{1t}와 ε_{2t}의 선형결합임이 분명하였다.

여기서 $\pi_2 = \pi_3 = \cdots = \pi_k = 0$ 이다.

X_t, X_t^2, …, X_t^k에 대한 Y_{2t}의 회귀로부터 Y_{2t}의 계산치를 아래와 같다고 하자.

$$\hat{Y}_{2t} = \hat{\pi}_0 + \hat{\pi}_1 X_t + \hat{\pi}_2 X_t^2 + \cdots + \hat{\pi}_k X_t^k. \qquad (8.29)$$

그러면, (8.28)과 모형 (8.12)와 (8.13)의 가정으로부터, $E(\hat{\pi}_2) = E(\hat{\pi}_3) = \cdots E(\hat{\pi}_k) = 0$이 된다.* 또한, 더 나아가 합당한 기술적 가정 아래에서 $\hat{\pi}_0$, $\hat{\pi}_1$, …, $\hat{\pi}_k$가 각각이 대응하는 모수의 일치추정량임을 보일 수 있다. 그러므로 $\hat{\pi}_2$, …, $\hat{\pi}_k$는 확률상 0으로 수렴한다. 이러한 사실이 함축하고 있는 의미는 다음과 같다. 만일 $g(Y_{2t}) = Y_{2t}$ 그리고 표본 크기가 무한대이면, $\hat{Y}_{2t}$는 $(\pi_0 + \pi_1 X_t)$에 수렴할 확률이 1이다. 그러므로, 만약 모형이고 또한 표본 크기가 무한대이면 Y_{2t}를 (8.29)의 $\hat{Y}_{2t}$로 대신한 TSLS절차는 두번째 단계에서 完全多重共線性을 가질 확률이 1이라는 것으로 특징지워 진다. 그 방식은 쓸모없게 되고, 따라서 그 절차도 일치추정량을 구하는 절차가 아니다. 그러나, 유한한 표본에서 $\hat{\pi}_2$, $\hat{\pi}_3$, …, $\hat{\pi}_k$는 일반적으로 0이 되지 않는다는 점을 명심하자. 그러므로, 만일 표본 크기가 유한하다면, TSLS를 수행할 수 있다. 그 이유는 (8.29)의 $\hat{Y}_{2t}$가 常數項과 X_t와 완전한 다중공선성을 갖지 않기 때문이다. 그러나, 그 절차는 불일치함을 강조하여 둔다. 일치성이란 속성은 무한한 표본의 경우와 관련하기 때문이다.

이제 $g(Y_{2t}) \neq Y_{2t}$인 비선형 사례를 보자. 이미 이 경우에 $g(Y_{2t})$를 다음과 같이 나타낼 수 있을 것이다. 〔(8.22)를 보라〕.

---

* (8.28)의 모형은 우리가 제4장에서 검토한 일반적인 선형모형에 상응한다. 제4장에서 검토한 모형에 대해서 모수추정량이 불편추정량임을 보였다.

$$g(Y_{2t}) = G(X_t) + V_t \qquad (8.30)$$

여기서 어떤 X값에 대응하는 V_t의 평균값은 0이다. 또한 평균 함수 $G(X_t)$는 전형적으로 X_t에서 비선형이 될 것임을 지적한 바 있다.

만일 $g(Y_{2t})$가 X_t의 1乘부터 k乘까지에 회귀하고, 또한 표본 크기가 무한하다면, 그 결과로 추정한 곡선,

$$\widehat{g(Y_{2t})} = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + \cdots + \hat{\alpha}_k X_t^k \qquad (8.31)$$

은 곡선 $G(X_t)$에 대하여 가장 근사한 k次의 다항식이 될 것이다.* 만

---

* 형식에 구애받는 독자들을 위하여 X_t의 k차 다항식을 $P(X_t, \alpha)$로 보이기로 한다. 여기서 α는 그 다항식의 모수이다. 그러면 최소자승 절차는 α를 잡아 다음을 최소화시키거나,

$$L = \sum_{t=1}^{n} [g(Y_{2t}) - P(X_t, \alpha)]^2$$

또는 표현상으로 L/n을 최소화한다. $V_t = g(Y_{2t}) - G(X_t)$임을 상기하면 L/n은 다음과 같이 나타낼 수가 있다.

$$\frac{L}{n} = \frac{\sum [g(Y_{2t}) - G(X_t) + G(X_t) - P(X_t, \alpha)]^2}{n}$$

$$= \frac{\sum V_t^2}{n} + 2\frac{\sum D_t V_t}{n} + \frac{\sum D_t^2}{n}$$

여기서 $D_t = G(X_t) - P(X_t, \alpha)$이다. D_t의 값이 X_t의 값에 의하여 결정된다는 점에 유의하자. (8.30)을 참조하여, 어떤 X값에 대응하여 V_t의 평균값은 0임을 알고 있다. 따라서 D_t 값에 대응하는 V_t 평균값은 0이다. 그러므로 제 6 장 3절의 논의는 V_t가 D_t와 相關이 없음을 가리킨다. 이러한 이유로 적절한 가정 아래에서 이상의 교적(cross-product)이 갖는 확률상의 極限은 0임을 보일 수 있다. 따라서 무한한 표본의 경우에 L/n은 α를 선택하여 $\sum D_t^2 / n$을 최소화함으로써 최소화될 것이 분명하다. 왜냐하면, $\sum D_t^2 / n$이 α를 포함하고 있는 L/n의 유일한 성분이기 때문이다.

일 $G(X_t)$가 비선형이면, 가장 근사한 다항식은 일반적으로 비선형 다항식이 되지 않을 것이다. 실제로, 이것은 더 높은 차수의 다항식이 고려됨에 따라 근사치가 향상되는 전형적인 경우이다. 따라서 만일 표본 크기가 무한하면, (8.31)의 $g(\widehat{Y_{2t}})$는 일반적으로 X_t의 선형 함수로 축소되지 않는다. 결과적으로 무한한 표본의 경우에 TSLS 절차가 쓸모가 없게 되지는 않을 것이다.

다. 非線型模型의 식별에 대한 규칙

이제 독자들은 선형모형의 식별에 대한 규칙을 아무런 변경없이 비선형 모형에 적용할 수 없음을 확인하게 된다. 비선형모형에 대응하는 규칙을 지금 제시하기로 한다. 그리고 나서 그 규칙이 설득력을 갖고 있음을 시사하는 예증을 제공할 것이다.

모수에서는 선형이지만 외생변수에서 비선형인 M개 방정식 모형을 검토하기로 하자. 이 모형의 각 방정식은 보통 주어진 경제변수와 관련이 있거나 또는 연관되어 있다. 이 변수는 전형적으로 방정식의 좌변에 있으며 그 계수는 묵시적으로 1이다. 이들 변수를 기본내생변수 (basic endogenous variable) 라고 한다. 예를 들어, (8.12)와 (8.13)에 있는 모형의 기본내생변수는 Y_{1t}와 Y_{2t}이다. 이 이외에 모형에 나타나 있는 다른 내생변수를 추가내생변수 (additional endogenous variable) 이라고 하자. 예를 들어, (8.12)와 (8.13)의 모형은 단 하나의 추가내생변수 $g(Y_{2t})$를 포함하고 있다. 제 8 장 1절의 논의는 만일 모형이 어떤 추가내생변수를 포함하고 있지 않다면, 그 모형은 추정하는 데에서 선형모형임을 의미한다.

모형의 교란항은 0의 평균을 갖고 자기상관을 갖고 있지 않으며 모형에 나타나는 모든 외생변수와 독립적이라고 가정하자. 또한, 모형의 기본

내생변수는 교란항, 외생변수, 시차 외생변수(만일 있다면), 그리고 추가내생변수로써 나타낼 수 있다고 가정한다. 예를 들어, (8.12)와 (8.13)으로 주어진 모형을 보자. 방정식 (8.12)는 기본내생변수가 우변에 없기 때문에 요구되어지는 형식(required form)으로 되어 있다. Y_{2t}에 대하여 비슷한 식은 단지 (8.13)의 Y_{1t}에 (8.12)를 대입만 하면 쉽게 얻을 수 있다.

마지막으로 외생변수, 시차 외생변수, 그리고 교란항에 의한 기본내생변수에 대한 모형의 해는 유일하다고 가정하자. 이 가정이 유지되지 않으면, 기본내생변수를 애초에 설명하려고 구성한 모형이 그 변수를 결정하거나 또는 설명하는 데 불완전 모형이 될 것이다.

이상의 조건과 약간의 기술적인 가정을 추가할 경우, 만일 다음과 같다면 주어진 모형의 방정식의 모수를 일치하게 추정할 수 있으므로, 식별된다. 예를 들어 i번째 방정식의 경우,

$$A_{1i} \geq A_{2i} \qquad (8.32)$$

여기서 A_{2i}는 i번째 방정식의 우변에 나타나는 기본내생변수의 숫자이고 A_{1i}는 모형에는 나타나나 i번째 방정식에는 없는 事前決定變數와 추가내생변수의 숫자이다. 우리가 현재 고려중인 비선형체계에서는 사전결정변수가 교란항의 經常値와 상관이 없이 모형에 있는 어떤 변수로 정의된다. 그러므로, 사전결정변수는 그 값이 동시에 존재하는 하나 또는 그 이상의 기본내생변수의 값에 종속하지 않고서 모형에 나타나는 어떤 변수가 될 것이다.(즉, 그 변수의 값이 오직 외생변수와 시차 내생변수에만 종속한다).

(8.32)의 계산 규칙만으로도 고려 대상이 되는 모형의 i번째 방정식을 식별하는 필요조건이 된다. 이는 만약 i번째 방정식이 식별되면 (8.32)가 성립하지만, (8.32)가 저절로 i번째 방정식이 사실상 식별

된다는 것을 보장할 수 없다는 사실을 의미한다. 非線型模型에서 주어진 모형의 방정식이 식별됨을 보장하기 위하여 만족시켜야 하는 "추가적인 기술적" 조건들은 결정하기 어렵고 실제로 거의 검토되지 않는다. 전형적인 절차는 (8.32)를 검사하고서 만약 (8.32)를 만족시키면, 더 나아가 식별이 성립되기에 충분한 "기술적 조건들"을 가정한다.

(8.32)로 주어진 계산규칙은 제 7 장의 결과가 시사한 규칙과는 다르다. 왜냐하면, 추가내생변수가 기본내생변수보다는 오히려 사전결정변수와 무리를 이루기 때문이다. (8.32)의 규칙을 정당화하기전에, 그 규칙은 (8.12)와 (8.13)의 모형에 적용하여 보자. 방정식 (8.12)에 대해서는 ($i=1$로 놓아서) $A_{11}=0$이다. 그 이유는 방정식에는 아무런 사전결정변수도 배제되어 있지않기 때문이다. 그리고 우변에는 기본내생변수가 없으므로, $A_{21}=0$이다. 그러므로 (8.32)는 ($0 \geq 0$)이 성립하여서 더 기술적인 조건을 가정하면, (8.12)는 식별된다. (8.13)에 대해서는 $g(Y_{2t})$와 X_t가 (8.13)에서 빠져있기 때문에 $A_{12}=2$이다. 그리고 Y_{1t}가 우변에 있기 때문에, $A_{22}=1$이다. 그러므로 $A_{12} \geq A_{22}$이다. 따라서 더 기술적인 조건을 가정하면, 방정식 (8.13)도 식별된다.

라. 규칙의 정당화

이제 추가내생변수가 식별의 목적 달성을 위하여 사전결정변수와 묶여지는 이유를 보도록 하자. 다시 (8.12)와 (8.13)의 단순한 모형을 검토하면, 이 모형은 추가내생변수 $g(Y_{2t})$때문에 비선형임을 알고 있다. 그러나, (8.30)에서 $g(Y_{2t})$를 두 성분의 합으로 나타낼 수 있음을 보였다. 첫번째 성분은 X_t의 비선형함수, 즉 $G(X_t)$이고, 두번째 성분, 즉 V_t는 어떤 주어진 X_t 값에 대응하고 0의 평균값을 갖는 교란항이다. 또한 예컨데 (8.31)에서 볼 수 있듯이, $G(X_t)$는 X_t의 멱에 대

한 $g(X_t)$의 다항회귀로써 접근될 수 있다는 사실을 보인 바 있다.

만일 (8.30)을 (8.12)에 대입하면, 2개 방정식 모형 (8.12)와 (8.13)은 다음과 같이 나타낼 수 있다.

$$Y_{1t} = a_0 + a_1 G(X_t) + a_2 X_t + w_t \qquad (8.33)$$

$$Y_{2t} = b_0 + b_1 Y_{1t} + \varepsilon_{2t} \qquad (8.13)$$

여기서 $w_t = \varepsilon_{1t} + a_1 V_t$이다. w_t는 바로 ε_{1t}와 V_t의 선형결합이기 때문에, 어떤 주어진 X_t 값에 대응하여 0의 평균값을 갖는 교란항으로 볼 수 있다. 그러므로, w_t는 X_t나 $G(X_t)$ 어느 것과도 相關이 없다.

방정식 (8.33)과 (8.13)은 종속변수 Y_{1t}와 Y_{2t}인 2개 방정식 선형모형으로 간주될 수 있으며, 常數項과 사전결정변수로서 $G(X_t)$와 X_t를 포함하고 있다. 이 모형의 원래의 모형 (8.12)와 (8.13)과 동일한 회귀모수, 즉 a_0, a_1, a_2, b_0와 b_1을 포함하고 있음에 유의하자. 또한 교란항을 제쳐 두고서라도 이 모형은 추가내생변수 $g(Y_{2t})$를 그것의 X_t로 이루어진 "평균 함수", 즉 $G(X_t)$로 대체하기만 하면, 원래의 모형에서 얻을 수 있다는 점도 주의하자. 만일 (8.33)과 (8.13)이 식별되면, (8.12)와 (8.13)도 반드시 식별된다는 사실은 확실하다. 이들 모형은 동일한 모수를 갖고 있다.

표현상의 편의를 위하여 $G(X_t)$에 대한 관찰치가 일단 X_t가 관찰되면 결정될 수 있다고 가정하여 논의를 계속하기로 한다. 이 가정은 필수적이지는 않으나 논의를 단순하게 한다. 이러한 가정을 합리화하는 한 방법은 $G(X_t)$가 X_t의 k차 다항식과 완전히 근사하게 될 수 있고, 이 다항식은 X_t의 冪으로 $g(Y_{2t})$를 회귀시킴으로써 일치하게 추정될 수 있다고 가정하는 것이다. 예를 들어 이러한 조건 아래에서 $G(X_t)$의 값은 X_t의 값으로 다음과 같이 결정될 것이다[(8.31)을 보라].

$$G(X_t) = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + \cdots + \hat{\alpha}_k X_t^k \qquad (8.34)$$

우리는 오직 大標本, $n=\infty$만 보기 때문에 k의 "매우 큰" 값을 허용할 수 있다.

$G(X_t)$의 관찰치가 이용가능하다면, (8.33)과 (8.13)의 모형은 제7장에서 고려한 선형모형의 분석틀에 직접 맞추어 진다. 즉, 더 기술적인 가정을 하면 (8.33)은 방정식의 우변에 사전결정변수만을 포함하고 있기 때문에, 식별됨을 알게 된다. 다른 말로 하면, 제7장의 표기에 따라서 K_1과 K_2가 모두 0이기 때문에 $K_2 \geq K_1$이 되므로 (8.33)은 식별된다. 여기서 K_2는 그 방정식에 빠져있는 사전결정변수의 숫자이고 K_1은 우변에 있는 내생변수의 숫자이다. 이러한 K_1과 K_2에 관한 결과가 이상의 A_{11}과 A_{21}에 관한 결과와 동일하고 원래의 방정식 (8.12)와 (8.13)에서 추가내생변수 $g(Y_{2t})$를 사전결정변수로 분류만 하면 그 결과를 얻을 수 있다는 사실을 명심하자. 마찬가지로 (8.13)에 관해서는, $K_2 = 2$〔$G(X_t)$와 X_t가 빠져 있다〕이고 $K_1 = 1$ (Y_{1t}가 포함되었기 때문이다) 이어서 $K_2 \geq K_1$이다. 이 결과는 단지 $g(Y_{2t})$를 원래의 비선형모형 (8.12)와 (8.13)의 사전결정변수로서 분류하기만 하면 얻을 수 있다는 점을 다시 명심하기 바란다.

마. 식별규칙의 정당화에 관한 일반화

이상의 결과는 일반화할 수 있다. 이를 보이기 위하여 먼저 보다 일반적인 모형에 대해서 (8.30)에 상응하는 결과를 보인다.

일반적으로 우리가 고려하는 m개 방정식의 비선형모형은 사전결정변수뿐만 아니라 많은 추가내생변수를 포함하고 있을 것이다. 매우 합리적인 가정 아래에서 각각의 추가내생변수는 두 성분의 合으로 나타낼 수 있다.

(앞에서의 $G(X_t)$와 유사한)한 성분은 주어진 사전결정변수의 값에 대응하는 추가내생변수의 평균값을 제공할 것이다. 다른 한 성분은 어떤 주어진 사전결정변수의 값에 대하여 평균이 0인 교란항이다. 설명을 위하여 $(Y_{1t}\,Y_{2t})$가 한 추가내생변수이고 X_{1t}, X_{2t}, X_{3t}가 비선형모형의 사전결정변수라고 가정한다. 그러면, 합리적인 가정 아래에서 다음과 같이 나타낼 수 있다.

$$(Y_{1t}Y_{2t}) = H(X_{1t}, X_{2t}, X_{3t}) + \Psi_t \qquad (8.35)$$

여기서 $H(X_{1t}, X_{2t}, X_{3t})$는 X_{1t}, X_{2t}와 X_{3t}의 함수이고, Ψ_t는 어떤 주어진 X_{1t}, X_{2t}, X_{3t} 값의 집합에 대응하여 평균이 0인 변수이다. 예를 들어, $H(X_{1t}, X_{2t}, X_{3t}) = (X_{1t}^2 + X_{2t})\,e^{X_{3t}}$이고 $X_{1t}=3$, $X_{2t}=5$ 그리고 $X_{3t}=0$이면, 대응하는 $(Y_{1t}\,Y_{2t})$의 평균값은 다음과 같을 것이다.

$$H(3, 5, 0) = (9 + 5)e^0 = 14 \qquad (8.36)$$

설명을 달리 하여 우리가 X_{1t}, X_{2t}, X_{3t}, Y_{1t}와 Y_{2t}에 대한 관찰치 표본을 갖고 있다고 하자. X_{1t}, X_{2t}와 X_{3t}에 대한 관찰치가 H_t에 대한 관찰치를 구성하는 데(앞서 설명한 것처럼) 사용된다고 가정하자. 여기서 H_t는

$$H_t = (X_{1t}^2 + X_{2t})e^{X_{3t}} \qquad (8.37)$$

이다.

따라서 만약 표본이 무한하면 (수직축상의) Y_{1t} Y_{2t}와 H_t 사이의 산포도는 원점을 지나는 45°선상에 있는 점들로 구성된다.

이제 보다 일반적인 m개 방정식의 비선형모형으로서 모수에서는 선형인 경우를 보자. 기본내생변수를 Y_{1t}, …, Y_{mt}라 하자. 추가내생변수는 $g_{1t}=$

$g_1(Y_{1t}, \cdots, Y_{mt}), \cdots, g_{rt} = g_r(Y_{1t}, \cdots, Y_{mt})$라 놓자.* 그리고 사전결정변수는 $X_{1t}, \cdots, X_{pt}$라 하자. 앞에서의 논의는 g_{it}를 합당한 가정 아래에서 다음과 같이 나타낼 수 있다는 점을 시사한다.

$$g_{it} = H_i(X_{1t}, \ldots, X_{pt}) + \psi_{it} \qquad i = 1, \ldots, r \qquad (8.38)$$

여기서 ψ_{it}의 평균은 어떤 주어진 $X_{1t}, \cdots, X_{pt}$ 값의 집합에 대해서 0이다. 이제 남은 논의는 명백할 것이다. 지금 고려중인 비선형모형은 단지 각각의 추가내생변수를 그 식인 (8.38)로 대신하고서 각 방정식 우변의 끝에 교란항을 모아 놓으면 선형모형으로 될 수 있다.** 그 결과인 선형모형은 내생변수로서 $Y_{1t}, \cdots, Y_{mt}$ 그리고 사전결정변수로서 $X_{1t}, \cdots, X_{pt}, H_{1t}, \cdots, H_{rt}$를 포함할 것이다. 여기서 H_{it}는

$$H_{it} = H_i(X_{1t}, \ldots, X_{pt}), \qquad i = 1, \ldots, r \qquad (8.39)$$

이다.

