經濟指標의 ARIMA模型 選定方法

1986. 7.

調査統計局 統計分析課

經濟指標의 ARIMA模型 選定方法

1986. 7.

調査統計局 統計分析課

目　　次

머 리 말

當局에서는 景氣綜合指數, 產業生產指數등의 月別 經濟指標들의 季節變動調整系列 作成을 위하여 일찌기 1973年1月부터 X－11을 導入 使用하였으며, 景氣綜合指數 開發過程에서 1979年9月부터 X－11－ARIMA方法을 導入, 使用하고 있읍니다.

X－11－ARIMA는 美國 商務省의 센서스局法 X－11方法을 補完한 것으로써 센서스局法이 數次에 걸친 移動平均에 따른 缺項處理方法의 未備로 인하여 最近年度의 季節要因 觀測值가 中央年度의 그것에 比하여, 精度가 낮은 점을 補完하기 위하여 Canada統計廳에서 2年間의 硏究를 거쳐 1975年에 開發한 方法으로써, 同方法은 原系列에 적합한 Box－Jenkins의 ARIMA模型을 選定하고, 이를 적용하여 原系列 兩端의 時系列을 擴張한 後, X－11方法으로 季節要因을 算出하는 方法입니다.

그런데 X－11－ARIMA프로그램에서는 個別指標의 ARIMA模型 選定을 3個의 模型 中에서 自動選定하도록 되어 있으며, 또한 必要時 利用者가 직접 個別指標別로 ARIMA模型을 選定하여 使用할 수도 있도록 설계 되어 있읍니다.

따라서 當局에서는 季節變動調整方法 개선방안의 하나로 우리나라에서는 처음으로 1983年부터 SAS Package를 활용, 個別指標의 特性에 맞는 模型을 別途 선정하여 X－11－ARIMA프로그램에 연결 使用하고 있읍니다.

이 資料는 ARIMA模型의 概念, 選定方法 및 ARIMA模型選定을 위한 SAS Package「ARIMA」부분의 利用方法에 관한 해설서로서 時系列 分析에 많은 도움이 되리라 믿고 발간합니다.

本 內容의 집필은 當課에서 이 業務를 擔當하고 있는 田 百根君이 담당하였읍니다.

1986. 7. 統計分析課長

金 民 卿

Ⅰ. 센서스局法Ⅱ X－11에 ARIMA模型의 導入背景

1. 센서스局法의 開發過程

時系列 資料에서 季節變動值를 抽出하고자 하는 노력은 1900年代 初부터 시작되었는데, 1910年 美Babson統計社가 눈금表에 依한 季節調整法을 고안하여 이에 의한 財界價格指數를 發表한데서 그 효시를 찾을 수 있다. 이어 1910年代에 부르크마이어經濟社, Harvard大學 等에서 連環比率法에 依한 季節調整景氣指數를 發表하였다. 그러나 이들 方法은 계절변동이 12개월을 주기로 一定한 Patten만을 反復한다는 固定季節變動을 전제로 하고, 經濟構成變化가 季節要素에 주는 影響을 고려하지 않았기 때문에 景氣豫測이 種種 問題가 되어왔다. 이에 따라 現實的인 경기예측을 爲하여는 可變的인 계절변동을 전제로 하는 계절조정법을 개발해야 한다는 硏究가 美國 NBER을 中心으로 이루어져, 1930年代 후반에 季節指標를 可變的으로 취하는 季節調整法이 實用化 되었다(1941年 H.C. Barton이 Federal Reserve Bulletin에 Adjustment for Seasonal Variation을 發表).

그후 Computer發達과 더불어 多樣한 이동평균계산이 容易해 짐에 따라 不規則變動의 조직적인 處理가 可能해 졌는데, NBER의 J. Shiskin이 美商務省 Census Bureau로 옮겨가 Census局 季節調整法 開發Group의 中心이 되어, 1955年 Census局法 Ⅰ을 公表하기에 이르렀다.

Computer의 發達과 더불어 局法Ⅰ의 活用이 늘어나면서 이의 未備點이 發見되자 Census局은 NBER의 協力을 얻어, 改良된 Census局法Ⅱ를 1957年에 發表했다. 그후 NBER, OECD統計局등의 協力을 얻어 特히 다음의

問題를 中心으로 硏究가 거듭되었다.

