米穀消費形態의 構造分析과 ARIMA 模型에 의한 米穀需要豫測

明 光 植*
成 明 煥**

Ⅰ. 序 論
Ⅱ. 米穀消費形態의 構造分析
Ⅲ. 米穀需要豫測
Ⅳ. 結 論

Ⅰ. 序 論

米穀은 우리나라의 주식으로서 農家所得에서 점하는 비중뿐만 아니라 非農家의 家計費에서 차지하는 비중도 높기 때문에 米穀政策은 農業政策 중에서 가장 중요시되어 왔다.

또한 그동안 경제의 高度成長과 더불어 國民生活 여건의 향상은 食品消費構造의 다양화와 고급화, 嗜好食品 및 便宜食品의 需要 증대를 가져와 食品에 대한 消費支出 구조의 변화를 초래하였다. 이에 따라 국민 1인당 米穀消費量은 점차 감소 추세에 있는 반면 國內 米穀生産量은 계속적인 生産基盤 확대와 新品種 개발 등으로 증가추세에 있다.

* 首席硏究員.
** 硏究員.

때문에 米穀需給이 감안된 長短期 米穀政策 수립이 필요한데, 米穀政策의 방향 설정을 위해서는 현재 米穀消費 형태의 깊이 있는 분석과 米穀 消費量의 예측이 선결과제이다. 따라서 本分析에서는 地域, 季節, 所得階層別로 米穀消費 형태가 어떤 차이가 있는지를 구조적으로 분석하고 Box-Jenkins 模型을 이용하여 국민 1인당 米穀의 消費量을 새로운 각도에서 예측하고자 하였다.

Ⅱ. 米穀消費形態의 構造分析

1. 分析模型의 構造

일반적으로 경제분석은 경제현상의 변동에 따른 변화를 주로 수량화하는 양적인 방법으로 설명한다 그러나 경제현상을 질적으로 구분할 필요가 있을 때 Dummy 변수를 사용케 된다. 즉 Dummy 變數는 性別, 地域, 季節 등과 같은 질적인 요소들에 대한 說明變數로서 질적인 變

數를 계량화하여 獨立變數와 從屬變數間의 관계를 추정함으로써 질적인 獨立變數가 從屬變數에 얼마나 중요한 영향을 미치고 있는지를 파악할 수 있다.

그룹간의 차이를 분석하는 데에는 일반적으로 두 가지 回歸分析方法이 이용되고 있다. 하나는 관측치를 각 그룹별로 구분하여 각각의 方程式을 추정하는 것이고 다른 하나는 Dummy 變數를 사용하여 單一方程式을 추정하는 방법이다.

Dummy 변수의 사용은 각 관측치들간에 질적인 차이가 구분될 때 獨立變數에 임의의 수치 1 또는 0을 부여한다. 일반적으로 하나의 獨立變數 X에 대한 從屬變數 Y의 回歸方程式은 다음과 같다.

(1) $Y=a+bX+u$
(u는 오차항)

여기에서 관측치가 질적인 두 그룹으로 구분될 수 있을 경우 그룹간의 效果를 추정하기 위해서 Dummy 변수를 사용하면 다음과 같은 식으로 나타낼 수 있다.

(2) $Y=a_0+a_1D_1+a_2D_2+bX+u$

관측치가 그룹 I에 포함되면 $D_1=1$이 되고 그렇지 않으면 $D_2=0$가 되며 반대로 그룹 II에 포함되면 $D_2=1$이 되고 $D_1=0$이 된다. 그러나 方程式(2)에서 Dummy 변수 D_1, D_2와 常數는 완전한 線型關係가 있기 때문에 特異性行列(singular matrix)이 되어 $\hat{b}$를 추정할수 없다.[1] 따라서 推定量 $\hat{b}$를 구하기 위해서는 하나의 Dummy 변수를 方程式으로부터 제거하여 다음과 같은 方程式을 설정한다.

(3) $Y=a_0+a_2D_2+bX+u$

식(3)에서 獨立變數 X에 대한 기울기 b는 그룹간에 동일하고 그룹 I의 관측치에 대한 절편항은 a_0가 되며 그룹 II의 관측치에 대한 절편항은 (a_0+a_2)가 된다.

가. 절편이 변화하는 경우

方程式 (2)와 (3)은 獨立變數의 기울기는 그룹간에 동일하고 절편항만이 그룹간에 차이가 있다고 假定한 回歸方程式이다. 절편항에 Dummy 변수 1과 0인 값을 부여하여 계수추정한 方程式(3)의 回歸結果는 〈그림 1〉과 같은 형태를 보여준다.

나. 기울기가 변화하는 경우

지금까지 기울기의 係數는 그룹간에 동일하다고 가정하였으나 그룹간에 기울기가 다른 경우에도 Dummy 變數를 사용하여 분석할 수 있다. 그 제약 조건은 절편항에서와 마찬가지로 기울기항중 하나의 Dummy 變數를 제거하여 다음과 같은 方程式으로 나타낼 수 있다.

(4) $Y=a_0+(b_0+b_2D_2)X+u$

따라서 式(4)에서 그룹 I에 대한 기울기는 b_0가 되며 그룹 II의 경우 (b_0+b_2)가 된다. 〈그림 2〉는 두 그룹간에 절편항은 동일하다고 가정하고 기울기만의 변화는 式(4)의 回歸結果의 形態를 보여준다.

또한 그룹간에 절편항과 기울거항이 동시에 變化하는 경우 다음과 같은 式으로 나타낼 수 있고 그 형태는 〈그림 3〉과 같다. 이때 그룹 I의

[1] 方程式(1)을 Vector와 行列의 式으로 나타내면 $Y=\beta X+u$로 나타낼 수 있다. 따라서 線型回歸方程式에서 OLS推定量 $\hat{\beta}$는 다음과 같다.
$\hat{\beta}=(X'X)^{-1}XY$
그러므로 推定量 $\hat{\beta}$가 구해지기 위해서는 $X'X$行列이 非特異性(nonsingular)이 되고 $(X'X)^{-1}$行列이 존재하는 非特異性行列(nonsingular matrix)이어야 한다.

