經營科學 第10卷 第2號
1993年 10月



# 推移確率의 推定을 위한 확장된 Markov Chain 模型

강정혁*

# An Extension of Markov Chain Models for Estimating Transition Probabilities

Jung-Hueek Kang*

## ABSTRACT

Markov chain models can be used to predict the state of the system in the future. We extend the existing Markov chain models in two ways. For the stationary model, we propose a procedure that obtains the transition probabilities by appling the empirical Bayes method, in which the parameters of the prior distribution in the Bayes estimator are obtained on the collaternal micro data. For non-stationary model, we suggest a procedure that obtains a time-varying transition probabilities as a function of the exogenous variables.

To illustrate the effectiveness of our extended models, the models are applied to the macro and micro time-series data generated from actual survey. Our stationary model yields reliable parameter values of the prior distribution. And our non-stationary model can predict the variable transition probabilities effectively.

## 1. 序 論

Markov Chain을 이용한 推移確率 推定技法은 정의된 특정기간에 있어서 상태(State)의 변동상황을 나타내는 추이확률을 과거에 있었던 변화를 토대로 推定함으로써, 시스템 內의 이동현상을 분석하게 한다. 여기서 Markov Chain은 시스템 內의 모든 개개 요소가 이산적인 시간에서 서로 독립적으로 행동한다는 가정과 상태간의 변동은 랜덤한 방식으로 발생하며, 현재 상태가 주어진 경우 목표상태에 도달한 확률은 과거에 어떠한 상태들을 통하여 도달하였는가에 無關하게, 현재와 목표의 상태에만 의존한다는데 근거를 두고 있다.

이로써, 단기적인 상황에서 시계열 및 경제적. 外生變數의 변화에 따른 변동추세를 파악함과 동시에 推移確率의 條件을 충족하는 예측치를 제시하게 된다. 이를테면, 各 狀態間의 推移確率 推定

* 한국농촌경제연구원 전산실

値를 구하여 전체 시스템에서의 내부적인 部門間의 동적 변화상태를 分析하고 豫測하게 되며, 아울러 분석치는 각종 정책대안을 계획하고 평가하는데 유용하게 사용될 수 있다. 그동안 이 분야에 관해서는 효율적인 모형개발을 위하여 많은 이론연구가 진행 되어 왔으며[6,12,13,14,17,19], 다양한 경제·사회분야에서 推移 構造를 파악하는데 광범위하게 응용되어 왔다[4,10,11,16,21,22]. 하지만 이용상의 충분한 잠재력이 있음에도 불구하고, 실제적인 실용성이 보다 다양한 모형개발과 추정기법의 활용에 따른 총괄적인 유용성에 관한 연구는 미약한 실정이다.

따라서, 본 연구에서는 개별변동시계열자료(Micro data)를 이용한 Stationary 추이확률의 추정과 총합시계열자료(Macro data)를 이용한 Non-Stationary 추이확률의 추정에 있어서, 기존연구에서 파악된 특성과 한계점을 근거로 하여 활용성측면에서 보다 확장된 모형을 구현하고자 한다.

Stationary 기법에서는 母數的 Empirical Bayes 기법을 추이확률의 추정에 적용함으로써, 기존 Bayes 기법과 같이 주관적인 확신도에 의존하여 사전분포의 母數値를 추정하는 대신 경험적 자료에 의하여 효율적으로 모수치를 추정할 수 있는 대안을 제시하였다. 그리고, Non-Stationary 모형에서는 실용적측면에서 기존연구를 확장하기 위해 시간의 변화에 따른 推移確率行列의 산출방법 개선, 그리고 관측치 및 진입/이탈자료에 관한 관계식을 부가하여, 분석절차를 단순화하면서 현실을 충분히 축약할 수 있도록 통합된 모형을 제시하였다. 또한 이러한 기법들은 산업인력의 이동에 관한 실제사례를 통하여 과거 부문간 이동패턴과 앞으로의 이동 잠재력의 推移에 관한 분석에 적용함으로써 예증되었다.

## 2. Markov Chain에 의한 推移確率 推定技法

본 장에서는 Markov Chain을 이용하는 기존 추이확률모형에 있어서의 문제점을 개선할 수 있는 확장된 Stationary와 Non-Stationary 추정기법의 접근방안을 제시하며, 이에따른 추정절차를 다룬다.

### 2.1 Stationary 추정량

Stationary 推移確率의 추정에 있어서 Anderson and Goodman[6]에 의한 최대우도 추정기법은 추정하고자 하는 未知의 母數가 고정된 것으로 가정하여 관련된 경험치 자료만을 이용하였다. 이에 반하여 Lee et al.[14]에 의한 Bayes 추정기법은 확률변수로 간주하고, 사전 분포를 선택함으로써 사전 및 현재의 정보가 통합될 수 있도록 하였다.

하지만 이러한 확장에도 불구하고, Bayes 기법은 사전 분포의 선택이 母數 특성치에 대한 주관적인 확신도에 의존함으로써, 수집되는 동일한 자료에서 조차도 의사결정자에 따라 각기 다른 사전분포의 선택에 의하여 상이한 결론에 도달할 수도 있다. 따라서 실제 적용에 있어서는 이러한 점이 사전에 충분히 고려되어야만 한다. 한편, 이에 대한 대안으로서 사전분포의 선택보다는 신뢰성 있는 과거 경험치자료를 직접 이용하는 Empirical Bayes 기법의 적용을 모색할 수 있다. 즉, Empirical Bayes 기법은 이론적으로는 推定 母數의 특성이 Bayes 기법과 동일하지만, 사전 분포를 실제 부대자료(collateral data)를 이용한 과거 推定値 형태의 사전 정보로부터 도출한다는 점에서 상이한 특성을 갖고 있다.

본 절에서는 사전분포를 특정화하면서 추정하

는 母數的 Empirical Bayes 技法을 推移確率 推定에 적용함으로써, 과거경험치를 이용한 moment 추정방식에 근거하여 사전분포의 모수치를 추정하는 대안적인 절차를 제시하고자 한다.

### 2.1.1 模型의 定立

과거경험치를 이용하여 Stationary 推移確率에서의 未知의 母數 P에 대한 사전정보를 도출하고자 한다면, 과거의 각 상태에 대한 종합시계열자료보다는 다음과 같은 일정한 시점에 있어서 각 상태간 이동되는 개별변동시계열자료가 필요하다.

| i \ j | 1 | 2 | 3 | ····· | r | 총 계 |
|---|---|---|---|---|---|---|
| 상태 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ | ····· | $n_{1r}$ | $n_{1.}$ |
| 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ | ····· | $n_{2r}$ | $n_{2.}$ |
| 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ | ····· | $n_{3r}$ | $n_{3.}$ |
| · | · | · | · | · | · | · |
| · | · | · | · | · | · | · |
| · | · | · | · | · | · | · |
| r | $n_{r1}$ | $n_{r2}$ | $n_{r3}$ | ····· | $n_{rr}$ | $n_{r.}$ |
| 총 계 | $n_{.1}$ | $n_{.2}$ | $n_{.3}$ | ····· | $n_{.r}$ | $n_{..}$ |

단, $n_{i.} = \sum_{j=1}^{r} n_{ij}$

$n_{.j} = \sum_{i=1}^{r} n_{ij}$

$n_{..} = \sum_{j=1}^{r} n_{.j} = \sum_{i=1}^{r} n_{i.}$

여기서 시간 t를 고려하면, $n_{ij}(t)$는 시간간격 (t, t+1)에서 상태 i로부터 상태 j로 이동하는 개개의 數로서, 주어진 초기상태 i에 대해서 $n_{.j}(t+1)=\sum_{i=1}^{r} n_{ij}(t)$ $(i,j=1,2,\cdots,r,\ \ t=1,2,\cdots,T)$이며, $n_{ii}(t)$은 상태 i에 계속 남아있는 數를 나타낸다.

