

정보발생확률함수 I : 표본에서 모집단으로*

김 학 은**

요약

시계열 자료를 이용하여 회귀분석을 하는 경우 '이론적인 측면'에서 보면 반복되는 한 개의 설명변수 관측치에 대하여 무한개의 피설명변수의 확률분포와 그에 대응하는 무한개의 오차의 확률분포가 존재한다고 가정한다. 그러나 실제로는 이론적인 확률분포 가운데 임의의 시점에서 실현된 (realized) 피설명변수의 표본은 한 개의 관측치만으로 이루어졌다. 본 논문은 실현된 한 개의 피설명변수 표본에서 실현되지 않은 나머지 피설명변수들을 발생시킬 수 있는 확률함수를 제시한다. 이 방법은 몬테카를로 방법과 전혀 다르다. 이 정보발생함수 존재의 핵심은 지금까지 숨겨져 있던 새로운 정보공간의 발견에 있다. 이 논문에서 우리는 일반적인 정보발생 이론은 뒤로 미루고 특수한 예로서 시계열 자본자산가격모형(capital asset pricing model)을 들어서 정보발생함수를 유도한다.

핵심주제어 : 회귀직선, 관측직선, 관측평행직선, 상위공간, 하위공간, 변수전환, 정보발생확률함수

I. 머리말

시계열 자료(time series data)를 이용하여 회귀분석을 하는 경우 '이론적인 측면'에서 보면 임의의 한 시점에서 반복되는 한 개의 설명변수

* 본 논문은 2003년 7월 5일 연세대학교에서 개최한 한국계량경제학회 하계대회에서 발표하였고 하계대회 전인 동년 6월 4일 연세대학교 계량경제학 세미나에서 발표한 바가 있다. 본 논문에 귀중한 조언을 해주신 유병삼, 정진욱, 최강식, 한순구 재 교수에게 감사로 표한다. 특히 논문을 세심하게 읽어주신 윤석범 교수에게 감사드린다. 조언에 기초하여 수정된 논문을 2004년 2월 12일 한국경제학회에서 주관하는 정기학회에서 발표하였다. 유익한 논평을 해주신 박주현 교수에게 감사로 드린다. 오류가 있다면 필자의 것이다.

** 연세대학교 상경대학 경제학과 교수, 서울특별시 서대문구 신촌동 134, 우편번호 120-749, hakun@base.yonsei.ac.kr

관측치에 대하여 무한개의 피설명변수의 확률분포와 그에 대응하는 무한개의 오차의 확률분포가 존재한다고 가정한다. 그러나 현실에서는 이론적인 확률분포 가운데 임의의 한 시점에서 실현된(realized) 피설명변수의 표본은 한 개의 관측치뿐이므로 실험자 혹은 연구자에게는 한 시점에서 한 개의 설명변수 관측치에 대하여 한 개의 피설명변수 표본만이 주어질 수밖에 없다. 추가적인 정보가 없는 상태에서 실현된 한 개의 관측치만으로는 실현되지 않은 나머지 피설명변수들을 파악할 수 없다는 것은 당연하게 생각할 수 있다.

본 논문은 임의의 한 시점에서 실현된 한 개의 피설명변수 표본에서 실현되지 않은 나머지 피설명변수의 모집단을 발생 복제시킬 수 있는 확률함수를 제시한다. 결론부터 말하면 여기서 제시하는 모집단발생확률함수는 $R_{t,n+1} = \frac{1}{1 + \theta_n \varepsilon_{t,n}} R_{t,n}$ 이다. t 시점의 n 번째 관측표본 $R_{t,n}$ 은 동일 시점의 $n+1$ 번째 표본 $R_{t,n+1}$ 을 발생시킨다. 이 때 $n=1, 2, \dots, N$ 으로 발생 복제 순서를 가리킨다. 여기서 θ_n 은 n 번째 정보이며 알려져 있고 $\varepsilon_{t,n}$ 은 n 번째 오차이다. 따라서 $\theta_n \varepsilon_{t,n}$ 은 확률변수이며 양수 또는 음수의 값을 취하는 할인자(discount factor) 또는 할증자(premium factor)이다. 이 방법은 실현된 관측치에 모집단이 감추어져 있다는 점에 착안한 것으로 모의실험으로 확률변수를 발생시키는 몬테칼로 방법(the Monte Carlo study)과 전혀 다르다.

이 정보발생 확률함수 존재의 핵심은 지금까지 숨겨져 있던 새로운 정보공간의 발견에 있다. 김학은(1998, 2001, 2003)은 이 정보공간을 이용하여 유통속도를 분리한 바 있다. 정보공간의 이름은 어떻게 칭하든 아무 상관없지만 이 논문에서는 편의상 이미 실현되어 알려진 정보공간을 상위공간(upper space)이라 칭하고, 새로운 정보공간을 하위공간(lower space)이라고 칭하겠다.

이 같은 주장을 직관적으로 받아들이기 힘들 것이므로 생물학의 예를 통하여 비교 설명하겠다. 이 예가 본 논문의 정보발생확률과정과 반드시 일치하지 않겠으나 직관은 전달할 수 있다고 믿는다. 하나의 표본(sample)에서 그 표본이 속한 모집단(population)을 발생시킬 수 있다는 발상은 단일의 배아세포(胚芽細胞, embryonic cell)에서 그 세포가 속한 전체 개체(whole body)를 복제할 수 있는 생물학적 복제기술에 비유할

수 있다. 하나의 배아세포 속에는 전체의 정보가 모두 포함되어 있기 때문이다. 특히 DNA의 이중나선(double helix)의 두 가닥은 나선계단의 발판처럼 한 쌍의 핵산에 의해서 연결되어 있다. 각각의 가닥에 두 개의 핵산이 들어 있어 모두 네 종류의 핵산이 나선계단을 따라서 나타나는 순서 속에 유전정보가 담겨 있고 이 정보를 이용해서 DNA 분자는 유기체를 조합하고 스스로를 복제한다. 본 논문은 상위공간에 두 개의 정보와 하위공간의 두 개의 정보가 들어 있어 모두 네 개의 정보가 확률적 이중나선(stochastic double helix)처럼 서로 엮물리면서 하나의 표본이 스스로 새로운 정보를 창출하고 모집단을 복제하는 것을 보일 것이다.

생물학에서 하나의 세포 속에 두 개의 가닥이 유기체를 복제하듯이 물리학에서 빛의 파동의 간섭으로 나타나는 홀로그램도 하나의 레이저에서 나온 빛이 두 개의 광선으로 갈라지면서 생성된다. 하나의 광선은 물체에 반사하여 감광판에 투사되고, 다른 하나의 광선은 렌즈를 통과하여 이미 물체에 투사된 첫 번째 빔과 충돌하면서 감광판 위에 간섭 패턴을 만들어 낸다. 현상판을 통해서 레이저를 비추면 원래 물체의 완전한 3차원 영상이 나타난다. 현상판은 2차원 면인데 이렇게 하여 3차원의 전체 영상을 재구성하는데 필요한 모든 정보를 갖고 있다. 이 사실은 3차원 세계 속에서 일어나는 모든 현상과 연관된 정보가 2차원에 저장될 수 있음을 뜻한다. 뒤집어 말하면 2차원의 정보로 3차원의 물체를 복제할 수 있다는 뜻이다.

이 같은 사실은 음악에서도 쉽게 찾아볼 수 있다. 동양음악에서는 임의의 한 시점에서 하나의 음밖에 발생시키지 못한다. 여러 사람이 혹은 여러 악기가 동시에 연주하여도 한 시점에서 하나의 음밖에 내지 않는다. 다시 말하면 화음(harmony)이 없다. 이 현상은 널리 알려진 사실이라서 새삼스러울 수 없겠으나 서양의 화음이 도입된 현재에도 화음이 없는 틀을 벗어나지 못하는 실정이다. 동양음악에서 한 시점에서 여러 가지 음을 발생할 수 있다는 발상은 일찍이 하지 못했다. 이 같은 발상은 동양음악에서는 불가능하다고 여기는 것이 당연할 것이다.

이와 대조적으로 서양음악에서는 대체로 10세기 이후 화음의 발견으로 동일 시점에서 여러 사람이 혹은 여러 악기가 이론적으로는 무한개의 음을 동시에 연주할 수 있게 되었다. 피아노 같은 하나의 악기에서는 동시에 10개의 음을 발생시킬 수 있다. 음과 음 사이의 조화를 발견한 것이

다. 가령 G음은 그 존재 자체가 이미 그의 화음에 합당한 C음이나 E음을 발생시킬 수 있다는 가능성을 암시한다. 여기에 더하여 두 개의 음자리(clef)의 발견이 조화음을 거의 무한대로 발생시킬 수 있다. 높은 음자리(treble clef)가 미리부터 알려진 상위정보공간(upper space)이라면 낮은 음자리(low clef)는 10세기까지 알려지지 않은 새로운 하위정보공간(lower space)에 비유할 수 있다. 하나의 음은 그것이 속한 화음의 집합을 모두 복제해 낼 수 있는 배아(embryo)를 품고 있음을 뜻한다. 수많은 변주곡은 바로 음의 동시 발생의 가능성을 갖고 있는 배아에 바탕을 두고 있다. 작곡가는 한 시점에서 하나의 음이 떠오르면 동시에 실현되지 않은 음을 발생시킬 수 있기 때문이다.¹⁾ 그러나 동일한 음에 기초를 한다고 하여도 작곡가에 따라 변주곡은 다르므로 실현된 음에서 실현되지 않은 화음을 발생시킬 수 있는 가능성은 확률에 의존한다고 말할 수 있다. 이 논문에서도 우리는 실현된 정보에서 실현되지 않은 정보를 확률적으로 발생시킨다. 이러한 점에서 현재의 회귀분석은 화음을 발견하지 못한 동양음악에 비유할 수 있고 본 논문이 제시하는 정보발생함수는 화음을 발견한 서양음악에 비유할 수 있다고 생각한다.

확률적 현상이 반드시 생물학적, 물리학적, 또는 음악적 현상과 일치하는 것은 아니겠으나 확률적 현상에 자연현상의 조화(harmony)가 잠재적으로 숨어 있다면, 실현된 단일의 사상(event)에서 그와 조화를 이루는 숨겨진 모집단의 사상을 모두 복제해 낼 수 있을 것이다. 이 논문에서 우리는 상위공간의 표본에서 하위공간의 표본이 발생하고, 다시 하위공간의 표본에서 상위공간의 표본이 발생하는 것을 보일 것이다. 본 논문에서 소개하는 두 개의 공간(상위공간과 하위공간)은 DNA의 두 핵산 가닥, 홀로그램의 두 개의 광선, 음악의 두 개의 음자리에 비유할 수 있다.

이 시점에서 강조해야 할 점이 하나 있다. 확률적 현상의 조화를 밝히고, 모집단을 복제하는 과정에서 우리는 선형회귀분석에서 사용하는 통상의 기본 가정 이외에 어떠한 가정도 추가하지 않는다. 다만 대문자공간과 소문자공간 사이의 수학적 관계만 이용할 뿐이다. 이러한 면에서 이 논문은 기술적 면에서 순수 수학적 절차를 따른다.

1) 음악가 Bach는 베이스 성부를 직접 작곡한 다음에 제자들로 하여금 알토와 테너 성부를 완성하게 하였다. 이것은 제자들 측면에서 보면 베이스 성부라는 알려진 정보에서 알토와 테너 성부라는 숨겨진 정보를 발생시키는 것으로 이해할 수 있다.

그러나 이 논문의 한계는 일반적인 선형회귀방정식에서 숨겨진 피설명 변수를 발생시킬 수 없다는 점이다. 모집단을 발생시킬 수 있는 배아(胚芽)에 해당하는 배아방정식(胚芽方程式, embryonic equation)만이 모집단을 발생시킬 수 있다. 배아방정식은 '산술평균'의 형태라는 것이 밝혀진다. 따라서 일반적인 정보발생 이론은 미래의 숙제로 미루고 이 논문은 산술평균의 예로서 잘 알려진 시계열 자본자산가격모형(time series capital asset pricing model)을 이용하여 정보발생을 설명하겠다.²⁾ 이 모형에 의하면 한 시점 t 에서 주어진 안전자산의 기대수익률과 시장포트폴리오 기대수익률의 산술평균으로 설명되는 임의의 하나의 위험자산 기대수익률의 관측치는 한 개이다. 이 한 개의 관측치 수익률에서 이 수익률이 속하는 모집단을 발생시킬 수 있다.