그림 1 절편항의 변화

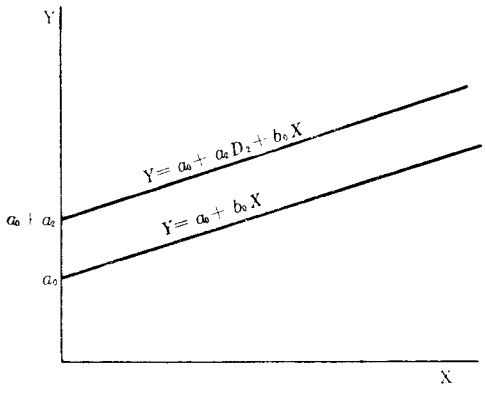


그림 2 기울기항의 변화

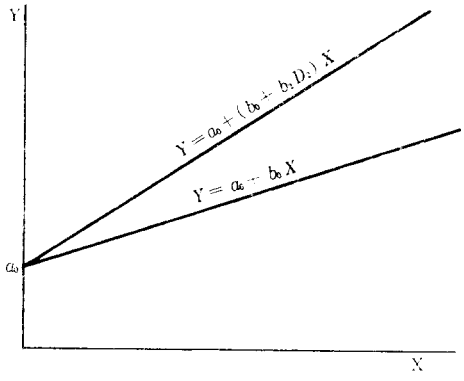
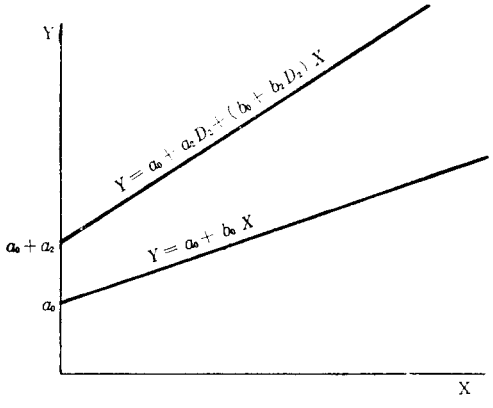


그림 3 절편 및 기울기항의 변화



절편과 기울기는 a_0와 b_0가 되며, 그룹Ⅱ의 경우에는 절편이 (a_0+a_2)가 되고 기울기는 (b_0+b_2)가 된다.

다. 2가지 이상의 質的變數로 된 경우

dummy 變數를 이용한 分析技法은 2가지 이상의 질적인 변수로 구성된 경우에도 확대 적용시킬 수 있다. 예를들어 從屬變數 Y가 獨立變數 X뿐만 아니라 性別, 地域, 季節 등과 같이 3가지의 質的變數에도 영향을 받는다고 가정하면 式(3)을 一般化시켜 다음과 같은 模型을 구성할 수 있다.

$$(5)\qquad Y_i=b_0+b_1D_{1k}+b_2D_{2l}+b_3D_{3m}+b_4X_i+u_i$$

여기에서 D_1, D_2, D_3는 각 질적인 變數에 대한 절편항의 Dummy 變數로서 이 경우 단순히 기울기는 동일하다고 가정하고 절편항의 변화만을 고려한 것이다.

$\hat{b}$의 推定方法은 앞에서와 마찬가지로 각 질적 변수에서 Dummy 變數를 하나씩 제거하여 $\hat{b}$를 최소자승법에 의해서 추정한다. 推定結果를 검정하여 통계적으로 유의적인 경우 각 質的 變數가 Y_i에 영향을 줌을 의미하며 Y_i에 주요 決定要素가 된다는 것을 의미한다.

2. 模型設定 및 分析構造檢定

가. 模型設定

일반적으로 소득을 獨立變數로 놓고 米穀 消費量을 從屬變數로 한 米穀의 消費函數는 다음과 같이 單純回歸模型으로 표시된다.

$$(6)\qquad Y_i=\beta_0+\beta_1X_i+u_i$$

Y_i : 米穀消費量
X_i : 所得

本分析에서는 이러한 단순모형에 Dummy 變數技法을 사용하여 米穀消費 형태의 구조 분석을 시도하였다. 구조 분석을 하기 위해 所得變數와 더불어 所得階層, 季節, 地域의 3가지 질적 변수를 포함시켜 所得階層別, 季節別, 地域別에 따라서 米穀消費 형태에 어떠한 차이를 보이며 각 질적 변수가 米穀消費量에 얼마나 중요한 영향을 미치고 있는지를 분석하였다. 이는 米穀消費量과 消費性向이 所得階層, 季節 및 地域에 따라 차이가 있다는 가정 하에 다음과 같은 米穀消費函數를 설정하였다.

$$(7)\quad Y_i=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_1X_6+\beta_8X_2X_6+\beta_9X_3X_6+\beta_{10}X_4X_6+\beta_{11}X_5X_6+u_i$$

위 式에서 X_6는 독립변수인 所得이고 $X_1\sim X_5$는 Dummy 변수이다. $\hat{\beta}$를 추정하기 위해서 分析模型의 이론상 각각의 質的變數에서 Dummy 변수를 하나씩 제거함으로써 所得階層 Dummy 변수 X_1, 季節 Dummy 변수 $X_2\sim X_4$, 地域 Dummy 변수 X_5만 남게 된다. 所得階層은 高所得과 低所得階層으로 구분하여 低所得階層 Dummy 변수를 제거하고, 季節의 경우에는 4계절로 구분하여 봄의 Dummy 변수를 제거하였으며 地域은 農家와 非農家로 구분하여 農家를 Dummy 변수에서 제거하였다.

나. 分析資料

分析에서 사용한 所得 및 米穀 消費量 자료는 현재 農林水産部에서 米穀消費量과 所得 또는 收入을 동시에 조사하고 있는 400호의 農家와 622호의 非農家를 대상으로 조사된 原資料(raw data)를 사용하였다. 分析對象 기간은 1985년 12월부터 1986년 11월까지 1개년의 橫斷面 資料를 사용하였다.

