

平率回歸分析和 豫測*

姜 錫 勳**

< 目 次 >

- I. 序 論
- II. 平率回歸分析
- III. 非母數平率回歸推定量
- IV. 例: 換率
- V. 結 論

I. 序 論

經濟理論을 연구하는 目的 中의 하나는 現實에 나타난 經濟現狀을 分析하고, 이에 根據하여 미래에 전개될 경제현상을 豫測하기 위함이다. 일반적으로 경제이론은 경제현상을 抽象化된 經濟模型을 통하여 설명하고자 하며, 이론에서 導出된 結論들은 경제현상이 만들어 낸 데이터를 통해 計量經濟學的인 방법으로 推定, 檢證되어지는 과정을 통해 그 現實性과 有用性이 판단되어지는 경우가 많다. 그러나 일단 경제이론이 계량경제학적인 방법을 이용한 實證分析의 단계에 접어 들게 되면, 경제이론과는 관계없이(또는 관계가 있더라도 미약한) 계량경제학에서 임의로 制約을 부과하여 분석을 행하거나, 理論의 本來의 目的과는 無關한 豫測技法을 사용하여 豫測을 행하는 경우가 있다.

分析의 측면에서 例를 들어 보자. 경제이론에서는 變數 Y가 變數 X의 함수라는 사실만이 도출되었음에도 불구하고, 실증분석의 단계에서는 두 변수사이의 함수관계가 線形이라든가, 두 변수의 結合(또는 條件附)確率密度函數가 正

* 한국경제학회 논문발표회에 유익한 논평을 해주신 이준행 박사님께 감사드립니다.

** 大宇經濟研究所 特殊研究室 先任研究員

規分布를 따른다는 식의 假定을 도입하는 경우가 많다. 普通最小自乘推定量(OLS) 또는 一般最小自乘推定量(GLS) 등은 변수 Y가 변수 X의 파라메터에 대한 線形函數라는 假定下에 도출된 것이며, 最尤推定量(MLE)은 두 변수의 確率分布가 有限次元의 파라메터를 제외하고는 알려져 있다는 假定下에 도출된 것이다. 물론 추상화된 경제이론을 현실경제에 그대로 適用하여 實證分析을 하기에는 計算量이라든지 이용가능한 模型 등이 制約되어 있는 것은 사실이나, 가능한 한 계산상의 便宜 또는 解釋上的 難易度を 回避하기 위해 계량경제학에서 임의로 부과하는 制約은 最小化하는 것이 좋다.

다음은 豫測의 측면에서 例를 들어보자. 豫測이라 함은 항상 豫測誤差를 前提로 하는 概念이며, 어떠한 종류의 豫測値도 豫測誤差에 의한 損失을 最小化하는 방법으로부터 도출된다. 豫測誤差란 豫測値과 실제 實現된 값의 差異를 말하는데, 예측을 행할 때에는 예측오차로부터 豫測目的과 관련하여 일정한 형태의 損失函數를 선택하고 이에 근거하여 豫測量을 선택해야 한다. 그러나 실제로 예측을 행할 때는 豫測目的과 관련하여 損失函數를 선택해야 한다는 사실이 看過되는 경우가 많다. 예를 들어 企業에서 換率을 예측하는 사람의 입장에서 보면 예측오차의 絕對的인 값도 중요하지만, 예측오차가 어떠한 符號를 띠게 될 것인가가 더욱 중요할 것이고, 따라서 그가 선택하는 손실함수는 이를 反映해야 한다. 그러나, 일반적으로 사용되는 最小自乘推定量을 도출하는 損失函數를 생각해보면 이는 陽의 손실이나 陰의 손실에 대해 對稱的으로 작용하여 그 절대값만 같으면 같은 量의 손실로 認識하므로 最小自乘推定量을 사용하여 예측을 행하는 것은 예측목적에 符合되지 않을 수도 있다. 다른 例는 어떤 設計士가 댐을 設計하는데 事前에 댐의 높이를 決定해야 하는 경우를 상정해보자. 이 경우 댐의 높이를 이제까지 내렸던 降雨量의 平均을 基準으로 결정한다면 (강우량의 확률분포가 對稱的이라면), 50%의 確率로 이 댐에서는 물이 넘치게 될 것이다. 이 설계사가 意圖하는 것은 적절한 費用으로 물이 넘치지 않을 정도의 댐을 建設하는 데 있을 것이므로 이제까지 내렸던 最大降雨量을 기준으로 댐의 높이를 결정하거나, 陽의 誤差(강우량이 댐의 높이를 초과하는 경우)에는 많은 加重値를 주고 陰의 誤差에는 적은 加重치를 주는 손실함수를 상정하고 이를 最小化하여 댐의 높이를 결정하는 방법이 타당할 것이다.

本稿에서는 實證分析에서 많이 사용되는 回歸分析(regression analysis)을

손실함수의 관점에서 再照明해 보고, 非對稱自乘損失函數를 最小化하는 平率回歸(expectile regression)를 提案한다. 비대칭 자승손실함수란 陽의 誤差의 제곱과 陰의 誤差의 제곱에 각기 다른 加重值를 부여함으로써 오차의 절대적인 값뿐만 아니라 그 符號도 함께 考慮하는 손실함수이며, 이에 근거한 平率回歸은 既存에 널리 사용되는 平均回歸(mean regression)를 특수한 경우로 포함하는 동시에 分位數回歸(quantile regression)의 성질도 함께 가진다. 한편 平率回歸函數를 推定함에 있어서는 함수형태를 線形이나 非線形 등으로 임의적인 制約을 加하지 않은 상태에서 추정할 수 있도록 非母數平率回歸推定量(nonparametric expectile regression estimator)을 제안한다. 비대칭 자승손실함수에 기초하여 平率회귀를 非母數的으로 추정하고 이에 근거하여 예측을 행하는 것은, 앞에서 提起한 두 가지 問題 즉 豫測技法이 豫測目的을 反映해야 한다는 점과 예측에 사용되는 임의적인 假定은 최소화하는 것이 좋다는 점을 모두 滿足시킬 수 있는 技法은 아니나, 두 가지 측면을 모두 考慮한 豫測技法이라고 할 수 있다.

本稿의 構成은 다음과 같다. 제Ⅱ절에서는 일반적으로 사용되는 回歸分析을 예측의 관점에서 再照明해 보고 예측에 사용되는 損失函數에 따라 回歸函數가 다르게 나타남을 보이며, 非對稱自乘損失函數를 사용하는 경우의 最適豫測量인 平率回歸를 제안한다. 제Ⅲ절에서는 平率回歸分析의 觀點下에서 回歸方程式의 非母數推定方法을 제안하며 이에 근거하여 제Ⅳ절에서는 우리나라 원화의 對美달리換率豫測을 例로 들었다. 제Ⅴ절은 結論을 敘述한다.