즉, 각 상태에서의 2개 이상의 일련의 관측치 벡타들

$(n_{111},n_{112},\cdots,n_{11T};\hat{p}_{11}),(n_{121},n_{122},\cdots,n_{12T};\hat{p}_{12}),\cdots,$

$(n_{211},n_{212},\cdots,n_{21T};\hat{p}_{21}),\cdots,(n_{2r1},n_{2r2},\cdots,n_{2rT};\hat{p}_{2r}),\cdots,$

$(n_{r11},n_{r12},\cdots,n_{r1T};\hat{p}_{r1}),\cdots,(n_{rr1},n_{rr2},\cdots,n_{rrT};\hat{p}_{rr})\ \ t\geq 2$

을 이용하게 된다면 Empirical Bayes 추정기법을 적용할 수 있다.

각 시점 t에서의 상태 i에서 상태 j로 이동하는 개개의 數와 總數로서 구성되어지는 과거경험치 자료들이 동일한 모집단(population)에 속한다고 가정하면, 변동관측치 $n_{ij}(t)(i,j=1,2,\cdots,r,\ t=1,2,\cdots,T)$의 분포는 상태 i로부터 상태 j로 이동의 數가 확률치 $p_{i1},p_{i2},\cdots,p_{ir}$를 갖는 다음의 다항분포

$$P_r(n_{ij}(t)/\mathbf{n(0)}',\ p_{i1},p_{i2},\cdots,p_{ir})$$

$$=\prod_{t,i}\frac{n_{i.}(t-1)!}{n_{i1}(t)!n_{i2}(t)!\cdots n_{ir}(t)!}\left[p_{i1}^{n_{i1}(t)}\cdot p_{i2}^{n_{i2}(t)}\cdots p_{ir}^{n_{ir}(t)}\right]$$

$$i,j=1,2\cdots,r\ \ t=1,2\cdots,T \qquad (1)$$

단, $\mathbf{n(0)}'=\{n_1(0),n_2(0),\cdots,n_r(0)\}'$ ; 시간 t=0에서의 각 상태에 있는 數들의 벡타

$n_{i.}(t-1)=\sum_j n_{ij}(t),\ \sum_{j=1}^{r} p_{ij}=1$

를 따른다고 볼 수 있다. 여기서 관측 불가능한 母數 P는 $\sum_{j=1}^{r} p_{ij}=1$과 $0\leq p_{ij}\leq 1(i,j=1,2,\cdots,r)$를 충족하는 사전확률분포로서, i번째 행에 대한 확률밀도함수는 다음형태의 다변량베타분포

$$g(p_{i1},p_{i2},\cdots,p_{ir}) = \frac{\Gamma(\alpha_{i1}+\alpha_{i2}+\cdots+\alpha_{ir})}{\Gamma(\alpha_{i1})\Gamma(\alpha_{i2})\Gamma(\alpha_{i3})\cdots\Gamma(\alpha_{ir})}$$

$$p_{i1}^{\alpha i1-1}\cdot p_{i2}^{\alpha i2-1}\cdots p_{ir}^{\alpha ir-1} \qquad (2)$$

단, $\alpha_{ij}>0\ \forall j,\ i=1,2,\cdots r$

를 가정한다.

따라서 상기의 조건분포와 사전분포에서 도출되는 사후확률분포는 母數 $n_{ij}+\alpha_{ij}$인 다변량베타분

포이다[14]. 이때 이차손실함수를 이용함으로써 사후확률분포의 평균인 Bayes 推定値는

$$\tilde{p}_{ij} = \frac{\alpha_{ij}+n_{ij}}{\alpha_{i.}+n_{i.}} \tag{3}$$

단, $n_{ij}=\sum_t n_{ij}(t)$, $n_{i.}=\sum_{j=1}^{r} n_{ij}$, $\sum_{j=1}^{r} \tilde{p}_{ij}=1$

이며, 다음과 같은 사전분포와 과거관측치 자료의 함수로서 표현될 수 있다.

$$\tilde{p}_{ij}=\left(\frac{\alpha_{i.}}{\alpha_{i.}+n_{i.}}\right)\frac{\alpha_{ij}}{\alpha_{i.}} + \left(\frac{n_{i.}}{\alpha_{i.}+n_{i.}}\right)\frac{n_{ij}}{n_{i.}} \tag{4}$$

따라서 推定値는

$$E[\tilde{p}_{i1},\tilde{p}_{i2},\cdots,\tilde{p}_{ir}/n_{i1},n_{i2},\cdots,n_{ir}]=Zu+(1-Z)\,\bar{n} \tag{5}$$

단, $Z = \dfrac{\alpha_{i.}}{\alpha_{i.}+n_{i.}}$, $u = \dfrac{\alpha_{ij}}{\alpha_{i.}}$

$$\bar{n} = \frac{1}{T}\sum_{t=1}^{T}\frac{n_{ijt}}{n_{i,t}} \quad i,j=1,2,\cdots,r \quad t=1,2,\cdots,T$$

으로, 집단적 자료인 사전분포의 평균치 u와 개별값의 특성을 반영하는 시간에 따른 T개의 개별 경험치 자료의 평균치 $\bar{n}$과의 가중된 선형함수이다. 이를테면, (5)式에서 적용된 加重値는 $n_{i.}$과 $\alpha_{i.}$의 상대적인 크기에 좌우된다. 즉, n이 증가함에 따라 加重値는 경험치에 근거한 고전적 추정치에 근접하게 되므로, 推定値의 상대적인 정확도는 u와 $\bar{n}$의 가중된 평균에 있어서 각각에 어느 정도의 加重値를 부여하는가에 의존하게 된다.

이러한 사전분포에서의 母數에 관한 추정치는 관측치에 의한 조건분포 내지는 母數 P에 비조건적인 주변분포에 의하여 도출될 수 있다.

사전분포 母數値의 추정방법으로서 최대우도추정은 큰 표본의 크기에서는 불편추정량의 산출이 가능하지만, 적은 표본의 크기에 있어서는 반드시 최적해가 산출된다고 단정하기 어렵다. 또한 추정이 복잡하므로 수치기법이 적용되어야 하며, 특히 0 이나 적은 확률치를 갖는 자료의 추정에는 부적합하다. 이에 반하여, moment 추정기법은 추정이 용이할 뿐아니라 母數推定에서 제한된 영역의 범위 內의 해를 산출할 수 있는 이점이 있다. 推移確率의 추정에 있어서는 적은 시행(trial)의 경우이거나 낮은 확률치를 추정하게 되는 경우가 많을 것이므로, moment 추정량을 이용하는 것이 합리적이다[18].

첫째, 사전분포의 母數値($\alpha_{ij}$, $\alpha_{i.}$)는 과거경험치 자료를 이용하여, 사전분포의 Matching Moment 방식에 의하여 추정할 수 있다. 여기서 분포에서의 주요한 특성치는 1차와 2차 모멘트를 이용함으로써 유도될 수 있으며, 집단통계량에 근거한 불편추정량(unbiased estimator)에 의하여 이러한 구조적 母數를 대체할 수 있다.

r변량 베타분포의 적률(moment), $\mu'_{v_1,\cdots,v_r}$은

$$\mu'_{v_1,\cdots,v_r} = \frac{\Gamma(\alpha_1+v_1)\cdots\cdots\Gamma(\alpha_{ir}+v_r)\Gamma(\alpha_{i.})}{\Gamma(\alpha_{i.}+v_1+v_2+\cdots+v_r)\Gamma(\alpha_{i1})\Gamma(\alpha_{i2})\cdots\cdots\Gamma(\alpha_{ir})} \tag{6}$$

단, $\sum_{j=1}^{r}\alpha_{ij}=\alpha_{i.}$, $i=1,2,\cdots r$

이며 1차, 2차 모멘트로부터 사전분포에 대한 평균과 분산은

$$\mu_{ij} = \alpha_{ij} / \alpha_{i.} \tag{7}$$

$$\sigma_{ij}^2 = \frac{\alpha_{ij}(\alpha_{i.}-\alpha_{ij})}{\alpha_{ij}^2(\alpha_{i.}+1)} \qquad i,j=1,2,\cdots\cdots,r \tag{8}$$