表 1 所得階層 및 地域別 1人當 月平均 米穀消費量 및 所得

單位 : kg, 원

區 分	低所得階層 (月 80,000원이하)		高所得階層 (月 80,001원이상)	
	農 家	非農家	農 家	非農家
平均消費量	13.20	9.05	14.60	8.85
平均所得	56,507	60,451	153,988	125,671

表 2 季節別 1人當 月平均 米穀消費量 및 所得

單位 : kg, 원, 명

區 分	봄	여 름	가 을	겨 울
平均消費量	10.94	10.82	10.89	11.19
農 家	13.91	13.93	14.20	14.42
非 農 家	9.03	8.83	8.77	9.13
平均所得	104,512	104,726	104,563	102,904
農 家	116,387	116,387	116,387	116,387
非 農 家	96,890	97,241	96,973	94,249
家口員數	4.55	4.58	4.58	4.63
農 家	4.60	4.59	4.60	4.71
非 農 家	4.52	4.58	4.56	4.57

여기에서 사용된 米穀 消費量은 同期間에 주식, 장유, 떡, 과자 및 기타에 지출된 1인당 平均 消費量을 사용하였고, 所得의 경우 非農家는 1인당 月收入額을 사용하였으나 농가의 경우에는 非農家처럼 일정한 收入이 없고 農作物 出荷期에 收入이 집중하는 경향이 있어 月平均 1인당 經常所得을 산출하여 이용하였다. 本分析에서 사용한 所得階層, 季節, 地域別에 따른 資料를 개략적으로 살펴보면 〈表 1, 表 2〉와 같다.

다. 分析構造의 檢定

米穀消費 형태에 관한 構造的 安定性檢定은 所得에 대한 米穀消費量 分布가 正規分布를 이룬다고 가정할 때 所得階層, 季節 또는 地域에 따른 米穀消費量 분포와의 分散의 差를 비교함으로써 假說의 채택여부를 결정하며 이때에 F-分布를 사용한다.

F값은 非制限模型(unrestricted model)인 式

(7)의 오차의 제곱합(ESS)과 검정을 하고자 하는 變數部分을 제외한 制限模型(restricted model)의 ESS를 가지고 다음과 같은 方法으로 계산한다.

$$(8) \qquad F=\frac{(\text{ESSrest}-\text{ESSunrest})/P}{\text{ESSunrest}/(N-K)}$$

P : 제외한 변수의 수
N : 관측치의 수
K : 비제한 모형의 설명변수 數

위의 方法에 의해 계산된 F-값과 일정한 有意水準에서 分母의 自由度가 $N-K$이고 分子의 自由度가 P인 F-分布表와 비교하여 假說檢定을 하게 되는데 계산된 F-값이 F-分布表보다 적을 때는 귀무가설을 채택하고 반대인 경우 기각하게 된다.

[1] 季節間의 절편(X_2, X_3, X_4) 및 기울기변수(X_2X_6, X_3X_6, X_4X_6) 檢定

계절간 절편에 대한 假說檢定은 우선 無歸假說 $\beta_2=\beta_3=\beta_4=0$ 對立假說 $\beta_2\neq\beta_3\neq\beta_4\neq0$을 설정하여 非制限模型의 ESS와 X_2, X_3, X_4의 變數를 제외한 制限模型의 ESS를 구하여 F-값을 계산한다. 계산된 F-값은 5.27으로서 5%의 有意水準에서 分子의 自由度가 3이고 分母의 自由度가 4073인 F-分布表의 값 2.60과 비교해 볼 때 계산된 F-값이 더 크므로 無歸假說은 기각된다. 따라서 米穀消費量은 계절에 따라 다르다고 할 수 있다.

기울기에 대한 假說檢定도 마찬가지로 귀무가설 $\beta_8, \beta_9, \beta_{10}=0$, 대립가설 $\beta_8, \beta_9, \beta_{10}\neq0$을 설정하고 F-값을 계산하면 4.02가 되어 $F_{3.4073}$-分布表의 2.60보다 크기 때문에 귀무가설이 기각되므로 계절에 따라 米穀消費性向이 다르다고 할 수 있다.

[2] 地域間 절편(X_5) 및 기울기변수(X_5X_6)檢定

지역간 절편에 대한 귀무가설 $\beta_5=0$, 대립가설 $\beta_5\neq0$을 설정하고 F-값을 계산하면 761.92가 된다. 반면 5%의 유의수준에서 $F_{1.4073}$에 값은 3.84이므로 귀무가설을 기각시킨다. 따라서 農家와 非農家間에 米穀消費量의 차이가 있음을 통계적으로 검정할 수 있다.

地域間 기울기에 대한 귀무가설 $\beta_{11}=0$, 대립가설 $\beta_{11}\neq0$을 설정하고 F-값을 계산하면 53.59가 되어 歸無假說을 기각한다. 이는 農家와 非農家間의 米穀消費性向에 차이가 있음을 의미한다.

[3] 所得階層間의 절편(X_1) 및 기울기변수(X_1X_6) 檢定

소득계층의 절편에 대한 歸無假說 $\beta_1=0$, 對立假說 $\beta_1\neq0$을 설정하고 非制限模型의 ESS와 制限模型의 ESS을 구하여 F-값을 계산하면 2.39가 된다. 이는 5%의 有意水準에서 $F_{1.4073}$ 값 3.84보다 작기 때문에 귀무가설을 채택한다. 따라서 소득계층간 米穀消費量의 차이는 통계적 유의성이 없다고 결론될 수 있다.

所得階層의 기울기에 대해서도 마찬가지로 歸無假說과 對立假說을 설정하여 F-값을 구하면 3.60이되어 F-分布表의 3.84보다 작기 때문에 歸無假說을 채택하고, 따라서 所得階層間 米穀消費 성향에도 차이가 없음이 통계적으로 검정되었다.

3. 分析結果

所得階層別, 季節別 및 地域別 米穀消費 형태의 構造分析 결과 米穀 消費量과 消費性向은 계절과 지역에 따라서는 차이가 있으나 소득계층

간에는 차이가 없는 것으로 나타났다. 계절과 지역에 따라서 米穀消費量 및 消費性向에 차이가 나타나는데 이는 社會·經濟·文化的인 측면에서 고려해 보면 그 타당성이 있다.

그러나 국민 1인당 미곡 소비량 및 소비성향이 所得階層과 관계없다는 사실은 의미있는 分析 結果라 하겠다. 일반적으로 所得이 증가함에 따라 1인당 米穀消費量이 감소한다는 견해가 지배적이었으나, 本分析結果는 1년간의 橫斷面 분석이기는 하나 현재 所得이 계속 증가하더라도 米穀 消費量은 크게 감소하지 않으리라는 점을 시사해 준다.