① 季節要因 및 趨勢·循環要因을 算出하는 경우에 使用하는 移動平均의 方法

② 移動平均에 依해 季節, 趨勢, 循環要素를 算出할 때 생기는 缺項의 處理方法

③ 月別資料에서 曜日變動要因의 推定方法

④ 特異項의 處理方法

이러한 硏究結果는 Census 局法Ⅱ에 改良型을 나타내는 X번호를 붙여 一連의 Program으로 發表되었다고 현재 X－11까지(1965.10)公表되었다.

現在 美國, 日本, OECD等 많은 國家 및 國際機構에서 이 Census 局法을 直接使用하거나, Census 局法의 原理를 利用, 計算過程을 一部 변형한 方法을 開發하여 使用하고 있다.

그런데, 移動平均技法의 原理를 利用하는 Census 局法은

① 原系列의 分解에 關한 明白한 模型이 없고,

② 移動平均을 반복함으로써 系列 兩端의 缺項에 對하여 편의적으로 補正함으로써 가장 最近年度의 觀測値에 對한 季節變動要因推定値는 中央의 觀測値의 계절요인 推定値보다 信賴度가 떨어진다는 問題가 있어 Canada 統計局에서, Dr. E.B.Dagum을 中心으로 1973年以來 Census 局法Ⅱ X－11의 缺點補完을 위한 硏究結果 1974年 X－11－ARIMA方法을 開發하였으며, 1975年1月 카나다統計局에서 勞動力等 經濟指標의 季節調整方法으로 公式 採擇되었다.

X－11－ARIMA方法은 基本的으로,

① 原系列과 잘 맞는 ARIMA模型(Box and Jenkins type)에 依하여 原

系列을 模型化

② ARIMA模型에 依해 原系列의 兩端 各各 1年을 擴張(forecasting, Backcasting)

③ Census 局法Ⅱ X－11에 依해 擴張된 系列의 季節調整의 計算過程으로 構成되어 있다.

X－11－ARIMA方法은 X－11에 比하여 계절변동요인의 推定値의 精度가 높다는 것이 증명되고 있다.

우리나라에서는 韓國銀行이 1968年 Census 局法Ⅱ X－10, X－11을 導入하여 使用하고 있고, 當局에서 1973年부터 X－11을 導入, 產業生產指數의 季節調整에 利用하여 왔으나 景氣綜合指數 開發을 爲한 時系列 分析過程에서 X－11－ARIMA를 導入하여 1979年 9月부터 使用하고 있다.

2. ARIMA模型의 背景

時系列變動의 構成要因을 推定하는 方法은 回歸法(Regression meth－od)과 移動平均法(Moving average technique, 線型平滑化過程이라고도 함)으로 크게 나눌 수 있다.

回歸法은 季節, 趨勢, 循環變動이 全系列에 걸쳐서 決定函數(deterministic function)임을 전제로 하고 있다.

移動平均方法에서는 時系列의 各 要因이 時間의 平滑函數(Smooth func－tion)이기는 하나, 단순한 함수에 의해 全期間의 要因을 推定할 수 없음을 전제로 하고 있다.

즉, 趨勢, 循環, 季節變動要因은 確率的(stochastic)이지, 決定的(deter－

ministic)이 아님을 의미하는 것이다.

센서스局法은 移動平均技法의 범주에 屬하는 것으로서 앞에서 본 바와 같이 數次의 移動平均반복을 통한 系列兩端의 缺項으로 因하여 가장 최근년도의 觀測値에 對한 계절변동요인 推定値는 中央의 觀測値에 對한 계절변동요인의 推定値에 比較하여 信賴性이 떨어지고, 아울러 現在의 原系列에 새로운 系列이 追加되면 계절요인의 수정폭이 크고, 그 修正에 依해 調整指數의 方向이 달라지는 경우도 있게 된다. 따라서 센서스局法Ⅱ X－11 program缺點補完을 위한 카나다統計局의 硏究作業은 어떠한 方法에 依하여 原系列을 擴張하느냐 하는 데에 優先順位가 주어졌는데, 系列擴張方法의 條件으로 다음 사항을 고려하였다.

첫째, 系列擴張方法은 가장 단순한 方法이라야 하며, 어떤 函數로 주어질 경우에도 說明變數를 包含하지 말아야 하고, 時系列은 그 自體의 過去의 觀測値(past value)와 時差를 가진 誤差項(lagged random disturbance) 에 依하여만 說明되어져야 한다.

둘째, 系列을 說明하는 것으로 識別된 模型은 1～2年의 系列을 追加하여도 變更이 되어서는 안되며, 模型에 依한 擴張値(豫測値)는 母數의 적은 變動으로는 크게 變動되어서는 안된다.