이다. 그리고 각 시점에서 $n_{i.}(n_{i.}=\sum_j n_{ij})$의 시행중 상태 i에서 상태 j로 이동하는 횟수를 $n_{ij}$값로 나타낸다면, 각 상태에 대한 推移確率値, $p_{ij}$의 推定値는 $n_{ij}/n_{i.}$이 된다. 이로써 관측된 경험치가 N개 있다고 가정하면, 이러한 관측치에 근거하여 설정된 조건분포인 다항분포의 평균과 분산은

$$\hat{\mu}_{ob} = (1/N)\sum^{N} n_{ij}/n_{i.} \tag{9}$$

$$\hat{\sigma}^2_{ob} = 1/(N-1)\sum^{N}((n_{ij}/n_{i.})-\hat{\mu}_{ob})^2 \qquad (10)$$

이다. 따라서 관측치 자료만을 사용한 평균과 분산을 사전분포로서 설정된 상기의 다변량베타분포의 평균과 분산의 방정식에 대응(matching)시켜, $\hat{\mu}_{ob}$와 $\hat{\sigma}^2_{ob}$ 으로서 정리하면

$$\hat{\alpha}_{ij} = \frac{\hat{\mu}^2_{ob}(1-\hat{\mu}_{ob})}{\hat{\sigma}^2_{ob}} - \hat{\mu}_{ob} \qquad (11)$$

$$\hat{\alpha}_{i.} = \frac{\hat{\mu}^2_{ob}(1-\hat{\mu}_{ob})}{\hat{\sigma}_{ob}} - 1 \qquad (12)$$

이다. 여기서 사전분포인 다변량베타분포 母數值의 추정치($\hat{\alpha}_{ij}$, $\hat{\alpha}_{i.}$>0)는 항상 양의 값으로 산출되어야만 한다. 이러한 조건을 유도하기 위하여 (9),(10)式을 (11)式에 대입하면,

$$\alpha_{ij} = \frac{\hat{\mu}_{ob}}{\hat{\sigma}^2_{ob}}\left[\frac{1}{N}\sum^{N}(n_{ij}/n_{i.}) - \frac{1}{N-1}\sum^{N}(n_{ij}/n_{i.})^2 + \frac{1}{(N-1)N^2}\left(\sum^{N}n_{ij}/n_{i.}\right)^2\right]$$

$$\alpha_{ij} > \frac{\hat{\mu}_{ob}}{\hat{\sigma}^2_{ob}} \cdot \frac{1}{N}\left[\sum^{N}(n_{ij}/n_{i.})\left(1-\frac{N}{N-1}(n_{ij}/n_{i.})\right)\right]$$

에서 $0 \le n_{ij}/n_{i.} \le (N-1)/N$이면 $\alpha_{ij}>0$ 이다. 단, 표본의 크기(N)가 충분히 커지거나 (10)式의 분산식이 $(1/N)\sum^{N}((n_{ij}/n_{i.})-\hat{\mu}_{ob})^2$으로 설정된다면, $n_{ij}/n_{i.}$에 대한 제약은 자동적으로 충족되게 된다. 추정치가 $\alpha_{ij}>0$ 이라면, (7)式에서 또한 $\alpha_{i.}=\alpha_{ij}/\hat{\mu}_{ob}>0$ 임을 알 수 있다.

둘째, 사전분포의 母數值는 사전분포와 조건분포를 P에 대하여 가중적분된 주변분포에 대한 Matching Moment 방식으로 추정할 수 있다. 주변분포는

$$h(n/\alpha_{ij}, \alpha_{i.}) = \int f(n/p)g(p/\alpha_{ij}, \alpha_{i.})dp \qquad (13)$$

$$= \binom{M}{n} \frac{\Gamma(\alpha_{i.})}{\Gamma(\alpha_{i1})\Gamma(\alpha_{i2})\cdots\Gamma(\alpha_{ir})} \cdot \frac{\Gamma(\alpha_{i1}+n_{i1})\Gamma(\alpha_{i2}+n_{i2})\cdots\Gamma(\alpha_{ir}+n_{ir})}{\Gamma(\alpha_{i.}+n_{i.})} \quad \text{단, } M=n_{i.}$$

로서, 0과 1사이의 크기로 나타낸 평균과 분산은

$$\hat{\mu}_{ob} \equiv E(n_{ij}/\alpha_{ij}, \alpha_{i.}) = \alpha_{ij}/\alpha_{i.} \qquad (14)$$

$$\hat{\sigma}^2_{ob} \equiv Var(n_{ij}/\alpha_{ij}, \alpha_{i.}) = 1/M(\alpha_{ij}/\alpha_{i.})(1-\alpha_{ij}/\alpha_{i.})\left(\frac{\alpha_{i.}+M}{\alpha_{i.}+1}\right) \qquad i,j=1,2,\cdots,r \qquad (15)$$

이다. 전술한 사전분포에 대한 Matching Moment 방식과 동일하게, 관측치에 근거한 다항분포에서 평균과 분산을 도출하여 상기의 주변분포의 평균과 분산식에 대응시킨다. 즉, (14)式의 관계식 $\alpha_{ij}=\hat{\mu}_{ob}\alpha_{i.}$를 (15)式에 대입하면

$$\alpha_{ij} = \left[\frac{M\hat{\mu}_{ob}(1-\hat{\mu}_{ob})-M\hat{\sigma}^2_{ob}}{M\hat{\sigma}^2_{ob}+\hat{\mu}^2_{ob}-\hat{\mu}_{ob}}\right]\hat{\mu}_{ob} \qquad (16)$$

$$\alpha_{i.} = \frac{M\hat{\mu}_{ob}(1-\hat{\mu}_{ob})-M\hat{\sigma}^2_{ob}}{M\hat{\sigma}^2_{ob}+\hat{\mu}^2_{ob}-\hat{\mu}_{ob}} \qquad (17)$$

으로 나타낼 수 있다. 단, 이러한 母數 推定值가 적용자료에 있어서의 확률치 특성으로 인하여 모두 양의 값을 산출하지 않는 경우에는, 추정절차가 보완되어야 한다. 이제 推定值($\hat{\alpha}_{ij}$, $\hat{\alpha}_{i.}$)들을 (5)式에 대입함으로써 Empirical Bayes 推定值가 산출되게 된다.

이로써, 이러한 추정절차는 과거 경험치에 대한 단순한 계산에 의하여 推定確率의 제약조건에 만족하는 추정치를 구할 수 있을 뿐아니라, 예측치의 산출식으로 사용될 수 있는 이점을 갖고 있다.

### 2.2 Non-Stationary 추정량

Non-Stationary 推移 構造에 관한 실제적인 상황을 폭넓게 다루려는 의도에서 기존연구 Lee et al. [13], Mellor[17], Multi-Logit[10], 강정혁 · 김여근[4]로부터 확장된 추정방안을 제시하였으며, 구체적인 내용은 다음과 같다.

첫째, 推移確率 模型에서의 피설명변수로는 실측치 내지는 비율치가 사용될 수 있는데, 대부분의 기존 연구가 비율치를 이용하고 있다. 의사 결정에 있어서 각각의 시점에 대하여 표본 크기에 무관한 상대적인 크기를 전제한다면, 비율치에 의한 함수의 설정이 적절할 것이다. 하지만 각 실측치가 모든 시점에서 독립적으로 할당된다면, 이때 실측치는 각기 할당된 總合値로서 구현되며, 이것은 각 시점에서의 큰 값의 실측치가 적은 값의 실측치보다 더 민감하게 반응하게 될 것이다. 또한 실측치에 의한 구현은 비율치를 사용함에 따라 전체총합치를 별개로 산출하여 추정하게 되는 번거러움을 줄일 수 있는 이점이 있다.