구체적으로 설명하면 일차적으로 우리에게 시계열 자본자산가격모형의 추정 회귀방정식이 주어졌다고 하면 회귀직선의 추정 기울기도 주어진다. 이 추정 기울기 하에서 시장수익률의 크기와 안전수익률의 크기가 주어지면 문제의 위험자산수익률의 한 개의 관측치를 이용하여 문제의 위험자산수익률의 독립적인 무한개의 수익률을 확률적으로 발생시킬 수 있다. 그러면 이 무한개의 수익률을 가지고 다시 추정한 회귀방정식의 추정 (estimated) 기울기는 실제(true) 기울기에 접근할 것이다.

이 논문에서 사용하는 자본자산가격모형은 하나의 특수한 예에 불과하며 자본자산가격을 추정하는 것이 목적이 아니라는 점을 먼저 밝혀 두어야 한다. 본 연구의 궁극적 목표는 본 논문의 방법을 일반적인 회귀방정식에 적용하는 것이 될 것이다.

II. 회귀직선

선형 회귀방정식은 일반적으로 피설명변수를 다음과 같이 다수의 설명변수와 하나의 오차로 표현한다.

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t \quad (1)$$

$$t = 1, 2, \dots, T$$

2) Huang and Litzenberger (1988), p. 311.

설명변수 x_{tk} ($k = 1, 2, \dots, K$)는 반복적으로 고정된(fixed repeated) 변수로서 다공선성이 없고(non-multicollinearity), 오차 ε_t 의 평균은 영이며, 유한의 분산을 갖는다고 가정한다. 여기에 추가하여 오차의 분산은 균등할 수 있고(homoskedastic), 불균등할 수도 있다(heteroskedastic). 또한 오차는 서로 독립일 수 있고(independent), 자기상관을 가질 수도 있다(autocorrelated). 이상은 선형회귀분석의 통상의 가정이며 아래에서 그 이외의 가정은 필요하지 않다.

현재로서는 다변량 선형회귀방정식 (1)의 형태로는 피설명변수의 모집단을 복제할 수 없으므로 복제할 수 있는 회귀방정식의 '기본형태'를 찾아야 한다. 식 (1)은 다수의 설명변수인 다변량 선형회귀 방정식(multiple linear regression equation)인데 등호의 오른쪽은 오차항 ε_t 와 설명변수 부분 $Ey_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_K x_{tK}$ 로 구성되어 있다. 즉,

$$y_t = Ey_t + \varepsilon_t, \quad t = 1, 2, \dots, T$$

오차항 ε_t 는 불확실한 부분인데 대하여 설명부분 Ey_t 는 확실 부분이다. 확실 부분 Ey_t 는 스칼라이므로 적당한 두 변수 f_t 와 z_t 의 일차선형 $Ey_t = (1 - \beta_y)f_t + \beta_y z_t$ 로 표현하는 것이 항상 가능하다. 따라서 다음과 같이 수학적으로 동등하게 쓸 수 있다.

$$y_t = (1 - \beta_y)f_t + \beta_y z_t + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (2)$$

여기서 오차에 대한 가정은 식 (1)과 동일하며 분할 $(1 - \beta_y)f_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_K x_{tK}$ 와 $\beta_y z_t = \beta_{k+1} x_{tk+1} + \beta_{k+2} x_{tk+2} + \dots + \beta_K x_{tK}$ 가 성립하는 상수의 가중치 β_y , 설명변수 f_t 와 z_t 가 존재한다. 반복되는 비확률 설명변수 x_{tk} 와 오차 ε_t 사이에 직교(orthogonality)가 성립하므로 새로운 설명변수 z_t 와 오차 ε_t 사이에도 직교가 성립한다.

식 (2)는 오차 ε_t 를 제외하면 비확률변수 f_t 와 z_t 의 '산술평균'의 형태로 표현되어 있다. 산술평균 형태 (2)는 회귀방정식 (1)과 수학적으로 동일(mathematically equivalent)하지만 아래에서 숨겨진 정보공간을 찾아

모집단을 복제하는데 '기초형태'임이 드러나므로 식 (2)의 산술평균 형태를 배아방정식(胚芽方程式, embryonic equation)이라고 부를 수 있다. 그러나 회귀방정식 (1)의 추정 결과를 일대 일로 배아방정식 (2)로 전환할 때 어려운 점은 $\beta_{k+1}x_{1k+1} + \beta_{k+2}x_{1k+2} + \dots + \beta_Kx_{1K}$ 의 추정치 덕분에 β, z_i 의 추정치는 알 수 있지만 가중치 β_y 와 비확률변수 z_i 의 정체를 분리할 수 없고, 마찬가지로 $(1-\beta_y)f_i$ 의 추정치는 알 수 있지만 f_i 를 알 수 없다는 점이다.

그러나 처음부터 배아방정식 형태 (2) 자체로서 추정이 가능하여 가중치와 비확률변수의 정체를 알 수 있는 대표적인 예로서 시계열 자본자산 가격모형(capital asset pricing model)이 존재한다. 아래에서 이 모형을 예로 삼겠다. 이 모형의 회귀방정식은 임의의 위험자산 $j=1, 2, \dots, J$ 의 수익률을 다음과 같은 일반 형태로 표현하면 식 (2)와 동일한 형태이다.³⁾

$$R_{jt} = \beta_{jm}R_{Mt} + (1-\beta_{jm})R_{Ft} + \varepsilon_{jt}, \quad t=1, 2, \dots, T \quad (3)$$

여기서 t 는 시간, R_j 는 j 자산의 수익률, R_M 은 시장포트폴리오 수익률, R_F 는 안전자산수익률을 가리킨다. 이 가운데 안전자산수익률 R_F 는 고정이라고 가정한다. 오차에 대해 통상적인 가정을 따른다. 아래에서 부호의 복잡성을 피할 목적으로 미리 부호를 단순화하여 $R_j = R$, $R_M = M$, $R_F = F$, $\varepsilon_j = \varepsilon$, $\beta_{jm} = \beta$ 로 줄이겠다. 그러면 식 (3)은

$$R_t = \beta M_t + (1-\beta)F + \varepsilon_t, \quad E\varepsilon_t = 0 \quad (4)$$

로 고쳐 표현할 수 있다. β 는 추정치가 아니고 모르는 실제 모수(母數, parameter)이며 기울기 역할을 한다. $(1-\beta)F$ 는 절편이므로 상수로 취급한다. 따라서 설명변수의 정보는 반복되는 비확률 열벡터(column vector)

3) Huang and Litzenberger (1988), p. 311의 식 (10.13.3)이다.

$$M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_T \end{pmatrix} \quad (5)$$

이다. 여기서 M_t ($t = 1, 2, \dots, T$)는 스칼라이다. 이에 대한 피설명변수의 정보는 '현실적으로는' 보통 열벡터(column vector)

$$R_{\bullet} = \begin{bmatrix} R_{1\bullet} \\ R_{2\bullet} \\ \vdots \\ R_{T\bullet} \end{bmatrix}$$

일 뿐이다. 여기서 $R_{t\bullet}$ ($t = 1, 2, \dots, T$) 역시 스칼라이다. 그러나 '이론적 또는 개념적으로는' 피설명변수의 정보는 행렬(matrix)

$$R = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{T1} & R_{T2} & \cdots & R_{TN} \end{pmatrix} = (R_1, R_2, \dots, R_N) \quad (6)$$

이라고 가정하는 것이 보통이다. 물론 자료가 이산적인(discrete) 경우이다.

행렬 (6)의 정보 R_{tn} ($t = 1, 2, \dots, T$, $n = 1, 2, \dots, \infty$)이 모두 주어지는 경우 이론적인 회귀계수는 말할 것도 없이

$$\beta = \frac{\sum_{n=1}^{\infty} \sum_{t=1}^T (R_{tn} - \bar{R})(M_t - \bar{M})}{\sum_{n=1}^{\infty} \sum_{t=1}^T (M_t - \bar{M})^2} \quad (7)$$

이다.

그러나 현실에서 실현된 피설명변수의 정보는 행렬 (6) 가운데 하나의 열벡터의 정보뿐이므로 그것을 편의상 행렬 (6)의 첫째 열(the first column)

$$R_1 = \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{T1} \end{pmatrix}$$

이라고 하자. 다른 열(another column)이어도 상관없다. 우리는 아래에서 피설명변수 행렬 (6)의 실현된 첫째 열벡터 R_1 에서 행렬 (6) 가운데 실현되지 않은 나머지 열벡터 R_2, R_3, \dots, R_n 을 찾아내고 열벡터 R_n 과 열벡터 R_{n-1} 사이의 관계가 확률적인 관계라는 특성까지 밝혀낸다. 앞서 언급한 대로 행렬 (6)의 각 행벡터(row vector)는 독립적이고 동일한 분포를 가질 수도 있고 그렇지 않을 수도 있다. 말을 바꾸면 각 행(row)의 오차 벡터가 독립적이고 동일한 분포를 가질 수도 있고 그렇지 않을 수도 있다.⁴⁾ 이에 대하여 선형회귀분석에서 각 열벡터(column vector) 사이에 확률적 특성은 추정에 전혀 영향을 미치지 않기 때문에 현시적으로나 암묵적으로 어떠한 가정도 하지 않고 있는 실정이다.⁵⁾ 과분한 탓인지 열벡터 사이의 확률적 특성에 대해서는 문헌조차 없는 실정이다. 따라서 행렬 (6)의 각 열벡터도 서로 독립적이고 동일한 분포를 가질 수도 있고 그렇지 않을 수도 있으며 다만 확률적인 관계이면 족하다.

먼저 우리의 출발점으로써 고정적으로 반복되는 설명변수 M 에 대하여 실현된 첫째 열 R_1 의 피설명변수의 회귀방정식을 추정하는 것으로 시작한다. 즉, 두 개의 열벡터

$$M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_T \end{pmatrix}, \quad R_1 = \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{T1} \end{pmatrix}$$

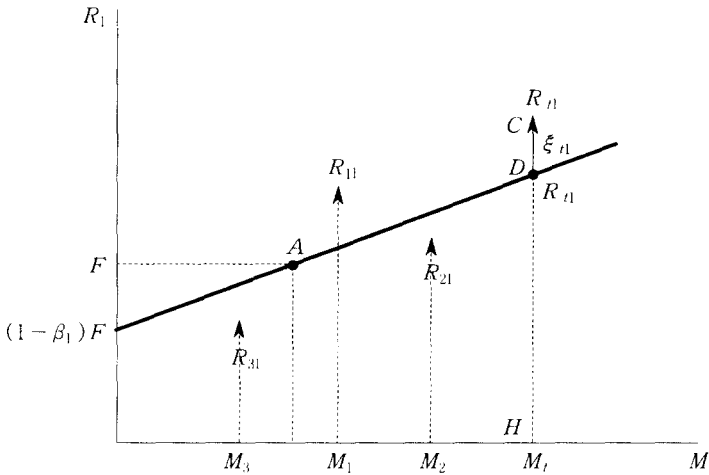
의 정보만으로 선형회귀방정식(linear regression equation)을 추정하면

$$\begin{aligned} \text{회귀방정식} \quad R_{t1} &= \hat{\beta}_1 M_t + (1 - \hat{\beta}_1) F + \hat{\varepsilon}_{t1}, \\ E \hat{\varepsilon}_{t1} &= 0, \quad t = 1, 2, \dots, T \end{aligned} \quad (8)$$

4) 사실상 자본자산가격모형은 오차의 독립적이고 동일한 분포를 가정하지 않는다.

5) 선형회귀분석의 다섯 가지 가정에 오차 열벡터의 독립성에 대해서는 아무 언급이 없다.