이 선형모형은 원래의 비선형모형과 동일한 회귀 모수를 포함할 것이다. 만일 제7장에서 제공된 식별에 관한 규칙을 이상의 선형모형의 각 방정

---

* 일반적으로 추가내생변수는 기본내생변수와 마찬가지로 사전결정변수의 함수일 것이다. 표기를 단순화하기 위하여 이를 다루지 않는다.

** 한 예로서, 첫번째 방정식은

$$Y_{1t} = a_0 + a_1 Y_{2t} + a_2 g_{1t} + a_3 g_{2t} + a_4 X_{1t} + \varepsilon_{1t}$$

이다.

그러면 (8.38)은 이 방정식을 다음과 같이 다시 쓸 수 있음을 의미하는 것이다.

$$Y_{1t} = a_0 + a_1 Y_{2t} + a_2 H_{1t} + a_3 H_{2t} + a_4 X_{1t} + (\varepsilon_{1t} + a_2 \psi_{1t} + a_3 \psi_{2t})$$

여기서 $(\varepsilon_{1t} + a_2\,\psi_{1t} + a_3\,\psi_{2t})$는 교란항으로 다룬다.

식에 적용하면, 그 결과는 분명 (8.32)를 적용하여 얻는 결과와 동일할 것이다.

여기에 우리가 이미 언급한 "더 나아간 기술적 가정"과 관련하여 있으며, 이상의 내용의 윤곽을 파악하는 데에 도움을 주는 점이 있다. 지금 막 얻은 결과는 사전결정변수 $X_{1t}, \cdots, X_{pt}, H_{1t}, \cdots, H_{mt}$가 多重共線性이 없다는 묵시적인 가정에 기초하고 있는 것이다. 만일 이들 변수가 다중공선성을 가지면, (8.32)의 결과는 성립하지 않는다. 예를 들어 다시 (8.38)을 위에서 지적한 비선형모형에 적용하여 얻은 결과인 선형모형을 검토하자. 그 모형은 첫번째 방정식의 우변에 3개의 기본내생변수 Y_{2t}, Y_{3t}, Y_{4t}와 상수항 그리고 사전결정변수 X_{1t}를 포함하고 있다 하자. 이 방정식은 X_{2t}, X_{3t} 그리고 H_{1t}를 배제하고 있다. 그러나 $H_{1t}=X_{2t}+X_{3t}$이다. "더 나아간 기술적 조건"의 제약 아래에서 이 방정식이 식별된다고 결론 내릴 수 있겠는가? 절대로 그럴 수 없다! 예를 들어 그 선형모형의 첫번째 방정식이 갖는 모수를 추정하는 데에 TSLS 절차를 적용하려 한다고 가정하자. 첫단계에서 각각의 변수 Y_{2t}, Y_{3t} 와 Y_{4t}를 상수항 X_{1t}, X_{2t}, X_{3t} 그리고 H_{1t}에 회귀시켜서 $\hat{Y}_{2t}, \hat{Y}_{3t}$ 그리고 $\hat{Y}_{4t}$를 계산하려 할 것이다. 그러나 이러한 첫번째 시도 노력은 선형관계 $H_{1t}=X_{2t}+X_{3t}$에 기인한 完全多重共線性 때문에 소용없게 된다. 분명히 그 방정식은 식별되지 않는다. 그 이유는 우변에 3개의 내생변수를 포함하고 있는 반면에 방정식에서 생략된 "非多重共線性의" 변수는 단지 2개이기 때문이다.

사전결정 변수 $X_{1t}, \cdots, X_{pt}, H_{1t}, \cdots, H_{mt}$의 집합이 다중공선성을 갖는지 여부를 연역하여 낼 수 있는 여러 조건이 있다. 이에 덧붙여서, 우리가 제시하였던 결과는 이러한 사례를 감안할 수 있도록 수정될 수 있다. 그러나 그 논의는 복잡하고, 완전다중공선성의 사례는 실제로 거의 접하기

힘들다. 그러므로 수준높은 독자들에게 다른 문헌*을 소개하는 것으로 결론을 맺기로 하고, 여타 독자들은 우리의 분석이 완전하지는 못하다는 점에 대하여 주의를 환기하기 바란다.

3. 이단계 최소자승 추정

이 節에서는 모수에서는 선형이지만 내생변수에서는 비선형인 계량경제학 모형을 추정하는 이단계 최소자승 절차의 윤곽을 잡기로 한다. 제시하는 절차는 8장 2절에서의 윤곽을 직접 일반화한 것이다.

가. 절차의 윤곽

우리가 고려한 종류의 계량경제학 모형에서 i번째 방정식이 식별된다고 가정하자. 그러면, 합당한 조건 아래에서 다음의 절차에 따라 그 방정식은 一致 推定될 수 있다.

첫번째, 방정식의 우변에 있는 각 기본내생변수를 모형에 있는 사전결정변수에 관하여 그리고 아마도 그 사전결정변수의 冪(즉, 제곱, 3제곱 등)에 대해 회귀시킴으로써, 그 기본내생변수의 계산치를 얻는다. 아래에서는 사전결정변수의 冪을 사용해야 하는지 여부를 자세히 설명할 것이다.

---

* F. Fisher, The Identification Problem in Econometrics (New York : McGraw-Hill, 1966)의 5장과 H. H. Kelejian, "Identification of Nonlinear Systems : An Interpretation of Fisher," Princeton University, Econometric Research Program, Research Paper No.22(Revised), 1970, 관련된 논점들을 모두 잘 재검토한 것은 S. Goldfeld and R. Qundt, Nonlinear Methods in Econometrics (Amsterdam : North Holland, 1972)의 제8장이다.

두번째, (1)에서 설명한 것과 같은 방식으로 추가내생변수의 계산치를 얻는다.

세번째, i 번째 방정식의 기본 및 추가내생변수를 그 계산치로 대체하고 최소자승으로 모수를 추정한다.

합당한 조건아래에서는 二段階 절차로 얻은 추정량이 一致推定量임을 보일 수 있다. 이제 이 절차의 약간 난해한 부분의 개요를 설명하기로 한다.

노트1. 전형적으로, 모형이 많은 사전결정변수를 포함하고 있는 경우 사전결정변수가 없어도 二段階에서 完全多重共線性이 초래되지 않는다면, 사전결정변수의 멱을 一段階에서 사용할 수 없다. 이 논지의 요점은 추정절차의 바람직한 속성, 즉 一致性은 大標本의 속성 ($n = \infty$)이라는 것이다. 만약 표본 크기가 무한하면, 적당한 조건 아래에서는 1段階에서 사전결정변수의 (더 낮은 次數 모두와 함께) 점점 더 높은 次數의 멱을 사용할수록 二段階에서 얻는 추정량의 분산은 더욱 더 작아질 것이다. 그 원리는 一段階의 다항 회귀가, 고려되는 멱의 차수가 높아질수록 대응하는 평균 함수에 더 좋은 근사치로 된다는 것이다. 〔(8.31) 과 그 밑의 각주를 보라〕. 그러나 실제로 표본 크기는 有限하다. 그러므로 一段階에서 고려되는 독립변수의 숫자 (부분적으로는 고려되는 멱의 차수에 달려 있다)는 제한되어야 한다.* 실제로, 일단계에서의 변수의 숫자가 관찰치의 숫자와 동일할 정도로 변수의 많은 멱이 고려된다면, 이단계 최소

---

* 예를 들어, 회귀모형 $Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{2t}^2 + b_4X_{2t}^2 + \varepsilon_t$ 는 (상수항을 포함하여) 5개의 독립변수를 포함하고 있다.

자승 절차는 통상 최소자승법(ordinary least squares)으로 축약되고,* 따라서 一致推定量을 발생시킨다. 그러므로, 우리는 곤경에 처하게 된다. 표본의 큰 분산을 줄이기 위해서는 일단계에서 변수의 숫자를 증가시켜야만 한다. 다른 한편으로, 표본 크기가 유한하면, 일단계에서 사용된 변수의 숫자가 표본에 접근함에 따라 二段階最小自乘推定量은 점점 더 最小自乘推定量과 같게 되는데, 이는 一致推定量이 아니다! 표본 크기와 일단계에서 사용된 변수의 숫자 사이의 최적 비율은 공공연한 문제이다. 그러나, 가능하다면 표본 크기와 일단계에서 사용된 변수의 숫자 사이의 차이가 적어도 20은 되어야 할 것이다.

노트2. w를 모형에서의 사전결정변수의 숫자로 하고, N을 표본 크기라 하자. 그러면, (1)에서 $(N-w)\geq 20$을 묵시적으로 가정하였기 때문에, 그 결과 사전결정변수의 멱수만이 일단계에서 제한되어야 한다. 그러나, 일부 대규모 모형에서는 w가 적어도 N의 크기가 되거나 그렇지 않아도 $(N-w)<20$이 되는 사례가 있을 것이다. 그러한 모형에 대해서는 일단계에 들어가는 선형의 사전결정변수의 숫자가 앞의 (1)에서와 같은 이유로 하여서 제한되어야만 한다. 일단계 회귀에 들어 가는 사전결정변수의 집합을 선정하는 다양한 방식이 있다. 그러나, 그 결과로 얻는 추정량이 일치추정량이 되기 위해서는 이러한 사전결정변수의 집합이 반드시 추정중인 방정식에 있는 모든 사전결정변수를 포함하여야 하고, 적어도 우변에 있는 내생 및 추가내생변수의 수만큼 방정식에 포함되지 않은 사전결정변수가 있어야 한다.

---

* 이단계 최소자승은 내생변수의 계산치가 대응하는 실제 값과 동일하다면 통상 최소자승으로 축약된다. 이론적인 난해점을 무시하고 보면, 이는 만약 표본 크기가 일단계에서 사용된 변수의 숫자(상수항 포함)와 같다면 틀림없이 발생하는 것이다. 이러한 결과는 명명백백하다. 즉, 만일 N개 변수로 설명할 예정인 한 변수의 N개 값이 있고, 따라서 N개 모수가 있다면, 그 설명은 완벽하게 됨이 분명하다.

비록 지금 고려하고 있는 모형이 비선형모형이지만, 이상의 이유는 선형모형의 사례에 대한 제7장 4절과 5절에서 제공된 근거와 똑 같다. 다음의 (3)에서 설명된 논점이 이 점을 명확하게 할 것이다.

노트3. (일단계의) 독립변수의 동일한 집합이 이단계에서 사용되는 변수의 계산치를 모두 얻는 데에 사용되어야 한다. 명백히 하기 위해서 i 번째 방정식을 다음과 같이 가정하기로 하자.

$$Y_{it} = b_0 + b_1 Y_{1t} + b_2(Y_{2t}Y_{3t}) + b_3 Y_{2t}^2 + a_1 X_{1t} + \varepsilon_{it} \qquad (8.40)$$

더 나아가 완전한 모형을 사전결정변수 X_{2t} 와 X_{3t} 도 포함한다고 한다. $\hat{Y}_{1t}$ 를 회귀로부터 얻도록 한다.*

$$Y_{1t} = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{1t}^2 + \alpha_5 X_{2t}^2 + V_{1t} \qquad (8.41)$$

$Z_{1t} = (Y_{2t}Y_{3t})$ 그리고 $Z_{2t} = Y_{2t}^2$ 이라 하자. 그러면 $\hat{Z}_{1t}$ 는 반드시 Z_{1t} 를 상수항 X_{1t}, X_{2t}, X_{3t}, X_{1t}^2, X_{2t}^2 에 회귀시켜서 얻어야 한다. 마찬가지로 $\hat{Z}_{2t}$ 는 Z_{2t} 를 반드시 동일한 변수의 집합에 회귀시킴으로써 얻어야만 한다. 만약 $\hat{Y}_{1t}$, $\hat{Z}_{1t}$ 와 $\hat{Z}_{2t}$ 를 정하는 데에 변수의 동일한 집합을 사용하지 않는다면, 이단계에서 얻은 추정량은 일치추정량이 아니게 된다. 그 이유가 무엇인지를 바로 아래에서 지적할 것이다.

노트4. (8.40)에 관한 절차를 논의하는 가운데, $Z_{1t} = (Y_{2t}Y_{3t})$ 와 $Z_{2t} = Y_{2t}^2$ 이 이단계에서 $\hat{Z}_{1t}$ 과 $\hat{Z}_{2t}$ 로 대체되어야만 한다고 하였다. 대신에 $\hat{Y}_{2t}$ 과 $\hat{Y}_{3t}$ 가 Y_{2t} 와 Y_{3t} 를 일단계 설명변수에 회귀시켜 얻은 것인데 만일 $(Y_{2t}Y_{3t})$ 와 Y_{2t}^2 을 $(\hat{Y}_{2t}\hat{Y}_{3t})$ 와 $(\hat{Y}_{2t})^2$ 으로 대체한다면, 이단계 회귀

---

* (8.41)이 X_{3t}^2 을 포함하지 않는다는 것을 유의하라. 이렇게 한 이유는 단지 일단계 회귀가 X_{1t} 와 X_{2t} 의 自乘을 포함하고 있기 때문에, X_{3t} 의 자승도 포함할 필요가 없다는 것을 보이기 위한 것이다.

로부터 얻은 회귀 모수의 결과적인 추정량이 불편추정량이 아니게 될 것이다. 더 일반적인 경우를 보기 위해서, $g_{1t} = g_1(Y_{1t}, \cdots, Y_{mt})$를 관심사가 되는 방정식의 설명변수인 추가내생변수라 하자. 그러면 회귀 모수의 일치 추정은 g_{1t}를 이단계에서 $\hat{g}_{1t}$으로 대체하는 것을 필요로 한다. 이 $\hat{g}_{1t}$는 g_{1t}를 사전결정변수(아마도 사전결정변수의 멱)으로 회귀시켜 얻는다. Y_{jt}를 사전결정변수(아마도 이것의 멱)에 회귀시켜 각각의 $\hat{Y}_{jt}$를 얻고서 이단계에서 g_{1t}를 $g_1(\hat{Y}_{1t}, \cdots, \hat{Y}_{mt})$로 대체한다면, 회귀 모수의 추정량은 일치추정량이 되지 않는다. 이러한 원리를 검토하기 전에, 모든 내생변수의 계산치를 결정하는 데에 왜 일단계에서와 동일한 설명변수의 집합이 사용되어야 하는가의 설명으로 되돌아 간다.

나. 일부 미묘한 내용의 정당화

방정식 (8.40)을 다시 보자. (8.40)이 일부가 되는 모형에서 사전결정변수가 常數項, X_{1t}와 X_{2t}라고 가정한다. 설명의 편의상, 일단계에서 사용되는 변수는 상수항, X_{1t}, X_{2t}, X_{1t}^2, X_{2t}^2이라 가정한다. 이들 변수에 대해 Y_{1t}를 회귀시켜 얻은 Y_{1t}의 계산치를 $\hat{Y}_{1t}$라고 한다. 그러면 $\hat{Y}_{1t}$는 다음과 같은 설명변수의 선형결합이 될 것이다.

$$\hat{Y}_{1t} = \hat{d}_0 + \hat{d}_1 X_{1t} + \hat{d}_2 X_{2t} + \hat{d}_3 X_{1t}^2 + \hat{d}_4 X_{2t}^2 \qquad (8.42)$$

여기서 $\hat{d}_0, \cdots, \hat{d}_4$는 일단계 회귀에서 추정한 모수이다.

$\hat{\phi}_{1t}$를 일단계 회귀로부터 추정한 殘差라고 하자.

$$\hat{\phi}_{1t} = Y_{1t} - \hat{Y}_{1t} \qquad (8.43)$$

그러면, 다음과 같은 조건에 기초한 일단계의 정규방정식들 때문에, 앞장들로부터 $\Sigma(\hat{\phi}_{1t}\hat{Y}_{1t})=0$이라는 결과를 알고 있다.

$$\begin{aligned}&\sum\hat{\phi}_{1t}=0, \quad \sum(\hat{\phi}_{1t}X_{1t})=0, \quad \sum(\hat{\phi}_{1t}X_{2t})=0\\&\sum(\hat{\phi}_{1t}X_{1t}^{2})=0, \quad \sum(\hat{\phi}_{1t}X_{2t}^{2})=0\end{aligned} \tag{8.44}$$

이제 $Z_{it}=(Y_{2t}Y_{3t})$라 하고, $\hat{Z}_{1t}$을 일단계 설명변수와 동일한, 즉 상수항, X_{1t}, X_{2t}, X_{1t}^{2} 와 X_{2t}^{2}에 Z_{1t}를 회귀시켜서 얻은 Z_{1t}의 계산치라 하자. $\hat{\phi}_{2t}$을 대응하는 推定 殘差라고 하면,

$$\hat{\phi}_{2t}=Z_{1t}-\hat{Z}_{1t} \tag{8.45}$$

$\hat{Z}_{1t}$를 계산한 회귀식이 다음과 같은 조건에 기초하였음에 유의한다.

$$\begin{aligned}&\sum\hat{\phi}_{2t}=0, \quad \sum(\hat{\phi}_{2t}X_{1t})=0, \quad \sum(\hat{\phi}_{2t}X_{2t})=0\\&\sum(\hat{\phi}_{2t}X_{1t}^{2})=0, \quad \sum(\hat{\phi}_{2t}X_{2t}^{2})=0\end{aligned} \tag{8.46}$$

또한 (8.46)의 조건이 $\Sigma(\hat{\phi}_{2t}\hat{Z}_{1t})=0$을 의미한다는 사실을 명심하자.

마지막으로 $Z_{2t}=Y_{2t}^{2}$이라 하고, $\hat{Z}_{2t}$를 동일한 설명변수의 집합에 Z_{2t}를 회귀시켜 얻은 Z_{2t}의 계산치라 하자. 이 회귀에서의 추정 잔차를 다음과 같다고 한다.

$$\hat{\phi}_{3t} = Z_{2t} - \hat{Z}_{2t} \tag{8.47}$$

그리고 $\hat{\phi}_{3t}$ 가 다음의 조건을 만족시킨다는 점에 유의하자.

$$\begin{aligned} &\sum(\hat{\phi}_{3t}) = 0, \quad \sum(\hat{\phi}_{3t}X_{1t}) = 0, \quad \sum(\hat{\phi}_{3t}X_{2t}) = 0 \\ &\sum(\hat{\phi}_{3t}X_{1t}^2) = 0, \quad \sum(\hat{\phi}_{3t}X_{2t}^2) = 0 \end{aligned} \tag{8.48}$$

또한 (8.48)의 조건이 $\Sigma(\hat{\phi}_{3t}\hat{Z}_{2t}) = 0$ 을 의미한다는 점을 명심하자. 더구나, $\hat{Y}_{1t}$, $\hat{Z}_{1t}$ 와 $\hat{Z}_{2t}$ 가 동일한 집합의 변수로 이루어진 선형결합이기 때문에, (8.44), (8.46)과 (8.48)은 $i = 1, 2, 3$에 대하여 다음과 같음을 의미한다.

$$\sum(\hat{\phi}_{it}\hat{Y}_{1t}) = 0, \quad \sum(\hat{\phi}_{it}\hat{Z}_{1t}) = 0, \quad \sum(\hat{\phi}_{it}\hat{Z}_{2t}) = 0 \tag{8.49}$$

즉, 한 일단계 회귀에서의 잔차와 다른 일단계 회귀에서의 변수의 계산치를 서로 곱하여 더한 합계는 0이다.

예비적인 결과가 하나 더 필요하다. (8.42)에서 Y_{1t} 의 계산치는 모수 추정량 $\hat{d}_0, \cdots, \hat{d}_4$ 에 경유하여 일단계 설명변수와 관련되어 있다. 만일 표본이 무한하면, 이들 추정량은 상수에 수렴한다. 이들 상수를 각각 $d_0, \cdots, d_4$ 라고 표시하자. 마찬가지로 大標本의 $\hat{Y}_{1t}$ 값을 Y_{1t}^m 으로 표시하는데, 여기서 Y_{1t}^m 은,

$$Y_{1t}^m = d_0 + d_1X_{1t} + d_2X_{2t} + d_3X_{1t}^2 + d_4X_{2t}^2 \tag{8.50}$$

이상과 대응하는 방식으로 대표본의 $\hat{Z}_{1t}$ 과 $\hat{Z}_{2t}$ 값을 각각 Z_{1t}^m, Z_{2t}^m 이라 하자.