둘째, 기존 연구에서는 조사기간 중의 각 상태에서의 새로운 增減(entry /exit)자료는 분석되지 않고, 전체 비율치 자체에 포함하여 고려되었다. 하지만 새로운 증감 정보도 특정식으로 도출되어 추계되어야할 필요가 있다. 따라서, 시간에 따라 새로운 증감이 외부적인 영향下에서 변동되는 것을 포괄한 통합된 분석을 목표로 한다.

세째, 각 년도에서의 상태별 총합비율치 자체만을 도출하는 기존결과치를 확장하여, 예측된 推移確率의 推定値가 외생변수의 변화에 따른 각각의 상태간 推移確率行列 형태로 산출되도록 하였다. 이에따라 내부적인 각 부분간의 움직임이 추적되므로 전체적인 변동에 대한 이해를 확대시킬 수 있을 뿐아니라, 외생변수와의 관계에 있어서 어떤 명기된 목표결과를 현실화하기 위하여 어떻게 변경되어야만 하는가를 일목요연하게 제시할 수 있는 이점을 갖게 된다.

#### 2.2.1 模型의 定立

1) 推移確率 函數의 設定

Non-Stationary 推移確率이 주어졌을 때, 특정시점에서 각 상태에 있을 확률은 바로 전기의 각 상태에 있을 확률에 의존된다는 가정에 근거하는 第1단계 Markov 과정을 이용하면

$$n_{jt}=\sum_{i=1}^{r}p_{ijt}n_{it-1}+u_{jt} \qquad j=1,2,\cdots,r \tag{18}$$

단, $n_{jt}$ : 시점 t상태 j에서의 총합 관측치

$n_{it-1}$ : 시점 t−1 상태 i에서의 총합 관측치

$p_{ijt}$ : Non-Stationary 推移確率 推定量

$u_{jt}$ : 시점 t에서 상태 j에 대한 disturbance 벡타

로서 나타낼 수 있다. Non-Stationary 推移確率値, $p_{ijt}$는 상태간 시간의 변화에 따른 외생변수치, $z_{ijt}$에 의하여 영향을 받는다.

하지만, 상기의 기본회귀식에서는 $0\leq p_{ijt}\leq 1$으로서 제약된 推移確率値가 산출되기 어렵기때문에 式의 변환과정이 요구된다. 여기서는 외생변수치, $z_{ijt}$가 증가함에 따라서 推移確率値가 1에 서서히 수렴하는 Logistic 함수를 따르는 것으로 가정한다.

이로써, 推移確率의 非陰 제약조건($0\leq p_{ijt}\leq 1$)과 row sum 제약조건($\sum_{j=1}^{r}p_{ijt}=1$)이 모든 시간에 결쳐서 다음 式에서 각각 자동적으로 충족될 수 있도록 하였다.

$$\tilde{p}_{ijt}=(1+\mathrm{Exp}(\beta_{io}+\beta_{i1}z_{ij1t},\cdots+\beta_{il}z_{ijlt}))^{-1} \tag{19}$$

단, $\beta_{io}$, $\beta_{i1},\cdots,\beta_{ik}$ : 절편 및 기울기

$$\tilde{p}_{iit}=1-\sum_{j=1,j\neq i}^{r}\tilde{p}_{ijt}\geq 0 \quad i,j=1,2,\cdots,r \tag{20}$$

또한 각 상태간 이동되는 외생변수치 $z_{ijt}(\forall i,j,t)$는 주어지지 않기때문에 확률개념을 도입한다.

즉, 상태간 이동에 관한 외생변수는 시간의 변화에 따른 상태간 상대적인 비율치를 고려하여,

$$z_{ijt} = \frac{z_{jt}}{z_{it}+z_{jt}} \tag{21}$$

단, $z_{it}$ : 시점 t에서 상태 i의 외생변수치

$z_{jt}$ : 시점 t에서 상태 j의 외생변수치

로서 정의하였다. 이를테면, 본 모형에서 3상태의 경우에 있어서 하나의 외생변수를 갖게 될 때

$$\begin{bmatrix} n_{1t} \\ n_{2t} \\ n_{3t} \end{bmatrix} = \begin{bmatrix} (1+\mathrm{Exp}(\beta_{10}+\beta_{11}z_{11t}))^{-1}, (1+\mathrm{Exp}(\beta_{20}+\beta_{21}z_{21t}))^{-1}, (1+\mathrm{Exp}(\beta_{30}+\beta_{31}z_{31t}))^{-1} \\ (1+\mathrm{Exp}(\beta_{10}+\beta_{12}z_{12t}))^{-1}, (1+\mathrm{Exp}(\beta_{20}+\beta_{22}z_{22t}))^{-1}, (1+\mathrm{Exp}(\beta_{30}+\beta_{32}z_{32t}))^{-1} \\ (1+\mathrm{Exp}(\beta_{10}+\beta_{13}z_{13t}))^{-1}, (1+\mathrm{Exp}(\beta_{20}+\beta_{23}z_{23t}))^{-1}, (1+\mathrm{Exp}(\beta_{30}+\beta_{33}z_{33t}))^{-1} \end{bmatrix}$$

$$\begin{bmatrix} n_{1t-1} \\ n_{2t-1} \\ n_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \tag{22}$$

으로 표현될 수 있다. 하지만 (22)式에서는 설명변수의 행렬이 선형결합관계를 갖게 됨으로써 심각한 다중공선성의 문제점을 야기할 수 있다. 따라서 다음의 2가지 제약조건을 부가하였다.

첫째, $\beta_{i0}=\beta_i$, $\beta_{ij}=\beta_i{'}$ $\forall j$를 가정함으로써 (22)式은 6개의 母數를 추정하는 式으로 줄이면서, 절편($a_1$, $a_2$, $a_3$)을 포함한

$$\begin{aligned} n_{1t} &= a_1+(1+\mathrm{Exp}(\beta_1+\beta_1{'}z_{11t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_2+\beta_2{'}z_{21t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_3+\beta_3{'}z_{31t}))^{-1}\, n_{3t-1}+u_{1t} \\ n_{2t} &= a_2+(1+\mathrm{Exp}(\beta_1+\beta_1{'}z_{12t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_2+\beta_2{'}z_{22t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_3+\beta_3{'}z_{32t}))^{-1}\, n_{3t-1}+u_{2t} \\ n_{3t} &= a_3+(1+\mathrm{Exp}(\beta_1+\beta_1{'}z_{13t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_2+\beta_2{'}z_{23t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_3+\beta_3{'}z_{33t}))^{-1}\, n_{3t-1}+u_{3t} \end{aligned}$$

으로 변환하였다. (23)

둘째, (23)式에 있어서 $\beta_i=\beta_0, \beta_i{'}=\beta_1$를 가정함으로써 2개의 母數와 절편($a_1$, $a_2$, $a_3$)을 갖는 다음과 같은 추정식으로 축소되었다

$$\begin{aligned} n_{1t} &= a_1+(1+\mathrm{Exp}(\beta_0+\beta_1 z_{11t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{21t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{31t}))^{-1}\, n_{3t-1}+u_{1t} \\ n_{2t} &= a_2+(1+\mathrm{Exp}(\beta_0+\beta_1 z_{12t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{22t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{32t}))^{-1}\, n_{3t-1}+u_{2t} \\ n_{3t} &= a_1+(1+\mathrm{Exp}(\beta_0+\beta_1 z_{13t}))^{-1}\, n_{1t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{23t}))^{-1}\, n_{2t-1} \\ &\quad +(1+\mathrm{Exp}(\beta_0+\beta_1 z_{33t}))^{-1}\, n_{3t-1}+u_{3t} \end{aligned} \tag{24}$$

여기서 推移確率 推定値는 推移確率의 제약조건($\sum_j p_{ijt}=1$)이 항상 충족되지만 예측치는 상황에 따라 충족되지 않는 경우가 있다. 이로써 예측치의 산출에서는 상태간의 이동되는 推移確率 豫測値($\tilde{p}_{ijt}$)를 우선 산출한 후, 1에서 제외된 값을 상태간의 이동이 없는 고정된 推移確率 豫測値($\tilde{p}_{iit}$)로서 고려하였다.