Ⅱ. 平率回歸分析

(1) 平率回歸分析

古典의 意味에서 確率變數 Y 에 대한 確率變數 X 의 回歸(regression)란, 주어진 X 에서 Y 의 條件附기대값을 의미하며, 回歸分析(regression analysis)이란 未知의 파라메터에 대하여 線形으로 假定된 조건부기대값의 파라메터를 最小自乘推定方法을 통해 추정함을 의미한다. 그러나 최근에는 회귀의 개념이 보다 一般化되어 X 의 함수로 표현되는 Y 의 X 에 대한 條件附確率分布의 性質을 의미한다(Manski, 1991). 條件附平均이나 條件附分位數 등은 모두 조건부

확률분포의 性質중의 하나이며 따라서 回歸의 하나의 종류가 된다. 일반적으로 이제까지는 조건부확률분포의 여러가지 성질중에 계산상의 便宜에 의해 조건부평균이 가장 많이 利用되었는데 이는 平均回歸(mean regression)라고 할 수 있다. 그러나 보다 一般화된 의미의 回歸의 觀點에서 보면 平均回歸은 여러가지 回歸分析技法 중의 하나에 지나지 않으며 平均이 가지는 여러가지 長點과 동시에 短點을 함께 지니고 있다. 최근에는 平均回歸分析뿐만 아니라 中央값回歸分析, 分位數回歸分析 등이 利用되기도 한다. Manski(1975, 1985)는 二進選擇模型(binary choice model)에서 조건부중앙값을 구하는 중앙값회귀(median regression)에 대하여 연구하였고, Powell (1984)은 Tobit모형에서 중앙값회귀에 대하여 연구하였다. Koenker & Bassett (1978)는 分位數回歸(quantile regression)을 제안하고 분위수회귀추정량을 이용하여 異分散性を 檢證할 수 있는 통계량을 구하였다. Lee(1991)는 최빈값회귀(mode regression)를 提案하였다. 대부분의 경우에 回歸은 어떤 특정한 損失函數下에서의 最適豫測量(best predictor)을 제공한다. 예를 들면 平均回歸($m(x)$)는 自乘損失函數下에서의 최적예측량이다. 즉

$$m(x) = \operatorname{argmin}_{m \in M} E[(y - m)^2 | x]$$

M은 적절하게 선택된 變數空間임.

같은 방법으로 중앙값회귀와 분위수회귀는 각각 絶對값損失函數와 非對稱絶對값損失函數下에서의 최적예측량이다.¹⁾

이러한 觀點에서 보면 Newey & Powell(1987)이 고려한 非對稱自乘損失函數를 최소화하는 최적예측량은 平率回歸(expectile regression)라고 할 수 있다.²⁾ 즉,

1) 중앙값회귀($l(x)$)와 분위수회귀($q(x)$)는 다음과 같은 損失函數하의 最適豫測量이다.

$$l(x) = \operatorname{argmin}_{l \in L} E[|y - l| | x]$$

$$q(x) = \operatorname{argmin}_{q \in Q} E[|\tau - 1(y < q)| \cdot |y - q| | x]$$

L과 Q는 適切하게 선택된 變數空間임.

2) Expectile은 平均을 의미하는 expectation과 分位數를 의미하는 quantile 合成語이다. 이는 後術하겠지만 平率이 평균을 특수한 경우로 포함하는 동시에 분위수의 성격도 內包하기 때문에 붙여진 이름이다.

$$t_\tau(x) = \operatorname{argmin}_{t_\tau \in T_\tau} E[|\tau - 1(y < t_\tau)| \cdot (y - t_\tau)^2 | x]$$

T_τ 는 적절하게 선택된 變數空間이며,

$1(\cdot)$ 는 方向函數(indicator function)이며,

τ 는 0과 1사이의 常數임.

이 非對稱自乘損失函數는 Aigner, Amemiya & Poirier(1976)가 非對稱의인 正規分布의 假定下에 生産可能曲線을 추정할 때 처음으로 다루었다. Newey & Powell(1987)은 平均回歸函數를 未知의 파라메터에 대해 線形으로 假定하고 이 손실함수를 다루었으며 이에 근거하여 파라메터를 추정하는 방법으로 非對稱最小自乘推定量(asymmetric least squares estimator, ALSE)을 제안하였다. 그들은 平率回歸函數가 分位數回歸函數와 類似한 방법으로 條件附確率分布의 성질을 나타내 주지만, 平率回歸函數의 파라메터추정은, 즉 非對稱最小自乘推定量은 反復再加重最小自乘推定方法(iterative reweighted least squares)을 사용하여 계산해낼 수 있기 때문에, 그리드방법이나 simplex방법 등의 非標準의인 계산방법을 사용하여 추정해야 하는 分位數回歸函數의 파라메터추정보다 계산하기가 容易함을 보였다. 또한 분위수회귀함수의 파라메터에 대한 추정량의 漸近的 分散行列과는 달리, ALSE의 점근적분산행렬은 계산하기가 매우 어렵다고 알려진 오차항의 確率密度의 계산없이 추정해낼 수 있다는 長點을 가짐을 보였다.

(2) 平率의 性質

確率變數 Y 의 τ -平率(τ -expectile)은 다음과 같은 損失函數를 最小化하는 값으로 定義된다(Newey & Powell, 1987).

$$\mu_\tau = \operatorname{argmin}_{m \in M_\tau} E[|\tau - 1(y < \mu_\tau)| \cdot (y - \mu_\tau)^2]$$

M_τ 는 적절하게 선택된 變數공간임.

Newey & Powell(1987)은 만약 확률변수 Y 의 기대값이 존재하면 μ_τ 는 다음 식을 만족한다는 것을 보였다.

$$\mu_\tau - E(Y) = [(2\tau - 1) / (1 - \tau)] \cdot \int_{\mu_\tau}^{\infty} (y - \mu_\tau) dF(y) \quad (1)$$

$F(y)$ 는 Y 의 累積分布函數를 나타냄.

식 (1)에서 보듯이 平率은 확률변수 Y 가 Y 의 확률분포의 末尾(tail)에 있을 조건하에서 Y 의 기대값의 성질에 의해 결정됨을 알 수 있다.

만약 $E(Y)$ 가 0으로 正規化되고 $yF(y)$ 가 y 가 陰의 無限大로 갈 때 0으로 간다고 가정하면, 部分積分을 써서 Y 의 平率 μ_r 가 다음을 만족함을 보일 수 있다.

$$\tau = \left[\int_{-x}^{\mu_r} F(y) dy \right] \cdot \left[2 \int_{-x}^{\mu_r} F(y) dy - \mu_r \right]^{-1} \quad (2)$$

식 (2)에서 약간의 계산을 통해 우리는 τ 가 0.5면 平率 μ_r 은 0이며 (0.5-平率은 평균을 의미하며 이는 正規化된 0임), τ 가 0.5보다 작으면 τ 는 0과 $2 \int_{-\infty}^{\mu_r} F(y) dy$ 사이에 있고, τ 가 0.5보다 크면 μ_r 은 0보다 작다는 사실을 알 수 있다. 식 (2)는 μ_r 가 0인 경우 τ 가 0.5임을 보이고 있는데 이 때에만 平率 이 Y 의 累積分布函數에 의존하지 않으며 그 이외의 경우에는 平率 이 누적분포함수에 따라 결정됨을 보여준다고 하겠다. 그림 1은 Y 의 확률밀도함수가 0에 대해 對稱의일 때 τ -平率은 面積 A 를 面積 $A+B$ 로 나눌 때의 값을 τ 로 하는 값을 보여준다. τ -平率과 τ -分位數를 도출해 내는 손실함수는 같은 加重值附與方法을 사용하면서 손실을 절대값으로 計測하느냐 또는 제곱으로 計測하느냐의 차이만 있으므로 이 두 가지를 비교해 보면 흥미롭다. τ -분위수 (θ_r)는 다음과 같이 쓰여질 수 있다.

$$\tau = \left[\int_{-x}^{\theta_r} f(y) dy \right] \cdot \left[\int_{-x}^{\infty} f(y) dy \right]^{-1} \quad (3)$$

$f(y)$ 는 Y 의 確率密度函數임.