이제 二段段最小自乘 推定量의 一致性은 일단계 회귀에서의 모든 설명변수의 동일한 집합을 요구한다는 사실을 보이기로 하자. 식 (8.43), (8.45)

와 (8.47)은 다음과 같이 재정리할 수 있다.

$$\begin{aligned} Y_{1t} &= \hat{Y}_{1t} + \hat{\phi}_{1t} \\ Z_{1t} &= \hat{Z}_{1t} + \hat{\phi}_{2t} \\ Z_{2t} &= \hat{Z}_{2t} + \hat{\phi}_{3t} \end{aligned} \tag{8.51}$$

(8.51)의 식들을 추정중인 식, 즉 (8.40)에 대입하면,

$$Y_{it} = b_0 + b_1\hat{Y}_{1t} + b_2\hat{Z}_{1t} + b_3\hat{Z}_{2t} + a_1X_{1t} + W_t \tag{8.52}$$

여기서 $W_t = b_1\hat{\phi}_{1t} + b_2\hat{\phi}_{2t} + b_3\hat{\phi}_{3t} + \varepsilon_{it}$ 이다. 제 7 장에서의 논의와 비슷한 방식으로, 교란항 W_t의 적합한 성분은 ε_{it}임을 명심하자. 그 이유는,

$$\begin{aligned} &\sum W_t = \sum \varepsilon_{it}, \qquad \sum (\hat{Z}_{2t}W_t) = \sum (\hat{Z}_{2t}\varepsilon_{it}) \\ &\sum (W_t\hat{Y}_{1t}) = \sum (\hat{Y}_{1t}\varepsilon_{it}), \qquad \sum (X_{1t}W_t) = \sum (X_{1t}\varepsilon_{it}) \\ &\sum (W_t\hat{Z}_{1t}) = \sum (\hat{Z}_{1t}\varepsilon_{it}) \end{aligned} \tag{8.53}$$

그러므로 (8.52)를 측정하는 것은 최소자승과 동일한, 아래의 주어진 조건의 〔(8.53)을 보라〕 전형적인 절차로 이루어짐을 암시한다.

$$\begin{aligned} &E[\sum \varepsilon_{it}] = 0 \text{ 이므로} && \sum \hat{W}_t = 0 \\ &E[\sum (Y_{1t}{}^m\varepsilon_{it})] = 0 \text{ 이므로} && \sum (\hat{W}_t\hat{Y}_{1t}) = 0 \\ &E[\sum (Z_{1t}{}^m\varepsilon_{it})] = 0 \text{ 이므로} && \sum (\hat{W}_t\hat{Z}_{1t}) = 0 \\ &E[\sum (Z_{2t}{}^m\varepsilon_{it})] = 0 \text{ 이므로} && \sum (\hat{W}_t\hat{Z}_{2t}) = 0 \\ &E[\sum (X_{1t}\varepsilon_{it})] = 0 \text{ 이므로} && \sum (\hat{W}_tX_{1t}) = 0 \end{aligned} \tag{8.54}$$

적당한 조건 아래에서 이상의 방식으로 얻어진 (8.52)의 모수가 갖는 추정량은 일치추정량이다.

만일 설명변수의 동일한 집합이 $\hat{Y}_{1t}$, $\hat{Z}_{1t}$와 $\hat{Z}_{2t}$의 계산에 사용되지 않았다면, (8.49)의 조건은 유지되지 않는다. 그러므로, 만약 우리의 통상적인 절차(최소자승 절차와 동일하다)로써 (8.52)를 추정하면, 그 결과의 추정

량은 그 모형의 특성과 "일치"하지 않는 정규방정식에 기초하게 될 것이다. 한 예로서 $\hat{Z}_{2t}$가 $\hat{Y}_{1t}$와 $\hat{Z}_{1t}$에 깔려 있는 설명변수와 동일하지 않은 설명변수에 기초하고 있다고 가정하자. 그러면 일반적으로 $\hat{\phi}_{1t}$와 $\hat{Z}_{2t}$ 또는 $\hat{\phi}_{2t}$와 $\hat{Z}_{2t}$의 서로 곱한 것의 合이 0이 될 이유가 없다. 일반적으로 $\Sigma(\hat{\phi}_{1t} Z_{2t}) \neq 0$ 그리고 $\Sigma(\hat{\phi}_{2t}\hat{Z}_{2t}) \neq 0$이 될 것이다. 따라서 이 경우에

$$\sum (W_t\hat{Z}_{2t}) = b_1[\sum (\hat{\phi}_{1t}\hat{Z}_{2t})] + b_2[\sum (\hat{\phi}_{2t}\hat{Z}_{2t})] + \sum (\varepsilon_{it}\hat{Z}_{2t}) \qquad (8.55)$$

이다.

이 경우에 모형의 가정은 $\Sigma(\hat{W}_t\hat{Z}_{2t}) = 0$으로 놓아 정규방정식을 얻도록 "시사하지" 않는다.* 예를 들어 (8.55)는 $\Sigma(\hat{W}_t \hat{Z}_{2t}) = b_1[\Sigma(\hat{\phi}_{1t}\hat{Z}_{2t})] + b_2[\Sigma(\hat{\phi}_{2t}\hat{Z}_{2t})]$라 놓은 채로 방정식을 추정한다는 것을 시사한다. 만일 그렇다면, 우리는 상이한(그리고 보다 복잡한) 추정 절차를 가질 것이다.

이제 (4)의 논점으로 되돌아 가자. $g_1 = g(Y_{1t}, \cdots, Y_{mt})$를 우리가 추정하고자 하는 한 방정식에 나타나는 추가내생변수라 한다. 이제, 만일 이 단계에서 $g(\hat{Y}_{1t}, \cdots, \hat{Y}_{mt})$로 $g(Y_{1t}, \cdots, Y_{mt})$를 대체하여 이단계 이상의 절차를 수행한다면, 그 결과로 나온 모수 추정량이 불일치 추정량이 됨을 볼 것이다.

다시 (8.40)의 추정을 하자. 그러나 이제는 Y_{2t}^2이 이단계에서 $\hat{Y}_{2t}^2$으로 대체된다고 가정한다. 여기서 $\hat{Y}_{2t}^2$은 Y_{2t}를 일단계 설명변수에 회귀시켜 얻은 것이다. 이 경우에 Y_{2t}를 다음과 같이 나타낼 수 있다.

---

* 수준있는 독자들을 위하여, 만일 일단계 회귀에서의 모든 설명변수와 동일한 집합이 사용되지 않으면, 이들 회귀에서의 殘差는 이단계 설명변수 모두와 直交(Orthogonal)하지 않을 것이라는 점을 밝혀 둔다. 이것이 불일치 추정량을 가져오는 이유는 만일 모형내 "포함된" 모든 사전결정변수가 일단계에서 사용되지 않으면 TSLS가 선형체계에서 불일치 추정량을 가져오는 이유와 같다.

$$Y_{2t} = \hat{Y}_{2t} + \hat{\phi}_{4t} \tag{8.56}$$

그 중에서도 특히

$$\sum (\hat{Y}_{2t}\hat{\phi}_{4t}) = 0 \tag{8.57}$$

(8.50)의 양변을 자승하면,

$$\begin{aligned} Y_{2t}^2 &= \hat{Y}_{2t}^2 + (\hat{\phi}_{4t}^2 + 2\hat{\phi}_{4t}\hat{Y}_{2t}) \\ &= \hat{Y}_{2t}^2 + \hat{\psi}_t \end{aligned} \tag{8.58}$$

여기서 $\hat{\psi}_t$는 (8.58)의 괄호 안의 항과 같다. $\hat{\psi}_t$가 (8.44), (8.46), (8.48)에서 주어진 조건과 비슷한 조건을 만족시키지 않는다는 것은 분명하다. 예를 들어, (8.57)과 (8.58)에 비추어 보면,

$$\sum \hat{\psi}_t = \sum \hat{\phi}_{4t}^2 \neq 0$$

또한 이단계 회귀의 교란항이 (8.53)에서와 같은 조건을 만족시키지 않는다는 것도 확실하다. 결과적으로, 얻어지는 모수 추정량은 일치추정량이 되지 않는다.

4. 大標本 분산

다행스럽게도 선형모형에서의 二段階 最小自乘推定量에 대하여 제 7 장에서 주어진 대표본 분산 공식이 비선형체계에서의 그 추정량에 대해서도 또한 성립한다. 예를 들어, Y_{1t}, $Z_{1t} = (Y_{2t}\, Y_{3t})$와 $Z_{2t} = Y_{2t}^2$를 각각 $\hat{Y}_{1t}$, $\hat{Z}_{1t}$와 $\hat{Z}_{2t}$로 대체하였을 때 이단계 추정 절차를 (8.40)에 다시 적용하여 보자. 그러면, 전형적인 가정 아래에서 $\hat{b}_1$의 대표본 분산에 관한 일치추정량은 아래와 같이 보일 수 있다.

$$\widehat{\text{var}(\hat{b}_1)} = \frac{\hat{\sigma}_i^2}{\sum \hat{q}_t^2} \tag{8.59}$$

여기서 $\hat{\sigma}_i^2$ 은 ε_{it} 의 분산에 관한 일치추정량이고 $\hat{q}_t$ 는 상수항, $\hat{Z}_{1t}, \hat{Z}_{2t}$ 와 X_{1t} 에 대한 $\hat{Y}_{1t}$ 의 회귀에 있어서 t 번째 잔차를 가리킨다. ε_{it} 의 분산에 관한 명백한 일치추정량은*

$$\hat{\sigma}_i^2 = \sum_{t=1}^{n} \frac{(Y_{it} - \hat{b}_0 - \hat{b}_1 Y_{1t} - \hat{b}_2 Z_{1t} - \hat{b}_3 Z_{2t} - \hat{a}_1 X_{1t})^2}{n-5} \tag{8.60}$$

여기서 n 은 표본의 크기이다. 마찬가지로, 제 7 장에서 처럼 정규분포를 근사치로 삼아 사용함으로써, 가설검정 또는 신뢰구간의 설정을 할 수 있을 것이다. 예를 들어, 이상의 사례에서, b_1 에 관한 추론은 아래의 가정에 근거하는 것이다. 즉,

$$\frac{(\hat{b}_1 - b_1)}{\sqrt{\text{var}(\hat{b}_1)}} \tag{8.61}$$

이 표준정규분포 변수라는 것이다. 다시 엄밀하게 보아서 선형의 경우에서와 마찬가지로 이 결과는 오직 표본 크기가 무한일 때만이 옳다.

5. 보 기

이제 이 장에서 얻은 결과를 설명하고 확장하는데 도움을 주는 예를 보자. 이 보기의 목적은 우리의 결과를 설명하려는 것이기 때문에, 모형의 사실성 또는 관련된 경제 관계의 미묘한 부분까지 지나치게 관심을 갖지 않을 것이다.

가. 모형

3개 방정식으로된 아래의 거시경제학적 경제모형을 보자.

$$C_t = a_0 + a_1 Y_t + a_2 Y_t^2 + a_3 Y_t^3 + a_4\left(\frac{1}{C_{t-1}}\right) + a_5 W_{t-1} + u_{1t} \tag{8.62a}$$

---

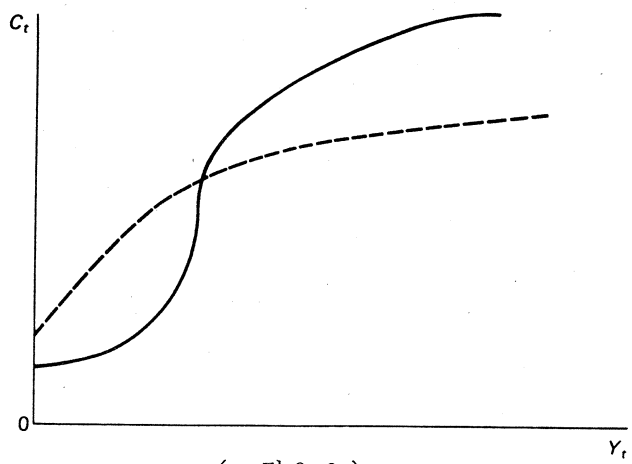
* 흥미로운 사항으로서, 만일 표본 크기가 무한하면, $n-5 = n = \infty$ 이다.

$$I_t = b_0 + b_1(Y_t Y_{t-1})^{1/2} + b_2 r_t + b_3 T_t + u_{2t} \qquad (8.62b)$$

$$Y_t = C_t + I_t + G_t, \qquad t = 1, \ldots, n \qquad (8.62c)$$

여기서 C_t는 t 시점의 총소비지출, Y_t는 t 시점의 총소득, W_{t-1}은 ($t-1$) 시점의 소비자의 富이고, I_t는 t 시점의 투자지출, r_t는 t 시점의 이자율, G_t는 t 시점의 정부지출, T_t는 시간추세 변수로서 $T_1 = 1$, $T_2 = 2$ 등등이며, u_{1t}와 u_{2t}는 t 시점의 교란항이 갖는 값이다. 우리의 통계학적 가정을 공식적으로 세우기전에, 이 모형이 갖는 본성을 간략하게 설명하기로 한다.

식 (8.62a)는 소비지출을 소득, 앞시기의 소비수준(일종의 관습 효과), 앞시기의 소비자가 처한 富의 위치와 관련시키는 소비함수이다. 물론 소득은 소비의 正의 효과를 미칠 것으로 기대한다. 그러나 이러한 正의 관계가 갖는 정확한 본성은 분명하지 않다. 특히 많은 기초 교재에서 그려지는 것과 같이 전형적인 線型이 아닐 것이다. 〈그림 8.3 〉은 포함된 "여타" 변수의 주어진 값에 대하여 소비지출과 소득 사이의 가능한 관계를 두 개 그린 것이다. (8.62a)에서 사용한 3제곱 형식은 〈그림 8.3 〉에서 묘사된 것과 같은 관계뿐만 아니라 선형($a_2 = a_3 = 0$)도 감안하기에 충분하게 신축적인 형식이다. 만일 신축성을 더욱 도모하고자 한다면, (8.62a)



〈그림 8.3 〉

에 소득 변수에 4차항을 더할 수 있다.

식 (8.62b)는 투자수준을 소득수준, 이자율, 그리고 시간 추세 변수로 설명하고 있다. 이 추세 변수는 시간이 흐름에 따라 꾸준히 증가하거나 ($b_3 > 0$이면), 감소 ($b_3 < 0$)하는 것으로 가정한 외생적인 투자의 힘이 갖는 純計를 나타내는 것으로 볼 수 있다. 항 $(Y_t Y_{t-1})^{1/2}$은 우리가 설명을 위하여 포함한 가속도원리 (accelerator principle)상의 변동이다.

식 (8.62c)는 소득을 지출의 합계로 설명하는 방정식이다. 이 식은 실제로 생산된 (공급된) 소득 (재화와 용역)이 시장의 대리인, 즉 소비자, 투자자, 정부에 의하여 수요된다는 것을 말하는 균형조건이다.

u_{1t}와 u_{2t}는 자기상관을 갖지 않으며 평균이 0이라고 가정한다. 즉, $E(u_{1t}) = E(u_{2t}) = 0$, 분산의 불변, $E(u_{1t}^2) = \sigma_1^2$, $E(u_{2t}^2) = \sigma_2^2$ 그리고 공분산의 불변 $E(u_{1t} u_{2t}) = \sigma_{12}$를 가정한다. 또한, 정부지출, 소비자의 富, 이자율은 외생적으로 발생하여서 그 결과로 교란항 u_{1t}와 u_{2t}는 G_t, r_t, W_{t-1}과 독립적이라고 가정한다. 추세 변수 T_t는 결정된 것이기 때문에 외생이며, 따라서 교란항 둘다 추세 변수와 독립적이다. 결국 (8.62)의 모형은 C_t, I_t와 Y_t를 C_{t-1}, W_{t-1}, r_t, T_t, G_t와 교란항으로 설명하는 3개 방정식 모형이다. 이보다 일반적인 모형에서는 그 중에서도 특히 이자율과 아마도 정부지출을 설명하려고 할 것이다.

나. 모형의 분석

(8.62)의 모형이 갖는 기본내생변수는 C_t, I_t와 Y_t이다. 추가내생변수는 Y_t^2, Y_t^3 그리고 $(Y_t Y_{t-1})^{1/2}$이다. 상수항에 보태서 5개의 사전결정변수 즉, $(1/C_{t-1})$, W_{t-1}, r_t, T_t, G_t가 있다. 이 5개 변수중에서 유일하게 $(1/C_{t-1})$만이 외생이 아니다. 이 모형은 C_{t-1}를 결정하고 나서 $t-1$ 시기의 그 역수를 얻게 된다.

식 (8.62a)를 보자. 이 방정식의 식별을 위한 필요조건, 즉 (8.32)는 만족된다. 왜냐하면 우변에는 하나의 기본내생변수 Y_t 만이 있으나, 추가내생변수 즉 $(Y_tY_{t-1})^{1/2}$ 과 사전결정변수(이 경우에는 외생변수) r_t, T_t, G_t 등 4개 변수를 제외하고 있기 때문이다. (8.62b)의 식별을 위한 필요조건 또한 충족된다. 이 식은 우변에 어떠한 기본내생변수도 포함하고 있지 않으며, 추가내생변수, 즉 Y_t^2 과 Y_t^3 또한 사전결정변수 $(1/C_{t-1})$, W_{t-1}, G_t 등 도합 5개 변수를 배제하고 있기 때문이다. (8.62)에 대한 식별문제는 생기지 않는다. 이 식은 추정하여야 할 모수를 포함하고 있지 않기 때문이다.

이제 (8.62a)를 추정하는 문제를 보자. TSLS 절차를 수행하기 위해서는 먼저 Y_t, $Q_{1t}=Y_t^2$ 과 $Q_{2t}=Y_t^3$ 의 계산치를 얻어야만 한다. 이들 계산치를 구하기 위하여 일단계 설명변수를 지정하여야 한다. 만약 표본 크기가 $n=50$ 이라면, 일단계의 설명변수는 아래와 같이 선정될 것이다.

$$\begin{array}{c}\text{상수항, } (1/C_{t-1}),\ W_{t-1},\ r_t,\ T_t,\ G_t,\ Y_{t-1},\\ (1/C_{t-1})^2,\ W_{t-1}^2,\ R_t^2,\ T_t^2,\ G_t^2,\ Y_{t-1}^2\end{array} \tag{8.63}$$

(8.63)에서는 단순히 모형의 모든 사전결정변수의 그 自乘値를 골랐다. 여기서 3제곱 또는 그 이상 차수의 항이나 교적항은 포함하지 않았다. 그 이유는 이러한 변수들이 (8.63)에 든 변수들과 극히 높은 상관 관계를 가질 것으로 예상되기 때문이다. 만약 이들 추가의 항들을 (8.63)에 보태면 그 결과로 추정량의 개선이 이루어질 것으로 예상할 수 있을 것이다. 그 이유는 표본 크기와 일단계에서 사용된 변수의 숫자 사이의 차이로 인한 손실을 줄임으로써 다항 회귀가 더 비중있는 평균 함수로 근접하기 때문이다 (8.3을 보라). 그러나 이렇다고 믿을 공식적인 근거는 없다! 그렇게 될 경우는 아마도 3제곱 및 더 높은 차수의 항과 서로

곱한 항의 추가가 회귀 모수 추정량의 개선을 가져오는 경우일 것이다.* 어쨌든, (8.63)에서의 선정은 (8.62a)의 모든 사전결정변수, 즉 상수항, $(1/C_{t-1})$, W_{t-1} 그리고 적어도 (8.62a)에는 포함되지 않는 많은 사전결정변수를 포함하고 있다. 즉, (8.62)의 우변에는 기본내생 및 추가내생변수가 3개인데 비하여 (8.63)에는 10개가 더 많다. 이에 덧붙여 (8.63)에서의 선정은 우리가 제시하였던 조건, 즉 표본 크기(시차 변수 때문에 첫번째 관찰치를 잃게 된 결과로 49가 될 것이다)와 일단계에서 사용되는 변수의 숫자(즉, 13) 사이의 차이가 적어도 20은 되어야 한다는 조건을 만족시킨다.