2) 새로운 增減(entry /exit)에 관한 推定

제반 환경의 변화에 따른 각 상태에서의 새로운 增減(entry /exit)에 관한 推移는 과거경험치와 외생요인의 로그변환(logarithmic transformation)을 통한 다음의 선형추정식

$$(\mathrm{entry}-\mathrm{exit})_{it} = \gamma_0+\gamma_1 \ln z_{i1t} \cdots\cdots + \gamma_k \ln z_{ikt}$$

$$i=1,2,\cdots\cdots,r \tag{25}$$

단, $\gamma_0, \gamma_1, \cdots, \gamma_k$ : 절편 및 기울기

의 분석을 고려함으로써, 외부현상과 관련된 상황

을 체계적으로 다룰 수 있도록 한다.

## 3. 事例硏究를 통한 實證分析

본 장에서는 앞에서 제시된 모형들을 산업인력에 관한 실제사례를 통하여 예증한다. Stationary 추이확률 추정치와 외생요인에 따른 Non-Stationary 추정 및 예측치를 산출하고, 이러한 결과치들의 유효성파악과 성능분석을 수행하고자 한다.

### 3.1 分析對象

특정시스템 內의 이동현상을 분석하기 위해서는, 전술한 바와 같이 경제, 사회, 경영학분야 등에서의 推移현상에 있어서의 다양한 응용을 고려해 볼 수 있다.

우선 제시된 모형들에 대한 사례연구를 통한 실증분석을 위해서는 개별변동시계열자료와 총합시계열자료가 요구된다. 이에대한 사례별 자료사용 가능여부를 검토해 본 결과, 산업의 인력이동에 관한 자료 外에는 관련자료가 구비되지 않고 있다.

특히 산업인력의 이동에 관한 산업, 직종별 인력이동자료는 경제기획원에서 산업 및 직종별로 전수조사와 표본조사에 의하여 각각 실측된 공신력있는 자료이다. 전술한 推移確率模型의 분석대상으로 우리나라 산업별 취업인력에 관한 과거 년도별 총합시계열자료[2]와 1983년부터 3년주기로 실시되고 있는 고용구조 특별조사에 의한 직종별 개별변동시계열자료[1]를 이용하였다. 이에 따라, 본 사례연구의 목적은 우리나라 노동력의 이동에 관한 사례를 통하여 확장된 모형에 있어서 실제적인 유용성을 검토하는데 있다.

### 3.2 Stationary 推移確率의 推定

Stationary 모형에 있어서 직종별 개별변동시계열자료[1]('84, '87, '90)를 이용하여 비경제활동인구와 실업자를 포함한 노동력의 직종간 이동을 분석하였다.

사전 및 주변분포를 이용한 Matching Moment 방식에 의하여 Empirical Bayes 기법이 적용되었다. 주변분포를 이용한 경우를 例로 하여 모두 陽의 값을 갖는 사전분포의 母數值($\alpha_{ij}, \alpha_{i.}$)에 관한 결과치가 〈표 1〉에 요약되어 있다. 여기서 (5)式에 의하여 산출된 Micro Empirical Bayes 추정치는 〈표 2〉와 같다.

1) 전문기술행정사무, 판매서어비스, 농림수산, 생산운수운전노무, 실업자, 비경제활동인구를 각각 상태 1에서 6까지 설정

〈표 1〉 주변분포의 Matching Moment기법에 의한 事前分布의 母數値(직종별)

| | 평균 | 표준편차 | $\alpha_{ij}$ | $\alpha_{i.}$ |
|---|---|---|---|---|
| $P_{11}$ | 0.9343 | 0.01550 | 237.78 | 254.50 |
| $P_{12}$ | 0.0133 | 0.00404 | 10.68 | 803.03 |
| $P_{13}$ | 0.0023 | 0.00115 | 3.99 | 1734.13 |
| $P_{14}$ | 0.0083 | 0.00208 | 15.78 | 1901.53 |
| $P_{15}$ | 0.0133 | 0.00252 | 27.47 | 2065.50 |
| $P_{16}$ | 0.0280 | 0.006608 | 20.59 | 735.24 |
| $P_{21}$ | 0.0057 | 0.00208 | 7.46 | 1308.98 |
| $P_{22}$ | 0.9397 | 0.01589 | 209.95 | 223.42 |
| $P_{23}$ | 0.0027 | 0.00153 | 3.10 | 1149.29 |
| $P_{24}$ | 0.0160 | 0.00458 | 11.99 | 749.56 |
| $P_{25}$ | 0.0127 | 0.00321 | 15.44 | 1215.87 |
| $P_{26}$ | 0.0233 | 0.00643 | 12.80 | 549.42 |
| $P_{31}$ | 0.0013 | 0.00058 | 5.02 | 3858.42 |
| $P_{32}$ | 0.0067 | 0.00058 | 132.54 | 19782.32 |
| $P_{33}$ | 0.9690 | 0.00458 | 1386.68 | 1432.04 |
| $P_{34}$ | 0.0123 | 0.00115 | 112.98 | 9185.17 |
| $P_{35}$ | 0.0017 | 0.00115 | 2.18 | 1282.26 |
| $P_{36}$ | 0.0087 | 0.00289 | 8.98 | 1031.59 |
| $P_{41}$ | 0.0063 | 0.00153 | 16.84 | 2673.32 |
| $P_{42}$ | 0.0190 | 0.00346 | 29.56 | 1555.93 |
| $P_{43}$ | 0.0060 | 0.00300 | 3.97 | 661.67 |
| $P_{44}$ | 0.9253 | 0.01498 | 284.09 | 307.02 |
| $P_{45}$ | 0.0200 | 0.00656 | 9.09 | 454.46 |
| $P_{46}$ | 0.0240 | 0.00346 | 46.94 | 1955.63 |
| $P_{51}$ | 0.1117 | 0.03027 | 11.98 | 107.29 |
| $P_{52}$ | 0.1597 | 0.01172 | 155.86 | 975.98 |
| $P_{53}$ | 0.0250 | 0.01778 | 1.90 | 76.10 |
| $P_{54}$ | 0.2727 | 0.01686 | 190.00 | 696.72 |
| $P_{55}$ | 0.3080 | 0.02551 | 100.57 | 326.52 |
| $P_{56}$ | 0.1223 | 0.02371 | 23.23 | 189.95 |
| $P_{61}$ | 0.0143 | 0.00252 | 31.73 | 2218.63 |
| $P_{62}$ | 0.0250 | 0.00100 | 609.35 | 24374.00 |
| $P_{63}$ | 0.0127 | 0.00929 | 1.83 | 144.29 |
| $P_{64}$ | 0.0247 | 0.00252 | 93.67 | 3792.45 |
| $P_{65}$ | 0.0137 | 0.00153 | 79.07 | 5771.27 |
| $P_{66}$ | 0.9097 | 0.00603 | 2054.27 | 2258.18 |

〈표 2〉 직종별 Micro Empirical Bayes 추정치

| | 전문기술 행정사무 | 판매 · 서어비스 | 농림 수산 | 생산운수 운전노무 | 실업자 | 비경제 활동인구 |
|---|---|---|---|---|---|---|
| 전문기술 · 행정사무 | 0.936 | 0.013 | 0.002 | 0.008 | 0.013 | 0.028 |
| 판매 · 서어비스 | 0.006 | 0.942 | 0.002 | 0.016 | 0.012 | 0.,022 |
| 농림수산 | 0.001 | 0.007 | 0.969 | 0.012 | 0.002 | 0.009 |
| 생산운수 · 운전노무 | 0.006 | 0.018 | 0.006 | 0.927 | 0.019 | 0.024 |
| 실업자 | 0.113 | 0.160 | 0.024 | 0.272 | 0.307 | 0.123 |
| 비경제 활동인구 | 0.015 | 0.025 | 0.013 | 0.025 | 0.014 | 0.910 |

이것은 실업자와 비경제활동인구를 포함한 표본조사를 이용한 분석결과로서, 특히 실업자와 비경제활동인구는 생산 · 운수 · 운전 · 노무직과 판매 · 서어비스직종으로 어느정도 진입되고 있음을 나타내고 있다.