식 (3)에서 對應의인 방법으로 생각하면 우리는 식 (3)에서 확률밀도함수가 累積確率密度函數로 代替된 확률분포의 要約統計量(summary statistics)을 상정해볼 수 있는데 이는 分母가 非限定的(unbounded)이기 때문에 定義되지 않음을 알 수 있다. 식 (3)은 분위수의 경우 尺度의 기준이 되는 分母가 항상 1인데 반하여 식 (2)의 平率의 경우에는 分母가 τ 에 따라 항상 변화한다는 사실을 보여준다.

平率에 관한 전반적인 理解를 돕기 위해 다음의 그래프를 살펴보자. 그림 2는 標準正規分布의 分位數와 平率을 보여준다. 이 그림에서 알 수 있듯이 平率函數는 분위수함수와 같이 τ 에 대해 單調增加函數이며, 平率의 기울기는 τ 가 0.5 부근에서는 분위수의 기울기보다 작지만 τ 가 0 또는 1 근방에서는 더

크다. 그림 3의 (가)는 분산이 모두 1이지만 평균이 0, 1, 2인 正規分布의 平率의 변화를 나타낸 그림이며, (나)는 평균은 모두 0이지만 분산이 1, 4, 9인 경우의 正規分布의 平率의 변화를 보여준다. 그림 3의 (나)는 τ 가 0.5보다 클 때는 平率이 분산의 增加函數이지만 τ 가 0.5보다 작을 때는 분산의 減少函數임을 보여준다.

平率의 몇가지 성질은 Newey & Powell(1987)의 定理 1에 요약되어 있다. 만약 확률변수 Y 의 평균이 존재하면, τ -平率は 唯一하며 τ 의 單調增加函數이다. 또한 平率は 분위수와 같이 位置와 尺度에 同一分散的(equivariant)³⁾이며 Y 의 확률분포를 결정한다. 여기서 주목할 점은 0.5-분위수는 τ -분위수 (단 $\tau \neq 0.5$)와 거의 類似한 성질을 가지고 있지만(중앙값의 중요하고 유용한 성질중의 하나가 單調變換에 대해 無關(invariant)하다는 것인데 분위수의 경우도 마찬가지이다), 0.5-平率 즉 평균은 τ -平率 ($\tau \neq 0.5$)과 다른 성질을 갖는다는 점이다. 즉 확률변수 Y 와 확률변수 X 의 0.5-平率의 합은 X 와 Y 의 합의 0.5-平率과 같지만, τ 가 0.5가 아닌 경우에는 이러한 관계가 성립하지 않는다.

定理 1. X 와 Y 가 平均이 有限한 確率變數라고 하면,

$$E_{\tau}(X+Y) \neq E_{\tau}(X) + E_{\tau}(Y). \text{ 단 } \tau \neq 0.5.$$

證明 : 反例가 이것을 증명하기에 충분하다. 확률변수 X_1 과 X_2 가 확률적으로 獨立인 標準正規分布를 따르고 $W = X_1 + X_2$ 라고 하자. X_1 의 τ -平率을 m 이라고 하면 X_2 의 τ -平率도 m 이다. 따라서 X_1 의 τ -平率과 X_2 의 τ -平率의 합은 $2m$ 이다. 한편 W 는 평균 0, 분산이 2인 正規分布를 따르며 $\sqrt{2} \cdot Z$ 로 표현될 수 있다 (Z 는 標準正規分布를 따르는 확률변수). 平率は 위치와 척도에 대해 同一分散的이므로 W 의 τ -平率は Z 의 τ -平率에 $\sqrt{2}$ 를 곱한 값이 되는데 이는 $2m$ 과는 다르다⁴⁾. 證明 끝

平率의 성질에 대한 深層的인 분석은 앞으로의 研究課題이다.

3) 函數 $k(\cdot)$ 가 $k(aX+b)=ak(X)+b$ 의 관계를 만족할 때 우리는 $k(\cdot)$ 가 位置와 尺度에 대해 同一分散的이라고 定義한다 (여기에서 a 는 陽의 常數이며 b 는 陰의 常數임).

4) 이 反例를 제공해 준 Goldberger 교수에게 감사드립니다.

Ⅲ. 非母數平率回歸推定量

(Nonparametric Expectile Regression Estimator, NPERE)

Newey & Powell(1987)은 平率回歸函數를 未知의 파라메터에 대해 線形으로 假定하고 파라메터에 대한 추정량을 제안하였다. 일반적으로 τ -平率回歸函數는 說明變數와 τ 의 함수이며 따라서 Newey & Powell(1987)이 사용한 선형함수는 미지의 平率회귀함수에 대한 線形近似量이라고 볼 수 있다. Newey & Powell(1987)의 定理 1은 平率回歸函數가 주어진 說明變數에 대한 被說明變數의 條件附確率分布를 特徵지움을 보이고 있지만, 線形近似量은 이러한 성질을 반드시 가질 필요가 없다. 따라서 平率회귀함수에 대해 사전에 임의의 制約을 가하지 않고 이 함수를 추정할 수 있다면 이는 보다 일반적인 경우라고 할 수 있다. 이를 위해 먼저 데이터 생성과정을 定義한다.

데이터 $(x_i, y_i, i=1, 2, \dots, n)$ 은 다음의 방정식에 의해 生成되었다고 가정하자.

$$y_i = g(x_i) + \varepsilon_i \quad (4)$$

$g(\cdot)$ 는 未知의 實數函數이고 x_i 는 p -次元의 獨立的인 확률변수이며 ε_i 는 1次元의 誤差項임.

식 (4)에서 τ -平率回歸函數, $E_\tau(y_i|x_i)$ 은 Newey & Powell (1987)의 정리 1에 의해 다음과 같다.

$$E_\tau(y_i|x_i) = g(x_i) + E_\tau(\varepsilon_i|x_i) \quad (5)$$

$E_\tau(\varepsilon_i|x_i)$ 는 주어진 x_i 에서 ε_i 의 τ -平率을 나타냄.

식 (5)에서 $g(x_i)$ 와 $E_\tau(\varepsilon_i|x_i)$ 는 개별적으로는 識別(identified)되지 않으나 주어진 x_i 에서 y_i 의 τ -平率回歸函數 $g_\tau(x_i)$ 는 識別되며 다음과 같이 쓰여질 수 있다.

$$y_i = g_\tau(x_i) + \eta_{\tau i} \quad (6)$$

$\eta_{\tau i} \equiv \varepsilon_i - E_\tau(\varepsilon_i|x_i)$ 이며 따라서 $E_\tau(\eta_{\tau i}|x_i) = 0$ 임.