米穀消費函數式 (7)에서 추정한 결과를 檢定한 결과 소득계층간에는 米穀消費量과 消費性向의 차이가 없었기 때문에 所得階層의 절편 및 기울기 Dummy 변수를 제외하고 계절과 지역구분 Dummy 변수만을 사용하여 최종적인 米穀消費函數를 다음과 같이 설정하였다.

$$(9)\quad y_i=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_1X_5+\beta_7X_2X_5+\beta_8X_3X_5+\beta_9X_4X_5+u_i$$

위 식에서 X_5는 所得變數이고 X_1, X_2, X_3는 봄의 Dummy 변수를 제거한 나머지 季節의 Dummy 변수이며 X_4는 농가를 제거시킨 地域區分 Dummy 변수로 구성된 모형으로서 최소자승법에 의해서 추정된 米穀消費 函數式은 다음과 같다.

$$(10)\quad y_i=13.0421-0.3796X_1-0.5535X_2-0.2324X_3-3.7070X_4+0.0867X_5+0.0252X_1X_5+0.0525X_2X_5-0.0440X_3X_5-0.1278X_4X_5$$
$$(78.980)\ (-1.890)\ (-2.965)\ (-0.866)\ (-22.932)\ (6.851)\ (1.546)\ (3.511)\ (1.783)\ (-9.534)$$

( )內는 t 値임. $R^2=0.54$

表 3 米穀消費形態의 構造分析結果

季節 \ 地域	農家		非農家	
	절편	기울기	절편	기울기
봄	13.0421	0.0867	9.3351	−0.0411
여름	12.6625	0.1119	8.9555	−0.0159
가을	12.4886	0.1392	8.7816	0.0114
겨울	12.8097	0.1307	9.1027	0.0029

構造分析結果를 계절별 지역별로 절편과 기울기를 요약하면 〈表3〉과 같다. 농가1인당 월평균 米穀消費量은 비농가보다 평균적으로 3.7kg 더 많이 소비하며, 농가 및 비농가 공히 所得에 관계없이 일정한 소비수준을 유지하고 있는 것으로나타났다.

Ⅲ. 米穀需要豫測

1. 分析模型

미래를 예측하는 分析技法에는 여러가지 있겠으나 과거 時系列의 형태가 미래에도 같은 형태로 반복하리라는 基本假定下에서 각 관측치의 相互關係를 분석하여 미래의 형태를 예측하는 時系列 分析方法이 있다. 이 分析方法 중에서 短期豫測에 가장 정확하고 轉換點(turning point)에 대한 예측도 가능하기 때문에 미래를 예측하는데 가장 유용한 統計的 方法이 Box-Jenkins 時系列分析方法인데 일명 ARIMA 分析이라고도 부른다.

時系列資料는 그 特性(μ, σ^2)에 따라 安定時系列 (stationary time series)과 不安定時系列(nonstationary time series)로 구분된다. 安定時系列이

란 時系列資料의 平均과 分散이 시간에 따라 불변인 統計的 均衡狀態를 말하며, 不安定時系列이란 그 특성이 시간의 변화에 따라 변화하는 時系列을 말한다. 不安定時系列중 原時系列에 적합한 差分計算으로 安定時系列로 변환될 수있는 時系列을 동질적인 不安定時系列(homogeneous non-stationary time series)이라 한다.

G.E.P. Box와 G.M. Jenkins에 의해서 고안된 ARIMA 模型은 현재치 y_t의 과거치 y_{t-1}, y_{t-2}…와 攪亂項(random disturbance) e_t, e_{t-1}…을 가지고 y_t의 현재치가 그의 과거치들에 관련된 일련의 母數를 추정하는 技法이다.

가. 安定時系列

[1] AR 模型

AR 型模(autoregressive model)은 時系列 y_t를 y_t의 과거치와 t 시점의 攪亂項 e_t로서 표시하며 y_t가 p기간의 과거치 y_{t-1}, y_{t-2}, …y_{t-p}에 의해서만 영향을 받을 때 y_t는 次數가 p인 AR(p) 過程이 되며 다음과 같은 식으로 표시한다.

$$(11)\quad y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+e_t$$

[2] MA 模型

MA 模型(moving-average model)은 時系列의 현재치 y_t를 과거의 攪亂項 $e_t, e_{t-1}\cdots e_{t-q}$로서 표시한 것으로서 次數가 q인 MA(q) 과정이 되며 다음과 같은 式으로 표시한다.

$$(12)\quad y_t=e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}\cdots\cdots-\theta_q e_{t-q}$$

[3] ARMA 模型

安定時系列이 AR 過程과 MA 過程의 特性을 모두 가지고 있는 경우 AR 모형과 MA 모형을 결합한 次數가 p, q인 ARMA(p, q) 과정이 되며 y_t로 표시하면 다음과 같다.

$$(13)\quad y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+e_t -\theta_1 e_{t-1}-\theta_2 e_{t-2}\cdots\cdots-\theta_q e_{t-q}$$

나. 不安定時系列

時系列의 대부분은 시간의 변화에 따라 時系列의 특성이 변화하는 不安定的 時系列이다. 不安定的 時系列이지만 差分計算에 의해서 安定的 時系列로 변형되어질 수 있는 동질적인 不安定的 時系列은 不安定的 ∇_1時系列 y_t에 대해서 ARIMA 模型(autoregressive integrated moving-average model)의 설정이 가능하며 일반적으로 ARIMA (p, d, q)로서 표시된다 ARIMA(p, d, q) 모형을 써서 y_t를 표시하면 다음과 같다.

$$(14)\quad \phi(B)\nabla^d y_t=\theta(B)e_t$$
$$\phi(B)=1-\phi_1 B-\phi_2 B^2\cdots-\phi_p B^p$$
$$\theta(B)=1-\theta_1 B-\theta_2 B^2\cdots-\theta_q B^q$$
$$\nabla^d=(1-B)^d$$

다. 季節性時系列

季節性이란 時系列이 일정한 週期로 비슷한 행동패턴이 반복되는 경우를 의미한다. 이 경우 계절간격을 주기로 강한 相關關係를 보이므로 이 상관관계를 제거하여 분석하여야 한다. 季節性 時系列에 대한 模型으로서 계절성 AR 模型, 계절성 MA 模型, 계절성 ARMA 模型, 계절성 ARIMA 模型이 있다.