### 3.3 Non－Stationary 推移確率의 推定

현실적으로 산업의 구조에 있어서 크기의 변화는 주로 외생적인 경제적요인 내지는 비경제적 요인들에 의하여 영향을 받는다고 가정할 수 있다. 이러한 요인들에 있어서 계량화할 수 있는 가장 중요한 자료로는 산업별 취업인력 數, 산업별 근로자 1인당 월평균 임금, 노동생산성을 나타내는 지표의 하나인 산업별 취업자 1인당 부가가치, 그리고 산업별 국내 총생산 등을 들 수 있다.

여기서, 외생변수는 가급적 확실하게 예측이 가능한 변수만을 이용하여야 하며, 자료의 사용에 무리가 없을뿐 아니라, 산업별 취업구조를 일관되게 설명해줄 수 있어야 한다는 조건을 충족시켜야 한다. 본 사례연구에서는 이러한 제약들을 고려한 변수의 선택으로서, 산업별 취업인력의 數[2] ('72～'91)를 피설명변수로 하고 설명변수로 해당 부문에서의 전년도 산업별 취업인력의 數와 산업별 국내총생산('72～'91)을 선택하였다. 이로써 노동력의 부문간 이동이 외생적인 요인에 의하여 어떻게 유발된 것인가한 실증분석을 수행하였다.

#### 3.3.1 Macro 推定值 및 豫測值

Non－Stationary 추이확률 추정절차는 제한된 총합시계열자료를 이용하면서 제약조건들을 동시에 만족하는 연립방정식 시스템을 풀어야 하므로, 해절차는 단순하지가 않다. 따라서 방정식이 상호종속인 시스템에서의 계수치를 추정하기 위하여, SAS 통계패키지를 이용하여 외견무관회귀(SUR) 기법으로서 계측되었다.

(23)式에서 절편($a_1, a_2, a_3$)을 제외한 추정식은 Ⅰ－1으로, 동일하게 (24)式에서 절편($a_1, a_2, a_3$)을 제외한 추정식은 Ⅱ－1로서 나타내었다.

각 산업별 상대적 소득이 산업별 취업분포의 전개에 어떠한 역활을 하고 있는가를 예증하고 있다. (19)式에서 양(음)의 $\hat{\beta}$ 値는 대응하는 변수가 推移確率値 $\tilde{p}_{ijt}$上에 시간의 변화에 따른 유한한 음(양)의 영향력을 갖는다는 것을 의미한다. 〈표 3〉의 추정결과에서, 추정식 Ⅰ은 산업간 상대

2) 농림어업, 광공업, 사회간접 및 기타부문을 각각 상태 1,2,3으로 설정

적 소득에 대해서 負의 유의한 관계를 나타냄으로써 해당 산업부문으로의 이동 확률을 증가시킬 것으로 보이며, 추정식 Ⅰ－1에서 $\hat{\beta}_2'$ 는 양의 유의한 관계로서 상대적 소득에 대해서 광공업부문에서의 이동확률은 감소되는 추세를 제시하고 있다. 결정계수 $R^2$에 의하면 양호한 모형의 설명력을 나타내고 있으며, 특히 추정식 Ⅰ에서 사회간접·기타부문과 타부문간의 상대적 소득에 따른 추정계수($\hat{\beta}_3$, $\hat{\beta}_3'$)의 t値는 5% 유의수준에서 유의하지 않음을 나타냄에 따라, 대체로 이러한 부문에서는 경쟁적인 관계가 없거나 적은 것으로 추론된다. 확장된 모형(Ⅱ)에서 추정식Ⅱ는 절편($\hat{a}_1$)의 t値가 5% 유의수준에서 유의하지 않은 것으로 나타났으나, 전반적으로 추정치들이 상대적 소득에 비례하는 현실적인 추이구조를 반영하고 있다.

〈표 3〉 확장된 모형(Ⅰ)에 대한 母數 推定値

| | Ⅰ (23)式 | | Ⅰ－1 |
|---|---|---|---|
| $\hat{\beta}_1$ | 3.983552 | $\hat{\beta}_1$ | 11.409722 |
| | ( 6.20 ) | | ( 7.75 ) |
| $\hat{\beta}_1'$ | －2.481708 | $\hat{\beta}_1'$ | －12.193498 |
| | ( －2.84 ) | | ( －7.24 ) |
| $\hat{\beta}_2$ | 4.116693 | $\hat{\beta}_2$ | －0.268962 |
| | ( 2.69 ) | | ( －0.72 ) |
| $\hat{\beta}_2'$ | －4.952949 | $\hat{\beta}_2'$ | 6.231174 |
| | ( －1.92 ) | | ( 4.50 ) |
| $\hat{\beta}_3$ | 17.388105 | $\hat{\beta}_3$ | 6.682752 |
| | ( 1.06 ) | | ( 3.87 ) |
| $\hat{\beta}_3'$ | －35.47068 | $\hat{\beta}_3'$ | －11.579918 |
| | ( －0.87 ) | | ( －2.76 ) |
| $\hat{a}_1$ | 533.237416 | | |
| | ( 1.23 ) | | |
| $\hat{a}_2$ | 951.044045 | | |
| | ( 3.56 ) | | |
| $\hat{a}_3$ | －1062.33 | | |
| | (－3.73 ) | | |
| $R^2(1)$ | 0.9735 | $R^2(1)$ | 0.9525 |
| $R^2(2)$ | 0.9759 | $R^2(2)$ | 0.9457 |
| $R^2(3)$ | 0.9971 | $R^2(3)$ | 0.9923 |

여기서 해당기간 동안에 있어서 각 부문에 진입하거나 이탈하는 취업자 수를 포함하여 동시에 추계되어야 하겠으나, 이용가능한 자료에 있어서의 제약으로 새로운 증감에 관한 파악은 수행되지 못하였다.

이러한 母數 推定値를 이용하여 년도별 추이확률 추정치를 산출하였다. 추정식Ⅰ과 Ⅰ－1에서는 기대한 바의 추정치가 산출되지 못하였다. 이에 반하여 추정식 Ⅱ와 Ⅱ－1에서의 년도별 추이확률 추정치는 대체로 사전기대치와 일치하는 결

〈표 4〉 확장된 모형(Ⅱ)에 대한 母數 推定値

| | Ⅱ (24)式 | | Ⅱ－1 |
|---|---|---|---|
| $\hat{\beta}_0$ | 8.789941 | $\hat{\beta}_0$ | 7.382750 |
| | ( 1.99 ) | | ( 6.04 ) |
| $\hat{\beta}_1$ | －6.86936 | $\hat{\beta}_1$ | －5.408974 |
| | ( －1.49 ) | | ( －3.65 ) |
| $\hat{a}_1$ | 97.107305 | | |
| | ( 0.55 ) | | |
| $\hat{a}_2$ | 151.777003 | | |
| | ( 2.74 ) | | |
| $\hat{a}_3$ | 173.063061 | | |
| | ( 1.22 ) | | |
| $R^2(1)$ | 0.9675 | $R^2(1)$ | 0.9399 |
| $R^2(2)$ | 0.9637 | $R^2(2)$ | 0.9453 |
| $R^2(3)$ | 0.9937 | $R^2(3)$ | 0.9895 |

註：1) ( )안은 t値임

2) $R^2(j)$는 j번째 상태에 대한 방정식의 $R^2$値임

과치로서, 각 산업부문에 있어서 계속 고정된 확률과 산업간 상대적소득에 비례하는 이동확률을 추정식 Ⅱ를 例로 하여 〈표 5〉에서 제시하고 있다.