有限次元의 파라메터로 平率回歸函數를 制約시키지 않는 범위내에서 이를

추정하기 위해서는 스무딩방법(smoothing method)을 이용해야 한다. 본래는 커널스무딩, nearest neighbor 방법 또는 시리즈확장방법(series expansion method) 등의 모든 스무딩방법이 비슷한 결과를 유도해야 하나 標準的인 Nadaraya-Watson의 平均回歸에 대한 非母數推定量과 比較分析하기 위해 커널 방법을 사용하였다. 고전적인 Nadaraya-Watson의 平均回歸函數에 대한 커널추정량은 다음과 같은 목적함수를 최소화하는 極端推定量(extreme estimator)으로 볼 수 있다.

$$g(x) = \operatorname{argmin}_{g \in G} \frac{1}{nh_n^p} \sum_{i=1}^n K \left[\frac{x-x_i}{h_n} \right] (y_i - g)^2 \quad (7)$$

G 는 적절히 선택된 파라미터공간, $K(\cdot)$ 는 커널함수, p 는 x_i 의 次數, h_n 은 n 이 無限大로 감에 따라 0으로 收斂해가는 구간넓이임.

식 (7)을 最小化하는 推定量은 우리가 $K(\cdot)$ 를 적절히 선택하면(예를 들어 다변량 · 표준정규분포의 확률밀도함수를 사용하면), 주어진 x 에서 x 주위의 값들에 대해 가까운 것은 큰 가중치를 주고 멀어질수록 작은 값의 가중치를 주어서 계산된 局所加重平均值임을 알 수 있다. 標本數가 증가함에 따라 구간넓이가 0으로 收斂하므로 x 주위의 아주 가까운 값들만의 局所加重平均值가 되고 이는 주어진 x 에서 x 의 函數값, 즉 $g(x)$ 에 收斂하게 됨을 짐작할 수 있다. 식 (7)을 최소화하는 一階條件으로부터 古典的인 Nadaraya-Watson의 平均回歸函數에 대한 다음과 같은 非母數推定量을 얻을 수 있다.

$$g(x) = \frac{\frac{1}{nh_n^p} \sum_{i=1}^n K \left[\frac{x-x_i}{h_n} \right] y_i}{\frac{1}{nh_n^p} \sum_{i=1}^n K \left[\frac{x-x_i}{h_n} \right]} \quad (8)$$

같은 방법으로 τ -平率回歸函數에 대한 非母數推定量은 다음의 목적함수를 최소화하는 값으로 定義되며 이를 非母數平率回歸推定量 (Nonparametric Expectile Regression Estimator, 이하 NPERE) 이라고 부르기로 한다.

$$g_\tau(x) = \operatorname{argmin}_{g_\tau \in G_\tau} \frac{1}{nh_n^p} \sum_{i=1}^n K \left[\frac{x-x_i}{h_n} \right] \cdot |\tau - 1(y_i < g_\tau)| \cdot (y_i - g_\tau)^2 \quad (9)$$

G_τ 는 적절히 선택된 파라미터공간임.

따라서 $\hat{g}_i(x)$ 는 다음을 滿足시키며, 母數的 推定方法과 마찬가지로 反復再加重 最小自乘推定方法에 의해 계산될 수 있다.

$$g(x) = \frac{\frac{1}{nh_n^p} \sum_{i=1}^n K\left[\frac{x-x_i}{h_n}\right] \cdot |\tau-1(y_i < \hat{g}_i(x))| \cdot y_i}{\frac{1}{nh_n^p} \sum_{i=1}^n K\left[\frac{x-x_i}{h_n}\right] \cdot |\tau-1(y_i < \hat{g}_i(x))|} \quad (10)$$

IV. 例 : 換率⁵⁾

이 절에서는 우리나라의 對美달러換率을 데이터로 하여 이제까지 論議된 내용에 대한 例를 들어보려고 한다. 사용된 데이터는 우리나라 對美달러換率 月別資料(月平均)이며 1980년부터 1992년 9월까지를 標本期間으로 상정하였다. 표 1의 첫번째 모형은 AR(1)을 가정하고 非對稱最小自乘推定方法을 통해 얻어진 결과이다.⁶⁾ 표 2의 두번째 모형은 우리나라의 대부분의 巨視經濟變數가 單位根을 가지고 있다는 既存의 연구에 따라 1차 차분한 환율예측모형을 가정하고 같은 분석을 실시하였다.⁷⁾ 사용된 가중치는 0.5, 0.25, 0.75 그리고 0.05이며, 係數 推定值에 대한 標準偏差는 White(1980)의 異分散性を 고려한 分散推定量(heteroscedasticity consistent covariance estimator)을 사용하였다. AR(1)을 假定한 모형에서는 기울기가 統計的으로 有意하다는 결과가 나왔으나 차분한 模型에서는 기울기가 통계적으로 有意하지 못하였다⁸⁾.

표 1은 각각의 加重值를 달리한 경우, 推定式의 기울기에 대한 推定量이 첫번째 모형에서는 τ 가 커질수록 커지나 두번째 모형에서는 τ 가 커질수록 작아진다는 사실을 보여준다. τ 의 가중치는 實際值가 豫測值보다 큰 경우, 즉 過小

5) 이 절은 단지 제 II과 제 III절에서 제안한 방법을 우리나라 對美달러換率에 適用한 例에 불과함을 強調한다.

6) 母數的 自己回歸模型을 구성할 때 최적사차수를 결정하기 위한 기준으로 Akaike정보기준이나 Schwarz기준이 널리 이용되나 非母數的인 자기회귀모형을 구성할 때의 최적사차수선택에 관한 한 아직까지 널리 연구된 바가 없는 것으로 보인다. 본고에서는 두가지 방법의 비교분석을 위해 時差數를 1로 상정하였으나 이에 관해서는 앞으로의 연구과제이다.

7) 비모수추정방법은 자기회귀함수가 非線形이라는 假定하에 회귀함수를 추정하는 방법이므로 單位根 문제와 직접적인 관련은 없는 것으로 보인다. 이에 관해서는 앞으로의 연구과제이다.

8) 通常의 OLS에서 사용되는 標準偏差를 사용하는 경우에는 1/2-平率回歸(즉 平均)를 추정하여 얻은 기울기가 통계적으로 有意하게 나타났다(t 값 : 9.24). 이는 차분한 환율시리즈가 異分散的이라는 사실을 시사한다고 하겠다.

豫測의 경우에 부과되는 加重値이므로 τ 가 커진다는 것은 過小豫測에 많은 가중치를 부여함을 의미한다. 따라서 첫번째 모형을 사용할 때 過小豫測이 문제시되면 예측치는 그렇지 않은 경우에 비해 前期의 환율에 더욱 依存하여 예측을 실행해야 함을 보여준다고 하겠다. 이것은 서로 다른 예측오차의 尺度를 통해 보면 더욱 明確해 진다. 표 1의 첫번째 모형에서 알 수 있듯이 τ 가 0.5일 때 最小自乘損失이 최소가 되지만, τ 가 0.05일때 陰의 損失이 최소화되며 τ 가 0.75일때 陽의 損失이 최소화됨을 알 수 있다⁹⁾. 표 1의 두번째 모형도 같은 방법으로 解釋할 수 있다.