남은 절차는 명백하다. Y_t, Q_{1t}, Q_{2t}를 최소자승 회귀로 $t=2, \cdots, n=50$에 걸쳐서 (8.63)의 변수에 회귀시킴으로써 $\hat{Y}_t$, $\hat{Q}_{1t}$, $\hat{Q}_{2t}$를 계산할 수 있다. $\hat{Y}_t$, $\hat{Q}_{1t}$, $\hat{Q}_{2t}$를 계산하는 데 동일한 일단계 설명변수의 집합을 사용하는 점을 명심하라. 그러면, (8.62a)의 추정에 대응하는 이단계 회귀는,

$$C_t = a_0 + a_1\hat{Y}_t + a_2\hat{Q}_{1t} + a_3\hat{Q}_{2t} + a_4(1/C_{t-1}) + a_5W_{t-1} + k_t \qquad t=2, \ldots, n=50 \tag{8.64}$$

이다.

여기서 k_t는 결과적인 오차항이다. 이어서 모수 a_0, a_1, a_2, a_3, a_4, a_5의 추정량은 (8.64)에 대응하는 최소자승 회귀로 얻게 된다. 특히 이 회귀에 대한 정규방정식은 다음의 조건으로 주어진다.

---

* TSLS 추정량의 속성인 일치성은 대표본의 속성이다. 有限한 표본에서는 TSLS 추정량은 편의를 갖는다. 그러므로 회귀 모수의 TSLS 추정량이 얼마나 좋은가는 바로 그 추정량의 편의와 분산 둘 다에 달려 있다. 보통, 이 두 가지의 추정량이 갖는 속성은 평균평방오차(mean square error)라고 불리는 것을 구하기 위하여 이 둘을 더함으로써 결합되는 것이다. 그러므로, 유한한 표본에서는 만약 한 추정량이 다른 추정량보다 작은 평균평방오차를 가진다면, 그 주어진 모수의 TSLS 추정량이 다른 것(이것은 일단계 설명변수의 집합을 다르게 가질 것이다)보다 "더 좋다"고 말할 수 있다.

$$\sum_{t=2}^{50} \hat{k}_t = 0, \qquad \sum_{t=2}^{50} (\hat{k}_t \hat{Y}_t) = 0, \qquad \sum_{t=2}^{50} (\hat{k}_t \hat{Q}_{1t}) = 0,$$
$$\sum_{t=2}^{50} (\hat{k}_t \hat{Q}_{2t}) = 0, \qquad \sum_{t=2}^{50} [\hat{k}_t (1/C_{t-1})] = 0, \qquad \sum_{t=2}^{50} (\hat{k}_t W_{t-1}) = 0 \tag{8.65}$$

여기서 $\hat{k}_t = C_t - \hat{a}_0 - \hat{a}_1 \hat{Y}_t - \hat{a}_2 \hat{Q}_{1t} - \hat{a}_3 \hat{Q}_{2t} - \hat{a}_4 (1/C_{t-1}) - \hat{a}_5 W_{t-1}$ 이고 $\hat{a}_i$, $i = 0, \cdots, 5$는 a_i의 추정량을 표시한다.

그러면 u_{1t}의 분산 σ_1^2 은 다음과 같이 추정될 것이다.

$$\hat{\sigma}_1^{\,2} = \sum_{t=2}^{50} \frac{\hat{u}_{1t}^{\,2}}{(49-6)} \tag{8.66}$$

여기서 $\hat{u}_{1t} = C_t - \hat{a}_0 - \hat{a}_1 Y_t - \hat{a}_2 Y_t^2 - \hat{a}_3 Y_t^3 - \hat{a}_4 (1/C_{t-1}) - \hat{a}_5 W_{t-1}$ 이다. 마지막으로, 대표본 분산을 a_2에 한정하여 보면, 다음과 같이 추정될 것이다.

$$\widehat{\text{var}(\hat{a}_2)} = \hat{\sigma}_1^{\,2} \left(\frac{1}{\sum_{t=2}^{50} \hat{Q}_{1t}^{\,2}} \right) \tag{8.67}$$

여기서 $\hat{Q}_{1t}$는 $\hat{Q}_1$를 상수항, $\hat{Y}_t$, $\hat{Q}_{2t}$, $(1/C_{t-1})$과 W_{t-1}에 대하여 최소자승 회귀를 하였을 때의 殘差이다.

(8.62b)의 추정에 관한 기계적인 절차는 (8.62a)에 대한 설명과 동일하다. 주의하여야 할 유일한 사항은 (8.62b)에 깔려 있는 일단계 설명변수가 (8.62a)의 추정에 사용된 설명변수와 반드시 동일할 필요가 없다는 것이다. 그 설명변수는 (8.63)의 집합일 필요가 없다. 다르게 설명하면, 어느 주어진 방정식의 추정에 일단계 설명변수로 무엇이든지 선정하더라도 그 방정식의 우변에 있는 기본 및 추가내생변수의 계산치를 결정하는 데에는 그와 동일한 집합을 사용하여야 한다. 모든 방정식에 대해서 동일한 집합을 꼭 사용하여야 하는 것은 아니다. 다른 한편으로는 방정식이 바뀔 때 일단계 설명변수를 변동시킬 유인이 전혀 없다. 왜냐하면, 표본크기는 전형적으로 모든 방정식에 대해서 같고, 비슷한 변수가 포함되기 때문이다.

부록. 內生變數와 母數 모두에서 비선형인 모형의 추정

이 부록에서는 앞에서의 결과를 내생변수와 모수 모두에서 비선형인 연립방정식 모형의 추정으로 확대할 것이다. 이에 대응하는 식별에 관한 논의는 고려하지 않을 것이다. 그 이유는 실제로 고려되는 상대적으로 간단한 규칙이 아직 존재하지 않기 때문이다.

제시할 추정 절차는 그 원리가 직관적으로 이해될 수 있는 것이다. 그러나, 수치해석과 컴퓨터프로그래밍에 익숙치 않다면, 이 절차를 선택사항으로 하고 있는 사용자 컴퓨터 프로그램을 이용하는 것이 그 절차를 실증적으로 수행하는 데 요구된다.*

가. 분석의 틀

내생변수 Y_{1t}, Y_{2t}, Y_{3t} 그리고 외생변수 X_{1t}, …, X_{kt}를 포함하고 있는 3개 방정식 모형을 보자. 이 모형의 첫번째 방정식을 다음과 같다고 한다.

$$Y_{1t} = a_0 + a_1 X_{1t} e^{a_2 Y_{2t}} + a_3 X_{2t} + a_4 Y_{3t}^2 + \varepsilon_{1t} \qquad (8A.1)$$

여기서 ε_{1t}는 교란항이다. a_2의 값을 몰라서 a_0, a_1, a_3, a_4의 값과 함께 추정하여야만 한다고 가정한다. 그러면 이제까지 고려하였던 것과 다르게, (8A.1)은 내생변수와 모수 모두에서 비선형이다. (8A.1)을 다음과 같이 모수에서 선형인 모형으로 변환시킬 수 없음에 유의하라.

$$Y_{1t} = a_0 + a_1 Z_t + a_3 X_{2t} + a_4 Y_{3t}^2 + \varepsilon_{1t} \qquad (8A.2)$$

---

* 이 절차는 다음에서 최초로 제안된 비선형 이단계 최소자승법이다. T.Amemiya, "The Nonlinear Two-Stage Least-Squares Estimator", Journal of Econometrics 2 (1974), pp.105-110. 우리의 논의는 Amemiya의 결과를 해석한 것이다.

여기서 Z_t는 $Z_t = X_{1t}\ e^{a_2 Y_{2t}}$인 추가내생변수이다. 그렇게 변환할 수 없는 이유는 a_2의 값을 알지 못하여서 Z_t에 관한 관찰치를 X_{1t}와 Y_{2t}에 대한 이용가능한 관찰치로부터 구성할 수 없기 때문이다.

내생변수와 모수 둘다 비선형인 모형의 또 다른 예는 아래의 2개 방정식 모형을 들 수 있다.

$$\log(Y_{1t}) = a_0 + a_1\left(\frac{Y_{2t}}{1 + b_2 X_{2t}}\right) + a_2\left(\frac{X_{1t}}{Y_{2t}}\right) + \varepsilon_{1t} \qquad \text{(8A.3a)}$$

$$(Y_{2t}{}^{x} X_{3t}) = b_0 + b_1 Y_{1t} + b_2 Y_{1t}{}^{2} + b_3 X_{2t} + \varepsilon_{2t} \qquad \text{(8A.3b)}$$

여기서 Y_{1t}와 Y_{2t}는 내생변수이고, X_{1t}, X_{2t}와 X_{3t}는 외생변수이며, ε_{1t}와 ε_{2t}는 교란항이다. 이 경우에 "모수의 비선형성"은 모수 α와 b_2 때문에 발생한다.

이제 (8A.1)과 (8A.3)의 모형을 사용하여 이 부록에서 고려하는 유형의 모형이 갖는 두가지 속성을 설명하기로 한다. 첫째로, 반드시(8A.1)과 (8A.3a)와 같을 필요없이, (8A.3b)에서 처럼 주어진 방정식의 좌변이 미지의 모수를 포함할 수가 있다. 둘째로, 각 방정식의 교란항은 더할 수가 있어야 한다. 아래에서 이러한 가정이 그리 제한적인 가정이 아님을 지적하기로 한다. 당분간 이러한 가정을 채택한다면, 그 의미는 교란항을 제외한 모든 항이 방정식의 좌변에 올 수 있으며, 그래서 교란항만이 우변에 있을 수도 있다는 것이 된다. 예를 들어, (8A.3a)는 아래와 같이 나타낼 수 있다.

$$\log(Y_{1t}) - a_0 - a_1\left(\frac{Y_{2t}}{1 + b_2 X_{2t}}\right) - a_2\left(\frac{X_{1t}}{Y_{2t}}\right) = \varepsilon_{1t} \qquad \text{(8A.4)}$$

이보다 더 일반화하기 위하여, Y_{1t}, …, Y_{mt}를 모형의 내생변수, X_{1t}, …, X_{pt}를 외생변수 그리고 u_{1t}, …, u_{mt}를 교란항이라 하자. 그러면, 우리가 추정하고자 하는 방정식을 i번째 방정식이라 할 때, 그 형식은 다음

과 같이 나타낼 수 있다.*

$$F_i(Y_{1t}, \ldots, Y_{mt}, X_{1t}, \ldots, X_{pt}) = u_{it} \tag{8A.5}$$

여기서 (8A.5)의 좌변은 변수 Y_{1t} , ⋯, Y_{mt} , X_{1t} , ⋯, X_{pt} 의 하나 또는 그 이상의 함수이며, 알려지지 않은 모수를 포함하고 있다. 한 예로서, 방정식(8A.3a)에 대해서 이 함수는 단지 (8A.4)의 좌변에 지나지 않는다.

방정식을 더할 수 있는 교란항으로 나타낼 수 있다는 가정은, 따라서 (8A.5)와 같은 형식으로 나타낼 수 있다는 가정으로서, 일반적으로 경제학자들이 고려하는 유형의 모형으로 미루어 볼때 그리 제한적인 것이 아니다. 예를 들어 이 가정은 단지 고려 대상의 모형이 갖는 특정한 방정식이 교란항에 대해서 풀 수 있어야 한다는 점만 요구한다. 한 예로서, 모형의 첫번째 방정식이 다음과 같은 형식을 갖는다고 하자.

$$Y_{1t} = a_0 X_{1t}^{a_1} Y_{2t}^{a_2} e^{u_{1t}} \tag{8A.6}$$

(8A.5)에 대응하는 형식은,

$$\log(Y_{1t}) - \log(a_0) - a_1 \log(X_{1t}) - a_2 \log(Y_{2t}) = u_{1t} \tag{8A.7}$$

이다.

또 다른 예로서, 다음과 같은 형식의 방정식을 보자.

$$Y_{1t} = a_0 + a_1 \left(\frac{e^{a_2 Y_{2t}}}{1 + u_{1t}} \right) + a_3 X_{1t} \tag{8A.8}$$

---

* 만일 모형의 모든 방정식을 추정할 예정이면, 우리의 분석에서는 모든 방정식이 (8A.5)의 형식으로 나타낼 수 있어야 한다.

여기서 교란항 u_{1t}가 취할 수 있는 값의 범위는 $1+u_{1t}>0$이다. 그러면, (8A.5)에 대응하는 형식은 아래에 의거하여 얻을 수 있다.

$$\left(\frac{Y_{1t}-a_0-a_3X_{1t}}{a_1e^{a_2Y_{2t}}}\right)=\frac{1}{1+u_{1t}} \tag{8A.9}$$

그래서

$$\left(\frac{a_1e^{a_2Y_{2t}}}{Y_{1t}-a_0-a_3X_{1t}}\right)-1=u_{1t} \tag{8A.10}$$

추정과 그 가설검정의 주제로 들어가기전에 예비적인 결과를 제시하기로 한다.

나. 예비적인 결과

표준적인 단일 방정식 회귀모형을 보자.

$$Y_t=b_0+b_1X_{1t}+\cdots+b_kX_{kt}+u_t, \qquad t=1,\ldots,N \tag{8A.11}$$

여기서 독립변수들은 다중공선성이 없으며, 교란항은 모든 표준적인 가정을 만족시킨다. 특히, u_t는 모든 시차, 현행, 미래의 독립변수와 독립적이며 0의 평균 $E(u_t)=0$과 일정한 분산 $E(u_t^2)=\sigma_u^2$를 가지며, 자기상관을 갖고 있지 않다.

(8A.11)과 같은 모형에서 회귀계수 b_0, b_1, …, b_k에 관한 기본가정이 이들 계수가 상수라는 것이었음을 상기하라. 즉, 이들 계수의 값은 t에 의존하지 않는다. 그러므로, 이들 계수의 일부 또는 전부가 0이 될 수 있다. 만일 이들 계수가 모두 0이면, 종속변수 $Y_t=u_t$이어서, Y_t가 모형의 설명변수와 독립적인 확률변수로 바뀐다는 사실은 분명하다.

$\hat{b}_i$를 代變數 技法으로 얻은 b_i의 추정량이라 하자. 이 기법은 최소자승법과 동등하다. 제 4 장에서는 $\hat{b}_i$이 불편추정량임을 보였기 때문에 E

$(\hat{b_i}) = b_i$ 이다. 이 결과는 $b_i = 0$의 여부와 관계없이 성립된다. 또한 4장의 부록에서 $\hat{b}_i$의 분산을 다음과 같이 나타낼 수 있다고 하였다.

$$\text{var}(b_i) = \frac{\sigma_u^2}{\sum_{t=1}^{N} \hat{v}_{it}^2} \tag{8A.12}$$

여기서 $\hat{v}_{it}$는 X_{it}를 (8A.11)의 모든 여타 설명변수 (상수항 포함)에 대하여 회귀시켰을 때의 잔차이다.

(8A.12)의 분모는 N개의 합이며, 그 각각의 항은 0보다 크거나 같다. 그러므로, 무엇보다도 먼저 var$(\hat{b_i})$의 값은 표본 크기 N에 달려 있다. 기술적인 가정을 더하면, N이 무한할 경우 (8A.12)의 분모 또한 무한하고 따라서 var$(\hat{b_i})$가 0이 된다는 것을 보일 수 있다.

이제 (8A.11)의 모든 회귀 계수가 0인 경우를 고려하자. 그러면 이 경우에 우리의 결과는 $E(\hat{b_i}) = 0$, $i = 0, \cdots, k$이고 또한 만약 $N = \infty$이면, var$(\hat{b_i}) = 0$, $i = 0, \cdots, k$임을 시사한다. 어느정도 직관적으로 보면, 회귀 모수가 0이고 표본 크기가 무한할 경우, 각 모수 추정량의 기대값은 0이고, 그 0으로부터의 편차의 자승(그 분산)도 0이 될 것이다. 그러므로, 이 경우에 각 회귀 모수 추정량의 값은 직관적으로 0이 될 것으로 기대된다. 이를 더 공식적인 설명으로는 다음과 같다. 즉, 만약 $E(\hat{b_i}) = 0$, $i = 0, \cdots k$이고 $\lim_{N \to \infty}$ Var$(\hat{bi}) = 0$이면, $\lim_{N \to \infty}$ Prob $(|\hat{b_i} - 0| > \delta) = \lim_{N \to \infty}$ Prob $(|\hat{b_i}|) \delta) = 0$이며, 여기서 δ는 양의 상수이나 상당히 작은 값이다. 말로 하면, 이상의 조건 아래에서 $\hat{b_i}$가 조금이라도 0과 다를 확률은 0이다.(즉, $\hat{b_i}$는 확률상 0에 수렴한다).

이상의 결론은, 만일 0의 평균과 일정한 분산을 갖고 또한 (교란항과 같이) 자기상관을 갖지 않은 어떤 변수를 그것과 독립적인 변수들의 집합에 회귀시킨다면, 회귀 모수 추정량은 더 나아간 기술적(그리고 합당한) 가정 아래에서 확률상 0에 수렴한다. 그러므로 예를 들어 앞의 사례의

변수의 계산치,

$$\hat{Y}_t = b_0 + b_1 X_{1t} + \cdots + b_k X_{kt} \qquad \text{(8A.13)}$$

도 $N \to \infty$에 따라 확률상 0에 수렴한다.

다. 추정 절차

이 節에서는 먼저 식(8A.3b)로 내생변수와 모수 모두에서 비선형인 방정식을 추정하는 절차를 설명할 것이다. 그러고 나서 이 결과를 일반화하기로 한다.

(8A.3)의 2개 방정식 모형의 변수, 즉 Y_{1t} , Y_{2t} , X_{1t} , X_{2t} 와 X_{3t} 에 관하여 N개 관찰치를 갖고 있다고 하자. 교란항에 관하여, 각각의 교란항이 0의 평균, 일정한 분산과 일정한 공분산을 갖고 있다고 하고, 각 교란항은 다른 시기의 여타 교란항의 값과 마찬가지로 다른 시기의 자신의 값과도 독립적이라고 가정한다. 또한, 두 교란항은 모두 3개의 외생변수의 시차, 현행, 미래의 값과 독립적이라 한다. 또한 우리가 설명할 추정절차가 갖는 일치성의 속성은 약간 기술적으로 더 나아간 가정에 달려 있으며, 그 가정의 일부는 직관적으로 보기 어려워서 실제로는 성립하는 것으로 가정하는 것이 보통이다. 그 가정의 제시와 이해는 이 책의 수준을 넘는 수학 및 통계학적 방법을 요구하기 때문에 실제로 이 조건을 자세히 하지 않고 가정하기만 한다.

식(8A.3b)는 (8A.5)의 형식을 나타낼 수 있다. 즉,

$$(Y_{2t} \cdot X_{3t}) - b_0 - b_1 Y_{1t} - b_2 Y_{1t}^2 - b_3 X_{2t} = \varepsilon_{2t} \qquad \text{(8A.14)}$$

(8A.14)의 좌변을 F_t로 표시한다. (8A.14)의 좌변에 있는 각각의 모수의 가상적인 값을 선정하여, 이들 값에 따라서 F_t에 대하여 "근사

한" 관찰치를 구성하였다고 하자. 예를 들어 α를 0.1, b_0를 3.7, b_1를 -1.5, b_2를 2.0, b_3를 -10이라고 했다고 하자. 그러면 "근사한" 관찰치는 다음과 같이 정하여 진다.

$$F_t^a = (Y_{2t}^{0.1} X_{3t}) - 3.7 + 1.5Y_{1t} - 2Y_{1t}^2 + 10X_{2t} \tag{8A.15}$$

(8A.14)에서 만일 모수의 진정한 값이 선정한 값과 같다면, 모든 t에 대하여 $F_t^a = F_t = \varepsilon_{2t}$이다. 다른 한편으로 선정한 값중에 하나 또는 그 이상이 대응하는 진짜 값과 같지 않다면, 모든 t에 대하여 $F_t^a \neq \varepsilon_{2t}$이다.