또한 외생변수치에 관한 예측치를 산출하기 위하여 기존연구[4]에서의 보정식을 이용하였다. 사전정보로서 최근의 각 산업별 국내총생산 성장율자료[3]('81~'91)에 있어서 과거 3개년도의 평균치를 가정하여, 과거자료에 대한 기대치를 포괄한 예측치가 산출되었다. 이에따라, 확장된 모형에서 추정식(Ⅱ)를 例로 하여 산출된 추이확률 요소의 예측결과는 〈표 6〉과 같다. 동일한 산업부문에 계속 종사할 확률이 0.874과 0.9978 사이에서 변화하고 있으며, 광공업과 사회간접·기타부문은 높은 고정확률을 유지하고 있다. 그리고 사회간접·기타부문은 상대적으로 타부문으로부터 높은 전입확률을 나타내고 있다.

이러한 결과는 과거역사로 볼 때, 산업 취업인력의 數가 산업별 국내총생산 격차의 증가에 따라 노동력 이동을 겪어왔다는 사실과 일치함으로써 합리적인 것으로 여겨진다.

〈표 5〉 확장된 모형(Ⅱ)에 대한 년도별 推移確率 推定値

| 년도 | $P_{11}$ | $P_{21}$ | $P_{31}$ | $P_{12}$ | $P_{22}$ | $P_{32}$ | $P_{13}$ | $P_{23}$ | $P_{33}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1973 | 0.9786 | 0.0065 | 0.0012 | 0.0034 | 0.9701 | 0.0009 | 0.0180 | 0.0234 | 0.9979 |
| 1974 | 0.9783 | 0.0058 | 0.0012 | 0.0038 | 0.9731 | 0.0010 | 0.0179 | 0.0212 | 0.9978 |
| 1975 | 0.9769 | 0.0050 | 0.0012 | 0.0044 | 0.9754 | 0.0011 | 0.0187 | 0.0196 | 0.9977 |
| 1976 | 0.9759 | 0.0042 | 0.0012 | 0.0052 | 0.9785 | 0.0013 | 0.0189 | 0.0173 | 0.9976 |
| 1977 | 0.9727 | 0.0035 | 0.0010 | 0.0063 | 0.9801 | 0.0013 | 0.0210 | 0.0164 | 0.9976 |
| 1978 | 0.9621 | 0.0022 | 0.0008 | 0.0100 | 0.9829 | 0.0015 | 0.0278 | 0.0149 | 0.9978 |
| 1979 | 0.9618 | 0.0021 | 0.0008 | 0.0105 | 0.9835 | 0.0015 | 0.0277 | 0.0144 | 0.9977 |
| 1980 | 0.9492 | 0.0015 | 0.0006 | 0.0145 | 0.9833 | 0.0014 | 0.0364 | 0.0151 | 0.9980 |
| 1981 | 0.9538 | 0.0016 | 0.0007 | 0.0135 | 0.9845 | 0.0016 | 0.0326 | 0.0139 | 0.9978 |
| 1982 | 0.9539 | 0.0017 | 0.0007 | 0.0133 | 0.9840 | 0.0015 | 0.0329 | 0.0144 | 0.9978 |
| 1983 | 0.9515 | 0.0015 | 0.0006 | 0.0145 | 0.9848 | 0.0016 | 0.0339 | 0.0137 | 0.9977 |
| 1984 | 0.9436 | 0.0012 | 0.0006 | 0.0187 | 0.9867 | 0.0018 | 0.0377 | 0.0122 | 0.9976 |
| 1985 | 0.9413 | 0.0011 | 0.0005 | 0.0196 | 0.9866 | 0.0018 | 0.0391 | 0.0123 | 0.9977 |
| 1986 | 0.9354 | 0.0009 | 0.0005 | 0.0231 | 0.9872 | 0.0020 | 0.0415 | 0.0111 | 0.9975 |
| 1987 | 0.9199 | 0.0007 | 0.0004 | 0.0309 | 0.9890 | 0.0021 | 0.0492 | 0.0103 | 0.9974 |
| 1988 | 0.9172 | 0.0007 | 0.0004 | 0.0325 | 0.9893 | 0.0022 | 0.0503 | 0.0100 | 0.9974 |
| 1989 | 0.9120 | 0.0006 | 0.0004 | 0.0339 | 0.9885 | 0.0020 | 0.0541 | 0.0109 | 0.9976 |
| 1990 | 0.8998 | 0.0005 | 0.0003 | 0.0393 | 0.9881 | 0.0019 | 0.0609 | 0.0113 | 0.9977 |
| 1991 | 0.8914 | 0.0005 | 0.0003 | 0.0429 | 0.9879 | 0.0019 | 0.0653 | 0.0116 | 0.9978 |

〈표 6〉 확장된 모형(Ⅱ)에서의 推移確率 要素에 대한 예측치

| 년도 | $P_{11}$ | $P_{21}$ | $P_{31}$ | $P_{12}$ | $P_{22}$ | $P_{32}$ | $P_{13}$ | $P_{23}$ | $P_{33}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1992 | 0.8881 | 0.0005 | 0.0003 | 0.0446 | 0.9877 | 0.0019 | 0.0673 | 0.0118 | 0.9978 |
| 1993 | 0.8825 | 0.0005 | 0.0003 | 0.0472 | 0.9870 | 0.0019 | 0.0703 | 0.0125 | 0.9978 |
| 1994 | 0.8784 | 0.0005 | 0.0003 | 0.0491 | 0.9866 | 0.0019 | 0.0725 | 0.0129 | 0.9977 |
| 1995 | 0.8740 | 0.0005 | 0.0003 | 0.0512 | 0.9862 | 0.0019 | 0.0748 | 0.0133 | 0.9976 |

### 3.4 性能 分析

축조된 推移確率行列로부터 산출된 추정치는 정확한 예측능력이 과거 동일한 기간에서의 관측된 실제치와 비교되었다.

Stationary 추정치에 관한 예측성능의 평가를 위하여, 평균자승오차(M.S.E.)를 이용하여 1985, 88, 91년에서의 예측치를 실적치와 對比한 성능비교치가 〈표 7〉에 제시되어 있다. 단지 한정적인 3개년도의 적은 자료만을 이용함에 따라 얻어진 결과이지만, 전반적으로 Empirical Bayes 추정치는 평균자승오차의 측면에서 Micro 최대우도 추정치보다 정밀도(precision)를 나타내고 있으며, 주변분포를 이용한 경우와 사전분포에 의한 경우가 유사한 성능을 나타내고 있음을 알 수 있다. 반면에, 주관적인 사전정보를 이용하는 Bayes 추정치와는 경험치자료에 의하여 객관적인 성능비교가 어려우므로, 추후 시뮬레이션에 의한 의태자료(Simulation data)를 생성하여 성능분석이 수행되어야할 것이다.

〈표 7〉 Stationary 추정치에 대한 M.S.E. 비교(직종별)

| 구 분 | Micro 최대우도 | Empir.Bayes(사전분포) | Empir.Bayes(주변분포) |
|---|---|---|---|
| 전문기술 · 행정사무 | 0.0000108 | 0.0000091 | 0.0000084 |
| 판매 · 서비스 | 0.0000966 | 0.0001011 | 0.0001018 |
| 농림수산 | 0.0001508 | 0.0001507 | 0.0001508 |
| 생산운수 · 운전노무 | 0.0000163 | 0.0000167 | 0.0000169 |
| 실업자 | 0.0000086 | 0.0000082 | 0.0000081 |
| 비경제활동 인구 | 0.0005067 | 0.0004904 | 0.0004897 |
| 전산업 | 0.0001588 | 0.0001564 | 0.0001558 |

하지만, 일반적으로 Empirical Bayes 추정치의 위험(risk)은 Robbins에 의하면, 많은 경험치를 이용하게 됨에 따라 최적인 Bayes 위험에 근접한다는 "접근적 최적"을 제시하였다[15]. 또한 기존 연구들에 있어서도 10-15개년의 경험치가 활용됨에 따라, Empirical Bayes 추정치가 Bayes 추정치보다 적은 평균자승오차를 갖게 됨을 실증한 바 있다[8].