표 2는 非母數平率回歸推定量(NPERE)을 이용, 같은 자료를 사용해서 얻은 추정결과이다. NPERE를 사용하려면 추정에 사용할 커널함수를 결정해야 하며 또한 구간넓이를 결정해야 한다. 커널함수는 NPERE의 漸近的 正規分布性を 부여하기 위해 다음과 같은 高次커널(higher order kernel)을 사용하였다.

$$K(u)=2f(u)-f(u/\sqrt{2})/\sqrt{2}$$

$K(\cdot)$ 는 커널함수이며 $f(\cdot)$ 는 標準正規分布의 確率密度函數임.

일반적으로 커널함수의 선택은 추정결과에 큰 영향을 미치지 못하지만 구간넓이의 선택은 추정에 큰 영향을 미치는 것으로 알려져 있다(Hardle, 1990). 지나치게 넓은 구간넓이를 사용하게 되면 과다스무딩(over-smoothing)을 하여 데이터가 가진 情報를 損失하게 되며 지나치게 작은 구간넓이를 사용하게 되면 추정된 방정식이 너무 거칠어서 전체적인 傾向에 대한 이해를 어렵게 한다. 여기에서는 Craven & Whaba(1979) 등에 의해 개발된 一般相互檢證(generalized cross validation)방법을 NPERE의 추정에 맞도록 補完한 修整一般相互檢證方法(modified generalized cross validation, MGCV)을 통해 적정구간넓이를 계산하였다. 修整一般相互檢證方法은 다음의 식을 최소화하는 구간넓이를 계산하여 이를 실제추정시 구간넓이로 사용하는 방법이다.

9) 陽의 損失은 豫測誤差가 陰일 때는 0으로 看做하고 陽일 때만의 경우를 합한 損失函數를 이용한 것이고 陰의 損失은 그 반대의 경우이다.

$$MGCV = \frac{\left[\sum_{j=1}^n |\tau - 1(y_j < \hat{g}_\tau(x_j))| \cdot (y_j - \hat{g}_\tau(x_j)) \right]^2}{\left[1 - \frac{1}{n} \sum_{i=1}^n w_{ii}^2 \right]^2}$$

$$\text{단 } w_{ii} \equiv \frac{\frac{1}{h_n^p} K(0) |\tau - 1(y_i < \hat{g}_\tau(x_i))|}{\frac{1}{n} \sum_{j=1}^n \frac{1}{h_n^p} K\left[\frac{x_i - x_j}{h_n}\right] |\tau - 1(y_j < \hat{g}_\tau(x_j))|}$$

그림 4는 구간넓이를 변화하였을 때 MGCV의 값의 변화를 그림으로 나타낸 것이다. 그림에서 알 수 있듯이 原系列을 기준으로 하여 非母數 AR(1) (nonparametric AR(1), NPAR(1))을 假定한 경우에는 MGCV가 구간넓이가 0.68일 때 최소화되며, 차분한 非母數 ARI(1,1)(nonparametric ARI(1, 1), NPARI(1,1))의 경우에는 최적구간넓이가 0.18일 때 최소화되었다¹⁰⁾. 표 2는 이 구간넓이를 사용하여 얻은 결과이다. 먼저 τ 값의 변화에 따른 損失量의 변화는 표 1과 마찬가지로 결과를 알 수 있다. 또한 표 2는 損失函數의 절대적인 값이 母數的 推定方法을 통한 예측보다 크게 작아짐을 보여주고 있다. 이는 母數的 推定方法이 非母數的 推定方法보다 많은 制約을 가하고 있으므로 손실함수의 값이 당연히 커짐을 보여주기도 하지만 특히 우리나라 환율 예측의 경우에는 모수적 추정방법과 비모수적 추정방법에 의한 추정방법 사이의 豫測誤差의 차이가 매우 크다는 점을 보여준다고 하겠다.¹¹⁾

V. 結 論

本稿에서는 實證分析에서 많이 사용되는 回歸分析을 손실함수의 관점에서 再照明해 보고, 非對稱自乘損失函數를 最小化하는 平率回歸(expectile regression)을 提案하였으며 平率回歸函數를 推定함에 있어서는 함수형태를 線形이나 非線形 등으로 임의적인 制約을 加하지 않은 상태에서 추정할 수 있다

10) τ 가 0.5인 경우를 계산하고 이를 τ 가 0.5가 아닌 경우에도 사용하였다. τ 가 0.5가 아닌 경우에 別途로 최적구간넓이를 계산하여 사용할 수도 있으나, 豫測誤差에 대한 比較시 서로 스무딩의 정도가 다른 경우를 비교한다는 것은 스무딩의 정도에 지나치게 의존하게 되는 短點이 있다.

11) 이와 같은 결과는 標本內 區間에서의 예측오차를 基準으로 나타난 것으로 標本 外 區間에서의 예측오차에 관해서는 시사하는 바가 없음을 留意해야 한다.

록 非母數平率回歸推定量을 제안하였다. 이는 豫測技法이 豫測目的을 反映해야 한다는 점과 예측에 사용되는 임의적인 가정은 최소화하는 것이 좋다는 점을 考慮한 豫測技法이라고 할 수 있다.

한편 실제로 豫測을 行함에 있어서는 가중치의 選定이 문제가 되나 이는 豫測目的을 考慮하여 예측자 스스로가 판단하여야 할 문제이다. 本稿의 제 IV절에서는 우리나라의 對美달리換率을 例로 들어, 서로 다른 가중치를 부여한 경우와 非母數의 推定方法을 사용하여 예측을 행하는 경우를 살펴 보았는데 이러한 시도는 앞으로 標本外에서의 豫測誤差檢討 등의 深層의인 분석이 필요할 것으로 보이며, 우리나라의 다른 主要 時系列資料에의 適用 또한 앞으로의 研究課題이다.

理論의으로는 平率의 統計學的 性格을 보다 명확히 糾明함으로써 일반적으로 無批判의으로 수용되는 平均의 概念을 보다 명확히 이해하는 동시에 서로 다른 가중치를 주는 경우의 解釋에도 도움을 줄 수 있을 것으로 판단된다.

參 考 文 獻

1. Aigner, D., T. Amemiya and D. Poirier (1976), "On the Estimation of Production Frontiers : Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function", *International Economic Review*, 17 : 377-396.
2. Craven, P. and G. Whaba (1979), "Smoothing Noisy Data with Spline Functions", *Numerische Mathematik*, 31 : 377-403
3. Hardle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press.
4. Koenker, R and G. Bassett (1978), "Regression Quantiles", *Econometrica*, 46 : 33-50
5. Lee, M. (1991), "Mode Regression", *Journal of Econometrics*, 42 : 337-349.
6. Manski, C. (1975), "The Maximum Score Estimation of Stochastic Utility Model of Choice", *Journal of Econometrics*, 3 : 205-228.
7. _____ (1985), "Semiparametric Analysis of Discrete Response :

- Asymptotic Properties of Maximum Score Estimator", *Journal of Econometrics*, 27 : 313-333.
8. _____(1988), *Analog Estimation Methods in Econometrics*, Chapman and Hill
 9. _____(1991), "Regression", *Journal of Economic Literature*, XXIX : 34-50
 10. Nadaraya, E. (1964), "On Regression Estimators", *Theory of Probability and its application*, 10 : 157-159.
 11. Newey, W. and J. Powell (1987), "Asymmetric Least Squares Estimation and Testing", *Econometrica*, 25 : 303-325
 12. Powell, J. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics*, 32 : 143-155.
 13. _____(1986), "Censored Regression Quantiles", *Journal of Econometrics*, 32 : 143-155.
 14. Watson, G., (1964), "Smooth Regression Analysis", *Sankhya, Series A*, 26 : 359-372
 15. White, H. (1980), "A Heteroscedasticity Consistent Covariance Estimator and a Direct Test for Heteroscedasticity", *Econometrica*, 48 : 817-838