세째, 系列擴張模型에 依한 豫測(擴張)値는 그 水準自體는 잘 맞지 않더라도 豫測한 年度內의 月別(分期間)움직임을 合理的으로 잘 나타내는 豫測値를 算出해야 한다. 왜냐하면, 이 擴張値는 그 數値自體가 豫測統計의 資料로서 利用되는 것이 아니고, 계절조정방법을 改善하기 위한 것이기 때문이다.

네째, 系列擴張方法은 最少平方和(MSE)를 갖는 最適擴張値를 算出해서

不完全한 資料蒐集으로 인한 잠정數值와 代替하여 Benchmark 數值로 利用될 수 있어야 한다.

다섯째, 系列擴張方法은 可能한 限 적은 數의 母數를 갖고 있어야 한다.

以上 위의 일련의 條件下에 Canada 統計局에서는 單一變數豫測法(Univariate method of forecasting)을 選擇하였고, 그 中에서도 Box & Jenkins type의 ARIMA 模型을 選擇하게 된 것이다.

이 ARIMA 模型은 系列擴張의 두가지 基本概念인 自己回歸(auto - regression)와 移動平均(moving average)으로 되어 있다.

ARIMA라는 略語는 AR(Auto - Reoressive)와 MA(Moving Average)와 I(Intergration)의 合成語로서 Intergration 部分은 ARIMA 模型의 중요한 부분인데, 階差를 가진 資料에 잘 適合하는 定常的인 模型들이 非定常(non - stationary)資料에 대하여 模型을 제공하기 위하여는 定常的(stationary)인 模型들이 統合(intergration) 또는 結合(summed)되어야 하기 때문이다.

Ⅱ. 時系列의 形態와 性質

1. 時系列 豫測의 諸方法

時系列이란 經濟現狀들을 時間의 흐름에 따라 觀測해 놓은 統計系列을 말한다. 예를 들면 GNP, 産業生産指數, 物價指數, 人口數, 特定地域의 降雨量등을 年度別, 分期別, 또는 月別 등과 같이 時間의 순서에 따라 整理해 놓은 觀測値들이다.

時系列 統計의 作成目的은 무엇보다도 現在까지의 實績資料를 分析함으로써 未來를 豫測할 수 있는 資料로 活用할 수 있다는데 있다. 즉 特定 時系列이 보여주는 過去의 資料에서 安定的인 時差模型을 設定할 수 있다면 그 模型에 있는 未知의 母數들을 統計的 方法으로 推定함으로써 未來를 豫測할 수 있는 것이다.

時系列 豫測方法으로는 指數平滑法, 要素分解法, 回歸法, Box－Jenkins의 時系列分析法 등을 利用할 수 있다.

指數平滑法은 以前의 豫測値의 比率에 誤差의 比率을 더하여 過去 誤差들을 平均化 하는 原則에 의하여 豫測値를 算出한다.

數學的으로 單純指數平滑法은

$$F_{t+1} = F_t + \alpha(X_t - F_t) \quad \cdots\cdots\cdots (2.1)$$

의 形態이다.

여기서 α : $0 < \alpha < 1$ 인 平滑常數

F_t : 豫測値

X_t : 觀測値 이다.

式(2.1)에 $F_t = F_{t-1} + \alpha(X_{t-1} - F_{t-1})$를 代入하면 다음과 같이 擴張시킬 수가 있다.

$$F_{t+1} = F_{t-1} + \alpha(X_{t-1} - F_{t-1}) + \alpha(X_t - F_t) \cdots\cdots\cdots\cdots (2.2)$$

式(2.2)의 첫번째 項에 F_{t-1}을 代入하면

$$F_{t+1} = F_{t-2} + \alpha(X_{t-2} - F_{t-2}) + \alpha(X_{t-1} - F_{t-1}) + \alpha(X_t - F_t) \cdots\cdots\cdots\cdots (2.3)$$

로 주어진다.

F_{t-2}에 임의 初期豫測值가 주어지면 새로운 豫測値는 初期豫測値에 의해 實測値와 豫測値間의 誤差$(X_{t-2} - F_{t-2})$의 比率을 더하여 구할 수 있다.

誤差中에는 "−"와 "+"가 있기 때문에 最終豫測値 F_{t+1}은 平均하여 볼때 資料의 實際類型에 近似하게 된다.