이러한 측면에서, 본 연구에서 제시된 Empirical Bayes 추정절차도 추후 충분한 경험치 자료가 축적되어 이용됨에 따라서, 최적의 Bayes 추정치에 근접하는 양호한 예측성능을 갖게될 수 있으리라 판단된다.

또한 Non-Stationary 모형에 있어서의 각 추정식에 대한 예측의 정확도는, 평균자승오차의 제곱근(R.M.S.E. ; Root Mean Square Error)을 성능 지표로 이용하여 〈표 8〉에서 비교되었다.

이러한 오차의 발생은 시계열자료나 외생변수의 특성에 따른 모형의 기능적인 선택에 의하여 달라짐을 볼 수 있다. 결론적으로 추정식(Ⅱ)가 전절의 추이확률 추정치에서도 사전기대치와 일치함을 나타내었을 뿐아니라, 적은 R.M.S.E. 측면에서도 가장 양호한 대안임을 나타내고 있다. 한편, 본 사례에서와 같이 외생요인이 시간의 변화에 따라서 일관되게 완만히 변화되는 경우라면, Non-Stationary 추정에 Bayes 기법의 확장된

〈표 8〉 Non-Stationary 모형의 추정식에 대한 RMSE 비교(1973-91)

| 구 분 | I | I-1 | II | II-1 |
|---|---|---|---|---|
| 농림어업 | 348.391 | 311.948 | 151.803 | 212.794 |
| 광공업 | 142.322 | 225.033 | 180.242 | 228.114 |
| 사회간접 및 기타 | 100.481 | 175.188 | 155.004 | 206.684 |
| 전산업 | 224.892 | 244.024 | 162.847 | 216.052 |

적용도 고려해 볼 수 있다. Bayes 추정방식은 推定値의 산출에 있어서 주관적 사전확률의 부가에 의하여 비계량적인 요인도 어느정도 반영시킬 수 있는 이점을 갖고 있다. 실제활용에 있어서는 시간의 경과에 따라 推移의 경향에 관한 더 최근의 이용 가능한 자료가 첨가됨으로써 갱신된 정보에 의하여 精度를 높여 갈 수 있게 될 것이다.

## 4. 結 論

본 연구에서는 Markov Chain 模型을 이용하여, 시스템 內의 구조적 변화를 분석하고 전망하기 위한 확장된 추이확률 추정모형이 개발되었다.

Stationary 모형의 확장을 통하여 과거 시계열자료를 이용한 사전정보가 도출될 수 있으며, 시계열 자체만의 구조적분석도 推移構造上의 잠재력의 파악이나 단기간 반응에서의 전망에 있어서 효과적임을 나타내고 있다. 더우기 충분한 과거시계열자료가 확보됨에 따라 신빙성있는 사전분포를 효율적으로 추정하게 되어, 상태간 推移에 관한 정교한 분석치의 산출에 기여할 수 있으리라 본다.

다음에 Non-Stationary 推移確率模型의 확장을 시도하여 시간의 변화에 따른 상태간의 변동파악에 촛점을 맞춰, 누가·어디로·왜 이동하는가하는 내부적인 부문간 움직임을 추적함으로써 전체적인 변화에 대한 이해를 증대시킬 수 있다는 점에서 유용함을 실증하고 있다.

이로써, 연구의 목표에 따른 상태의 범주는 이동적인 것뿐 아니라 거의 연속적인 속성에서도 적절하게 선택 활용될 수 있으며, 정책수립에서의 구조적 변화에 대한 동적 내지는 양적인 파악이 요구되는 많은 상황에 있어서 폭넓게 일반화 될 수 있다.

그리고 본 사례연구는 산업 및 직종별 인력이동의 추이에 관한 예증으로서, 특정 분석대상 자료에 있어서의 성능분석 결과라는 한계점을 갖고 있다. 추후 예측의 성능측도를 구체적으로 파악하기 위해서는, 정확도에 영향을 미치는 요인에 따라서 타기법보다 언제 어떠한 특정상황에서 우월하는가 하는 점이 고려되어야 한다. 이러한 측면에서 시계열자료에서의 내부적인 특성치와 관련된 이를테면, 표본의 크기나 랜덤성 등의 다양한 측도에서 몬테칼로 시뮬레이션기법을 통한 이론적인 성능분석도 고려될 수 있을 것이다.

## 參 考 文 獻

1. 경제기획원 조사통계국, 『고용구조 특별조사 결과보고』, 현재 및 1년전 활동상태 및 직업별 14세이상인구-전국, 1984. 1987. 1990.
2. 경제기획원 조사통계국, 『경제활동인구연보』,

산업별 취업자추이, 1971－91.
3. 통계청, 『주요경제지표』, 경제활동별 국내총생산(1985년 불변가격), 1972－91.
4. 강정혁, 김여근, "Non－Stationary 推移確率模型에 의한 농작물의 植付體系에 관한 연구"『경영과학』, 제8권 제1호, pp. 3－11,1991.
5. 박명수, 『중단기 노동력 수급전망』, 한국노동연구원, 1991. 8
6. Anderson, T. W. and Goodman, L. A. , "Statistical Inference about Markov Chains", The Annals of Mathematical Statistics, Vol. 28,pp. 89－110, 1957.
7. Bhat, U. N. ,"Elements of Applied Stochastic Processes", John Wiley and Sons Inc. , New York, 1972.
8. Canavos, G. C. ,"Performance Comparison of Empirical Bayes and Bayes Estimators of the Weibull and Gamma Scale Parameter", Naval Research Logistic Quarterly,Vol. 30, pp. 465－470,1983.
9. Casella,G. , "An Introduction to Empirical Bayes Data Analysis", The American Statistician,Vol. 39,pp. 83－87,No. 2,1985.
10. Durham, S. E. and David,R. Lee, "An Evaluation of Alternative Approaches to Market Share Analysis with Application to the Kuwaiti Poultry Market", Journal of Agricultural Economics, Vol. 38, pp. 85－97,1987.
11. Harary, F. and Lipstein, B. , "The Dynamics of Brand Loyalty : A Markovian Approach", Operations Research,Vol. 10, pp. 19－40,1962.
12. Kalbfleisch,J.D., Lawless,J.F. , Vollmer, W. M. , "Estimation in Markov Models from Aggregate Data", Biometrics,Vol. 39,pp. 907－919,1983.
13. Lee, T. C, Judge, G. G. , Zellner, A. , Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North Holland, New York, 1977.
14. ________ , "Maximum likelihood and Bayasian Estimation of Transition Probabilities", Journal of the American Statistical Association, Vol.63, pp. 1162－1179, 1968.
15. Maritz, H. F. and Waller,R. A. , Bayesian Reliability Analysis, John Wiley and Sons, 1982.
16. McCarthy, C. and Ryan, T. M., "Estimates of Voter Transition Probabilities from the British General Elections of 1974", Journal of the Royal Statistical Society Series A. ,Vol.140, pp.78－85,1977.
17. Mellor, C. J. , "An Application and Extension of the Markov Chain Model to Cereal Production", Journal of Agricultural Economics,Vol. 35, pp.203－215,1984.
18. Shultis, J. K. , Tillman, N. D. , Eckhoff, D. G. , Bayesian Analysis of Component Failure Data, Technical Report, U. S. Nuclear Regulatory Commision, Washington,1979.
19. Van Der Plas, Adrian P. , "On the Estimation of the Parameters of Markov Probability Models Using Macro Data", the Annals of Statistics, Vol. 11,No. 1,pp. 78－85,1983.
20. Wilks,S. S. , Mathematical Statistics,

John Wiley and Sons, New York. London, 1962.

21. Wilson, W. , Koo, W. , Carter, A. , "Importer Loyalty in the International Wheat Market", Journal of Agricultural Economics, Vol. 41, pp. 94−102, 1990.
22. Zanakis, S. H. and Maret, M. W. , "A Markov Chain Application to Manpower Supply Planning", Journal of the Operational Research Society, Vol. 31, No. pp. 1095−1101, 1980.