〈表 1〉 非對稱最小自乘推定量에 의한 豫測

(1) 模型 : $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$

항 목	$\tau=0.05$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$
係數推定値				
常數項	13.0218	14.6875	15.8446	17.2376
기울기	0.9743	0.9791	0.9808	0.9821
標準偏差 (1)				
常數項	120.7836	59.6855	56.1190	72.8614
기울기	0.1591	0.0795	0.0741	0.0950
豫測誤差				
對稱自乘損失函數	14769.3	6632.72	5712.17	6605.23
陽의 損失	1235.9	560.26	338.03	184.04
陰의 損失	-65.0	-186.75	-338.03	-552.12

(2) 模型 : $\Delta y_t = \alpha + \beta \Delta y_{t-1} + \varepsilon_t$

항 목	$\tau=0.05$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$
係數推定値				
常數項	-5.1314	-1.4040	0.5322	2.5773
기울기	0.6474	0.6125	0.6040	0.5936
標準偏差 ¹⁾				
常數項	10.0715	5.6143	5.3517	6.6820
기울기	1.1640	0.9037	0.9953	1.2964
豫測誤差				
對稱自乘損失函數	5250.77	2316.04	2486.43	3869.42
陽의 損失	676.92	258.14	129.58	60.67
陰의 損失	-36.49	-173.28	-335.47	-573.08

주 : 1) 標準偏差는 White(1980)의 異分散性을 고려한 分散推定量 (heteroscedasticity-consistent covariance estimator)을 사용하였음.

〈表 2〉 非母數平率回歸推定量에 의한 豫測

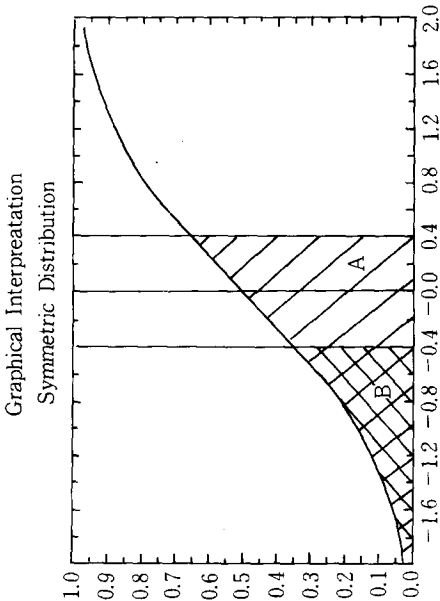
(1) 模型 : $y_t = g(1, y_{t-1}) + e_t$, 구간넓이 : 0.68

항 목	$\tau=0.05$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$
豫測誤差				
對稱自乘損失函數	3321.22	1707.16	1239.82	7845.33
陽의 損失	323.83	205.11	122.90	63.58
陰의 損失	-16.48	-60.26	-121.76	-214.65

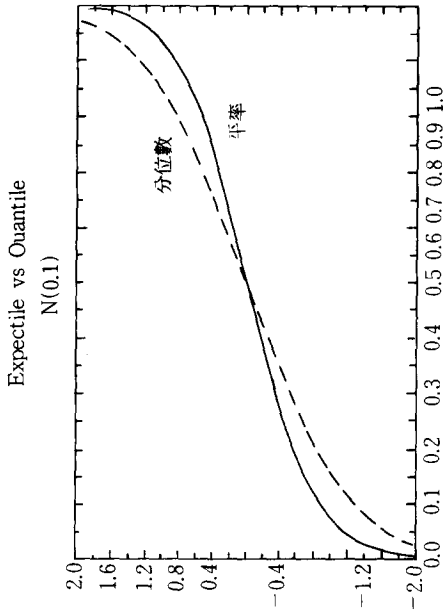
(2) 模型 : $\Delta y_t = g(1, \Delta y_{t-1}) + e_t$, 구간넓이 : 0.18

항 목	$\tau=0.05$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$
豫測誤差				
對稱自乘損失函數	5636.10	1625.41	1284.03	1738.46
陽의 損失	595.82	240.41	139.37	70.96
陰의 損失	-28.62	-73.03	-145.10	-249.62

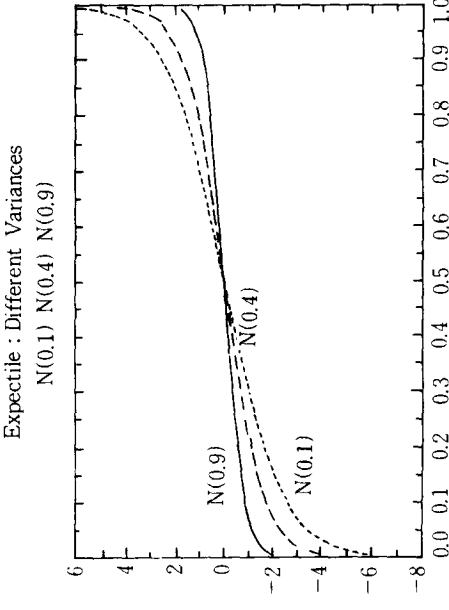
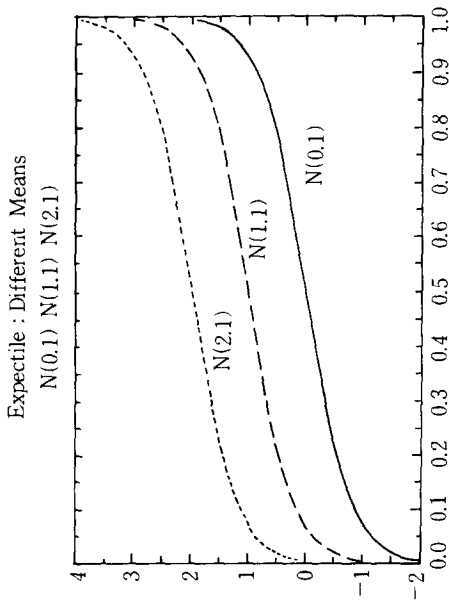
〈그림 1〉 平率의 圖表的 理解