時系列 要素分解法은 時系列을 季節性, 趨勢 循環 및 不規則性(임의성)의 各 成分을 "分解시키는" 原理에 根據를 두고 있다. 따라서 各 成分을 別個로 豫測하여(豫測할 수 없는 不規則(任意)性은 제외함) 이 豫測値를 재결합함으로써 時系列을 豫測하게 된다. 指數平滑法과 要素分解法은 時間의 函數로써 이 豫測値를 表示하게 된다.

回歸法은 관측값을 모두 이용하여 관심있는 變數(종속변수)의 變動이 豫測하고자 하는 것이다.

多重 回歸에서 原因 또는 說明模型(Model)의 形態는

$$Y = a + b_1 X_1 + b_2 X_2 \cdots\cdots\cdots\cdots + b_k X_k + \mu \cdots\cdots (2.4)$$

이다.

式(2.4)에 X_1 , X_2,⋯⋯, X_k 를 G. N. P , 産業生産, 物價, 通貨 等과 같이 任意의 要因으로 나타낼 수 있다.

따라서 이들 變數를 $X_1 = Y_{t-1}$, $X_2 = Y_{t-2}$, $X_3 = Y_{t-3}$,⋯⋯⋯⋯, $X_k = Y_{t-k}$ 을 代入하면

$$Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + \cdots\cdots + b_k Y_{t-k} + \mu_t \quad \cdots\cdots (2.5)$$

이 된다.

式(2.5)은 回歸方程式이지만 式(2.4)의 右邊에 獨立變數와는 다르다. 단. 式(2.5)의 右邊값들은 從屬變數 Y_t 의 以前값이다. 즉 간략히 從屬變數의 時差값이라 하며 , 式(2.5)의 形態에 대한 方程式이나 模型을 說明하기 위하여 自己回歸(Auto- Regressioniar)를 利用한다.

式(2.5)은 單純指數平滑法 等의

$$F_{t+1} = \alpha X_t + \alpha(1-\alpha) X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \alpha(1-\alpha)^3 X_{t-3} + \cdots\cdots$$

과 類似한 形態이다.

指數平滑法에 의한 豫測에 있어서 過去値는 係數(母數) α , $\alpha(1-\alpha)$ $\alpha(1-\alpha)^2$, $\alpha(1-\alpha)^3$ 를 利用하여 加重된 값이다.

式(2.5)을 살펴볼때 한가지 疑問이 생기는데 이것은 時系列(즉 AR) 에 대하여 應用된 回歸法이 왜 原因模型에서 利用된 回歸模型과 다르게 취급되어지는가이다.

그러나 그 理由를 보면,

式(2.5)의 獨立變數가 대개 다른 變數에 의존하기 때문에 AR에 있어서 殘差의 獨立性에 대한 基本的인 假定이 쉽게 무너질 수 있고,

式(2.5)에 포함시키기 위한 Y_t 의 過去項의 數를 決定한다는 것이

아주 쉬운 것이 아니기 때문이다.

매우 一般的이고 높은 정확도를 갖는 時系列模型을 갖게 하기 위해선 효율적으로 AR模型을 MA (Moving Average , 이동평균) 模型과 結合시킬 수 있다.

이때 이 模型을 結合된 ARMA (Mixed Autoregressive Moving Average) 模型이라 한다.

2. 時系列의 特性 判別

올바른 豫測方法을 選擇하기 위해서는 時系列의 特性이 判別되어야 한다. 定常性 (Stationary), 季節性 (Seasonal) 等과 같은 系列의 特性을 判別한다는 것은 체계적인 接近方法이 必要하다.

이러한 過程을 時系列 分析이라 부르며 豫測하고자 하는 變數의 相異한 時差에 대한 自己相關係數를 利用하려는 것이다.

式 (2.5) 은 從屬變數 Y_t 와 右邊項의 k 差 變數 Y_{t-1} , Y_{t-2} ……… Y_{t-k} 로 構成되어 있는데 모든 變數의 過去値로 되어 있다.

Y_t 와 Y_{t-1} , Y_t 와 Y_{t-2} , Y_t 와 Y_{t-3} 또는 任意의 Y_t 와 Y_{t-k} 間에 單純相關關係는 回歸分析에서 알 수 있다. 이러한 相關關係는 같은 (自己) 變數로 생각될 수 있지만 相異한 時間 (時差) 의 變數이기 때문에 이것들을 自己 相關關係라 한다.

이들의 意味는 정확히 相關關係의 意味와 같다.