그림 1 회귀직선



이다. 여기서 모자(hat)는 모수의 추정치를 가리킨다. 회귀방정식 (8)의 회귀계수는

$$\text{회귀계수} \quad \hat{\beta}_1 = \frac{\sum_{t=1}^T (R_{t1} - \bar{R}_1)(M_t - \bar{M})}{\sum_{t=1}^T (M_t - \bar{M})^2} \quad (9)$$

이다. 회귀계수 $\hat{\beta}_1$ 은 모수 β 의 여러 추정치 가운데 하나이다. <그림 1>에 각각의 관측치의 좌표가 점 A, B, C, D, E로 표시되어 있는데 이 가운데 문제의 R_{t1} 의 좌표는 점 $C = (M_t, R_{t1})$ 이다.

회귀방정식 (8)에서 표본과 오차의 차이를 추정치 $\hat{R}_{t1} = R_{t1} - \hat{\varepsilon}_{t1}$ 로 정의하면 식 (8)은 기울기가 $\hat{\beta}_1$ 이고 절편이 $(1 - \hat{\beta}_1)F$ 인 회귀직선으로 표현될 수 있다.

$$\text{회귀직선} \quad \hat{R}_{t1} = \hat{\beta}_1 M_t + (1 - \hat{\beta}_1)F, \quad t = 1, 2, \dots, T \quad (10)$$

이다. <그림 1>에서 회귀직선이 45도선과 교차하는 점의 좌표는 $A = (F, F)$ 인데 이 좌표 역시 주어져 있다. 가로축 상에 점 H의 좌표는

$H = (M_t, 0)$ 이다. 점 H 에서 수직으로 세운 직선이 회귀직선과 만나는 곳이 점 D 이고 이 점의 좌표는 $D = (M_t, \hat{R}_{t1})$ 이다. <그림 1>에서 편 의 상 $M_t > F$ 을 상정하였다. 회귀직선의 식 (10)에서 계수 $\hat{\beta}_1$ 는 평균치 \hat{R}_{t1} 과 절편 $(1 - \hat{\beta}_1)F$ 를 통과하는 회귀직선의 기울기이다. 즉

$$\hat{\beta}_1 = \frac{\hat{R}_{t1} - F}{M_t - F}, \quad t = 1, 2, \dots, T \quad (11)$$

이다. 기울기 식 (11)은 회귀직선 (10)의 또 하나의 표현에 불과하다.

III. 관측직선

<그림 2>는 <그림 1>의 재생이지만 추가적인 정보를 포함하고 있다. <그림 2>에서 설명변수가 $M = M_t$ 로 주어져서 관측치 크기가 R_{t1} 일 때 추정오차 $\hat{\varepsilon}_{t1}$ 은 선분길이 $CD = \|R_{t1} - \hat{R}_{t1}\|$ 로 측정된다. 여기서 실제 관측치 R_{t1} 의 좌표 $C = (M_t, R_{t1})$ 과 좌표 $A = (F, F)$ 를 통과하는 직선의 기울기는

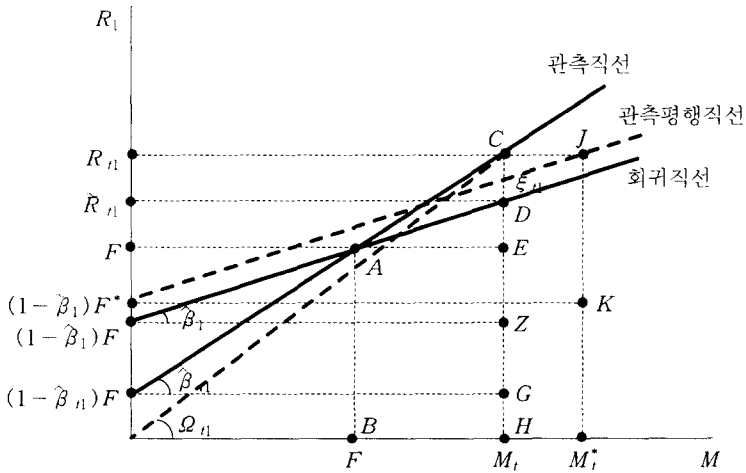
$$\begin{aligned} \hat{\beta}_{t1} &= \frac{R_{t1} - F}{M_t - F} \\ &= \frac{\hat{\varepsilon}_{t1} + \hat{R}_{t1} - F}{M_t - F} \\ &= \hat{\beta}_1 + \frac{\hat{\varepsilon}_{t1}}{M_t - F}, \quad t = 1, 2, \dots, T \end{aligned} \quad (12)$$

이다. 이 기울기는 확률변수인데 매기 알려진 관측치 (M_t, F, R_{t1}) 에서 구할 수 있으므로 이미 우리에게 알려져 있다. 확률변수 (12)를 재정리하면

$$\text{관측직선} \quad R_{t1} = \hat{\beta}_{t1}M_t + (1 - \hat{\beta}_{t1})F, \quad t = 1, 2, \dots, T \quad (13)$$

이다. 식 (13)을 식 (10)의 회귀직선(regression line)과 비교하기 위하여

그림 2 관측직선



주 : $R_{1t} = \hat{R}_{1t} + \xi_{1t}$, $\xi_{1t} = (\beta_{1t} - \hat{\beta}_1)(M_t - F)$, $\Omega_{1t} = \frac{R_{1t}}{M_t} = \frac{\hat{\beta}_{1t}}{\hat{\alpha}_{1t}}$

편의상 관측직선(observation line)이라고 부르겠다. $\hat{\beta}_{1t}$ 은 관측직선의 기울기이므로 확률계수이다. 하나의 회귀직선 (10)에 대하여 T 개의 관측직선이 존재한다. 각각의 관측치 $R_{1t} (t = 1, 2, \dots, T)$ 에 대하여 각각 하나씩 확률기울기 $\hat{\beta}_{1t}$ 가 존재하고 각각 하나씩 확률절편 $(1 - \hat{\beta}_{1t})F$ 가 존재하기 때문이다. 즉, $\hat{\beta}_{1t}$ 는 확률계수이고 따라서 관측직선 (13)은 회귀방정식 (8)의 동일한 내용을 다른 형태로 표현한 Hildreth-Houck (1968)의 확률모수 방정식(random parameter equation)의 특별한 예에 속한다.⁶⁾

회귀계수 (11)과 확률계수 (12)의 중요한 차이는 $M_t = F$ 인 좌표

6) 확률모수 회귀방정식 가운데 가장 많이 사용되는 대표적인 예로서 Hildreth-Houck (1968)를 참조할 수 있다. 이들의 특별한 형태가 관측직선 (13)이다. 구체적으로 표현하면 이들은 비확률모수 방정식 $Y_t = \beta X_t + \gamma Z_t + \epsilon_t$ 를 확률모수 방정식 $Y_t = \beta_t X_t + \gamma_t Z_t$ 로 전환하였는데 $\epsilon_t = X_t \mu_t + Z_t \nu_t$, $\beta_t = \beta + \mu_t$, $\gamma_t = \gamma + \nu_t$ 이다. 여기에 $Z_t = Z$, $\gamma = 1 - \beta$, $\gamma_t = 1 - \beta_t$ 로 대체하면 $\nu_t = -\mu_t$ 가 된다. 따라서 $\epsilon_t = (\beta_t - \beta)(X_t - Z)$ 이 되는데 이 결과를 비확률모수 방정식에 대입하면 관측직선 (13)이 되며 오차항 ϵ_t 는 오차항의 정의 (17)과 일치한다.

$A = (F, F)$ 에서 드러난다. 임의의 시점이 $t = t'$ 일 때 $F = M_{t'}$ 이면 $F = M_{t'} = \hat{R}_{t',1}$ 이지만 그렇더라도 좌표 A 에서 식 (11)에 의하면 $\hat{\beta}_{t',1}$ 은 회귀계수로서 불변이다. 그러나 좌표 A 에서 $F = M_{t'} \neq R_{t',1}$ 이므로 식 (12)에 의하면 확률계수 $\hat{\beta}_{t',1}$ 은 무한대가 된다.

정리 1 임의의 시점 $t = t'$ 일 때 $F = M_{t'} \neq R_{t',1}$ 이 되면 확률계수 $\hat{\beta}_{t',1}$ 은 무한대가 된다.

IV. 기울기

회귀직선의 비확률 기울기 (11)과 관측직선의 확률 기울기 (12)를 비교하면 $\hat{R}_{t,1} \neq R_{t,1}$ 이므로

$$\hat{\beta}_{t,1} \neq \beta_{t,1} \tag{14}$$

이 됨을 알 수 있다. <그림 2>를 보면 식 (12)는 새로운 절편 $(1 - \hat{\beta}_{t,1})F$ 에서 시작하여 좌표 $A = (F, F)$ 를 거쳐 실제 관측치 $R_{t,1}$ 의 좌표 $C = (M_t, R_{t,1})$ 을 통과하는 관측직선의 기울기 $\hat{\beta}_{t,1}$ 을 의미한다. 가로축 상의 M_t 의 좌표 $H = (M_t, 0)$ 에서 수직으로 세운 직선이 세로축 상의 새로운 절편 $(1 - \hat{\beta}_{t,1})F$ 에서 수평으로 연결한 직선과 만나는 좌표를 $G = (M_t, (1 - \hat{\beta}_{t,1})F)$ 라고 표기하면 관측직선의 확률 기울기 $\hat{\beta}_{t,1}$ 는 기하학적으로 선분길이 $CG = R_{t,1}G$ 와 선분길이 $OH = OM_t$ 의 비율이다. 즉, 관측직선의 확률 기울기 (12)는 확률계수로서 기하학적으로는

$$\hat{\beta}_{t,1} = \frac{CG}{OH} \tag{15}$$

로 정의할 수 있다.

$R_{t,1}$ 의 관측직선의 식 (13)을 이용하면 $R_{t,1}$ 의 회귀방정식 (8)은 다음

과 같이 수학적으로 동등하게(mathematically equivalent) 표현될 수 있다.

회귀방정식

$$R_{t1} = \hat{\beta}_1 M_t + (1 - \hat{\beta}_1)F + (\hat{\beta}_{t1} - \hat{\beta}_1)(M_t - F) \\ t = 1, 2, \dots, T \quad (16)$$

식 (16)을 식 (8)과 비교하면 추정오차의 길이 $\hat{\varepsilon}_{t1} = CD = \|R_{t1}\hat{R}_{t1}\|$ 을 '사후적으로' 다음과 같이 정의할 수 있다.⁷⁾

추정오차

$$\hat{\varepsilon}_{t1} = (\hat{\beta}_{t1} - \hat{\beta}_1)(M_t - F) \\ = CD, \quad t = 1, 2, \dots, T \quad (17)$$

따라서 다음의 정리가 성립한다.

정리 2

추정오차는 관측직선의 확률 기울기와 회귀직선의 비확률 추정 기울기의 차이와 두 개 설명변수의 차이의 곱으로 정의된다.

<정리 2>는 아래에서 계속 중요한 역할을 하게 되므로 추정오차 (17)이 기하학적 표현과도 일치한다는 사실을 추가적으로 설명하겠다. <그림 2>에 새로운 좌표 $E = (M_t, F)$ 를 추가하면 추정오차 $\hat{\varepsilon}_{t1}$ 는 $CD = \|R_{t1}\hat{R}_{t1}\|$ 의 선분길이인데 이것을 다시 두 개의 선분길이의 차이 $CD = CE - DE$, 즉 $\|R_{t1}\hat{R}_{t1}\| = \|R_{t1}E\| - \|\hat{R}_{t1}E\|$ 로 나눌 수 있다. 그런데 구성 요소 가운데 $\|R_{t1}E\| = \hat{\beta}_{t1}(M_t - F)$ 이고 $\|\hat{R}_{t1}E\| = \hat{\beta}_1(M_t - F)$ 이므로 오차 $\hat{\varepsilon}_{t1}$ 은 $CD = (\hat{\beta}_{t1} - \hat{\beta}_1)(M_t - F)$ 가 되므로 <정리 2>는 기하학적으로도 성립한다.

식 (17)은 추정오차가 설명변수와 기울기의 차이에 의해서 결정되는 것

7) 오차항의 정의 (17)은 각주 6)에서 유도한 결과와 일치한다.