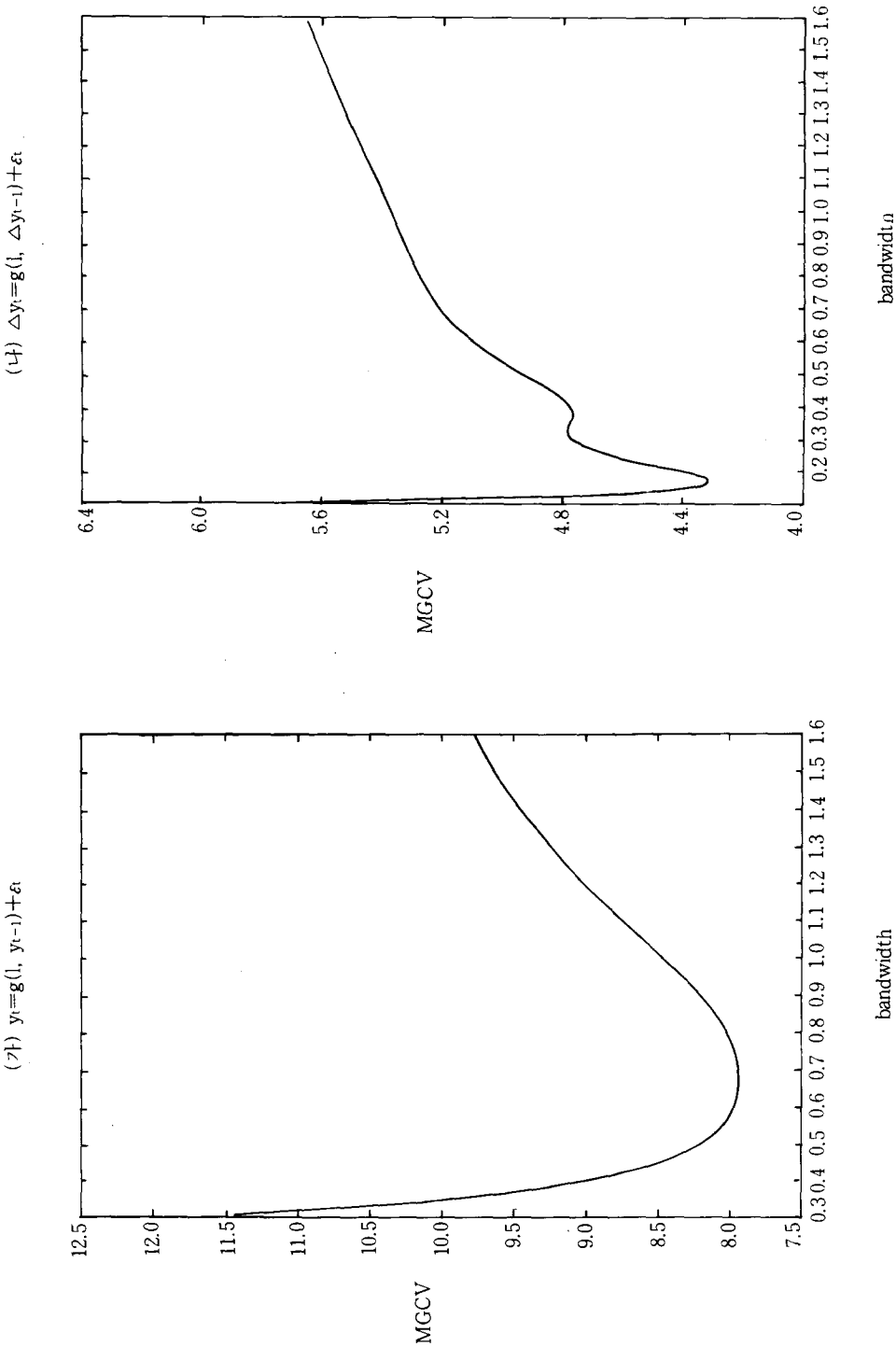
〈그림 2〉 標準正規分布의 平率과 分位數



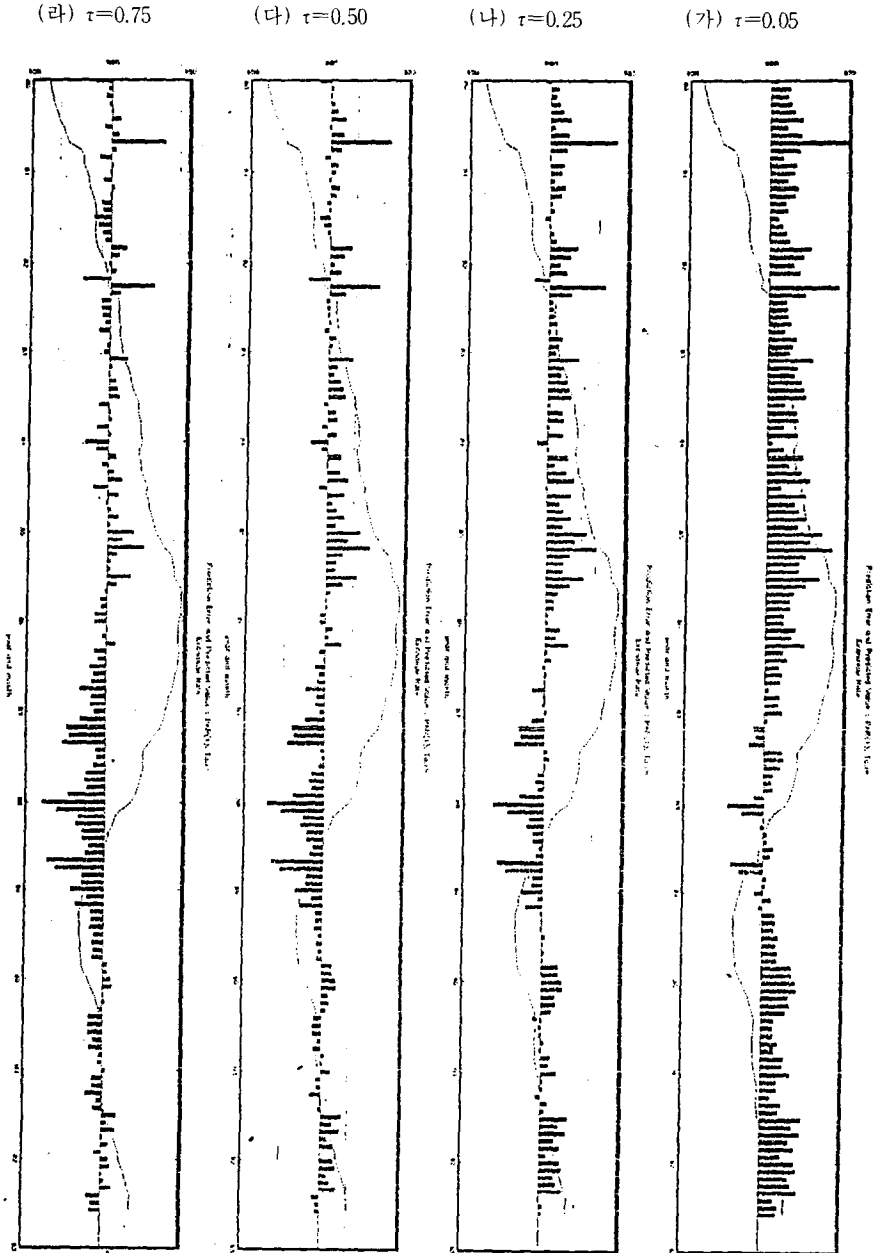
〈그림 3〉 平率의 變化



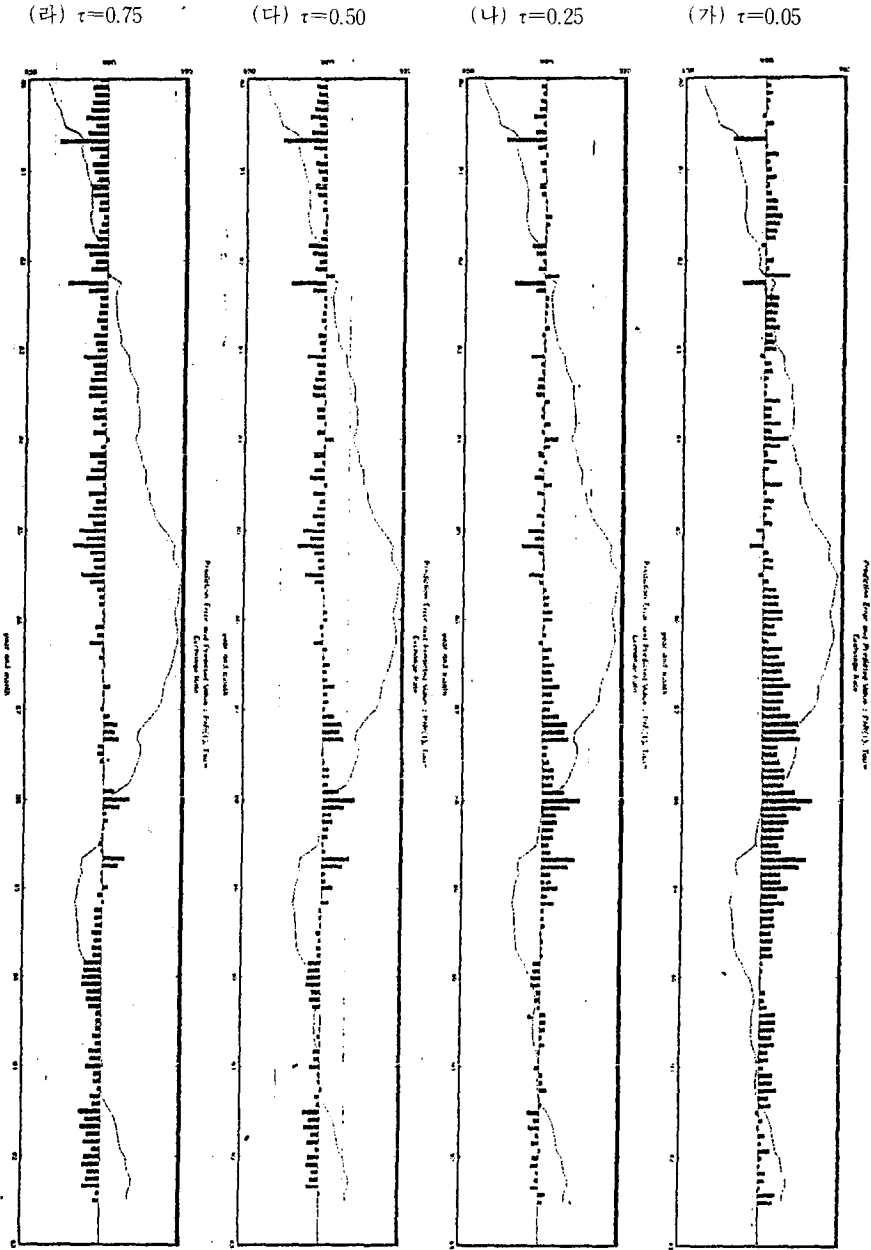
〈그림 4〉修整一般相互檢證方法에 의한 最適區間 넓이 選擇