당분간, 우리가 고른 모수값이 진짜 값과 같아서 $t = 1, \cdots, N$에서 $F_t^a = F_t = \varepsilon_{2t}$라고 가정하자. 더 나아가 이제부터 우리가 구성한 F_t 변수라고 부르는 것을 모형의 외생변수와 그 자승, 즉 X_{1t}, X_{2t}, X_{3t}, X_{1t}^2, X_{2t}^2, X_{3t}^2과 상수항에 일단계 회귀를 한다고 가정하자. 아래에서는 일단계 설명변수의 선정에 관한 논점을 논의할 것이다. 우리의 분석으로 되돌아 가서 $\hat{F}_t$를 F_t의 일단계 회귀에서 얻은 계산치라고 하자. 그러면, 앞의 節에서 본 예비적 결과는 다음을 의미한다. 즉, $F_t = \varepsilon_{2t}$는 모든 표준적인 가정을 만족하는 교란항이고, 또한 $F_t = \varepsilon_{2t}$가 세 외생변수와 모두 독립적이기 때문에, $\hat{F}_t$는 $N \to \infty$에 따라서 확률상 0에 수렴한다. 다음의 총합을 보자.

$$S = \sum_{t=1}^{N} \frac{\hat{F}_t^2}{N} \tag{8A.16}$$

S도 $N \to \infty$에 따라 확률적으로 0에 수렴할 것이라는 점은 명백하다.

이제 우리가 선정한 모수값이 진정한 값이 아니라고 가정하자. 그러면, 이미 지적하였듯이 우리가 구성한 변수 F_t^a는 $F_t^a \neq \varepsilon_{2t}$, $t = 1, \cdots, N$이

게 된다. 이 경우에 F_t^a의 값은 단지 방정식의 변수, 즉 Y_{1t}, Y_{2t}, X_{2t}와 X_{3t}의 함수에 불과하다. 이제 위의 사례에서 처럼, F_t^a를 일단계 설명변수, 즉 X_{1t}, X_{2t}, X_{3t}, X_{1t}^2, X_{2t}^2, X_{3t}^2과 상수항에 대하여 회귀시킨다고 하자. 그 결과인 F_t^a의 계산치를 $\hat{F}_t^a$라 하자. 그러면, 계산치 자승의 평균 합계인 아래의 식이 확률상 0에 수렴하지 않는다.

$$S^a = \sum_{t=1}^{N} \frac{(\hat{F}_t^a)^2}{N} \tag{8A.17}$$

이것이 바로 보다 기술적인 가정을 취하였을 때의 경우이다. S^a는 자승의 합이기 때문에 이상의 기술적 가정 아래에서는 S^a가 확률상 陽의 수로 수렴할 것이라는 점을 보일 수 있다.

이상의 논의에서 주된 사항은, $(\hat{F}_t^a)^2$항의 평균이 만일 모수로 선정된 값이 옳은 값일 경우에는 확률상 0에 수렴하지만, 그렇지 않을 경우는 陽數로 수렴한다는 것이다. 따라서 이것이 바로 뒤따르는 추정 절차 그 자체이다. 계산된 변수의 자승$(\hat{F}_t^a)^2$의 평균을 최소화하는 값의 집합을 구하기 위하여 회귀모형의 모수값의 가능한 집합을 탐색하는 것이다. 우리는 단지 그러한 값의 집합을 추정치(더 공식적으로는 추정량)로 삼으면 된다. 이러한 추정량은, 만약 표본 크기가 무한하다면 계산한 변수의 자승$(\hat{F}_t^a)^2$의 평균이 진정한 모수값에 의해서 (0으로) 최소화되기 때문에 일치추정량이 된다.

라. 일단계 설명변수의 선정

일단계 설명변수의 선정에 관한 논의는 제8장 3절에서의 논의와 매우 유사하다. 예를 들어, 만일 표본 크기가 무한하면, 회귀 모수 추정량의 분산은 외생변수에서의 일단계 회귀의 次數와 반비례한다는 것이 일반적이다. 그러므로, 이들 변수의 (더 낮은 冪과 함께) 冪을 더욱 더 높이는 것

이 일단계 회귀에서 고려되어야 한다. 그러나 실제로는 표본 크기가 무한하여서, 일단계 회귀에서 사용되는 항의 수도 제한되어야 한다. 실제로, 만일 일단계 회귀에서 사용되는 항의 수 p가 표본 크기 N과 같다면, 그 추정량은 일치추정량이 아님을 보일 수 있다.* 제8장 3절에서 처럼 우리는 곤경에 빠지게 되어서 다시 p를 $(N-p) \geq 20$이 되도록 선택하게 한다. 물론 일치성의 속성은 $(N-p)$가 무한할 것을 요구한다. 또한, 증명은 않겠지만, 추정 절차에서 일치성의 속성은 p가 적어도 추정하는 방정식의 모수의 숫자만큼은 되어야 한다는 사실에 유의하자.

마. 절차의 재검토와 일반적인 윤곽

일반적으로 내생변수와 모수 모두에서 비선형인 방정식을 추정하는 절차는 다음과 같은 순서로 정리된다.

1) 방정식을 (8A.5)형식으로 나타내고 좌변을 F_t로 표시한다.

2) F_t^a를 방정식에 있는 모수의 값의 집합을 선정하여 결정하고 F_t의 근사치로 삼는다.

3) 일단계 설명변수를 결정할 외생변수의 다항 형식을 결정한다. 일단계 설명변수의 숫자 p는 적어도 추정하는 방정식에 있는 모수의 숫자만큼은 되어야 한다는 점을 명심하라. 표본크기 N은 실제로 유한하기 때문에, $(N-p)$가 적어도 20은 되도록 p를 선택한다.

4) F_t^a를 일단계 설명변수에 회귀시켜 얻은 F_t^a의 계산치를 $\hat{F}_t^a$라 한다.

---

* 수준높은 독자들을 위해서, 만일 $p=N$이면 $\hat{F}_t^a = F_t^a$라는 것을 밝혀둔다. 그러므로 우리는 $\sum_{t=1}^{N} (F_t^a)^2 / N$을 최소화하는 모수값을 선정할 수 있으며, 이것은 추정한 교란항의 자승의 합계를 최소화하는 것과 같다(즉, 비선형 최소자승 절차). 그러나, 만일 회귀모형이 우선 방정식의 우변에 내생변수를 갖는다고 한다면, 최소자승 절차는 일치추정량을 낳지 않는다.

5) $\sum_{t=1}^{N} (\hat{F}_t^a)^2 / N$을 최소화시키는 값의 집합을 발견하기 위하여 가능한 모수의 값들을 탐색한다. 그러한 값을 대응하는 모수의 추정치로 취한다.

이러한 절차를 실증적으로 사용하기 위해서는 수치 해석과 컴퓨터 프로그래밍의 지식이나 이 절차를 선택 사항으로 포함하고 있는 사용자용 컴퓨터 프로그램의 이용이 가능하여야 한다.

바. 가설검정, 신뢰구간 그리고 대표본 분산 : 논평

불행히도 수준이 높은 독자들을 제외하고는 추정량의 대표본 분산에 대해서 설명할 수가 없다. 왜냐하면 그것에 대한 소개는 우리가 이 책에서 정한 수준을 넘는 수학적 방법을 요구하기 때문이다.* 그러나, 이 절차를 선택사항으로서 포함하고 있는 사용자용 컴퓨터 프로그램이 이용 가능하다면, (무엇보다도 먼저) 모수의 추정치와 대응하는 대표본 분산의 추정치가 자동으로 출력될 것이다. 보다 기술적인 가정 아래에서는 만약 표본 크기가 무한하면, 모수 추정량이 정규분포한다는 사실을 보일 수 있다. 그러므로, 컴퓨터 프로그램의 작동 결과에 의거하여 다시 정규분포를 근사치로 이용하여 단일 모수의 값과 관련된 가설을 검정하거나 신뢰구간을 설정할 수 있다. 예를 들어, a_2가 모수중의 하나이고 컴퓨터 결과가 $\hat{a}=10$

---

* 수준높은 독자들을 위해서, 추정하는 방정식이 (8A.5)의 형식으로 되어 있다고 가정하고 그 좌변을 F_t라고 표기한다. 그 방정식은 모수 $a_0, \cdots a_k$를 포함한다고 하자. 또한 $f_{it} = (\partial F_t / \partial a_i)$, $i = 0, \cdots, k$라 한다. 일반적으로 f_{it}는 모수 $a_0, \cdots a_k$중의 하나 또는 그 이상을 포함할 것이다. 관련된 모수를 일치추정량으로 대체함으로써 각각의 f_{it}에 관한 N개 관찰치를 얻는다. 이제, f_{it}를 모든 일단계 설명변수에 회귀시켜 얻은 f_{it}의 계산치를 $\hat{f}_{it}$라 한다. 마지막으로 $\hat{f}_{it}$·를 $\hat{f}_{it}$의 k개 변수에 회귀시켜 얻은 t번째 잔차를 $\hat{f}_{it}$·라 한다. 여기서 $j \neq i$, 그러면, $\hat{a}_i$의 대표본 분산은 $\hat{\sigma}^2 / \sum_{t=1}^{N} f_{it}^2$·과 같이 일치 추정된다. 여기서 $\hat{\sigma}^2$은 방정식이 갖는 교란항의 분산에 관한 일치추정량이다.

그리고 $\hat{\sigma}^2_{\hat{a}_2} = 16$ 이라 가정하자. 그러면, a_2에 대한 근사한 95% 신뢰구간은 대표본 정규분포에 기초하여 10 ± 4 (1.96) 또는 10 ± 7.84 가 될 것이다.

〈附　　錄〉

1. 통　　계　　표

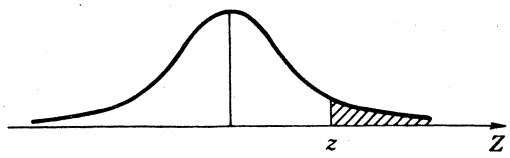
2. 해　　답

〈附錄 1 〉

통 계 표

〈표 1 〉 표준화정규분포

$$Z = \frac{X - \mu}{\sigma}$$



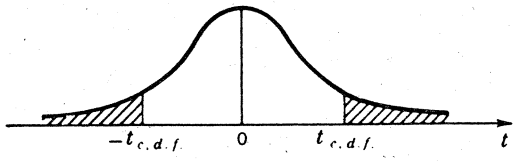
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

표는 누적확률 $Z \geq z$ 을 나타낸다.

출전 : Reprinted from Edward J. Kane, *Economic Statistics and Econometrics: An Introduction to Quantitative Economics*, New York: Harper & Row, Publishers, 1968.

〈표 2〉

스튜던트 t 분포



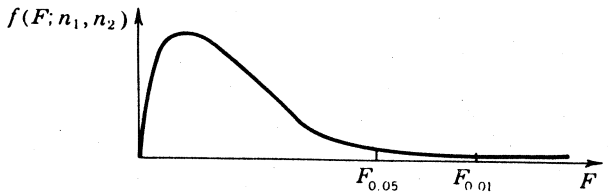
자유도	표의 값보다 절대치가 클 확률					
	0.01	0.02	0.05	0.1	0.2	0.3
1	63.657	31.821	12.706	6.314	3.078	1.963
2	9.925	6.965	4.303	2.920	1.886	1.386
3	5.841	4.541	3.182	2.353	1.638	1.250
4	4.604	3.747	2.776	2.132	1.533	1.190
5	4.032	3.365	2.571	2.015	1.476	1.156
6	3.707	3.143	2.447	1.943	1.440	1.134
7	3.499	2.998	2.365	1.895	1.415	1.119
8	3.355	2.896	2.306	1.860	1.397	1.108
9	3.250	2.821	2.262	1.833	1.383	1.100
10	3.169	2.764	2.228	1.812	1.372	1.093
11	3.106	2.718	2.201	1.796	1.363	1.088
12	3.055	2.681	2.179	1.782	1.356	1.083
13	3.012	2.650	2.160	1.771	1.350	1.079
14	2.977	2.624	2.145	1.761	1.345	1.076
15	2.947	2.602	2.131	1.753	1.341	1.074
16	2.921	2.583	2.120	1.746	1.337	1.071
17	2.898	2.567	2.110	1.740	1.333	1.069
18	2.878	2.552	2.101	1.734	1.330	1.067
19	2.861	2.539	2.093	1.729	1.328	1.066
20	2.845	2.528	2.086	1.725	1.325	1.064
21	2.831	2.518	2.080	1.721	1.323	1.063
22	2.819	2.508	2.074	1.717	1.321	1.061
23	2.807	2.500	2.069	1.714	1.319	1.060
24	2.797	2.492	2.064	1.711	1.318	1.059
25	2.787	2.485	2.060	1.708	1.316	1.058
26	2.779	2.479	2.056	1.706	1.315	1.058
27	2.771	2.473	2.052	1.703	1.314	1.057
28	2.763	2.467	2.048	1.701	1.313	1.056
29	2.756	2.462	2.045	1.699	1.311	1.055
30	2.750	2.457	2.042	1.697	1.310	1.055
∞	2.576	2.326	1.960	1.645	1.282	1.036

출전 : Sir Ronald A. Fisher, *Statistical Methods for Research Workers*, 13th edition. Oliver & Boyd Ltd., Edinburgh, 1963, 의 표Ⅳ에서 재인용

〈표 3〉

F 분 포 의 임 계 치

F분포에 대한 5 %(**Roman Type**)와 1 %(**Bold Face Type**)의 점



n_2	n_1자유도 (평균평방이 클 경우)																								n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	1
	4,052	**4,999**	**5,403**	**5,625**	**5,764**	**5,859**	**5,928**	**5,981**	**6,022**	**6,056**	**6,082**	**6,106**	**6,142**	**6,169**	**6,208**	**6,234**	**6,258**	**6,286**	**6,302**	**6,323**	**6,334**	**6,352**	**6,361**	**6,366**	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	2
	98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.34**	**99.36**	**99.38**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53	3
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63	4
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.45**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36	5
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.27**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67	6
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23	7
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93	8
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71	9
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54	10
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**	**4.78**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40	11
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.66**	**3.62**	**3.60**	
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30	12
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.98**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**	
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21	13
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**	

〈표 3〉

F 분포의 임계치 (계속)

n_2	n_1 자유도 (평균평방이 클 경우)																							n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	15
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.48**	**3.36**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.61**	**3.55**	**3.45**	**3.37**	**3.25**	**3.18**	**3.10**	**3.01**	**2.96**	**2.89**	**2.86**	**2.80**	**2.77**	**2.75**	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.45**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.86**	**2.79**	**2.76**	**2.70**	**2.67**	**2.65**	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	18
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**	**3.44**	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.91**	**2.83**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.70**	**2.63**	**2.60**	**2.54**	**2.51**	**2.49**	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**	**3.30**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.77**	**2.69**	**2.63**	**2.56**	**2.53**	**2.47**	**2.44**	**2.42**	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.63**	**2.58**	**2.51**	**2.47**	**2.42**	**2.38**	**2.36**	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.37**	**2.33**	**2.31**	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.53**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.25**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.36**	**2.33**	**2.27**	**2.23**	**2.21**	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.21**	**3.13**	**3.05**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.32**	**2.29**	**2.23**	**2.19**	**2.17**	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.17**	**3.09**	**3.02**	**2.96**	**2.86**	**2.77**	**2.66**	**2.58**	**2.50**	**2.41**	**2.36**	**2.28**	**2.25**	**2.19**	**2.15**	**2.13**	

〈표 3〉

F 분포의 임계치 (계속)

n_2	n_1 자유도 (평균평방이 클 경우)																							n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**	
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**	
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	42
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**	**2.70**	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**	**1.91**	**1.85**	**1.80**	**1.78**	
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48	44
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**	**1.88**	**1.82**	**1.78**	**1.75**	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46	46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**	**1.86**	**1.80**	**1.76**	**1.72**	
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	48
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**	**1.84**	**1.78**	**1.73**	**1.70**	

〈표 3〉

F 분포의 임계치 (계속)

n_2	n_1 자유도 (평균평방이 클 경우)																							n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	x	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35	70
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.46**	**1.40**	**1.37**	
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	x
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**	

출전 : George W. Snedecor, *Statistical Methods*, 5th edition, Ames, Iowa: The Iowa State University Press, 5th edition, 1956, pp. 246–249.

지수 $2Z$ 를 가진 함수 $F = e$ 는 부분적으로 Fisher의 표 Ⅵ(7)에서 계산. 추가된 값들은 내삽법에 의해 대부분 도식으로 구함.

〈표 4〉 양측검정하의 d_1과 d_2의 5퍼센트 신뢰점

n	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
16	0.98	1.24	0.86	1.40	0.75	1.59	0.64	1.80	0.53	2.03
17	1.01	1.25	0.90	1.40	0.79	1.58	0.68	1.77	0.57	1.98
18	1.03	1.26	0.93	1.40	0.82	1.56	0.72	1.74	0.62	1.93
19	1.06	1.28	0.96	1.41	0.86	1.55	0.76	1.72	0.66	1.90
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
21	1.10	1.30	1.01	1.41	0.92	1.54	0.83	1.69	0.73	1.84
22	1.12	1.31	1.04	1.42	0.95	1.54	0.86	1.68	0.77	1.82
23	1.14	1.32	1.06	1.42	0.97	1.54	0.89	1.67	0.80	1.80
24	1.16	1.33	1.08	1.43	1.00	1.54	0.91	1.66	0.83	1.79
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
26	1.19	1.35	1.12	1.44	1.04	1.54	0.96	1.65	0.88	1.76
27	1.21	1.36	1.13	1.44	1.06	1.54	0.99	1.64	0.91	1.75
28	1.22	1.37	1.15	1.45	1.08	1.54	1.01	1.64	0.93	1.74
29	1.24	1.38	1.17	1.45	1.10	1.54	1.03	1.63	0.96	1.73
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
31	1.26	1.39	1.20	1.47	1.13	1.55	1.07	1.63	1.00	1.72
32	1.27	1.40	1.21	1.47	1.15	1.55	1.08	1.63	1.02	1.71
33	1.28	1.41	1.22	1.48	1.16	1.55	1.10	1.63	1.04	1.71
34	1.29	1.41	1.24	1.48	1.17	1.55	1.12	1.63	1.06	1.70
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
36	1.31	1.43	1.26	1.49	1.20	1.56	1.15	1.63	1.09	1.70
37	1.32	1.43	1.27	1.49	1.21	1.56	1.16	1.62	1.10	1.70
38	1.33	1.44	1.28	1.50	1.23	1.56	1.17	1.62	1.12	1.70
39	1.34	1.44	1.29	1.50	1.24	1.56	1.19	1.63	1.13	1.69
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.62	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
75	1.53	1.58	1.50	1.61	1.47	1.64	1.45	1.67	1.42	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
85	1.56	1.60	1.53	1.63	1.51	1.65	1.49	1.68	1.46	1.71
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
95	1.58	1.62	1.56	1.65	1.54	1.67	1.52	1.69	1.50	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72

출전 : J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, vol. 38 (1951), pp. 159–177.

〈附錄 2〉

해 답

제 1 장

1. $\bar{X} = 3, \sum_{t=1}^{n} (X_t - \bar{X}) = -3 + 2 + 3 - 2 = 0$

2. $$\begin{aligned}\sum_{t=1}^{n} (aX_t + bY_t + cZ_t) &= (aX_1 + bY_1 + cZ_1) + \cdots + (aX_n + bY_n + cZ_n) \\ &= (aX_1 + \cdots + aX_n) + (bY_1 + \cdots + bY_n) \\ &\quad + (cZ_1 + \cdots + cZ_n) \\ &= \sum_{t=1}^{n} aX_t + \sum_{t=1}^{n} bY_t + \sum_{t=1}^{n} cZ_t \\ &= a\sum_{t=1}^{n} X_t + b\sum_{t=1}^{n} Y_t + c\sum_{t=1}^{n} Z_t\end{aligned}$$

3. $$\begin{aligned}\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y}) &= \sum_{t=1}^{n} [X_t(Y_t - \bar{Y}) - \bar{X}(Y_t - \bar{Y})] \\ &= \sum_{t=1}^{n} X_t(Y_t - \bar{Y}) - \sum_{t=1}^{n} \bar{X}(Y_t - \bar{Y})\end{aligned}$$

그러나 $\bar{X}$는 常數이기 때문에 마지막 항을 $\bar{X}\sum_{t=1}^{n}(Y_t - \bar{Y}) = 0$ 으로 쓸 수 있다.