으로 표현되었지만 이론적인 측면에서 '사전적으로는' 오차가 무엇에 의해 결정되는지에 대해서는 알려진 바가 없다. 식 (17)이 의미하는 것은 다만 '사후적인' 정의뿐이다. 마치 회귀분석에서 최소자승법에 의해 기울기의 계수가 추정되면 추정오차를 '사후적으로' 설명변수와 추정계수에 의해 추정하여 정의하는 것과 같은 이치이다. 특히 식 (17)은 기울기의 차이로 오차를 정의하고 있다.

V. 파생기울기

피설명변수 관측치 R_{1t} 의 관측직선의 확률 기울기 (12)에서 다음의 계수를 파생적으로 정의한다.

$$\hat{\alpha}_{1t} = \hat{\beta}_{1t} \frac{M_t}{R_{1t}}, \quad t = 1, 2, \dots, T \quad (18)$$

이 계수의 대수적 의미는 다음과 같다. 관측직선의 식 (13)에서 관측치 R_{1t} 은 두 부분으로 구성되어 있다. 첫째 부분 $\hat{\beta}_{1t} M_t$ 와 둘째 부분 $(1 - \hat{\beta}_{1t})F$ 이다. 관측치 R_{1t} 에 대한 첫째 부분 $\hat{\beta}_{1t} M_t$ 의 비율이 $\hat{\alpha}_{1t}$ 이고, 둘째 부분 $(1 - \hat{\beta}_{1t})F$ 의 비율은 $1 - \hat{\alpha}_{1t}$ 이다. 즉, 관측직선 (13)의 양변을 관측치 R_{1t} 으로 나눈 결과가

$$\begin{aligned} 1 &= \hat{\beta}_{1t} \frac{M_t}{R_{1t}} + (1 - \hat{\beta}_{1t}) \frac{F}{R_{1t}} \\ &= \hat{\alpha}_{1t} + (1 - \hat{\alpha}_{1t}) \end{aligned} \quad (19)$$

이므로 $\hat{\alpha}_{1t}$ 을 식 (18)처럼 정의할 수 있다. 이 계수의 기하학적 의미도 <그림 2>를 빌려 설명할 수 있다. 먼저 <그림 2>에서 원점을 출발하여 관측치 R_{1t} 을 통과하는 점선의 기울기의 정의는

$$\Omega_{1t} = \frac{CH}{OH} = \frac{R_{1t}}{M_t} \quad (20)$$

이다. 이것 역시 확률변수이다. 따라서 정의 (18)은

$$\hat{\alpha}_{t1} = \hat{\beta}_{t1} \frac{M_t}{R_{t1}} = \frac{\hat{\beta}_{t1}}{\Omega_{t1}} = \frac{CG}{OH} \frac{OH}{CH} = \frac{CG}{CH} \quad (21)$$

이다. $\hat{\alpha}_{t1}$ 은 대수적으로는 두 개의 기울기 $\hat{\beta}_{t1}$ 과 Ω_{t1} 의 비율이며, 기하학적으로는 선분길이 CH 에 대한 선분길이 CG 의 비율이지만, 기울기 $\hat{\beta}_{t1}$ 에서 파생되었으므로 $\hat{\beta}_{t1}$ 의 파생기울기라고 부르자. 기울기 $\hat{\beta}_{t1}$ 이 확률변수이므로 파생기울기 $\hat{\alpha}_{t1}$ 역시 확률변수이다.

이 파생기울기를 사용하여 식 (18)을 재정리하면

$$\frac{\hat{\beta}_{t1}}{R_{t1}} = \frac{\hat{\alpha}_{t1}}{M_t} \quad (22)$$

이다. 이번에는 식 (18)을 식 (13)에 대입한 결과를 정리하면

$$\frac{1 - \hat{\beta}_{t1}}{R_{t1}} = \frac{1 - \hat{\alpha}_{t1}}{F} \quad (23)$$

을 얻는다. 식 (22)에 식 (23)을 더하면 직각쌍곡선

$$\frac{1}{R_{t1}} = \frac{\hat{\alpha}_{t1}}{M_t} + \frac{1 - \hat{\alpha}_{t1}}{F}, \quad t = 1, 2, \dots, T \quad (24)$$

를 얻는다. 직각쌍곡선 (24)에서 R_{t1} 은 M_t 와 F 의 '조화평균'이고 그 가중치가 파생기울기 $\hat{\alpha}_{t1}$ 이다. 관측직선 (13)에서 R_{t1} 이 M_t 와 F 의 '산술평균'이며 그 가중치가 기울기 $\hat{\beta}_{t1}$ 인 것과 대조적이다.

VI. 상위공간과 하위공간

실현된 한 개의 표본에서 실현되지 않은 나머지 표본을 확률적으로 드러나게 할 수 있는 근본 이치는 숨겨진 제2공간의 발견에 있다. 여기서

추가적인 정보를 얻을 수 있기 때문이다. 이것은 머리말에서 소개한 비유에서 밝혔듯이 두 개의 핵산 가닥, 두 개의 레이저 광선, 두 개의 음자리의 발견이 추가적인 정보를 제공하여 복제를 가능하게 만든 것에 비유할 수 있다.

이제 변수전환 방식 $r_{11}R_{11} = 1$, $m_iM_i = 1$, $fF = 1$ 에 의해 새로운 소문자 변수 r_{11} , m_i , f 를 정의할 수 있다. 즉, 대문자변수와 소문자변수는 서로 역수의 관계이다. 지금까지 모든 변수를 대문자로 표현하였으므로 상위공간의 방정식체계라고 이름할 수 있다. 대문자를 변수 전환하면 소문자로 표현되는 하위공간의 함수체계를 얻을 수 있다. 공간의 이름은 아무래도 상관없고, 다만 상위공간은 높은 음자리에 해당하고, 하위공간은 낮은 음자리에 해당한다.

변수전환은 단순한 수학적 개념일 수 있다. 허수(imaginary number)는 경제학적으로 의미가 부여되지 않지만 경기변동을 설명하는 수학적 도구(mathematical instrumental)로 사용되듯이 변수전환도 그와 같은 차원에서 단지 수학적 도구로만 생각할 수 있다.⁸⁾ 그러나 다행스러운 것은 역의 변수전환은 단순한 수학적 도구에 그치지 않고 경제학적으로도 의미가 있다는 점이다. 대표적인 예가 Fisher의 유통속도 V 와 Marshall의 k 사이의 관계 $Vk = 1$ 이다. 사회적 후생함수의 Negishi 구매력 계수 θ 와 개인 소득의 그림자가격 λ 사이에도 $\theta\lambda = 1$ 의 관계가 성립한다. Leontieff의 생산함수 $Q = \sigma K$ 에서는 자본의 평균생산 σ 는 가속도 v 의 역수이므

8) 물리학에서는 실시간(real time)에 대하여 허시간(imaginary time)을 사용하여 물리적 현상을 설명한다. 실시간 t 와 허시간 T 사이의 관계는 직각이다. 허시간은 수학적 구성물이기 때문에 물리적인 현현(顯現)을 필요로 하지 않는다. "… 양자이론이 시간과 공간을 어떻게 형성하는지 기술하기 위해서는 허시간이라는 개념을 도입할 필요가 있다. 허시간은 … 훌륭하게 정의된 수학적 개념이다. 그것은 허수라고 불리우는 것으로 측정되는 시간이다. … 물리적인 현현을 필요로 하지 않는다. … 허시간은 실시간에 대하여 직각방향이기 때문에 실시간에 비해서 훨씬 풍부한 가능성의 범위를 가진다. 실시간은 오직 시작과 끝을 가지거나 원을 그리고 돌 수 있을 뿐이다. 시간이 형태를 가지는 것은 이러한 허수적인 의미에서이다." (스티븐 호킹 (김동광 옮김), 『호두껍질 속의 우주』, pp. 57~60.) "… 우리가 느끼는 실시간의 수직방향에 위치하는 허시간이 있다. 실시간에 있어서의 우주의 역사는 허시간에서의 역사를 결정하며, 그 역도 성립한다. 그러나 두 종류의 역사는 매우 다를 수 있다. 특히 우주는 허시간에서 시작과 끝을 가질 필요가 없다. 허시간은 공간 속에서 또 다른 방향처럼 움직인다. 따라서 허시간에서의 우주의 역사는 공, 비행기, 그리고 안장의 표면처럼 휘어져 있지만 2차원이 아닌 4차원인 것으로 생각될 수 있다." (스티븐 호킹 (김동광 옮김), 『호두껍질 속의 우주』, p. 83).

로 $\sigma v = 1$ 이다. 이 논문에서 분석의 예로서 택한 자본자산가격이론에서는 자산의 수익 Y 는 자산의 가격 P 에 수익률 R 을 곱한 것이므로 수익률 R 은 자산의 수익 Y (자본이득 + 배당금)를 가격 P 로 나눈 것이다. 즉, $Y = P \times R$ 이므로 $\frac{P}{Y} \times R = 1$ 이다. 여기서 $r = \frac{P}{Y}$ 로 표기하면 $rR = 1$ 이 되므로 소문자 r_{t1} , m_t , f 는 각각 자산의 가격-수익비율(asset price-returns ratio)을 가리킨다. 사실상 자본자산가격모형(capital asset pricing model)이라는 명칭에 합당하는 것은 자산의 수익률 사이의 관계가 아니라 자산가격 사이의 관계이다. 따라서 자산의 수익률이 아니라 자산가격 사이의 관계를 연구하는 것이 오히려 원래의 명칭에 어울리는 이론이 될 것이다. 이런 점에서 이 논문이 다루는 자본자산가격모형에서는 변수전환이 의의를 가지며, 이에 기초하여 상위공간(upper space)을 수익공간(return space), 하위공간(lower space)을 가격공간(price space)이라고도 부를 수 있다.

변수전환 $r_{t1}R_{t1} = 1$, $m_t M_t = 1$, $fF = 1$ 을 상위공간의 직각쌍곡선(24)에 적용하면

$$r_{t1} = \hat{\alpha}_{t1} m_t + (1 - \hat{\alpha}_{t1})f$$

인데 이것을 재정리하면

하위공간 회귀방정식

$$r_{t1} = \hat{\alpha}_1 m_t + (1 - \hat{\alpha}_1)f + \hat{u}_{t1}, \quad t = 1, 2, \dots, T \quad (25)$$

의 방정식을 얻게 된다. 여기서

하위공간 회귀오차

$$\hat{u}_{t1} = (\hat{\alpha}_{t1} - \hat{\alpha}_1)(m_t - f), \quad t = 1, 2, \dots, T \quad (26)$$

이다. 식 (25)는 상위공간의 직각쌍곡선 (24)와 수학적으로 동일(mathematically equivalent)하고 하위공간에서 표현한 것에 지나지 않지만, 하위공간의 회귀방정식의 형태를 갖추고 있다. 결과적으로 식 (25)는 식 (8)의 상위공간의 시계열 자본자산수익률(rate of return)을 하위공

간의 자산가격수익비율(price-return ratio)로 바꾸어 표현한 것에 불과하다. 이 같은 변수전환의 몇 가지 중요한 특성을 지적할 수 있다.

첫째, 상위공간의 식 (18)과 하위공간의 식 (25)~(26)에 공통적으로 나타나는 확률계수 $\hat{\alpha}_{11}$ 의 정체를 규명하기 위하여 식 (25)~(26)을

$$\hat{\alpha}_{11} = \frac{r_{11} - f}{m_1 - f} \tag{27}$$

로 고쳐 쓸 수 있다. 식 (27)은 <그림 3>의 하위공간 좌표 (r_1, m) 에서 새로운 절편 $(1 - \hat{\alpha}_{11})f$ 와 r_{11} 을 통과하는 하위공간 관측직선의 확률기울기 $\hat{\alpha}_{11}$ 을 의미한다. 특기할 점은 확률계수 $\hat{\alpha}_{11}$ 이 하위공간에서 (27)에 의해 기울기로 정의되지만 상위공간에서는 식 (19)에 의해 파생기울기로 정의된다는 것이다.