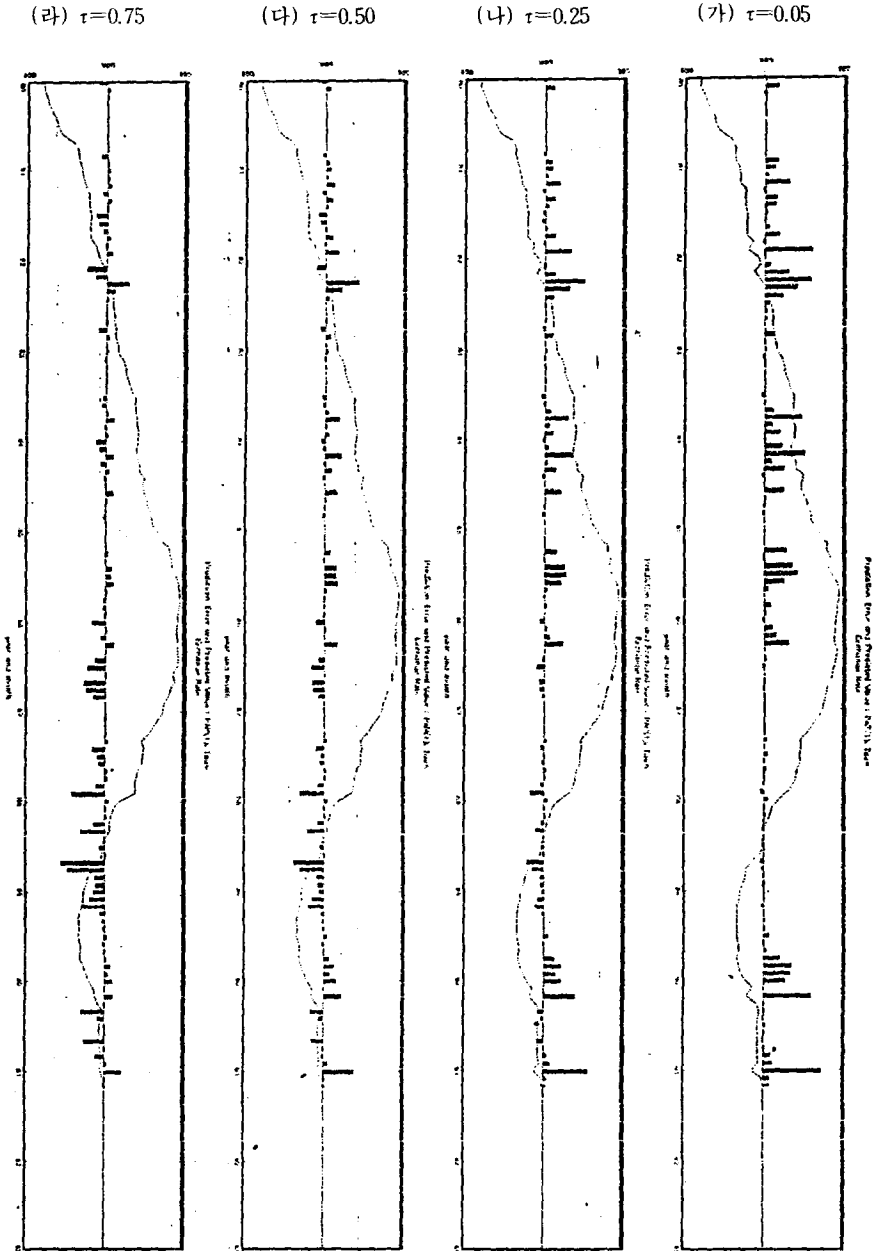
〈그림 5〉 豫測量과 豫測誤差 : 非對稱最小自乘推定方法에 의한 AR(1) 豫測



〈그림 6〉 豫測量과 豫測誤差 : 非對稱最小自乘推定方法에 의한 $ARI(1, 1)$ 豫測



〈그림 7〉 豫測量과 豫測誤差 : 非母數平率回歸推定方法에 의한 AR(1) 豫測



〈그림 8〉 豫測量과 豫測誤差 : 非母數率回歸推定方法에 의한 $ARI(1, 1)$ 豫測