제 2 장

1. $$\hat{a} = \bar{Y} - \hat{b}\bar{X} = \sum_{t=1}^{n} \frac{Y_t}{n} - \bar{X}\sum_{t=1}^{n} \frac{(X_t - \bar{X})Y_t}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

$A = \sum_{t=1}^{n} (X_t - \bar{X})^2$ 이라 놓으면,

$$\hat{a} = \sum_{t=1}^{n} \frac{Y_t}{n} - \frac{\bar{X}}{A}\sum_{t=1}^{n} (X_t - \bar{X})Y_t = \sum_{t=1}^{n} \left[\frac{1}{n} - \frac{\bar{X}}{A}(X_t - \bar{X})\right] Y_t \text{ 임.}$$

문제에서 $W_t = (X_t - \bar{X})/A$이기 때문에, $\hat{a} = \sum_{t=1}^{n} \left(\frac{1}{n} - \bar{X}W_t\right) Y_t$

2. a. 정규방정식은

$$\sum Y_t = n\hat{a}_0 + \hat{b}\sum X_t$$

$$\sum X_tY_t = \hat{a}_0\sum X_t + \hat{b}\sum X_t^2$$

X_t 와 Y_t 의 관찰치를 계산하면 다음과 같다.

$$\sum_{t=1}^{5} Y_t = 30, \quad \sum_{t=1}^{5} X_t = 12, \quad \sum_{t=1}^{5} X_t^2 = 34, \quad \sum_{t=1}^{5} X_t Y_t = 74, \quad N = 5$$

b. 정규방정식은 아래와 같이 된다.

$$30 = 5\hat{a} + 12\hat{b}$$
$$74 = 12\hat{a} + 34\hat{b}$$

윗식을 풀면,

$$\hat{a} = 5.076, \qquad \hat{b} = 0.385$$
$$\hat{\sigma}_u^2 = \frac{\sum (Y_t - \hat{Y}_t)^2}{n-2} = 3.077$$

3. 그의 주장은 틀리다. 왜냐하면 陽과 陰의 값을 취하나 그 기대값 $E(u_t)$가 0인 교란항 u_t의 존재를 무시하였기 때문이다. $C_t = a + b\,Y_t$ 관계는 정확한 관계가 아니라 오히려 평균 관계이다.

4. 먼저 μ_Y를 도출한다.

$$E(Y) = E(5 - 3X) = 5 - 3E(X) = 5 - 3\mu_X = \mu_Y$$

그러므로, 공분산은

$$\sigma_{X,Y} = E(Y - \mu_Y)(X - \mu_X) = E(5 - 3X - 5 + 3\mu_X)(X - \mu_X)$$
$$= E[-3(X - \mu_X)^2] = -3\sigma_X^2$$

Y의 분산은 $E(Y - \mu_X)^2 = 9\,\sigma_X^2$ 이다. 따라서 $\sigma_Y = 3\,\sigma_X$. 그러므로 상관계수는,

$$\rho_{X,Y} = \frac{-3\sigma_X^2}{3\sigma_X\sigma_X} = -1$$

5. 확률변수의 합이 갖는 분산에 대한 근본적인 관계를 이용한다. 그 관계는 만일 $Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ 이고 $X_1, \cdots, X_n$ 이 독립적이라면 다음과 같다고 하는 것이다.

$$\text{var}(Y) = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \cdots + a_n^2\sigma_{X_n}^2$$

위의 관계를 이 문제에 응용하면,

$$\text{var}(Y) = 4 + 27 + 500 = 531$$

6. *a*. 문제의 주장은 다음과 같이 정식화될 수 있다.

$$Y_i = a + b(T_{ci} - T_{si}) + u_i$$

여기서

Y_i = i 번째 도시의 평균 가구소득

T_{ci} = i 번째 도시의 조세율

T_{si} = i 번째 도시 교외의 조세율

그리고 u 는 교란항

b. 또 다른 정식화는

$$Y_i = a + b\frac{T_{ci}}{T_{si}} + u_i$$

두 경우 모두 $b < 0$ 이다. 즉, T_c 가 T_s 보다 상대적으로 높다면, 중류 및 상류 소득 가구는 교외로 이주할 것으로 기대된다. 따라서, 도시에 남아있는 가구의 평균 소득 Y_i 는 낮을 것이다.

7. (2)를 (1)에 대입하면,

$$Y_t = (a_1 + a_2) + (b_1 + b_2)X_t + \varepsilon_t$$

즉, X_t가 Y_t의 평균에 미치는 영향은 $(b_1 + b_2)$가 될 것이다. 이 문제는 우리가 (2)에서처럼 설명변수에 선형으로 선형으로 관련된 또 다른 모형으로부터 도출되는 것으로서 표준적인 이변수 회귀모형을 관찰할 수 있음을 보여준다.

8. 위배된다. 이를 보기 위하여, 교란 관계를 문제 7의 (1)에 대입하면,

$$Y_t = (a_1 + a_2) + b_1 X_t + (b_2 X_t^2 + \varepsilon_t)$$

즉, Y_t를 X_t에 관련시키는 모형은 교란항을 $W_t = b_2 X_t^2 + \varepsilon_t$로 갖는다. W_t가 0의 평균을 가지지 않을 것임은 분명하다. 더구나 X_t와 X_t^2은 분명히 관련되어 있기 때문에 W_t는 설명변수 X_t와 상관이 있다. 만일 상이한 시점의 X_t 값들이 관련되어 있다면, W_t도 그럴 것이다.

그러므로, cov $(W_t, W_s) \neq 0$이다. 마지막으로 W_t의 값이 부분적으로 X_t의 값에 달려 있기 때문에, W_t의 분산은 매 시기마다 동일하지 않다.

9. A_t를 t 시점의 어린이의 나이라 하고 H_t를 인치(inch)로 잰 어린이의 키라 한다. 그러면 다음과 같이 가정할 수가 있다.

$$H_t = a + bA_t + u_t$$

여기서 우리는 $b > 0$을 기대할 것이다. 이러한 관계의 단점은 그 관계가 오직 제한된 年數에만 유지됨이 분명하다는 사실이다. 즉, 어린이가 해를 거듭하여 나이를 먹어갈지라도, 그 키는 분명 한계에 도달한다!

10. *a*. 회귀모형은 다음과 같다.

$$Y_t = a + bX_t^m + (u_t - b\varepsilon_t)$$

b. 있다. 교란항이 설명변수 X_t^m과 상관이 있게 된다. 이를 보기 위해서 다음에 유의하자.

$$E[X_t^m(u_t - b\varepsilon_t)] = E(X_t^m u_t) - bE(X_t^m \varepsilon_t)$$
$$= 0 - b\sigma_\varepsilon^2 \neq 0$$

왜냐하면 $(X_t^m \varepsilon_t) = X_t \varepsilon_t + \varepsilon_t^2$ 이고, X_t와 ε_t는 독립적이기 때문이다.

제 3 장

1. S를 I, Q의 모집단 평균이라 하자. 그러면, $H_1 : S > 100$에 대한 $H_0 = 100$을 검정한다. 우리는 오직 $S \geq 100$에만 관심이 있기 때문에 $(100 - 1 = 99)$의 자유도를 갖는 단측 t검정을 사용할 수 있다. S의 좌측 임계치는 $\hat{S} - (t_{n-1;\,0.975})\hat{\sigma}_s = 110 - 1.65(2)$이다. 좌측 임계치가 100 이상이기 때문에 $H_0 : S = 100$을 기각하고 $H_1 : S > 100$을 채택한다.
2. $\overline{X}_{20}$과 $\overline{X}_{30}$을 각각 표본 크기 20과 30에 기초한 표본 평균이라 하자. 만일 두 표본이 모두 동일한 모집단에서 추출되었다면, $E(\overline{X}_{20}) = E(\overline{X}_{30}) = \mu$이다. 여기서 μ는 모집단의 평균이다. 따라서 두 추정량은 모두 불편 추정량이다. 그러나 $\overline{X}_{30}$이 더 선호될 것이다. 왜냐하면 그 분산이 더 작기 때문이다. 그러므로 $\overline{X}_{30}$의 사용은 신뢰구간을 더 좁혀서 보다 신뢰할 수 있는 가설검정을 할 수 있도록 한다.
3. 결과로부터 $\hat{\alpha} = 0.125$임을 안다. $H_1 : \alpha \neq 0$에 대한 $H_0 : \alpha = 0$의 가설검정은 $H_1 : (1 - \alpha) \neq 1$에 대한 $H_0 : (1 - \alpha) = 0$을 검정함으로

써 문제의 회귀식을 이용할 수 있다. t 비율의 절대값은 아래와 같이 바로 얻는다.

$$\left|\frac{0.875-1.00}{0.15}\right|=\frac{0.125}{0.15}\doteq 0.83$$

이는 2보다 상당히 작다. 그러므로 귀무가설을 채택하고, 노동시간의 변동은 40으로부터의 이탈에 비례하지 않는다고 결론내린다.

4. h를 평균키라 하자. 그러면 $H_1: h < 70$에 대한 $H_0: h = 70$을 검정한다. 오직 $h \leq 70$에만 관심이 있기 때문에 단측 정규분포 검정을 사용할 수 있다. 따라서 h값의 우측 임계치는 $(\hat{h} + 1.65\,\sigma_h)$ $= 68 + 1.65\,(2) = 71.3$이다. $70 < 71.3$이기 때문에 $H_0: h = 70$을 채택한다.

5. 정규성의 가정은 신뢰구간의 구성과 가설검정을 용이하게 한다. 이 점은 본문에서 보였다. 그러나, 정규성의 가정은 신뢰구간 또는 가설검정에 필수적이지는 않다. 그러나 정규성의 가정없이는 이 두 작업이 더 어려워 진다.

6. 귀무가설을 서부 사람들의 평균키가 67인치라고 하자. 대립가설은 서부 사람들의 평균키가 67인치를 넘는다가 될 것이다. 만일 서부 사람들이 사실 키가 더 크지 않는데도 더 크다고 믿게 된다면 제1종의 과오를 범한 것이다. 그 결과로 불필요한 재장비가 이루어 질 것이다. 더구나 틀린 크기의 코트를 생산할 것이다. 제2종의 과오는 서부 사람들이 사실 더 큼에도 더 크지 않다고 믿는 경우 범하게 된다. 그 결과는 잘못된 크기의 코트가 생산되는 것이다. 이상은 이 경우에 제1종 오류에 따른 비용이 제2종 오류의 비용보다 더 많음을 시사한다.

7. 회귀모형은 얼핏 보이는 것처럼 그렇게 제한적이지 않다. 왜냐하면 다양한 변환을 현명하게 이용함으로써 매우 많은 비선형 관계를 선형으로 변환할 수 있다. $Z_t = 1/(1-X_t)$라 놓으면, 관찰치 행렬은 다음과 같다.

Y_t	X_t	Z_t
1	0	1
10	0.1	1.11
12	0.5	2

8. *a*. (5% 유의수준에서) 가설검정은 단지 t 비율의 절대값이 그를 "상당한" 정도로 초과하는지 여부를 관찰함으로써 이루어 진다. 여기서는 그렇기 때문에, 귀무가설을 기각한다.

b. 표준편차는

$$\hat{\sigma}_{\hat{a}} = \frac{15}{3.1} \doteq 4.84, \qquad \hat{\sigma}_{\hat{b}} = \frac{0.81}{18.7} \doteq 0.043$$

c. 95% 신뢰구간은,

$$0.81 \pm (t_{n-2;\,0.975})\,0.043 = 0.81 \pm 0.091$$

9. 복지에 대한 수요 D_t가 복지수혜율 B_t에 관련되있다는 가설은 다음과 같이 표현할 수 있을 것이다.

$$D_t = a_1 + a_2 B_t + u_{1t}$$

여기서 $a_2 > 0$을 기대할 것이다. 수혜율이 정치적 압력을 통해서 복지에 대한 수요에 관련된다는 가설은 만약 정치적 과정에 기인한 시차를 가정한다면, 다음과 같이 표현될 수 있을 것이다.

$$B_t = b_1 + b_2 D_{t-1} + u_{2t}$$

10. N_t를 어떤 주에 입지한 기업의 수라 하자. 그러면 기업의 입지모형은 다음과 같이 나타낼 수가 있을 것이다.

$$N_t = a_1 + b_1\left(\frac{T_{1t}}{T_{2t}}\right) + u_{1t}$$

여기서 T_{1t}는 그 주의 조세율이며, T_{2t}는 인접한 주들의 평균·조세율이다. 여기서 $b_1 < 0$일 것으로 기대된다. 마찬가지로, P_t를 공해의 어떤 측정 수단이라 하자. 그러면, 공해 관계는 다음과 같이 표현될 것이다.

$$P_t = a_2 + b_2 N_t + u_{2t}$$

여기서는 $b_2 > 0$으로 기대된다.

11. X_t에 대해서 풀면,

$$X_t = \frac{5 - Z_t}{3}$$

윗식을 회귀모형에 대입하면,

$$Y_t = a + b\left(\frac{5}{3} - \frac{Z_t}{3}\right) + u_t$$

$$= (a + \tfrac{5}{3}b) - \frac{b}{3}Z_t + u_t$$

따라서 절편은 $(a + \frac{5}{3}b)$이고, 기울기는 $-b/3$이다.

제 4 장

1. 모형의 가정은 다음과 같다.

a. 1. 교란항의 기대치는 0이다. 곧, $E(u_t) = 0$.

2. 교란항의 분산은 일정하다. 곧, $E(u_t - 0)^2 = E(u_t^2) = \sigma_u^2$

3. 하나의 관찰치에 대한 교란항의 값은 여타의 관찰치의 값과는 무관하다. 그러므로 어떠한 두 관찰치에 대한 교란항, u_s와 u_t 사이의 공분산은 0이다. 곧, cov $(u_t, u_s) = 0$.

4. 교란항은 각각의 독립변수와 그것들의 모든 시차변수와 무관하다. 따라서 cov $(u_t, X_{it}) = 0$.

5. 독립변수중에서 여타변수의 선형결합인 것은 아무것도 없다.

b. 정규방정식은 다음과 같다.

1. $\sum Y_t = n\hat{a}_0 + \hat{a}_1 \sum X_{1t} + \hat{a}_2 \sum X_{2t}$

2. $\sum X_{1t}Y_t = \hat{a}_0 \sum X_{1t} + \hat{a}_1 \sum X_{1t}^2 + \hat{a}_2 \sum X_{1t}X_{2t}$

3. $\sum X_{2t}Y_t = \hat{a}_0 \sum X_{2t} + \hat{a}_1 \sum X_{2t}X_{1t} + \hat{a}_2 \sum X_{2t}^2$

첫번째 정규방정식은 $E(u_t) = 0$이라는 가정에 따라 $\Sigma \hat{u}_t = 0$으로 놓음으로써 도출된다. 두번째 방정식은 $E(X_{1t}\, u_t) = \text{cov}\,(X_{1t}, u_t) = 0$이라는 가정에 따라 $\Sigma(X_{1t}\hat{u}_t) = 0$으로 놓음으로써 도출된다. 세번째 방정식은 $E(X_{2t}\, u_t) = \text{cov}\,(X_{2t}, u_t) = 0$이라는 가정에 따라 $\Sigma(X_{2t}\, u_t) = 0$으로 놓음으로써 도출된다.

c. $10 = 100\ \hat{a}_0$

$30 = 35\ \hat{a}_1$

$20 = 3\ \hat{a}_2$

$\hat{a}_0 = 1/10$

$\hat{a}_1 = 6/7$

$\hat{a}_2 = 20/3$

2. 추정할 수 없는 모수는 a_1, a_2와 a_3이다. 이는 세번째 설명변수, $(X_{1t} - X_{2t})$가 X_{1t}와 X_{2t}의 선형결합이고, 따라서 완전다중공선성이 존재하기 때문이다. 이는 해당되는 정규방정식이 X_{1t}와 X_{2t}에 해당되

는 정규방정식의 선형결합임을 의미한다. 그러므로 일반적으로 모수추정량을 구할 수 없다. 유의할 점은 X_1X_2가 선형결합이 아니며, 따라서 문제는 나타나지 않는다. 방정식을 다음과 같이 써보자.

$$Y_t = a_0 + (a_1 + a_3)X_{1t} + (a_2 - a_3)X_{2t} + a_4X_{1t}X_{2t} + \varepsilon_t$$

그러면 $\hat{a}_0$, ($\widehat{a_1 + a_3}$), ($\widehat{a_2 - a_3}$)와 $\hat{a}_4$을 구할 수 있음을 알게 된다.

3. 정규방정식은 다음과 같다.

$$\begin{aligned}
\sum Y_t &= nb_0 + b_1 \sum X_{1t} + b_2 \sum X_{2t} \\
\sum X_1Y_t &= b_0 \sum X_{1t} + b_1 \sum X_{1t}^2 + b_2 \sum X_{1t}X_{2t} \\
\sum X_2Y_t &= b_0 \sum X_{2t} + b_1 \sum X_{2t}X_{1t} + b_2 \sum X_{2t}^2
\end{aligned}$$

X_t와 Y_t의 관찰치로 계산하면 다음과 같다. $n = 5$, $\Sigma X_1 = 8$, $\Sigma X_2 = 9$, $\Sigma Y_t = 30$, $\Sigma X_{1t} X_{2t} = 12$, $\Sigma X_{1t}^2 = 16$, $\Sigma X_{2t}^2 = 19$, $\Sigma X_{1t} Y_t = 50$, $\Sigma X_{2t} Y_t = 51$이다. 이 계산치를 정규방정식에 대입하면 다음을 얻게 된다.

$$\begin{aligned}
30 &= 5b_0 + 8b_1 + 9b_2 \\
50 &= 8b_0 + 16b_1 + 12b_2 \\
51 &= 9b_0 + 12b_1 + 19b_2
\end{aligned}$$

4. 회귀모형은 다음과 같이 나타낼 수 있을 것이다.

$$Y_i = a + b_1(T_{ci} - T_{si}) + b_2(C_{ci} - C_{si}) + b_3(H_{ci} - H_{si}) + b_4D_{ci} + u_i$$

여기에서

Y_i = i 번째 도시의 평균소득

T_{ci} = i 번째 도시의 세율

T_{si} =해당되는 i번째 근교의 세율

C_{ci} =i번째 도시의 범죄율

C_{si} =해당되는 i번째 근교의 범죄율

H_{ci} =i번째 도시의 거주비용

H_{si} =해당되는 i번째 근교의 거주비용

D_{ci} =i번째 도시의 밀도(1평방마일당 인구)

u_i =교란항

횡단면자료와 시계열자료의 상대적인 중요성은 문제의 속성에 달려있을 것이다. 위의 회귀모형에서는 횡단면자료가 아마 보다 유용할 것이다. 왜냐하면 세율, 범죄율, 거주비용과 밀도는 도시내에서 시간에 걸쳐 변동하기보다는 도시간에 더욱 변화하는 경향이 있기 때문이다. 이러한 모형에서 모든 모수들은 음수일 것으로 예상된다.

5. a. 완전선형관계가 다음과 같다.

$$\bar{P}_t = \frac{\sum_{i=1}^{k} P_{it}}{k}$$

이는 $(k+1)$번째 정규방정식이 처음의 k개 정규방정식의 선형결합임을 의미한다. 추정할 모수는 $(k+3)$인 반면, 독립적인 정규방정식은 단지 $(k+2)$개 뿐이다. 그러므로 일반적으로 모든 모수추정량에 대한 유일한 해는 없다.

b. 관계 $P_t = \sum_{i=1}^{k} P_{it} / k$ 를 수요방정식을 대입하여 항을 정리하면 다음과 같아진다.

$$D_{it} = a_0 + P_{1t}\left(a_1 + \frac{b}{k}\right) + P_{2t}\left(a_2 + \frac{b}{k}\right) + \cdots + P_{kt}\left(a_k + \frac{b}{k}\right) + cY_t + u_{1t}$$

그러므로 $i=1, \cdots, k$에서 a_0, $(a_i + b/k)$와 C를 추정할 수 있을 것이다.