둘째, 하위공간의 추정오차 식 (26)은 상위공간의 추정오차 식 (17)과 동일한 형태이므로 <정리 2>에 부합한다. 즉, 기울기의 차이와 설명변수의 차이의 곱으로 되어 있다.

셋째, 오차 (26)에서 $\hat{\alpha}_{11}$ 은 확률계수이고 $\hat{\alpha}_1$ 은 비확률 회귀계수이므로 식 (14)처럼

$$\hat{\alpha}_{11} \neq \hat{\alpha}_1 \tag{28}$$

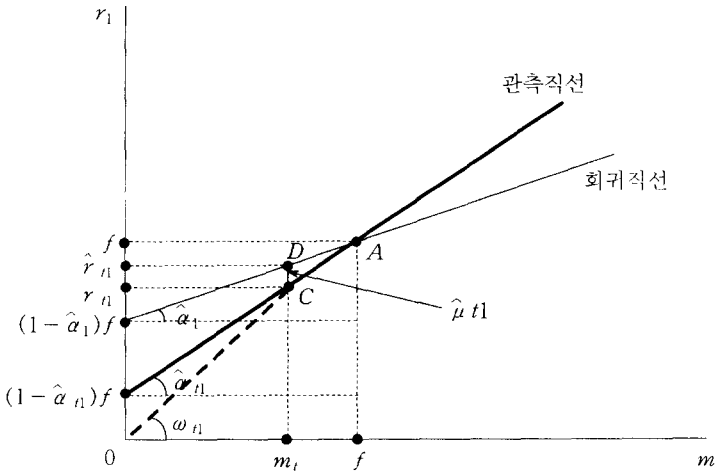
이다. 따라서

$$\hat{u}_{11} \neq 0 \tag{29}$$

이 성립한다. 거꾸로 하위공간의 회귀방정식 (25)의 추정오차가 영이 아니므로 식 (28)이 성립한다.

넷째, 상위공간의 오차 식 (17)이 기하학적으로 표현되듯이 <그림 3>의 하위공간의 좌표 (r_1, m) 에서 오차 식 (26) 역시 기하학적으로 $\hat{u}_{11} = CD$ 이며 그 부호는 $\hat{\varepsilon}_{11}$ 의 부호와 반대이다(이것은 식 (63)에서 대수적으로 증명됨).

그림 3 하위공간의 회귀직선과 관측직선



주 : $\hat{\mu}_{n1} = (\hat{\alpha}_{n1} - \hat{\alpha}_1)(m_t - f)$, $\tau_{n1} = \hat{\tau}_{n1} - \hat{\mu}_{n1}$.

$$\omega_{n1} = \frac{\tau_{n1}}{m_t} = \frac{M_t}{R_{n1}} = \frac{\hat{\alpha}_{n1}}{\hat{\beta}_{n1}} = \frac{1}{\Omega_{n1}}$$

다섯째, 상위공간의 직각쌍곡선 (24)는 하위공간에서는 회귀방정식 (25)로 역할이 바뀐다.

여섯째, <그림 3>의 하위공간에서 원점과 관측치 r_{1t} 을 연결하는 점선의 기울기를 ω_{1t} 이라고 표기하면

$$\omega_{1t} = \frac{r_{1t}}{m_t} = \frac{M_t}{R_{1t}} = \frac{\hat{\alpha}_{1t}}{\hat{\beta}_{1t}} = \frac{1}{\Omega_{1t}} \tag{30}$$

이 된다. 따라서 식 (30)에서

$$\omega_{1t} \Omega_{1t} = 1 \tag{31}$$

이 성립한다. 이것은 상위공간과 하위공간 사이에 또 하나의 변수전환의 역관계를 의미한다.

일곱째, 식 (30)에 의하여 하위공간에서

$$\hat{\beta}_{11} = \frac{\hat{\alpha}_{11}}{\omega_{11}} \quad (32)$$

이 성립한다. 즉, 상위공간의 기울기는 하위공간에서는 파생기울기가 된다. <정리 3>이 성립한다.

정리 3 상위공간의 파생기울기는 하위공간에서 기울기로 역할이 바뀌고, 상위공간의 기울기는 하위공간의 파생기울기로 전환된다.

VIII. 관측평행직선

<그림 2>의 상위공간의 추정오차 $\hat{\varepsilon}_{11} = CD$ 를 다시 생각한다. 이것은 선분길이 CD 로 측정되는데 이 길이는 좌표 C 에서 좌표 D 까지의 수직 길이이다. 점 D 의 좌표는 $D = (M, R_{11} - \hat{\varepsilon}_{11})$ 이다. 이번에는 좌표 C 에서 수직 대신 수평으로 수직길이와 동일한 수평길이 $CJ = CD = \hat{\varepsilon}_{11}$ 에 해당하는 새로운 좌표 J 를 선택한다. 새로운 좌표는 $J = (M_t + \hat{\varepsilon}_{11}, R_{11})$ 이다. 즉, 좌표 D 는 좌표 C 에서 M_t 는 그대로 둔 채 R_{11} 에서 오차 $\hat{\varepsilon}_{11}$ 을 뺀 좌표이고, 좌표 J 는 좌표 C 에서 R_{11} 을 그대로 둔 채 M_t 에서 오차 $\hat{\varepsilon}_{11}$ 을 더한 좌표이다. 그러면 회귀직선 (10)을 기울기 $\hat{\beta}_1$ 은 그대로 둔 채 평행이동시켜 새로운 절편 $(1 - \hat{\beta}_1)(F + \hat{\varepsilon}_{11})$ 과 새로운 좌표 $J = (M_t + \hat{\varepsilon}_{11}, R_{11})$ 을 통과하는 점선으로 만들 수 있다. 이 점선의 식은

회귀방정식

$$\begin{aligned} R_{11} &= \hat{\beta}_1 M_t + (1 - \hat{\beta}_1)F + \hat{\varepsilon}_{11} & (8) \\ &= \hat{\beta}_1 M_t + (1 - \hat{\beta}_1)F + [\hat{\beta}_1 \hat{\varepsilon}_{11} + (1 - \hat{\beta}_1) \hat{\varepsilon}_{11}] \\ &= \hat{\beta}_1 (M_t + \hat{\varepsilon}_{11}) + (1 - \hat{\beta}_1)(F + \hat{\varepsilon}_{11}) \\ &= \hat{\beta}_1 M_t^* + (1 - \hat{\beta}_1)F_t^*, \quad t = 1, 2, \dots, T \quad (\text{관측평행직선}) \quad (33) \end{aligned}$$

이다. 여기서 $M_t^* = M_t + \hat{\varepsilon}_{t1}$ 이고 $F_t^* = F + \hat{\varepsilon}_{t1}$ 인 새로운 확률변수이다. 회귀방정식 (8)에서 피설명변수 R_{t1} 의 확률부분은 추정오차 $\hat{\varepsilon}_{t1}$ 에서 기인한다. 이 추정오차 $\hat{\varepsilon}_{t1}$ 을 두 부분으로 분리할 수 있다. 즉, $\hat{\varepsilon}_{t1} = \hat{\beta}\hat{\varepsilon}_{t1} + (1-\hat{\beta})\hat{\varepsilon}_{t1}$ 이다. 각 부분을 각 설명변수에 더해 주면 각 설명변수는 확률변수가 되어 식 (33)이 된다. 식 (33)에서 피설명변수의 확률부분은 설명변수의 확률부분에서 기인한다. 따라서 직선의 식 (33)은 회귀방정식 (8)과 수학적으로 동일(mathematically equivalent)하지만 피설명변수의 추정치 \hat{R}_{t1} 대신 실제 관측치 표본 R_{t1} 을 통과하고 회귀 직선과 평행을 이루므로 R_{t1} 의 관측평행직선(observed parallel line)이라고 부르겠다. 그러면 하나의 회귀직선 (10)에 대하여 T 개의 관측평행 직선이 존재한다. 가로축의 점 L 의 좌표 $L = (M_t^*, 0)$ 에서 새로운 수직 길이 $JL = \|R_{t1}M_t^*\|$ 을 그릴 수 있다. 이 새로운 수직길이 JL 은 앞서의 수직길이 $CH = \|R_{t1}M_t\|$ 와 나란하며 $CH = JL$ 이다.

VIII. 정보발생에 있어서 두 단계 수학적 동등

관측평행직선 (33)은 두 부분으로 구성되어 있다. 첫째 부분은 $\hat{\beta}_1 M_t^*$ 이고, 둘째 부분은 $(1-\hat{\beta}_1)F_t^*$ 이다. 첫째 부분은 관측치 R_{t1} 의 일부분을 이루고, 둘째 부분은 관측치 R_{t1} 의 나머지 부분을 이룬다. 관측치 R_{t1} 에 대한 첫째 부분 $\hat{\beta}_1 M_t^*$ 의 비율이 $\hat{\alpha}_{t2}$ 이고, 둘째 부분 $(1-\hat{\beta}_1)F_t^*$ 의 비율은 $(1-\hat{\alpha}_{t2})$ 이다. 즉, 관측평행직선 (33)의 양변을 R_{t1} 으로 나누면

$$\begin{aligned} 1 &= \frac{\hat{\beta}_1 M_t^*}{R_{t1}} + \frac{(1-\hat{\beta}_1)F_t^*}{R_{t1}} \\ &= \hat{\alpha}_{t2} + (1-\hat{\alpha}_{t2}) \end{aligned} \quad (34)$$

이다. 그 결과 상위공간의 식 (11)의 기울기 $\hat{\beta}_1$ 에서 다음의 새로운 파생 기울기를 정의할 수 있다.

$$\hat{\alpha}_{i2} = \hat{\beta}_1 \frac{M_i^*}{R_{i1}} \quad (35)$$

식 (35)를 식 (18)과 비교하면 $(\hat{\alpha}_{i1} - \hat{\alpha}_{i2})R_{i1} = (\hat{\beta}_{i1} - \hat{\beta}_1)[(1 - \hat{\beta}_1)M_i + \hat{\beta}_1 F]$ 이므로

$$\hat{\alpha}_{i2} \neq \hat{\alpha}_{i1} \quad (36)$$

이다. 따라서 $\hat{\alpha}_{i2}$ 는 $\hat{\alpha}_{i1}$ 과 전혀 다른 새로운 확률계수이다. 확률계수인 까닭은 식 (35)의 분자와 분모가 모두 확률변수이기 때문이다. 식 (35)의 의미를 설명하기 위하여 <그림 2>의 절편 $(1 - \hat{\beta}_1)F_i^*$ 에서 수평으로 연결한 수평길이가와 가로축의 점 M_i^* 에서 수직으로 세운 수직길이가 만난 좌표를 $K = (M_i^*, (1 - \hat{\beta}_1)F_i^*)$ 라 정의한다. 식 (35)는 상위공간의 가로축의 한 점 M_i^* 에서 수직길이 JL 에 대한 길이 JK 의 비율로서 기울기 부분 JK 가 전체길이 JL 에서 차지하는 비율을 의미한다. 즉, 식 (35)는 $\frac{\hat{\beta}_1}{\hat{\alpha}_{i2}} = \frac{JK}{OL}$ 인데 이것은 다시 $\hat{\alpha}_{i2} = \frac{JK}{JL}$ 이다. 그러므로 식 (35)는 상위공간의 기울기 $\hat{\beta}_1$ 을 상위공간의 파생기울기 $\hat{\alpha}_{i2}$ 로 전환시키는 식이다. 뿐만 아니라 식 (35)는

$$\frac{\hat{\alpha}_{i2}}{M_i^*} = \frac{\hat{\beta}_1}{R_{i1}}$$

이며 이것을 식 (33)에 대입하여 정리하면

$$\frac{1 - \hat{\alpha}_{i2}}{F_i^*} = \frac{1 - \hat{\beta}_1}{R_{i1}}$$

가 되므로 이 둘을 더하여 상위공간에서 새로운 직각쌍곡선

$$\begin{aligned} \frac{1}{R_{i1}} &= \frac{\hat{\alpha}_{i2}}{M_i^*} + \frac{1 - \hat{\alpha}_{i2}}{F_i^*} \\ &= \frac{\hat{\alpha}_{i2}}{M_i + \hat{\varepsilon}_{i1}} + \frac{1 - \hat{\alpha}_{i2}}{F + \hat{\varepsilon}_{i1}}, \quad t = 1, 2, \dots, T \end{aligned} \quad (37)$$