P, D, Q인 계절성 ARIMA$(P, D, Q)_s$ 模型은 다음의 式으로 표시된다.

$$(15)\quad \Phi(B^S)\nabla_S^D y_t=\Theta(B^S)e_t$$
$$\Phi(B^S)=1-\Phi_1 B^{1\cdot S}-\Phi_2 B^{2\cdot S}\cdots-\Phi P B^{P\cdot S}$$
$$\Theta(B^S)=1-\Theta_1 B^{1\cdot S}-\Theta_2 B^{2\cdot S}\cdots-\Theta_Q B^{Q\cdot S}$$
$$\nabla_S^D=(1-B^S)^D$$

그러나 위 式은 계절성의 영향만을 제거한 것이기 때문에 계절성 ARIMA 模型의 攪亂項 e_t는 계절성이 제거된 상태일 뿐이며 white noise라고 볼 수 없다. 따라서 e_t는 일반적인 ARIMA 過程인 式(14)에 의해서 생성되고, y_t는 계절성 ARIMA 模型인 式(15)에 의해서 만들어 진다고 가정하여 다음과 같은 乘法 季節性 ARIMA (p,d,q) $(P,D,Q)_S$ 模型을 만들었다. 이를 時系列 y_t로 표시하면 다음과 같다.

$$(16) \qquad \phi(B)\Phi(B^S)\nabla^d\nabla_S^D y_t=\theta(B)\Theta(B^S)e_t$$

2. 模型定立

일반적으로 설정한 Box-Jenkins 模型이 時系列에 적합한 모형으로 정립되기 위해서는 잠정적인 模型의 識別(model identification), 母數의 推定(parameter estimation), 模型의 적합성 검토(diagnostic checking)의 3단계 과정을 거쳐 수행된다.

가. 模型의 識別

模型의 識別이란 ARIMA(p,d,q) $(P,D,Q)_S$ 모형의 승수 p,d,q,P,D,Q의 값을 결정하는 것으로 識別의 절차는 먼저 原資料의 自己相關函數와 偏自己相關函數를 구하여 時系列의 안정성 여부를 검토한다. 不安定的 時系列인 경우 安定時系列이 될 때까지 原資料를 差分計算하여 d값을 정하고 季節性을 띠고 있는 경우에는 D값도 함께 정한다. 差分計算의 승수 d값은 時間差 k가 증가함에 따라 自己相關函數가 급격히 감소하여 0에 수렴하도록 정한다.

d번의 差分計算을 한 후 얻은 時系列 $\nabla^d y_t$가 安定時系列이라면 $\nabla^d y_t$의 自己相關函數와 偏自己相關函數을 분석하여 ARIMA 模型의 次數 p, q를 정한다.

AR(p) 모형의 自己相關函數는 시간차 k가 증가함에 따라 점점 감소하는 반면에 偏自己相關函數는 k가 p보다 크면 0이다. 반대로 MA(q) 모형의 자기상관함수는 k가 q보다 크면 0이고 편자기상관함수는 k가 증가함에 따라 점점 감소한다. ARMA(p,q) 모형의 자기상관함수는 k가 q까지는 불규칙적이고 그 다음 時間差부터는 AR(p) 과정처럼 점차적으로 소멸된다. 반면에 偏自己相關函數는 時間差 k가 p까지는 불규칙적이고 그 다음 時間差부터는 MA(q)과정처럼 점차적으로 소멸된다.

〈그림 4, 그림 5〉는 時系列에 많이 사용되고 있는 AR, MA, ARMA 모형의 自己相關函數와 偏自己相關函數를 나타낸 것으로 次數 p,q를 결정하는 데 매우 유용하다.

나. 母數의 推定

잠정적인 模型이 識別되면 最小自乘法을 사용하여 母數를 추정한다. 일반적으로 ARIMA(p,d,q) 模型의 母數를 추정할 때는 먼저 e_t를 가장 적게 하는 ϕ,θ의 推定値 $\hat{\phi},\hat{\theta}$를 구한다.

式(14)에서

$$(17) \qquad e_t=\theta^{-1}(B)\phi(B)\nabla^d y_t$$

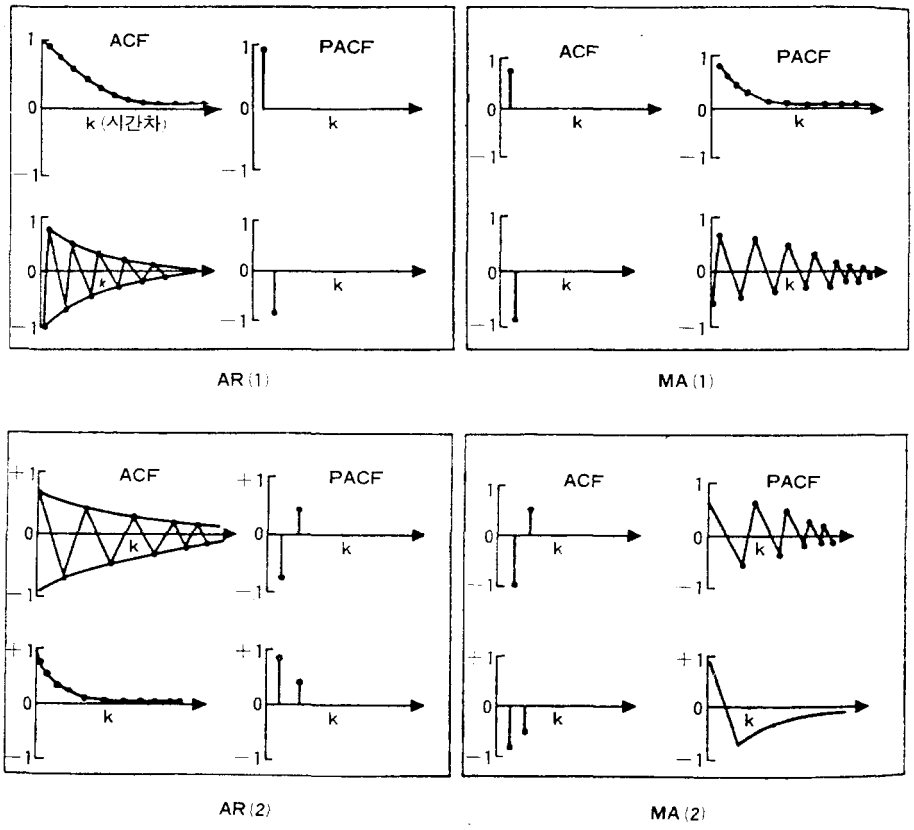
이때 e_t는 殘差項으로서 white noise임을 가정하여 殘差제곱의 合

$$(18) \qquad S=(\hat{\phi},\hat{\theta})=\sum_t e_t^2$$

을 최소화시키는 $\hat{\phi},\hat{\theta}$을 추정하는 것인데 式(17)은 非線型이기 때문에 $\hat{\phi},\hat{\theta}$를 구하기 위해서는 非線型推定法을 사용한다.