6 . *a*. X_t와 X_t^2이 완전상관이 아니므로 방정식은 완전다중공석성을 갖지 않는다. 정규방정식은 다음과 같다.

$$\sum Y_t = n\hat{b}_0 + \hat{b}_1 \sum X_t + \hat{b}_2 \sum X_t^2$$
$$\sum Y_t X_t = \hat{b}_0 \sum X_t + \hat{b}_1 \sum X_t^2 + \hat{b}_2 \sum X_t^3$$
$$\sum Y_t X_t^2 = \hat{b}_0 \sum X_t^2 + \hat{b}_1 \sum X_t^3 + \hat{b}_2 \sum X_t^4$$

이러한 세 방정식은 선형독립관계에 있으며 따라서 $\hat{b}_0$, $\hat{b}_1$과 $\hat{b}_2$의 해를 구할 수 있음을 유의해야 한다.

b .
$$4 = 4\hat{b}_0 + 8\hat{b}_1 + 30\hat{b}_2$$
$$23 = 8\hat{b}_0 + 30\hat{b}_1 + 134\hat{b}_2$$
$$107 = 30\hat{b}_0 + 134\hat{b}_1 + 642\hat{b}_2$$

제 5 장

1. 로그변환을 이용하면 다음과 같다.

$$Q_t' = B + aL_t' + bK_t' + u_t$$

여기에서

$$Q_t' = \log Q_t$$
$$B = \log(1/A)$$
$$L_t' = \log L_t,$$
$$K_t' = \log K_t$$

그러면 B, a, b를 추정하여 $\hat{A} = e^{-\hat{B}}$로 놓으면 된다. $\hat{A}$은 편의가 있지만 일치성을 가짐에 유의해야 한다.

2. 회귀모형은 다음과 같다.

$$I_t = a_0 + b_1 r_t + b_2 D_t + u_t$$

여기에서

$D_t = 0$ (t 시점에 대통령이 민주당원이면)

D_t= 1 (그렇지 않고 공화당원이면)

r_t= 이자율

u_t= 교란항

3. 모형은 다음과 같다.

$$C_t = a_0 + a_1 Z_{1t} + a_2 Z_{2t} + a_3 Z_{3t} + u_t$$
$$Z_{1t} = F_t Y_t$$
$$Z_{2t} = Y_t^{1/2}$$
$$Z_{3t} = \frac{1}{A_t}$$

4. 함수의 가능한 이동을 고려하기 위해 가변수를 도입한다. 그러면 회귀모형은 다음과 같다.

$$C_i = a + bY_i + cD_i + u_i$$

여기에서

$D_i = 0$ (i번째 소비자가 도시지역에 거주하는 경우)

$D_i = 1$ (그렇지 않은 경우)

그러므로 만일 i번째 소비자가 농촌주민인 경우 회귀는 $C_t = (a + c) + bY_i + u_i$이다. 그리고 만일 i번째 소비자가 도시지역에 거주하면 함수는

$$C_i = a + bY_i + u_i$$

5. a. 회귀모형은 다음과 같다.

$$I_t = b_0 + b_1 r_t + b_2 \Pi_t + b_3 \Delta S_t + u_t$$

여기에서

I_t =투자지출

π_t =이윤율

$\triangle S_t$=판매액의 변화

r_t =이자율

u_t =교란항

b. 위의 회귀의 추정상의 문제는 완전다중공선성에 의해 야기된다. 분명히 말하면 이자율이 각 시점에 15퍼센트이면 b_0와 b_2를 추정할 수 없을 것이다.

6. a. 제약이 없는 형태의 모형은 다음과 같다.

$$I_t = a_0 + a_1 r_t + b_0 S_t + b_1 S_{t-1} + \cdots + b_7 S_{t-7} + u_t$$

b. 모형의 알몬형태는 $b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$ 을 이용하면 다음과 같다.

$$I_t = a_0 + a_1 r_t + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + u_t$$

여기에서 $Z_{1t} = \sum_{i=0}^{7} S_{t-i}$, $Z_{2t} = \sum_{i=1}^{7} iS_{t-i}$, $Z_{3t} = \sum_{i=1}^{7} i^2 S_{t-i}$ 이다.

c. 정규방정식은 다음과 같다.

$$\sum_{t=0}^{n} I_t = n\hat{a}_0 + \hat{a}_1 \sum_{t=0}^{n} r_t + \hat{\alpha}_0 \sum_{t=0}^{n} Z_{1t} + \hat{\alpha}_1 \sum_{t=0}^{n} Z_{2t} + \hat{\alpha}_2 \sum_{t=0}^{n} Z_{3t}$$

$$\sum_{t=0}^{n} r_t I_t = \hat{a}_0 \sum_{t=0}^{n} r_t + \hat{a}_1 \sum_{t=0}^{n} r_t^2 + \hat{a}_0 \sum_{t=0}^{n} r_t Z_{1t} + \hat{\alpha}_1 \sum_{t=0}^{n} r_t Z_{2t} + \hat{\alpha}_2 \sum_{t=0}^{n} r_t Z_{3t}$$

$$\sum_{t=0}^{n} Z_{1t} I_t = \hat{a}_0 \sum_{t=0}^{n} Z_{1t} + \hat{a}_1 \sum_{t=0}^{n} Z_{1t} r_t + \hat{\alpha}_0 \sum_{t=0}^{n} Z_{1t}^2 + \hat{\alpha}_1 \sum_{t=0}^{n} Z_{1t} Z_{2t} + \hat{\alpha}_2 \sum_{t=0}^{n} Z_{1t} Z_{3t}$$

$$\sum_{t=0}^{n} Z_{2t} I_t = \hat{a}_0 \sum_{t=0}^{n} Z_{2t} + \hat{a}_1 \sum_{t=0}^{n} Z_{2t} r_t + \hat{\alpha}_0 \sum_{t=0}^{n} Z_{2t} Z_{1t} + \hat{\alpha}_1 \sum_{t=0}^{n} Z_{2t}^2 + \hat{\alpha}_2 \sum_{t=0}^{n} Z_{2t} Z_{3t}$$

$$\sum_{t=0}^{n} Z_{3t} I_t = \hat{a}_0 \sum_{t=0}^{n} Z_{3t} + \hat{a}_1 \sum_{t=0}^{n} Z_{3t} r_t + \hat{\alpha}_0 \sum_{t=0}^{n} Z_{3t} Z_{1t}$$

$$+ \hat{\alpha}_1 \sum_{t=0}^{n} Z_{3t}Z_{2t} + \hat{\alpha}_2 \sum_{t=0}^{n} Z_{3t}^2$$

7. a. b_2 의 추정량은 다음과 같다.

$$\hat{b}_2 = \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2 + 8\hat{\alpha}_3 + 16\hat{\alpha}_4$$
$$= 1 + 6 + 20 + 32 - 160 = -101$$

b. 원래의 모형에서 b 를 대체하면 다음과 같다.

$$Y_t = a + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + \alpha_3 Z_{4t} + \alpha_4 Z_{5t} + u_t$$

여기에서

$$Z_{1t} = \sum_{i=0}^{6} X_{t-i}, \qquad Z_{2t} = \sum_{i=0}^{6} iX_{t-i}$$

$$Z_{3t} = \sum_{i=1}^{6} i^2 X_{t-i}, \qquad Z_{4t} = \sum_{i=1}^{6} i^3 X_{t-i}$$

$$Z_{5t} = \sum i^4 X_{t-1}$$

$$\sum_{i=0}^{6} b_i = 7\alpha_0 + \sum_{i=0}^{6} i\alpha_1 + \sum_{i=0}^{6} i^2\alpha_2 + \sum_{i=0}^{6} i^3\alpha_3 + \sum_{i=0}^{6} i^4\alpha_4 = 1$$

이므로 다음과 같이 α_0 의 해를 구할 수 있다.

$$\alpha_0 = \frac{(1 - 21\alpha_1 - 91\alpha_2 - 441\alpha_3 - 2275\alpha_4)}{7}$$

위의 회귀에서 α_0 를 치환하면 다음과 같게 된다.

$$Y_t = a + \alpha_1 Q_{1t} + \alpha_2 Q_{2t} + \alpha_3 Q_{3t} + \alpha_4 Q_{4t} + u_t$$

여기에서

$$Y_t^* = \left(Y_t - \frac{Z_{1t}}{7}\right)$$

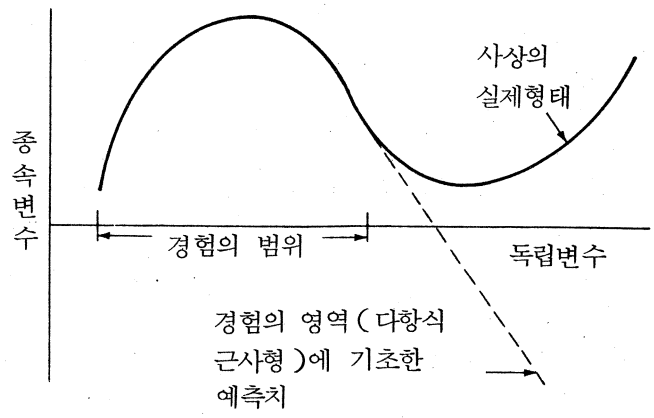
$$Q_{1t} = Z_{2t} - \frac{21Z_{1t}}{7}$$

$$Q_{2t} = Z_{3t} - \frac{91Z_{1t}}{7}$$

$$Q_{3t} = Z_{4t} - \frac{441Z_{1t}}{7}$$

$$Q_{4t} = Z_{5t} - \frac{2275Z_{1t}}{7}$$

8. 경제학자들이 생각하는 대부분의 함수는 다항식으로 근사화할 수 있다. 다항식의 차수는 경험의 영역에 의해 아니면 포함된 변수의 범위에 의해 결정될 것이다. 경험의 영역을 벗어나는 사상(events)에 대해서는 여러 차수의 다항식이 적당할 것이다. 그러므로 그러한 예측을 목적으로 추정방정식을 이용하는 것은 부적당할지도 모른다. 이는 다음의 그래프에서 증명되고 있다.



9. 방정식에 시차를 주고, 전체에 λ를 곱한 다음 원래의 방정식에서 그것을 제하면 다음과 같다.

$$Y_t = (a_0 - \lambda a_0) + a_1 X_t - a_1 \lambda X_{t-1} + \lambda Y_{t-1} + b_0 Z_t + u_t - \lambda u_{t-1}$$

10. $b_5 = 3$이면 $\alpha_0 + 5\,\alpha_1 + 25\,\alpha_2 = 3$이다. 그러므로 $\alpha_0 = 3 - 5\,\alpha_1 - 25\,\alpha_2$의 해를 구할 수 있다. 알몬모형에서 제한이 없는 형태는 다음과 같다.

$$Y_t = b + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + u_t$$

여기에서

$$Z_{0t} = \sum_{i=0}^{10} X_{t-i}, \quad Z_{1t} = \sum_{i=1}^{10} iX_{t-i}, \quad Z_{2t} = \sum_{i=1}^{10} i^2 X_{t-i}$$

α_0 를 대입하면 다음과 같은 제한이 없는 형태를 얻게 된다.

$$Y_t^* = b + \alpha_1 Q_{1t} + \alpha_2 Q_{2t} + u_t$$

여기에서

$$Y_t^* = Y_t - 3Z_{0t}, \quad Q_{1t} = (Z_{1t} - 5Z_{0t})$$

$$Q_{2t} = (Z_{2t} - 25Z_{0t})$$

11. 다음과 같다고 하자.

$$D_t = \begin{cases} 1 & (\text{만약 } r_t > 0.05) \\ 0 & (\text{그렇지 않으면}) \end{cases}$$

그러면 회귀모형은 다음과 같다.

$$C_t = a + b_1 Y_t + b_2(D_t r_t) + u_t$$

12. 다음과 같다고 하자

$$\log Y_t = Y_t^*$$

$$e^{X_{1t}} = Z_{1t}$$

$$\frac{1}{1 + X_{1t}X_{2t}} = Z_{2t}$$

그러면 회귀모형을 다음과 같이 쓸수 있을 것이다.

$$Y_t^* = a_0 + a_1 Z_{1t} + a_2 Z_{2t} + u_t$$

정규방정식은 다음과 같다.

$$\sum Y_t^* = n\hat{a}_0 + \hat{a}_1 \sum Z_{1t} + \hat{a}_2 \sum Z_{2t}$$

$$\sum Z_{1t}Y_t^* = \hat{a}_0 \sum Z_{1t} + \hat{a}_1 \sum Z_{1t}^2 + \hat{a}_2 \sum Z_{1t}Z_{2t}$$

$$\sum Z_{2t}Y_t^* = \hat{a}_0 \sum Z_{2t} + \hat{a}_1 \sum Z_{1t}Z_{2t} + \hat{a}_2 \sum Z_{2t}^2$$

제 6 장

1. 제 1 단계. $\hat{b}$ 과 $\hat{a}$ 을 계산한다.

$$\hat{b} = \frac{\sum_{t=1}^{15} (X_t - \bar{X})Y_t}{\sum_{t=1}^{15} (X_t - \bar{X})^2} = \frac{255}{280} \doteq 0.91$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = -0.28$$

그러므로 $\hat{Y}_t = -0.28 + 0.91X_t$ 이며, $\hat{u}_t = Y_t - (-0.28 + 0.91X_t)$ 이다.

제 2 단계. $\hat{u}_t^2$ 과 $(\hat{u}_t - \hat{u}_{t-1})^2$ 을 계산한다.

Y_t	$\hat{u}_t^2$	$(\hat{u}_t - \hat{u}_{t-1})^2$
0.63	1.876	—
1.54	0.211	0.828
2.45	0.203	0.828
3.36	5.570	3.648
4.27	1.612	1.188
5.18	0.032	1.188
6.09	0.008	0.008
7.00	1.000	0.828
7.91	4.368	9.548
8.82	1.392	0.828
9.73	0.073	0.828
10.64	1.850	1.188
11.55	11.903	4.369
12.46	6.052	34.928
13.37	5.617	0.008
	41.767	60.213

$$\sum \hat{u}_t^2 = 41.767$$

$$\sum (\hat{u}_t - \hat{u}_{t-1})^2 = 60.213$$

그러므로 $d = 60.213/41.767 \fallingdotseq 1.44$ 이며, 이는 통계표 4 에서 보면 상한 1.23 보다 훨씬 크다. 따라서 자기상관이 있다는 가설을 기각한다.

2. a. 이분산성이 있다고 하는 것이 의미있을 것이다. 왜냐하면 산출량은 시간에 걸쳐 증가함에도 그 요소의 분산과 교란항은 그렇

지 않다고 믿기는 어렵기 때문이다.

b. 회귀에서 k_t를 빠뜨리면 a와 b의 추정량은 편의를 가지게 될 것이다. 왜냐하면 그에 따른 방정식의 교란항은 $u_t^* = ck_t + u_t$ 일 것이기 때문이다. 일반적으로 이러한 교란항은 평균이 0이 아니며, 설명변수 L_t와 상관이 있을 것이다. $\hat{b}$에서 양수의 편의가 있을것으로 예상된다. 왜냐하면 L_t는 산출량에 대한 L_t와 k_t 전부의 영향을 자기의 것으로 인정받게 될 것이기 때문이다.

3. a. 방정식(1)을 총합하여 N으로 나누면 다음을 얻게 된다.

$$\sum_{i=1}^{N} \frac{C_{it}}{N} = a + b_1 \sum_{i=1}^{N} \frac{Y_{it}}{N} + b_2 \sum_{i=1}^{N} \frac{Y_{it}^2}{N} + \frac{\sum u_{it}}{N}$$

정의를 이용하면 다음과 같다.

$$C_t = a + b_1 Y_t + b_2 \sum_{i=1}^{N} \frac{Y_{it}^2}{N} + u_t$$

이제

$$\sum_{i=1}^{N} Y_{it}^2 = \sum_{i=1}^{N} (Y_{it} - Y_t + Y_t)^2$$

$$= \sum_{i=1}^{N} (Y_{it} - Y_t)^2 + \sum_{i=1}^{N} Y_t^2 + 2 \sum_{i=1}^{N} Y_t (Y_{it} - Y_t)$$

Y_t는 Y_{it}의 평균이므로 마지막항 $2\sum_{i=1}^{N} Y_t(Y_{it} - Y_t) = Y_t \sum_{i=1}^{N} (Y_{it} - Y_t)$

$= 0$이다. 또한 $\sum_{i=1}^{N} Y_t^2 = NY_t^2$ 임에 유의해야 한다. 그러므로 N으로 나누게 되면 거시모형은 다음과 같을 것이다.

$$C_t = a + b_1 Y_t + b_2 Y_t^2 + b_2 s_t^2 + u_t$$

여기에서

$$S_t^2 = \sum_{i=1}^{T} \frac{(Y_{it} - Y_t)^2}{N}$$

이 S_t^2항은 사람들사이의 소득변동의 측정치이다. 그러므로 거시모

형은 방정식(3)에서 나타나듯이 규정할 수 없는 것이다.

b. 관찰치행렬을 두가지 기본적인 방법으로 쓰게 될 것이다.

t	C_{it}	Y_{it}
1	C_{11}	Y_{11}
1	C_{21}	Y_{21}
1	C_{31}	Y_{31}
2	C_{12}	Y_{12}
2	C_{22}	Y_{22}
2	C_{32}	Y_{32}

t	C_{it}	Y_{it}
1	C_{11}	Y_{11}
2	C_{12}	Y_{12}
1	C_{21}	Y_{21}
2	C_{22}	Y_{22}
1	C_{31}	Y_{31}
2	C_{32}	Y_{32}

c. 일반적으로 총계편의가 있을 것이다. 왜냐하면 변수의 비선형형태의 평균이 변수의 평균의 비선형형태와 같지 않을 것이기 때문이다. 예를 들면, 위에서 $\Sigma Y_{it}^2/N \neq Y_t^2$ 임을 보았다. 여기에서 Y_t 는 Y_{it} 의 평균이다. 보다 일반적으로 말하면 $\sum_{i=1}^{N} f(X_{it})/N \neq f(X_t)$ 일 것이다. 여기에서 X_t 는 사람들에 대한 X_{1t} 의 평균치이다.

4. a. 정규방정식은 다음과 같다.

$$(1)\ \sum M_{dt} = n\hat{b}_0 + \hat{b}_1 \sum i_t + \hat{b}_2 \sum i_{(t-1)} + \hat{b}_3 \sum \Delta i_t$$

$$(2)\ \sum i_t M_{dt} = \hat{b}_0 \sum i_t + \hat{b}_1 \sum i_t^2 + \hat{b}_2 \sum i_t i_{(t-1)} + \hat{b}_3 \sum i_t \Delta i_t$$

$$(3)\ \sum i_{(t-1)} M_{dt} = \hat{b}_0 \sum i_{t-1} + \hat{b}_1 \sum i_t i_{(t-1)} + \hat{b}_2 \sum i_{(t-1)}^2 + \hat{b}_3 \sum i_{t-1} \Delta i_t$$

$$(4)\ \sum \Delta i_t M_{dt} = b_0 \sum \Delta i_t + \hat{b}_1 \sum \Delta i_t i_t + \hat{b}_2 \sum \Delta i_t i_{t-1} + \hat{b}_3 \sum \Delta i_t^2$$

$\triangle i_t = i_t - i_{t-1}$ 임을 상기하면 방정식(4)를 다음과 같이 다시 쓰게 될 것이다.

$$\begin{aligned}(5)\ \sum (i_t - i_{t-1}) M_{dt} = {} & \hat{b}_0 \sum i_t - \hat{b}_0 \sum i_{t-1} + \hat{b}_1 \sum i_t^2 - \hat{b}_1 \sum i_t i_{t-1} \\ & + \hat{b}_2 \sum i_t i_{t-1} - \hat{b}_2 \sum i_{t-1}^2 + \hat{b}_3 \sum i_t^2 \\ & + \hat{b}_3 \sum i_{t-1}^2 - 2\hat{b}_3 \sum i_t i_{t-1}\end{aligned}$$

간단히 살펴보면 방정식(5)=방정식(2)−방정식(1)임이 분명하다. 이는 4번째 방정식이 독립이 아님을 뜻한다. 해를 구할 모수추정량은 4

개, 곧 $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ 와 $\hat{b}_3$ 이지만, 단지 세개의 독립적인 방정식이 있다. 그러므로 모든 모수를 추정할 수는 없다.

5. 생산함수를 로그변환하면 다음과 같게 된다.

$$\log Q_t = \log A + \alpha_1 \log L_{1t} + \alpha_2 \log(10{,}000 - L_{1t}) + \alpha_3 \log K_t$$

그러면 통상적인 절차에 의해 이 방정식을 추정할 수 있게 된다. 왜냐하면 $\log L_1$ 과 $\log(10{,}000 - L_1)$ 은 완전공선성을 갖지 않기 때문이다.