을 얻을 수 있다. 식 (37)에 변수전환 $r_{11}R_{11} = 1$, $m_t M_t = 1$, $fF = 1$ 을 적용하여 하위공간에서 재정리하면

$$y_t = \hat{\alpha}_{12} m_t + (1 - \hat{\alpha}_{12}) f \tag{38}$$

이다. 여기서 피설명변수는

$$y_t = r_{11} \{ 1 + [\hat{\beta}_1 m_t + (1 - \hat{\beta}_1) f] \hat{\epsilon}_{11} \} \tag{39}$$

이므로 식 (38)을 다시 표현하면

$$\hat{\alpha}_{12} = \frac{y_t - f}{m_t - f} \tag{40}$$

이다. 식 (40)에 의하면 $\hat{\alpha}_{12}$ 는 설명변수 m_t 에 대한 피설명변수 y_t 의 확률계수가 되므로 식 (38)은 관측직선이다. 따라서 식 (38)을 다음과 같이 수학적으로 동등하게 표현할 수 있도록 하는 비확률 계수 $\hat{\alpha}_2$ 가 존재한다.

$$y_t = \hat{\alpha}_2 m_t + (1 - \hat{\alpha}_2) f + (\hat{\alpha}_{12} - \hat{\alpha}_2)(m_t - f) \tag{41}$$

확률계수 $\hat{\alpha}_{12}$ 는 확률변수이므로 비확률 기울기 $\hat{\alpha}_2$ 와 다르고, 또 하나의 비확률 기울기 $\hat{\alpha}_1$ 와 다르다. 즉

$$\hat{\alpha}_{12} \neq \hat{\alpha}_2, \quad \hat{\alpha}_{12} \neq \hat{\alpha}_1 \tag{42}$$

이다.⁹⁾ 더욱이 중요한 사실은

$$\hat{\alpha}_2 \neq \hat{\alpha}_1 \tag{43}$$

9) 식 (35)에 식 (22), 식 (17), 식 (26)을 대입하여 정리하면 $\hat{\alpha}_{12} = \hat{\alpha}_1 + \hat{\alpha}_1 \frac{\hat{\epsilon}_{11}}{M_t} + \frac{FM_t \hat{u}_{11} + F\hat{u}_{11} \hat{\epsilon}_{11} + \hat{\epsilon}_{11}}{F - M_t}$ 이므로 $\hat{\alpha}_{12} \neq \hat{\alpha}_1$ 이다.

이라는 점이다.¹⁰⁾ 식 (41)의 등호의 마지막 항은 <정리 2>에 의하여 새로운 오차항 \hat{u}_{i2} 가 되어야 한다. 즉

$$\hat{u}_{i2} = (\hat{\alpha}_{i2} - \hat{\alpha}_{i1})(m_i - f) \quad (44)$$

따라서 식 (41)은

$$y_i = \hat{\alpha}_{i2}m_i + (1 - \hat{\alpha}_{i2})f + \hat{u}_{i2} \quad (45)$$

이다. 이렇게 되면 등호의 왼쪽 항 y_i 는 r_{i2} 가 되어야 한다. 즉

$$y_i = r_{i1}\{1 + [\hat{\beta}_1 m_i + (1 - \hat{\beta}_1)f]\hat{\varepsilon}_{i1}\} = r_{i2} \quad (46)$$

이다. 종합하면 식 (38)은

하위공간의 회귀방정식

$$\begin{aligned} r_{i2} &= \hat{\alpha}_{i2}m_i + (1 - \hat{\alpha}_{i2})f + \hat{u}_{i2} \\ \hat{\alpha}_{i2} &\neq \hat{\alpha}_{i1}, \quad \hat{\alpha}_{i2} \neq \hat{\alpha}_{i1}, \quad \hat{\alpha}_{i2} \neq \hat{\alpha}_{i1}, \quad \hat{\alpha}_{i2} \neq \hat{\alpha}_{i1} \\ t &= 1, 2, \dots, T \end{aligned} \quad (47)$$

이어야 한다. 가장 중요한 사실은 식 (46)에서 $\hat{\varepsilon}_{i1} \neq 0$ 이므로

$$r_{i2} \neq r_{i1} \quad (48)$$

이라는 점이다. 즉, r_{i2} 는 r_{i1} 과 전혀 다른 새로운 표본이다. 이 표본은 추정치가 아니다. 회귀방정식 (47)에서 추정치 $\hat{\alpha}_{i2}m_i + (1 - \hat{\alpha}_{i2})f$ 와 오차 \hat{u}_{i2} 의 합으로 되어 있기 때문이다.

10) 식 (38)을 정리하면 $r_{i1}[\hat{\beta}_1 m_i + (1 - \hat{\beta}_1)f]\hat{\varepsilon}_{i1} = \hat{u}_{i2} - \hat{u}_{i1} + (\hat{\alpha}_1 - \hat{\alpha}_2)(m_i - f)$ 인데 여기에 식 (25)를 대입하면 $A\hat{\varepsilon}_{i1} + B\hat{\varepsilon}_{i1}^2 = \hat{u}_{i2} - \hat{u}_{i1} + (\hat{\alpha}_1 - \hat{\alpha}_2)(m_i - f)$ 이다. $A = \hat{\beta}_1^2 + \hat{\beta}_1(1 - \hat{\beta}_1)M_i f + \hat{\beta}_1(1 - \hat{\beta}_1)m_i F + (1 - \hat{\beta}_1)^2$ 와 $B = \hat{\beta}_1 m_i + (1 - \hat{\beta}_1)f$ 이다. 여기에 수학적 기대부호 E 를 적용하면 $E(\hat{\varepsilon}_{i1}) = E(\hat{u}_{i1}) = E(\hat{u}_{i2}) = 0$ 이므로 $BE(\hat{\varepsilon}_{i1}^2) = (\hat{\alpha}_1 - \hat{\alpha}_2)(m_i - f) \neq 0$ 이다. 따라서 $\hat{\alpha}_1 \neq \hat{\alpha}_2$ 이다.

식 (47)은 하위공간에서 동일 시점 t 에 m_1 에 대한 r_{12} 의 새로운 회귀방정식이다. 이 회귀방정식의 회귀계수는 $\hat{\alpha}_{12}$ 이고 이에 대한 확률계수는 $\hat{\alpha}_{12}$ 이며 오차는 \hat{u}_{12} 이다. 이 때 다음 식이 성립한다.

$$\hat{u}_{12} \neq 0 \tag{49}$$

확률계수 $\hat{\alpha}_{12}$ 는 상위공간에서는 파생기울기였는데 하위공간에서 어떠한 역할을 하는지 규명하기 위하여 식 (40)을 식 (44)~(45)의 도움으로

$$\hat{\alpha}_{12} = \frac{r_{12} - f}{m_1 - f} \tag{50}$$

로 고쳐 쓸 수 있다. 식 (50)은 하위공간 (r_1, m) 의 좌표에서 절편 $(1 - \hat{\alpha}_{12})f$ 와 r_{12} 를 통과하는 관측직선의 기울기 $\hat{\alpha}_{12}$ 를 의미한다. 상위공간에서는 앞서 소개한 대로 식 (35)에 의해서 수평축의 한 점 M_1^* 에서 전체 수직길이 R_{11} 에 대한 부분 수직길이 JK 의 비율로서 기울기부분이 전체 길이에서 차지하는 파생기울기를 의미하였지만 하위공간에서는 식 (50)에 의해서 기울기 그 자체가 된다.

결국 새로운 표본이 발생하는 핵심은 상위공간의 회귀방정식 (8)은 관측평행직선 (33)과 수학적으로 동등하고(mathematically equivalent). 다시 관측평행직선 (33)은 변수 전환하여 하위공간에서 회귀방정식 (47)과 수학적으로 동등(mathematically equivalent)하다는 점이다. 즉

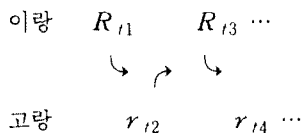
$$\begin{aligned} \text{상위공간 회귀방정식 (8)} &= \text{상위공간 관측평행직선 (33)} \\ &= \text{하위공간 회귀방정식 (47)} \end{aligned}$$

인데, 이 두 단계의 수학적 동등 과정에서 상위공간 회귀방정식 (8)의 R_{11} 이 하위공간 회귀방정식 (47)의 r_{12} 를 발생시켰다. <정리 4>가 성립한다.

정리 4 상위공간에서 배아방정식 형태의 회귀방정식은 상위공간의 관측평행직선과 수학적으로 동등하며, 상위공간의 관측평행직선은 하위공간에서 배아방정식 형태의 회귀방정식과 수학적으로 동등하다. 두 단계 수학적 동등에서 새로운 표본이 발생한다.

IX. 확률적 이중나선

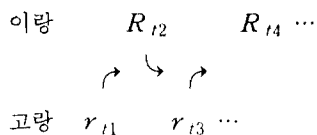
〈정리 4〉가 본 논문의 핵심이다. 〈정리 4〉의 동일한 수학적 동등단계를 하위공간의 r_{12} 에 적용하면 상위공간에서 R_{13} 를 발생시키고 이 과정은 계속되며 다음과 같다.



하나의 표본이 다른 표본을 발생할 때 그 발생 순서는 두 가지 특성치로 정의된다. 하나는 위상이고 다른 하나는 진폭이다. 위상은 이랑(R_{1n})인가 고랑(r_{1n})인가를 결정하고 진폭은 위상의 크기를 확률적으로 결정한다. 이랑은 상위공간(이자울공간)을 가리키고 고랑은 하위공간(가격공간)을 가리킨다. 이 구조는 흡사 나선의 흠처럼 생겼다. 따라서 다음의 정리가 성립한다.

정리 5 상위공간의 R_{11} 이 하위공간의 r_{12} 를 낳고, 하위공간의 r_{12} 는 상위공간의 R_{13} 를 낳으며, 상위공간의 R_{13} 는 하위공간의 r_{14} 를 낳고, 이 나선의 과정은 한없이 계속된다.

〈정리 5〉에 의하면 상위공간의 R_{11} 에서 시작하면 하위공간의 도움을 받아서 발생하는 상위공간의 표본은 홀수 표본이다. 이번에는 반대로 하위공간의 r_{11} 에서 시작하면 발생 순서는



이다. 이 현상 역시 나선의 흠처럼 생겼다. 또 하나의 정리가 성립한다.

정리 6 하위공간의 r_{i1} 이 상위공간의 R_{i2} 를 낳고, 상위공간의 R_{i2} 는 하위공간의 r_{i3} 를 낳으며, 하위공간의 r_{i3} 는 상위공간의 R_{i4} 를 낳고, 이 나선의 과정은 계속된다.

〈정리 6〉에 의하면 하위공간의 r_{i1} 에서 시작하면 상위공간의 도움을 받아서 발생하는 상위공간의 표본은 짝수이다. 홀수와 짝수의 표본을 합하여 상위공간의 표본과 하위공간의 표본의 발생이 완성된다.