다. 模型의 檢定

그림 4 AR(1), AR(2), MA(1), MA(2) 模型의 自己相關函數(ACF)와 偏自己相關函數(PACF)



模型定立의 세번째 단계는 설정된 模型의 적합성과 타당성을 檢定하는 것이다. 잠정적으로 설정한 ARIMA 模型이 검정을 통해서 적합한 것으로 판정되면 그 模型은 여러가지 豫測이나 分析에 사용된다.

模型의 檢定은 殘差分析에 의한 실제치와 예측치간의 차이를 분석함으로써 할 수 있다. 특히 잠정적인 模型의 적합성을 측정할 수 있는 효과적인 방법으로서 殘差의 自己相關關係를 나타내는 統計量을 가지고 模型의 적합성 여부를 판단한다. 이 統計量을 Box-Pierce χ^2 統計量이라고 하는데 일반적으로 Q로 표시하고 산출 공식은 다음과 같다.

(19) $\quad Q=(n-d)\sum^{K} r_l^2(e)$

여기서

n : 原時系列資料의 관측치수

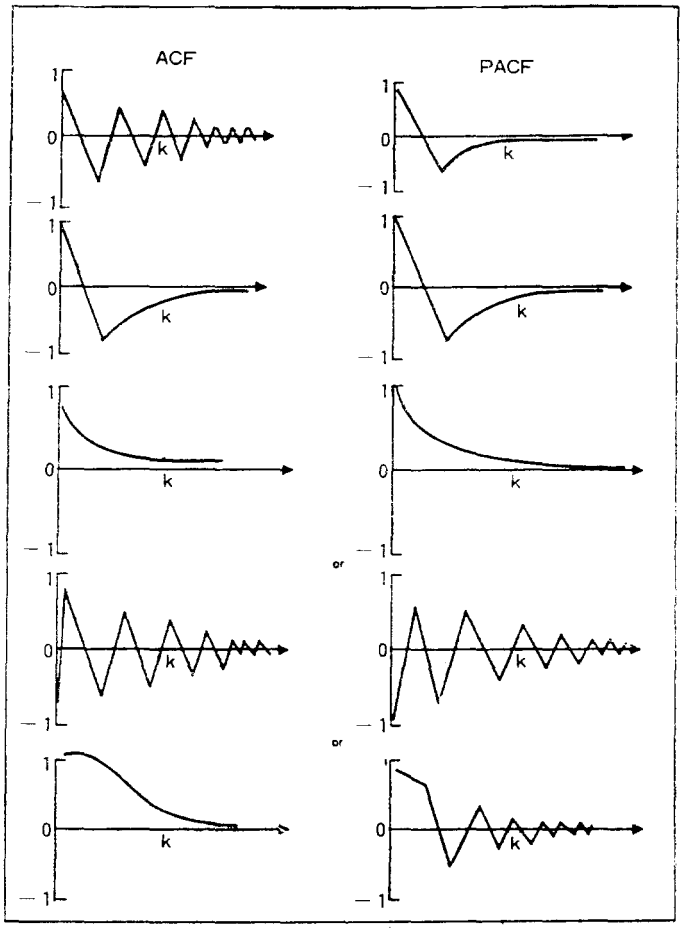
d : 原時系列資料에서 安定時系列로 전환하기 위해 사용한 差分計算數

$r_l(e)$: 時間差 l에서 殘差의 標本自己相關關係

K : Q를 계산하는데 사용된 殘差의 自己相關關係數

式(19)에서 殘差의 自己相關이 적을수록 統計量 Q도 적어지며, 殘差의 自己相關이 클수록 Q 값도 커진다. 계산된 統計量이 自由度 $K-n_p$를 갖는 χ^2 分布와 비교하여 模型의 假說을 檢定하는데 일반적으로 Q 값이 有意水準 α에서 χ_α^2 보다 적으면 模型이 적합하다고 판단한다. 여기에서 n_p는 母數(parameter)의 數이다.

그림 5 ARMA(1,1) 模型의 自己相關函數와 偏自己相關函數



3. 模型分析結果

가. 分析資料

米穀消費豫測分析에 이용한 時系列資料는 農林水産部에서 조사하는 農家 및 非農家의 月別 1인 1일당 米穀消費量資料로서 1969년 11월부터 1986년 10월까지 17개년을 分析對象期間으로 하였다. 조사된 米穀消費量은 주식, 장유, 떡, 과자 및 기타에 지출된 純飮食用 消費量이다.

나. 分析結果 및 檢定

農家 및 非農家의 米穀消費量을 예측하기 위하여 乘法 季節性 ARIMA 模型인 ARIMA$(p, d, q)(P, D, Q)$를 적용하여 3단계의 分析模型定立 과정을 거친 分析結果를 農家와 非農家로 구분하여 살펴본다.

非農家의 月別 1인 1일당 米穀消費量 原資料에서 自己相關函數와 偏自己相關函數를 추정한 결과 불안정성을 내포하고 있으며 12개월마다 돌출을 보이므로 계절성을 나타내고 있다. 불안정성을 제거하기 위해서 1차 差分計算을

하였으나 時間差 12와 24에서 여전히 自己相關函數가 큰 값을 가지고 있어 季節的 不安定性이 존재한다고 판단하여 다시 12차 差分計算을 하였다. 季節的 不安定性을 제거하고 최종적으로 ARIMA$(1,1,1)(1,1,1)_{12}$ 模型을 설정하여 母數를 추정한 결과 다음과 같다.

$$(20)\quad \underset{(-3.17)}{(1-B)(1-B^{12})}(1+\underset{(2.60)}{0.3607B})(1+\underset{(-2.01)}{0.2165B^{12}})y_t=(1-\underset{(5.71)}{0.5327B^{12}})e_t$$

( )內는 T ratio 임.