6. 본문에서 다음의 사실을 알고 있다.

$$\hat{b} = b + \frac{W_1 u_1}{A} + \cdots + \frac{W_n u_n}{A} \qquad (1)$$

여기에서 $W_t = X_t - \bar{X}$이고 $A = \Sigma (X_t - \bar{X})^2$ 이다. (1)에서 다음과 같다.

$$(\hat{b} - b)^2 = \frac{W_1^2 u_1^2}{A^2} + \cdots + \frac{W_n^2 u_n^2}{A^2} + \frac{W_1 W_2 u_1 u_2}{A^2}$$
$$+ \cdots + W_i W_j u_i u_j + \cdots$$

그러므로

$$\sigma_{\hat{b}}^2 = E(\hat{b} - b)^2 = \frac{\sigma_u^2}{\sum (X - \bar{X})^2} + E(\text{모든 교적항})$$

자기상관 때문에 교적항의 기대치는 더 이상 0이 아니며, 따라서 분산식도 더 이상 유지될 수 없다.

7. 방정식에서 불균등분산을 제거하기 위해 Y_t 로 각 변을 나눈다. 곧,

$$\frac{C_t}{Y_t} = b_0 \frac{1}{Y_t} + b_1 + b_2 \frac{A_t}{Y_t} + u_t^*$$

여기에서 $u_t^* = u_t / Y_t$ 이다. 정규방정식은 다음과 같다.

$$\sum \frac{C_t}{Y_t^2} = \hat{b}_0 \sum \frac{1}{Y_t^2} + \hat{b}_1 \sum \frac{1}{Y_t} + b_2 \sum \frac{A_t}{Y_t^2}$$

$$\sum \frac{C_t}{Y_t} = b_0 \sum \frac{1}{Y_t} + nb_1 + b_2 \sum \frac{A_t}{Y_t}$$

$$\sum \frac{C_t A_t}{Y_t^2} = b_0 \sum \frac{A_t}{Y_t^2} + b_1 \sum \frac{A_t}{Y_t} + b_2 \sum \frac{A_t^2}{Y_t^2}$$

유의해야할 점은 이 방정식이 상수항을 가지며 따라서 $\Sigma \hat{u}_t^* = 0$ 이라는 조건을 이용한다는 것이다.

8. 제 1 단계. 통상적인 방법에 의해 a_0, a_1 과 a_2 를 추정한다.

제 2 단계. 추정계수를 이용하여 교란항에 대한 일련의 추정치를 구한다. 곧,

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = Y_t - \hat{a}_0 - \hat{a}_1 \Delta Y_t - \hat{a}_2 r_t$$

제 3 단계. $\hat{\varepsilon}_t$에 대한 값을 다음의 관계에 대입한다.

$$\hat{\varepsilon}_t = \rho_1 \hat{\varepsilon}_{t-1} + \rho_2 \hat{\varepsilon}_{t-2} + u_t$$

$$\Sigma(\hat{u}_t \hat{\varepsilon}_{t-1}) = 0 \text{ 과 } \Sigma(\hat{u}_t \hat{\varepsilon}_{t-2}) = 0$$

으로 놓음으로써 ρ_1 과 ρ_2 를 추정한다.

제 4 단계. 원래의 모형을 다음과 같이 변환한다.

$$I_t^* = a_0^* + a_1 \Delta Y_t^* + a_2 r_t^* + u_t$$

여기에서

$$\begin{aligned} I_t^* &= I_t - \hat{\rho}_1 I_{t-1} - \hat{\rho}_2 I_{t-2} \\ a_0^* &= a_0 - \hat{\rho}_1 a_0 - \hat{\rho}_2 a_0 \\ \Delta Y_t^* &= \Delta Y_t - \hat{\rho}_1 \Delta Y_{t-1} - \hat{\rho}_2 \Delta Y_{t-2} \\ r_t^* &= r_t - \hat{\rho}_1 r_{t-1} - \hat{\rho}_2 r_{t-2} \end{aligned}$$

제 5 단계. 통상적인 방법에 의해 a_0^*, a_1 과 a_2 를 추정한다. 그리하여 다음을 얻는다.

$$\hat{a}_0 = \hat{a}_0^* / (1 - \hat{\rho}_1 - \hat{\rho}_2)$$

제 7 장

1. a. 내생변수는 C_t, Y_t 와 I_t 이다. 사전결정변수는 C_{t-1}, Y_{t-1} 과 r_t 이다.

   b. 방정식(1)의 문제는 Y_t 가 ε_{1t} 과 상관이 있다는 것이다. 그리하여 절차는 Y_t 를 $\hat{Y}_t$ 으로 대체하는 것이다. 모든 사전결정변수, C_{t-1}, Y_{t-1}, r_t 에 대해 Y_t 를 회귀함으로써 $\hat{Y}_t$ 을 얻는다. 그러므로 $\hat{Y}_t$ 은 다음과 같을 것이다.

$$\hat{Y}_t = \hat{\gamma}_0 + \hat{\gamma}_1 C_{t-1} + \hat{\gamma}_2 Y_{t-1} + \hat{\gamma}_3 r_t$$

   그러면 Y_t 를 $\hat{Y}_t$ 으로 대체한 뒤에 통상적인 방법에 의해 방정식(1)을 추정한다. 곧 정규방정식은 다음의 합들을 0 으로 놓음으로써 얻게 된다. 곧, $\Sigma\, \hat{\varepsilon}_1^* = 0$, $\Sigma(\hat{\varepsilon}_1^* C_{t-1}) = 0$, $\Sigma(\hat{\varepsilon}_t^* \hat{Y}_t) = 0$ 이다.

   c. 방정식(3)을 추정하기 위해 방정식(1)을 추정할 때와 동일한 절차를 이용한다. 정규방정식은 다음과 같은 합들을 0 으로 놓음으로써 얻게 될 것이다. 곧,

$$\Sigma\, \hat{\varepsilon}_2^* = 0, \Sigma(\hat{\varepsilon}_2^* Y_{t-1}) = 0, \Sigma(\hat{\varepsilon}_2^* r_t) = 0, \quad \Sigma(\hat{\varepsilon}_2^* \hat{Y}_t) = 0$$

2. a. 첫번째 방정식은 식별된다. 왜냐하면 첫번째 방정식에서 나타나지 않는 모형의 사전결정변수 $\dot{M}_t$ 의 수는 첫번째 방정식에서 나타나는 내생설명변수($\dot{P}_t$) 의 수보다 크거나 같기 때문이다. 이는 두번째 방정식의 경우에는 타당하지 않다. 그러므로 그것은 식별되지 않는다.

   b. 첫번째 방정식에 대한 추정절차는 다음과 같다.

   제1단계. 곧 모든 사전결정변수, $\dot{M}_t$ 와 UN_t 에 대해 $\dot{P}_t$ 을 회귀함으로써 $\hat{\dot{P}}_t$ 을 얻는다.

   그러면 $\hat{\dot{P}}_t$ 은 다음과 같게 된다.

$$\hat{\dot{P}}_t = \hat{\gamma}_0 + \hat{\gamma}_1 \dot{M}_t + \gamma_2 UN_t$$

제 2 단계. 첫번째 방정식에 $\dot{P}_t$ 을 $\hat{\dot{P}}_t$ 으로 대체하고 다음과 같은 합들을 0 으로 놓음으로써 제 2 단계의 정규방정식을 통례대로 만들게 된다. 곧,

$$\sum \hat{\varepsilon}_{1t}^{*} = 0, \sum (\hat{\varepsilon}_{1t}^{*} UN_t) = 0, \sum (\hat{\varepsilon}_{1t}^{*} \hat{\dot{P}}_t) = 0$$

3. a. 축약형방정식을 얻기 위해 L_t 와 W_t 에 대해 방정식(1)과 (2)를 푼다. 축약형방정식은 다음과 같을 것이다.

$$L_t = a_0^{*} + a_1^{*} P_t + a_2^{*} S_t + v_t$$
$$W_t = b_0^{*} + b_1^{*} S_t + b_2^{*} P_t + \varepsilon_t$$

여기에서

$$a_0^{*} = \frac{a_0 + a_1 b_0}{1 - a_1 b_1}, \quad a_1^{*} = \frac{a_1 b_2}{1 - a_1 b_1}, \quad a_2^{*} = \frac{a_2}{1 - a_1 b_1}$$
$$v_t = \frac{a_1 u_{2t} + u_{1t}}{1 - a_1 b_1}, \quad b_0^{*} = \frac{b_0 + b_1 a_0}{1 - a_1 b_1}, \quad b_1^{*} = \frac{b_1 a_2}{1 - a_1 b_1}$$
$$b_2^{*} = \frac{b_2}{1 - a_1 b_1}, \quad \varepsilon_t = \frac{u_2 + b_1 u_{1t}}{1 - a_1 b_1}$$

b. 방정식(1)에서의 문제는 W_t 가 u_{1t} 와 상관이 있다는 것이다. 그리하여 절차는 W_t 를 $\hat{W}_t$ 으로 대체하는 것이다. 사전결정변수 S_t 와 P_t 에 대해 W_t 를 회귀함으로써 $\hat{W}_t$ 을 구한다. 그러면 $\hat{W}_t$ 은 다음과 같은 형태일 것이다.

$$\hat{W}_t = \hat{\gamma}_0 + \hat{\gamma}_1 S_t + \hat{\gamma}_2 P_t$$

그러면 방정식(1)은 W_t 를 $\hat{W}_t$ 으로 대체한 뒤에 통상적인 방법에 의해 추정한다. 곧, 정규방정식은 다음과 같은 합들을 0 으로 놓음으로써 얻게 된다. 곧,

$$\sum \hat{u}_{1t}^{*} = 0, \sum (\hat{u}_{1t}^{*} \hat{W}_t) = 0, \sum (\hat{u}_{1t}^{*} S_t) = 0$$

4. a. $D_{i(t-1)}$ 이 u_t 와 상관이 있음을 알기 위한 직관적인 방법은 다음과 같다. 곧 방정식(1)에서 $D_{i(t-1)}$ 이 $u_{i(t-1)}$ 과 상관이 있음을

아는 것이다. 하지만 $u_{it} = \rho u_{i(t-1)} + \varepsilon_{it}$ 이므로 또한 그것은 $u_{i(t-1)}$에 의존한다. 그러므로 $D_{i(t-1)}$ 과 u_{it}는 공통적인 요소를 가지고 있으므로 상관이 있는 것이다.

b. 만일 기본방정식을 다시 풀면, D_{it}가 P_t와 그것의 모든 시차치에 의존하는 것이 분명해질 것이다. 보다 명확하게 하면 다음과 같다.

$$D_{it} = a_0 + a_0a_2 + a_0a_2^2 + \cdots + a_1P_t + a_1a_2P_{t-1} + a_1a_2^2P_{t-2} + \cdots + u_{it} + a_2u_{i(t-1)} + a_2^2u_{i(t-2)} + \cdots$$

게다가 D_{it}는 또한 u_{it}와 그것의 모든 시차치에 의존함도 알고 있다. 그러므로 D_{it}의 평균부분은 P_t와 그것의 시차치와 상관이 있을 것이다. 만일 D_{it}와 P_t와 u_{it}의 값을 관련시키는 위의 방정식을 축약형방정식으로 간주하면, 모든 사전결정변수가 알려져 있지 않거나 그에 대한 관찰치가 없는 모형에 대해 본문에서 서술한 TSLS 기법을 수행할 수 있다. 달리 말하면 P_t와 $\hat{D}_{i(t-1)}$에 대해 D_{it}를 회귀함으로써 TSLS 절차를 수행할 수 있다. 여기에서 $\hat{D}_{i(t-1)}$은 P_t와 그것의 시차치의 몇개, 가령 3개에 대해 $D_{i(t-1)}$을 회귀함으로써 계산된다.

5. (1)을 (2)에 대입하면 다음을 얻게 된다.

$$X_{2t} = b_0^* + b_1^*X_{1t} + v_t^* \qquad (1')$$

여기에서

$$b_0^* = \frac{c_0 + c_1b_0}{1 - c_1b_2}, \quad b_1^* = \frac{c_1b_1}{1 - c_1b_2}, \quad v_t^* = \frac{c_1u_{1t} + u_{2t}}{1 - c_1b_2}$$

(1′)에 u_{1t}를 곱하고 기대치를 구하면 다음과 같다.

$$E(X_{2t}u_{1t}) = b_0^*E(u_{1t}) + b_1^*E(u_{1t}X_{1t}) + \frac{c_1E(u_{1t}^2)}{1 - c_1b_2} + \frac{E(u_{1t}u_{2t})}{1 - c_1b_2}$$

$E(u_{1t}) = 0, E(u_1X_{1t}) = 0,$; $E(u_{1t}u_{2t}) = \text{cov}(u_1, u_2)$ 임을 기억하면 다음과 같게 된다.

$$E(X_{2t}u_{1t}) = \frac{c_1\sigma_1^2}{1 - c_1b_1} + \frac{\text{cov}(u_1, u_2)}{1 - c_1b_2} \neq 0$$

다만 다음과 같을 때 성립한다. 곧,

$$\frac{c_1\sigma_1^2}{1 - c_1b_2} = -\frac{\text{cov}(u_1, u_2)}{1 - c_1b_2}$$

6. a. 첫번째 방정식을 추정할 때 TSLS 기법이 붕괴됨을 보이기 위해 다음과 같이 진행한다. 곧, 우선 모형의 모든 사전결정변수에 대해 $\dot{P}_t$을 회귀한다. 이는 다음과 같게 된다.

$$\hat{\dot{P}}_t = \hat{c}_0 + \hat{c}_1(UN_t)$$

다음으로 첫번째 방정식에서 $\dot{P}_t$ 을 $\hat{\dot{P}}_t$으로 대체한다. 이는 다음과 같을 것이다.

$$\dot{W}_t = a_0 + a_1\hat{\dot{P}}_t + a_2(UN_t) + \varepsilon_{1t}^*$$

$\hat{\dot{P}}_t$과 UN_t가 완전상관관계에 있으므로 a_1과 a_2를 추정할 수 없음에 유의해야 한다. 그러므로 첫번째 방정식을 추정하려 한다면 TSLS 기법은 붕괴된다.

b. 두번째 방정식을 추정할 때 TSLS 기법은 붕괴되지 않는다. 왜냐하면 완전다중공선성의 문제에 직면하지 않기 때문이다. 우선 UN_t에에 대해 $\dot{W}_t$을 회귀함으로써 $\hat{\dot{W}}_t$을 구하고 두번째 방정식에서 $\dot{W}_t$을 $\hat{\dot{W}}_t$으로 대체한 다음 다음과 같은 합들을 0으로 놓음으로써 곧, $\sum \hat{\varepsilon}_{2t}^* = 0, \sum(\hat{\varepsilon}_{2t}^*\hat{\dot{W}}_t) = 0$ 으로 하여 통례대로 정규방정식을 구하게 된다. 정규방정식은 다음과 같다.

$$\sum \dot{P}_t = nb_0 + b_1 \sum \hat{\dot{W}}_t$$
$$\Sigma(\dot{P}_t) = b_0 \sum \hat{\dot{W}}_t + b_1 \sum \hat{\dot{W}}_t^2$$

이러한 방정식들은 선형독립이며, 따라서 그것들을 풀어서 $\hat{b}_0$와 $\hat{b}_1$을 구할 수 있다.

7. a. 방정식은 식별된다. 왜냐하면 그 방정식에 나타나지 않는 체계의 사전결정변수의 수가 내생설명변수의 수보다 크기 때문이다.

   b. TSLS 기법에 의해 방정식을 추정할 수 없다. 왜냐하면 그 방정식에 나타나지 않는 적어도 2개의 사전결정변수에 대한 관찰치가 필요하기 때문이다. 그런데 그러한 하나의 변수(X_{2t})에 대한 관찰치만 있는 것이다. 이러한 자료정보하에서 이 문제의 회귀방정식을 추정하는 방법은 없음을 보일 수 있다.

8. a. 두 방정식 모두 식별된다. 왜냐하면 각각의 배제된 사전결정변수의 수가 내생설명변수의 수와 같기 때문이다. 유의해야할 점은 X_t 와 X_t^2이 완전상관관계에 있지 않으며, 따라서 X_t^2를 두번째 방정식에 포함되지 않는 사전결정변수로 간주할 수 있다는 것이다.

   b. 축약형을 얻기 위해 첫번째 방정식에서 Y_{2t}를, 두번째 방정식에서 Y_{1t}을 대체한다. 항을 다시 정리하면 다음과 같다.

$$Y_{1t} = a_1^* + b_1^* X_t^2 + c_1^* X_t + v_{1t}^*$$
$$Y_{2t} = a_2^* + b_2^* X_t^2 + c_2^* X_t + v_{2t}^*$$

여기에서

$$a_1^* = \frac{a_1 + c_1 a_2}{1 - c_1 c_2}$$

$$b_1^* = \frac{b_1}{1 - c_1 c_2}$$

$$c_1^* = \frac{c_1 b_2}{1 - c_1 c_2}$$

$$v_{1t}^* = \frac{c_1 \varepsilon_{2t} + \varepsilon_{1t}}{1 - c_1 c_2}$$

$$a_2^* = \frac{a_2 + c_2 a_1}{1 - c_1 c_2}$$

$$b_2^* = \frac{c_2 b_1}{1 - c_1 c_2}$$

$$c_2^* = \frac{b_2}{1 - c_1 c_2}$$

$$v_2^* = \frac{c_2 \varepsilon_{1t} + \varepsilon_{2t}}{1 - c_1 c_2}$$

c. 첫번째 방정식을 추정할 때는 TSLS 기법을 이용한다. 그러므로 모든 사전결정변수에 대해 Y_{2t}를 회귀하면 가령 다음과 같은 것을 얻게 된다. 곧,

$$\hat{Y}_{2t} = \hat{a}_2^* + \hat{b}_2^* X_t^2 + \hat{c}_2^* X_t$$

다음으로 첫번째 방정식에서 Y_{2t}를 $\hat{Y}_{2t}$으로 대체하고 통상적인 절차에 의해 그 방정식을 추정하게 된다. 다음과 같은 합들을 0으로 놓음으로써, 즉 $\sum \hat{\varepsilon}_{1t}^* = 0$, $\sum (\hat{\varepsilon}_t^* X_t^2) = 0$, $\sum (\hat{\varepsilon}_{1t}^* \hat{Y}_{2t}) = 0$으로 놓음으로써 정규방정식을 도출하게 된다.

9. a. 첫번째 방정식은 식별되지 않는다. 왜냐하면 r_t는 횡단면분석에서 상수이기 때문이다. 곧 이미 첫번째 방정식에서 상수가 있기 때문에 r_{it}의 대해 그것은 배제된 사전결정변수의 기능을 할 수 없는 것이다. 두번째 방정식은 식별된다. 왜냐하면 판매변수가 배제되기 때문이다. 이제 모형의 변수에 대해 가령 T기간동안의 시계열 자료가 있다고 가정하자. 또한 r_t가 외생변수로 취급될 수 있도록 그러한 N개의 기업이 경제의 작은 부분을 나타낸다고 가정하자. 그러면 첫번째 방정식은 식별될 것이다. 왜냐하면 r_t는 배제된 사전결정변수로 간주될 수 있기 때문이다. 이러한 경우에 투자방정식은 다음과 같이 추정될 것이다. 우선 T시계열관찰치를 이용하여 해당되는 판매변수와 r_t 각각에 대해 r_{it}를 회귀함으로써 다음을

얻게된다.

$$\hat{r}_{it} = \hat{\gamma}_{0i} + \hat{\gamma}_{1i} S_{i(t-1)} + \hat{\gamma}_{2i} r_t, \qquad i = 1, \ldots, N$$

이제 각각의 r_{it}를 $\hat{r}_{it}$으로 대체하여 통상적인 절차에 의해 첫번째 방정식을 추정하게 된다. 이 가운데에서 유의할 점은 이 제2단계에서 관찰치는 NT 개라는 사실이다.

b. 첫번째 방정식에서 r_{it}를 대체함으로써 I_{it}에 대한 축소형을 구한다. 항을 다시 정리하면 다음을 얻게 된다.

$$I_{it} = a^* + b_1^* r_t + b_2^* S_{it-1} + v_{it}^*$$

여기에서

$$a^* = \frac{a}{1 - b_1 b_3}, \qquad b_1^* = \frac{b_1}{1 - b_1 b_3}$$

$$b_2^* = \frac{b_2}{1 - b_1 b_3}, \qquad v_{it}^* = \frac{b_1 \varepsilon_{it} + u_{it}}{1 - b_1 b_3}$$

이다.