상위공간과 하위공간은 역의 관계 $rR = 1$ 이므로 직각쌍곡선을 이룬다. 그러므로 상위공간의 홀수 표본에서 하위공간의 짝수 표본으로 이동하는 과정(〈정리 5〉)은 두 공간 사이에서 나선의 모습을 형성한다. 즉, 식 (46)은 $r_{i2}R_{i1} = 1 + [\hat{\beta}_1 m_i + (1 - \hat{\beta}_1)f] \hat{\epsilon}_{i1}$ 인데 $\hat{\epsilon}_{i1} = 0$ 이면 직각쌍곡선 $r_{i2}R_{i1} = 1$ 이 되지만 일반적으로 $\hat{\epsilon}_{i1} \neq 0$ 이므로 R_{i1} 과 r_{i2} 사이는 확률적인 직각쌍곡선이고 확률적 나선이다. 반대로 하위공간의 홀수 표본에서 상위공간의 짝수 표본으로 이동하는 과정(〈정리 6〉)에서 또 하나의 확률적 직각쌍곡선 $r_{i1}R_{i2} = 1 + [\hat{\alpha}_1 M_i + (1 - \hat{\alpha}_1)F] \hat{u}_{i1}$ 이 성립하므로 r_{i1} 과 R_{i2} 사이에서 또 하나의 확률적 나선을 형성한다. 두 개의 확률적 나선이 엉켜서 모집단을 만들어 낸다. 확률적 이중나선(stochastic double helix)이라고 부를 수 있다. 즉

$$\begin{array}{cccc}
 R_{i1} & R_{i2} & R_{i3} & \cdots \\
 \curvearrowright & & \curvearrowright & \\
 r_{i1} & r_{i2} & r_{i3} & \cdots
 \end{array} \quad (51)$$

상위공간의 표본계열과 하위공간의 표본계열 사이에 두 개의 홈이 파진 나선 모양이 형성되며 각 홈의 크기는 확률적으로 결정된다. 두 개의 나선은 각각 이차율 나선과 가역 나선이다. 표본의 발생은 퇴보하지 않는다(non-degeneration).¹¹⁾

11) 증명은 생략한다.

X. 정보발생확률함수

상위공간의 표본은 반드시 하위공간을 거쳐 상위공간의 또 다른 표본을 발생하는 구조가 하위공간의 표본을 낳고 이를 매개로 다시 상위공간의 표본을 낳는 하나의 확률적 나선현상과 그 반대의 확률적 나선현상은 서로 소거되어 이량과 고량이 없는 단일의 확률적 직선현상처럼 보인다. 즉, R_{11} 에서 r_{12} 를 거치지 않고 직접 R_{12} 를 발생하는 것처럼 보인다. 이 내용을 식 (46)에서 확인할 수 있다. 식 (46)은 r_{11} 과 r_{12} 의 관계를 나타내는데 여기에 변수 사이의 역관계 $r_{11}R_{11} = 1$, $r_{12}R_{12} = 1$ 을 적용하면¹²⁾ 식 (46)은

$$\begin{aligned} R_{12} &= \frac{1}{1 + [\hat{\beta}_1 m_t + (1 - \hat{\beta}_1)f] \hat{\varepsilon}_{t1}} R_{11} \\ &= \frac{1}{1 + [f + m_t - f m_t \hat{R}_{t1}] \hat{\varepsilon}_{t1}} R_{11}, \\ & \quad t = 1, 2, \dots, T \end{aligned} \quad (52)$$

와 동등하다. 식 (52)에서 $\hat{\varepsilon}_{t1} \neq 0$ 이므로

$$R_{12} \neq R_{11} \quad (53)$$

이 성립한다. 동일 시점에서 전혀 새로운 피설명변수의 크기이다. 즉, 동일 시점 t 에서 피설명변수의 새로운 표본 R_{12} 가 식 (52)에 의해 이미 실현된 피설명변수의 표본 R_{11} 에서 발생한다. 이 발생은 $\hat{\varepsilon}_{t1}$ 의 존재 때문에 확률적이다. 식 (53)은 다음과 같이 요약된다.

$$R_{12} = \frac{1}{1 + \theta_1 \hat{\varepsilon}_{t1}} R_{11} \quad (54)$$

여기서 $\theta_1 = \hat{\beta}_1 m_t + (1 - \hat{\beta}_1)f > 0$ 이다. 따라서 R_{12} 와 R_{11} 의 관계

12) $r_{11}R_{11} = 1$ 은 정의식이지만 $r_{11}R_{11} = 1$ 일 때 $r_{12}R_{12} = 1$ 이 성립하는 증명은 생략한다.

는 오차가 양수이면 할인(discount)의 관계이고 오차가 음수이면 할증(premium)의 관계이다. 즉, 동일 시점 t 에서 상태(state) 1의 R_{t1} 과 상태(state) 2의 R_{t2} 사이의 할증 또는 할인의 관계이다.

정리 7 정보발생의 과정은 할인 또는 할증의 과정이다.

이 결과는 하위공간을 점검함으로써 재확인할 수 있다. 재확인을 위하여 이상의 동일한 방식을 이번에는 하위공간에 적용하여 하위공간의 식 (25)를

하위공간 관측평행직선

$$\begin{aligned} r_{t1} &= \hat{\alpha}_1 m_t + (1 - \hat{\alpha}_1) f + \hat{u}_{t1} \\ &= \hat{\alpha}_1 m_t^* + (1 - \hat{\alpha}_1) f_t^*, \quad t = 1, 2, \dots, T \end{aligned} \quad (55)$$

로 표현한다. 여기서 $m_t^* = m_t + \hat{u}_{t1}$ 이고 $f_t^* = f + \hat{u}_{t1}$ 이다. 식 (55)의 양변을 r_{t1} 으로 나누면

$$1 = \frac{\hat{\alpha}_1 m_t^*}{r_{t1}} + \frac{(1 - \hat{\alpha}_1) f_t^*}{r_{t1}} \quad t = 1, 2, \dots, T$$

인데 이것을 다시 구성비율로 표현하면

$$\begin{aligned} 1 &= \frac{\hat{\alpha}_1 m_t^*}{r_{t1}} + \frac{(1 - \hat{\alpha}_1) f_t^*}{r_{t1}} \\ &= \hat{\beta}_{t2} + (1 - \hat{\beta}_{t2}) \end{aligned}$$

이다. 따라서

$$\hat{\beta}_{t2} = \hat{\alpha}_1 \frac{m_t + \hat{\mu}_{t1}}{r_{t1}} = \hat{\alpha}_1 \frac{m_t^*}{r_{t1}}, \quad \hat{\beta}_{t2} \neq \hat{\beta}_1, \quad \hat{\beta}_{t2} \neq \hat{\beta}_{t1} \quad (56)$$

이다. 식 (56)의 두 개의 조건은 상위공간의 두 개의 조건 (40), (36)에 상응한다. 이것으로 $\hat{\beta}_{t2}$ 가 하위공간에서 파생기울기를 뜻하는데 상위공

간으로 전환되면 기울기 자체가 되는 것을 알 수 있다. 결국 식 (55)~(56)에서

$$\frac{\hat{\alpha}_1}{r_{t1}} = \frac{\hat{\beta}_{12}}{m_t^*} \tag{57}$$

$$\frac{1 - \hat{\alpha}_1}{r_{t1}} = \frac{1 - \hat{\beta}_{12}}{f_t^*}$$

가 성립하므로 이 둘을 더하면 직각쌍곡선

$$\frac{1}{r_{t1}} = \frac{\hat{\beta}_{12}}{m_t^*} + \frac{1 - \hat{\beta}_{12}}{f_t^*}, \quad t = 1, 2, \dots, T \tag{58}$$

가 된다. 식 (58)은 r_{t1} 을 m_t^* 과 f_t^* 의 '조화평균'으로 정의하고 있으며 가중치는 $\hat{\beta}_{12}$ 이다. 이와 대조적으로 식 (55)는 r_{t1} 을 m_t^* 과 f_t^* 의 '산술평균'으로 정의하고 있고 가중치는 $\hat{\alpha}_1$ 이다. 식 (58)에 변수전환을 적용하여 정리하면 상위공간에서

회귀방정식

$$R_{t2} = \hat{\beta}_2 M_t + (1 - \hat{\beta}_2) F + \hat{\varepsilon}_{t2}, \quad t = 1, 2, \dots, T \tag{59}$$

가 되는데 이 때

$$R_{t2} = R_{t1} \{ 1 + [\hat{\alpha}_1 M_t + (1 - \hat{\alpha}_1) F] \hat{\mu}_{t1} \} \tag{60}$$

$$\hat{\varepsilon}_{t2} = (\hat{\beta}_{12} - \hat{\beta}_2)(M_t - F) \tag{61}$$

이다. 오차 (61) 역시 전과 동일한 형태로서 기울기의 차이로 정의되었음은 변함이 없고 <정리 2>에 부합한다. 식 (29)가 성립하므로 식 (60)에서

$$R_{t2} \neq R_{t1} \tag{62}$$

이다. 따라서 식 (62)에 의해 동일 시점 t 에서 피설명변수의 전혀 새로운 정보 R_{t2} 가 이미 실현된 피설명변수 R_{t1} 에서 발생한다. 식 (56)이 성립

하므로 식 (61)에서

$$\hat{\varepsilon}_{t2} \neq 0 \quad (63)$$

이다.

하위공간의 식 (46)과 상위공간의 식 (60)에서 변수전환 $r_{t1}R_{t1} = 1$ 과 $r_{t2}R_{t2} = 1$ 이 성립해야 하므로

$$\begin{aligned} & [\hat{\alpha}_1 M_t + (1 - \hat{\alpha}_1) F] \hat{\mu}_{t1} \\ &= - \frac{[\hat{\beta}_1 F + (1 - \hat{\beta}_1) M_t] \hat{\varepsilon}_{t1}}{[\hat{\beta}_1 F + (1 - \hat{\beta}_1) M_t] \hat{\varepsilon}_{t1} + F M_t} \end{aligned} \quad (64)$$

이 된다. 식 (64)의 결과는

$$\begin{aligned} \hat{\varepsilon}_{t1} > 0 \text{이면 } \hat{\mu}_{t1} < 0 \\ \hat{\varepsilon}_{t1} < 0 \text{ 이면 } \hat{\mu}_{t1} > 0 \end{aligned}$$

을 뜻한다. 상위공간의 오차가 양수(음수)이면, 하위공간의 오차는 음수(양수)이다. 식 (64)를 식 (60)에 대입하여 정리하면

$$R_{t2} = \frac{1}{1 + [\hat{\beta}_1 m_t + (1 - \hat{\beta}_1) f] \hat{\varepsilon}_{t1}} R_{t1} \quad t = 1, 2, \dots, T \quad (65)$$

가 된다. 이 결과는 식 (52)와 일치한다. 따라서 임의의 시점 t 에서 설명변수 M_t 와 상수 F (따라서 각각의 역수 m_t 와 f)가 일정한 크기로 주어지고 추정기울기 $\hat{\beta}_1$ 가 주어질 때 한 개의 수익률의 관측치 R_{t1} 에서 또 하나의 수익률 R_{t2} 가 도출된다. 즉, 피설명변수의 행렬 (6)의 첫째 열 R_{t1} 에서 동일 시점 t 의 둘째 열 R_{t2} 의 정체가 드러난다. R_{t2} 는 R_{t1} 에서 발생하지만, 식 (65)의 등호의 오른쪽 항의 분모에 등장하는 오차항 $\hat{\varepsilon}_{t1}$ 때문에 '확률적인 관계'이다. 이 사실은 아래에서 발생정보의 확률적 특성을 조사할 때 더욱 분명해진다. 뿐만 아니라 회귀방정식 (59)를 보면

R_{12} 는 추정치가 아니라 표본의 형태가 되어 있음을 알 수 있다. 말을 바꾸면 R_{12} 는 실현되지 않고 숨겨진 표본이다.

식 (63)은 회귀방정식 $R_1 = \hat{\beta}_1 M + (1 - \hat{\beta}_1)F + \hat{\epsilon}_1$ 에서 설명변수가 $M = M_1$ 이고 피설명변수가 $R_1 = R_{11}$ 일 때 R_{12} 를 발생시키는 식이다. 결과적으로 피설명변수의 행렬 (6)의 실현된 첫째 열

$$R_1 = \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{T1} \end{pmatrix}$$

에서 둘째 열

$$R_2 = \begin{pmatrix} R_{12} \\ R_{22} \\ \vdots \\ R_{T2} \end{pmatrix}$$

가 '확률적으로' 발생한다.