이를 분해하여 y_t로 정리하면 다음과 같다.

$$\begin{aligned}(21)\quad y_t=&0.6393y_{t-1}+0.3607y_{t-2}+0.7835y_{t-12}\\&-0.5009y_{t-13}-0.2826y_{t-14}\\&+0.2165y_{t-24}-0.1384y_{t-25}\\&-0.0781y_{t-26}+e_t-0.3026e_{t-1}\\&-0.5327e_{t-12}+0.1612e_{t-13}\end{aligned}$$

설정한 模型의 적합성 검정을 하기 위해서 Q 통계량을 계산한 결과 11.42였다. 이를 5% 有意水準에서 自由度$(K-n_p=12-4)$ 8인 χ^2_5값 15.50과 비교하면 Q값이 더 적기 때문에 假說이 받아들여져 설정된 模型은 적합하다고 판단한다.

農家의 米穀消費量豫測模型의 設定에 있어서도 非農家와 동일한 方法을 통하여 잠정적인 乘法 季節性 ARIMA$(1,0,1)(1,1,1)_{12}$ 模型을 적용하여 母數를 추정한 결과는 다음과 같고

$$(22)\quad \underset{(35.69)}{(1-B^{12})}(1-\underset{(11.81)}{0.9618B})(1+\underset{(-3.25)}{0.3887B^{12}})y_t=(1-\underset{(1.80)}{0.7517B})(1-0.2319B^{12})e_t$$

( )內는 T ratio 임.

이를 다시 분해하여 y_t로 정리하면 다음과 같다.

$$\begin{aligned}(23)\quad y_t=&0.9618y_{t-1}+0.6113y_{t-12}\\&-0.5879y_{t-13}+0.3887y_{t-24}\\&-0.3739y_{t-25}+e_t-0.7517e_{t-1}\\&-0.2319e_{t-12}+0.1743e_{t-13}\end{aligned}$$

주어진 時系列에 대해서 잠정적으로 설정한 模型이 적합한지의 여부에 대한 判斷基準으로서 時系列의 최근 3개년에 대한 觀測値와 模型에 의한 豫測値間의 오차가 5% 이하여야 하며 模型에 의한 殘差系列의 임의성 검정에서 χ^2 확율이 10이상이어야 한다. 이 기준을 통과한 모형중 최근 3개년의 豫測誤差가 가장 적은 것을 最適模型으로 결정한다.

農家에 적용한 式(22)의 模型에서 계산한 Q 統計量은 22.48로서 自由度 8인 χ^2_5 값 15.5보다 크기 때문에 假說을 기각되어야 한다. Q값이 높은 이유는 분석에 사용된 204개의 時系列 觀測値중 몇 개의 불규칙적인 변동으로 적합한 模型의 設定은 곤란하다.

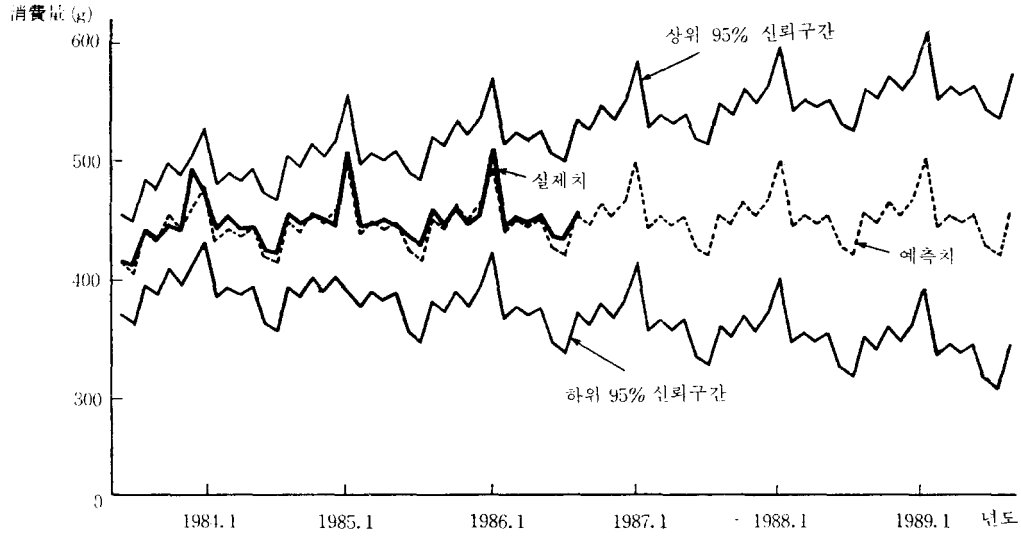
그러나 本分析에서 설정한 模型의 豫測値와 최근 3개년에 대한 時系列 觀測値間의 豫測誤差가 1.0%에 불과하므로 農家에 대한 ARIMA$(1,0,1)(1,1,1)_{12}$ 模型은 타당하다고 간주하여 농가의 米穀消費量 豫測에 사용하였다.

4. 米穀需要豫測

模型定立의 3단계를 거쳐 時系列에 적합한 모형이 선정되면 그 모형으로부터 미래치를 예측하거나 분석에 사용된다. 農家 및 非農家의 月別 1인 1일당 米穀消費量豫測은 式(21)과 (23)에서 攪亂項 $e_t, e_{t-1}\cdots$을 제거하고 $y_{t-1}, y_{t-2}, \cdots$의 過去値에 대한 변수들만으로 y_t를 예측한다.