이제 동일한 방식을 둘째 열에 적용하기 위하여 정보발생확률함수 (65)에 의해 구한 둘째 열 R_2 와 첫째 열 R_1 을 합친 표본을 고정적으로 반복되는 설명변수 M 에 대하여 회귀방정식을 추정할 수 있으므로 두 개의 열벡터 사이에

회귀방정식

$$R_{12} = \hat{\beta}_2 M_t + (1 - \hat{\beta}_2)F + \hat{\epsilon}_{12}, \quad t = 1, 2, \dots, T \quad (66)$$

를 구할 수 있다. 둘째 열의 추정 회귀계수는 $\hat{\beta}_2$ 이고 $\hat{\beta}_2 \neq \hat{\beta}_1$ 이며 추정 오차는 $\hat{\epsilon}_{12}$ 이고 $\hat{\epsilon}_{12} \neq \hat{\epsilon}_{11}$ 이다.

첫째 열의 회귀방정식 (8)에 적용하여 둘째 열의 정보 (65)를 구한 것과 동일한 방법을 둘째 회귀방정식 (66)에 대하여 적용하면 셋째 열의 정보발생은 확률적으로

$$R_{t3} = \frac{1}{1 + [\hat{\beta}_2 m_t + (1 - \hat{\beta}_2)f] \hat{\varepsilon}_{t2}} R_{t2} \quad (67)$$

$$t = 1, 2, \dots, T$$

를 따른다. 식 (67)에 의해 피설명변수의 행렬 (6)의 첫째 열과 둘째 열

$$R_2 = \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{T1} \\ R_{12} \\ R_{22} \\ \vdots \\ R_{T2} \end{pmatrix}$$

에서 셋째 열

$$R_3 = \begin{pmatrix} R_{13} \\ R_{23} \\ \vdots \\ R_{T3} \end{pmatrix}$$

가 확률적으로 발생하였다.

이상의 분석을 일반화하면 첫째 열 R_1 의 실현된 회귀방정식이

$$R_{t1} = \hat{\beta}_1 M_t + (1 - \hat{\beta}_1)F + \hat{\varepsilon}_{t1}$$

$$E\hat{\varepsilon}_{t1} = 0, \quad t = 1, 2, \dots, T \quad (68)$$

와 같이 주어졌을 때 실현되지 않은 $n+1$ 번째 열의 피설명변수의 정보는 다음의 공식

$$R_{tn+1} = \frac{1}{1 + [\hat{\beta}_n m_t + (1 - \hat{\beta}_n)f] \hat{\varepsilon}_{tn}} R_{tn} \quad (69)$$

$$R_{tn} = \hat{\beta}_n M_t + (1 - \hat{\beta}_n)F + \hat{\varepsilon}_{tn}, \quad E\hat{\varepsilon}_{tn} = 0 \quad (70)$$

$$t = 1, 2, \dots, T \quad (71)$$

$$n = 2, \dots, \infty \quad (72)$$

에서 '확률적으로' 연속 발생한다. 참고로 식 (69)는 식 (54)의 축소형으로 표현하면 <정리 7>에 의하여

$$R_{t,n+1} = \frac{1}{(1 + \theta_1 \hat{\epsilon}_{t1})(1 + \theta_2 \hat{\epsilon}_{t2}) \cdots (1 + \theta_n \hat{\epsilon}_{tn})} R_{t1} \quad (73)$$

이 된다. 여기서 $\theta_n = \hat{\beta}_n m_t + (1 - \hat{\beta}_n)f > 0$ 이다. 정보발생의 과정은 서로 다른 상태(state) 사이의 복리할인 또는 복리할증의 과정이다.

식 (70)에 의해 각 열에서 추정된 회귀계수는

$$\hat{\beta}_n = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_N) \quad (74)$$

인데 (7)과 (9)의 도움으로 다음이 성립한다.

$$\beta = \frac{1}{N} \sum_{n=1}^{\infty} \hat{\beta}_n \quad (75)$$

즉, 실제 모수 β 는 추정치 $\hat{\beta}_n (n = 1, 2, \dots, \infty)$ 의 평균이다. 이 추정치의 평균을 이용하면 실제 회귀방정식을 구할 수 있는데 그것이 식 (4)이고 실제오차까지 구할 수 있다. 즉, 식 (75)를 식 (4)에 대입하여 실제오차

$$\epsilon_t = R_t - \beta M_t + (1 - \beta)F, \quad E\epsilon_t = 0 \quad (76)$$

을 구할 수 있다. 다른 식으로 표현하면 식 (70)에 의해 각 열에서 구한 추정오차는

$$\hat{\epsilon}_{tn} = (\hat{\epsilon}_{t1}, \hat{\epsilon}_{t2}, \dots, \hat{\epsilon}_{t\infty}) \quad (77)$$

인데

$$\epsilon_t = \frac{1}{N} \sum_{n=1}^{\infty} \hat{\epsilon}_{tn} \quad (78)$$

으로 구할 수 있다.

XI. 오차의 특성

오차 (17)은 모든 $n = 1, 2, \dots, N$ 에 대하여 적용되므로 $\hat{\varepsilon}_{t,n} = (\hat{\beta}_{t,n} - \beta_n)(M_t - F)$ 이고 여기에 기대부호 E 를 적용하면

$$\begin{aligned}
 E(\hat{\varepsilon}_{t,n}) &= E(R_{t,n} - \hat{R}_{t,n}) \\
 &= E(\hat{\beta}_{t,n} - \beta_n)(M_t - F) \\
 &= (M_t - F)E(\hat{\beta}_{t,n} - \beta_n) \\
 &= (M_t - F)E(\hat{\beta}_{t,n} - \beta) \\
 &= 0
 \end{aligned} \tag{79}$$

이다. 이 과정에서 M_t 는 F 와 더불어 반복되는 비확률변수이므로 기대부호와 관계가 없이 처리하였고 오차의 평균이 영이 되는 것은 통상의 기본 가정 (4)에 의존한다. 넷째 등호는 식 (75) 덕택에 성립한다. 따라서

$$E(\hat{\beta}_{t,n}) = \beta \tag{80}$$

이다. 관측직선의 확률 기울기의 평균은 회귀직선의 비확률 기울기이다.

한편 오차 $\hat{\varepsilon}_{t,n}$ 의 분산은 식 (76)에 의하면 다음과 같다.

$$\sigma^2(\hat{\varepsilon}_{t,n}) = (M_t - F)^2 \sigma^2(\hat{\beta}_{t,n}), \quad t = 1, 2, \dots, T \tag{81}$$

만일 확률 파라메타 $\beta_{t,n}$ 의 분산이 균등하면 오차 $\hat{\varepsilon}_{t,n}$ 의 분산은 불균등하고, 오차의 분산이 균등하면 확률 파라메타의 분산이 불균등하다. 오차의 특성을 규명할 수 있는 한 가지 단서는 $F = M_t \neq R_{t,1}$ (따라서 $F = M_t \neq R_{t,n}$, $n = 1, 2, \dots, N$)인 좌표 $A = (F, F)$ 에서 찾을 수 있다. <정리 1>에 의하면 좌표 A 에서는 각각의 유한의 오차 $\hat{\varepsilon}_{t,n} = R_{t,n} - \hat{R}_{t,n}$ ($n = 1, 2, \dots, N$)에 대하여 확률계수 $\hat{\beta}_{t,n}$ ($n = 1, 2, \dots, N$)는 무한대이므로 오차의 분포에 상관없이 확률계수의 분산은 무한대이다. 이것은 확률계수 $\hat{\beta}_{t,n}$ 의 분산이 불균등(heteroskedastic)함을 나타낸다.

따라서 오차의 분산은 균등하다고(homoskedastic) 가정할 수 있다. 즉

$$\sigma^2(\hat{\varepsilon}_{t_n}) = \sigma^2, \quad t = 1, 2, \dots, T \quad (82)$$

이다.

식 (69)를 일반적인 경우에 재정리하여 기대부호를 적용하면 다음과 같다.

$$\begin{aligned} & [\hat{\alpha}_n M_t + (1 - \hat{\alpha}_n) F][\hat{\beta}_n F + (1 - \hat{\beta}_n) M_t] E(\hat{\varepsilon}_{t_n} \hat{\mu}_{t_n}) \\ & + [\hat{\alpha}_n M_t + (1 - \hat{\alpha}_n) F] F M_t E(\hat{\mu}_{t_n}) \\ & + [\hat{\beta}_n F + (1 - \hat{\beta}_n) M_t] F M_t E(\hat{\varepsilon}_{t_n}) = 0 \end{aligned} \quad (83)$$

따라서 $E(\hat{\varepsilon}_{t_n}) = 0$ 이므로 $E(\hat{\mu}_{t_n}) = 0$ 이면 $E(\hat{\mu}_{t_n} \hat{\varepsilon}_{t_n}) = 0$ 이고 $E(\hat{\mu}_{t_n}) \neq 0$ 이면 $E(\hat{\mu}_{t_n} \hat{\varepsilon}_{t_n}) \neq 0$ 이다.

식 (83)을 재정리하면 다음이 성립한다.

$$\begin{aligned} & \hat{\varepsilon}_{t_{n+1}} - \hat{\varepsilon}_{t_n} + [\hat{\beta}_n m_t + (1 - \hat{\beta}_n) f][\hat{\varepsilon}_{t_{n+1}} + \hat{R}_{t_n}] \hat{\varepsilon}_{t_n} = 0 \\ & n = 1, 2, \dots, \infty \end{aligned} \quad (84)$$

식 (84)에 기대부호를 적용하여 정리하면

$$\begin{aligned} & f E[\hat{\varepsilon}_{t_{n+1}} \hat{\varepsilon}_{t_n}] + (m_t - f) \{ E[\hat{\beta}_n \hat{\varepsilon}_{t_{n+1}} \hat{\varepsilon}_{t_n}] \\ & + E[\hat{\beta}_n \hat{R}_{t_n} \hat{\varepsilon}_{t_n}] \} = 0 \\ & n = 1, 2, \dots, \infty \end{aligned} \quad (85)$$

이 된다. 둘째 항 대괄호 속에 스타인의 정리(the Stein's lemma)를 적용하면

$$E[\hat{\varepsilon}_{t_{n+1}} \hat{\varepsilon}_{t_n}] = 0 \quad (86)$$

이다. 즉, 동일 시점 t 의 오차발생에서 오차 사이의 관계는 독립이다. 이것은 다시 동일시점의 정보발생에서 정보 사이에 독립적인 관계를 의미한다.

XII. 맺는 말

본 논문은 실현된 하나의 표본에서 실현되지 않은 표본을 복제하는 방법을 제시하였다. 그 배경은 숨겨진 정보공간의 발견과 그 활용에 있었다. 상위공간과 하위공간은 서로 긴밀한 관계를 맺으며 흡사 확률적 이중나선(stochastic double helix)의 역할을 하며 새로운 정보를 창출한다. 각각 고유의 정보를 갖고 있는 두 공간이 서로 엇물리면서 새로운 정보를 창출하는 과정을 통하여 하나의 표본에서 모집단을 복제할 수 있는 것이다.

본 논문에서는 단일선형의 자본자산가격 모형을 이용하였지만 일반적인 회귀분석에 적용하는 일이 앞으로 남은 문제이다. 선형회귀분석은 일반적으로 다변량 변수를 포함하는데, 추정된 다변량 선형회귀방정식 $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \dots + \hat{\beta}_K x_{tK} + \hat{\epsilon}_t$ 을 일대 일로 단일 변수의 '산술평균 형태'의 배아방정식 $y_t = (1 - \hat{\beta}_y)z + \hat{\beta}_y w_t + \hat{\epsilon}_t$ 으로 전환시킬 수 있다면 본 논문이 제시한 정보발생의 일반화는 가능하다. 전환된 방정식은 모집단을 복제할 수 있는 배아(embryo)에 해당하므로 배아방정식(embryonic equation)이다. 배아방정식은 다음의 연구과제이다.

◆ 참고문헌 ◆

- 김학은(1998), "유통속도의 부문별 분할 I", 『연세경제연구』, 제V권 제2호, pp. 101~128.
- _____(2001), "화폐유통속도의 수학적 분할," 『經濟學研究』, 49, pp. 273~303.
- _____(2003), "유통속도의 부문별 분할 II", 『연세경제연구』, 제X권 제1호, pp. 53~96.
- Hildreth, C. and J. Houck (1968), "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, vol. 63, pp. 584~595.
- Huang, C. and R. H. Litzengerger (1988), *Foundations for Financial Economics*, New York: North-Holland.