그림 6 農家의 月別 1人 1日當 米穀消費量 推移 및 豫測

그림 7 非農家의 月別 1人 1日當 米穀消費量 推移 및 豫測



豫測模型에 의한 농가 및 비농가 月別 1인 1일당 米穀消費量 推移 및 豫測圖는 〈그림 6〉과 〈그림 7〉에 나타나 있다. 이를 米穀年度를 기준으로 國民 1인당 연평균 米穀消費量 豫測과 總

表 4 國民1人當 米穀消費量 豫測

單位 : kg

區分 / 米穀年度	農家		非農家		國民1人當		總米穀*
	실제치	예측치	실제치	예측치	실제치	예측치	消費量
1984	164.0	161.7	120.1	119.6	129.9	129.0	5226千㎏
1985	164.0	162.7	117.9	117.3	127.5	126.7	5202
1986	164.8	163.5	118.2	116.3	127.4	125.6	5221
1987	—	164.6	—	114.8	—	124.5	5276
1988	—	164.9	—	113.7	—	123.3	5302
1989	—	164.6	—	111.9	—	121.3	5292
1990	—	164.8	—	110.5	—	119.7	5298
1991	—	164.9	—	109.0	—	118.7	5288

1) 전년 11월부터 익년 10월까지임.
2) 예측치임.

消費量 豫測을 한 것이 〈表4〉에 나타나 있다. 農家의 1인당 연평균 米穀消費量은 향후 5년 동안 현재의 소비량 수준을 유지할 것으로 예상되며, 非農家의 경우 아주 완만하게 감소하고 국민1인당 米穀消費量은 향후 5년 동안 연평균 1% 내외의 감소가 예상된다. 계속적인 인구 증가를 고려하더라도 1991년까지 매년 약 530萬㎏ 수준의 純飮食用 米穀需要가 있으리라 예상된다.

本分析에서 예측한 米穀需要와 1991년도 米穀生産 전망과를 비교해 보면 〈表5〉와 같다. 현재의 植付面積推移를 감안할 때 1991년에는 總米穀植付面積 1,224千ha 中 新品種 植付 비율이 16%가 될 것으로 가정하여 代案別 米穀生産量을 산출해서 1991년 豫想米穀需要量과 비교하면 加工用을 고려하더라도 米穀의 完全自給이 가능할 뿐만 아니라 過剩在庫의 우려도 예상된다.

表 5 代案別 豫想米穀自給率, 1991

단수 / 대안별	품종별 식부면적(千ha)		품종별 단수(kg)		미곡생산량(千㎏)			예상미곡 자급률
	재래종	신품종	재래종	신품종	재래종	신품종	계	
A	1028	196	424	491	4359	962	5321	100.6%
B	1028	196	434	496	4462	972	5434	102.8
C	1028	196	446	504	4585	988	5573	105.4
D	1028	196	468	496	4811	972	5783	109.4
E	1028	196	468	500	4811	980	5791	109.5

A : 평년작 단수적용
B : 최근 3개년간 평균단수 적용
C : 최근 5년간 최고단수 적용
D : 재래종은 추세식, 신품종은 최근 3개년 평균단수 적용
E : 재래종, 신품종 추세식 적용

Ⅳ. 結 論

經濟成長에 따라 農業部門에서의 米穀 비중이 점차 감소하고 있으나 아직도 米穀政策은 農業政策 중에서 가장 중요한 부분으로 인식되고 있다. 과거 일본이나 대만경제가 경험했듯이 國民生活與件의 변화에 따른 米穀의 消費形態가 변화하고 있는 현 시점에서 바람직한 米穀政策方向의 설정이 필요한데 무엇보다 소비 행위에 대한 정확한 규명이 절실하다.

米穀消費형태의 構造分析 결과를 볼 때 所得에 대한 米穀消費量이 地域과 季節에 따라 차이가 있다는 것은 일반적으로 믿어 왔던 사실이기는 하나 이 分析結果로서 통계적으로 검정할 수 있었다. 또한 1986년 橫斷面資料를 이용한 米穀消費의 構造分析 결과 所得變化에 대한 米穀의 消費形態에는 차이가 없음이 통계적으로 檢定되었다. 즉, 米穀消費가 所得과는 무관하기 때문에 1인당 米穀消費量은 앞으로 크게 감소하지 않으리라는 것을 시사해 준다. 이를 國內總需要 側面에서 볼 때 國民所得이 증가하더라도 안정적인 米穀需要가 있음을 의미한다.

이러한 結果는 ARIMA 模型을 이용하여 예측한 1인당 米穀消費量의 豫測結果와도 일치한다. 豫測結果에 따르면 국민 1인당 米穀消費量은 아주 완만한 감소가 예상된다. 그리고 1인당 米穀消費減少率이 人口增加率과 비슷한 수준이기 때문에 總需要 측면에서도 현 수준과 큰 차이가 없을 것으로 판단되었다. 따라서 급격한 자연적인 이변으로 생산 쪽에 차질이 없는한 향후 5년

간은 적어도 米穀의 完全自給에 큰 무리가 없으리라 판단된다. 그렇지만 현재의 米穀生産基盤 확대 및 品種開發 등으로 米作生産 조건이 개선되어 현재보다 높은 生産量 증대가 이루어질 경우 生産過剩의 염려도 없지 않다고 보겠다. 미곡이 農家所得源의 큰 비중을 차지하고 점차 늘어나는 生産量의 판로확보라는 차원에서 米穀의 消費促進과 加工食品의 개발도 고려되어야 하겠다. 이것은 곧 米穀의 需給安定 뿐만 아니라 지속적인 農家所得向上과 農業成長을 보장하는 길이라고 하겠다.

參 考 文 獻

呂運邦 外, 「乘法 ARIMA 模型의 構造識別方法」, 韓國開發研究院, 研究報告 85-03, 1985.

柳志星, 「計量經濟學原論」, 博英社, 1987.

農林水産部, 「糧穀消費量 調査結果報告」.

———, 所得 및 米穀消費量資料.

韓國銀行, 「X-11 ARIMA METHOD에 關한 研究」, 1980.

Box. G.E.P. and G.M. Jenkins, *Time Series Analysis: forecasting and control*, Holen-Day Inc., 1976.

Huang, D.S., *Regression and Econometric Methods*, John Wiley & Sons, 1970.

Makridakis, S. and S.C. Wheelwright, *Forecasting-Methods and Applications*, John Wiley & Sons, 1978.

Makridakis, S., et al, *The Forecasting Accuracy of Major Time Series Methods*, John Wiley & Sons, 1984.

O'Connell, R.T. and B.L. Bowernman, *Forecasting and Time Series*, Duxbury Press, 1979.

Pindyck, R.S. and D.L. Rubinfeld, *Econometric Models and Economic Forecasting*, McGraw-Hill, 1976.

Thomopoulos, N.T., *Applied Forecasting Methods*, Prentice-Hall, 1980.