예를 들면 Y_t 와 Y_{t-1} 의 自己相關關係는 變數 Y_t 와 Y_{t-1} 가 어떻게 다른 變數와 聯關되어 있는가를 나타낸다. 만약에 하나의 系列이 완

전히 任意系列을 갖고 Y_t 와 Y_{t-1}의 相關關係를 計算하였다면 時系列의 各 값은 다른 값과 아무런 關係를 갖지 않기 때문에 그 相關關係는 "0"에 近似해질 것이다.

自己相關係數가 "0"에 접근한다는 것은 連續的인 값이 다른 값과는 關聯이 없다는 時系列을 意味한다. 반면에 雜音(noise : 不規則變動)이 없는 時系列 1, 2, 3, 4 ············ 20의 連續값 Y_t 와 Y_{t-1}의 自己相關關係는 아주 높은 것으로 期待되는데 이것은 連續的인 값 사이의 높은 정도의 從屬性이 存在하기 때문이다.

期間 1보다 많은 時差에 대한 自己相關을 살펴본다는 것은 주어진 時系列의 값이 어떻게 聯關되어 있는가에 관한 追加的인 정보를 제공해 주게 된다. 月別 時系列 資料를 利用한다면 그 系列이 季節性이 있는지 없는지를 알고자 한다고 假定하자. 만약에 그 系列이 季節的이라면 12個月로 分離되는 期間에 대한 資料는 다른 것과 關聯이 있는 것이다.

예를 들면 季節的 資料에서 모든 1月은 平均보다 높고 모든 8月은 平均보다 낮다면 Y_t 와 Y_{t-12} 間의 自己相關關係는 높은 값을 갖는다.

만약에 時差 12의 自己相關關係가 "0"으로 접근한다면 그 資料는 每年 같은 月間의 관계가 缺如되어 있음을 뜻하므로 季節類型이 存在하지 않는다. 비슷한 方法으로 다른 時差의 自己相關關係는 資料에 對해 다음과 같은 것을 判別할 수 있도록 利用된다.

1) 그資料가 任意的인 것인가?

2) 資料가 定常的인가?

3) 만약에 非定常的이라면 어떠한 水準에서 資料가 定常的이 되는가?

4) 資料가 季節的인가?

5) 만약에 季節的이라면 季節性의 길이는 얼마나 되는가?

위의 特性들을 自己相關分析을 利用하여 一般的인 方法으로 決定할 수 있다.

3. 自己相關關分析

3.1 自己相關係數

自己相關係數의 意味는 Graphic (Plot)를 利用하여 數學的으로 檢討하면 아주 잘 判別할 수 있다.

例에서와 같이 變數 Y_t를 製品 A에 대한 需要量이고 지난 10年間에 대하여 觀測된 값이 〈表 2.1〉의 第2列에 있는 값이라고 假定하자.

〈表 2.1〉 製品 A 需要量

時 間 t	原系列 Y_t	時差 1 系列 Y_{t-1}	時差 2 系列 Y_{t-2}
1	13	8	15
2	8	15	4
3	15	4	4
4	4	4	12
5	4	12	11
6	12	11	7
7	11	7	14
8	7	14	12
9	14	12	—
10	12	—	—

〈表2.1〉의 資料에 근거로 Y_t를

$$Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + \mu_t \quad \cdots\cdots\cdots\cdots (2.6)$$

로 표현할 수 있다.

式(2.6)은 두개의 以前값의 線形結合으로써 Y_t를 표현한 AR(自己回歸)時系列 模型이다.

變數 Y_{t-1}과 Y_{t-2}는 各各 1期間 또는 2期間 前의 값으로 移動하여 구할 수가 있다.

Y_{t-1}는 1개의 값, Y_{t-2}는 2개의 값을 잃게 된다.

Y_t와 Y_{t-1}, Y_t와 Y_{t-2} 間의 自己相關關係는 쉽게 계산할 수 있다. Y_t와 Y_{t-1}의 자기상관관계는 같은 變數의 連續値間에 어떤 관계가 있는지 또 Y_t와 Y_{t-2}의 自己相關關係는 期間(時差) 2인 連續値間에 어떤 關係가 있는가를 나타내는 것이다.

Y_t와 Y_{t-1}間의 單純相關關係는

$$\gamma_{Y_t, Y_{t-1}} = \frac{\sum_{i=1}^{n-1} (Y_t - \overline{Y}_t)(Y_{t-1} - \overline{Y}_{t-1})}{\sigma_{Y_t} \cdot \sigma_{Y_{t-1}}} \quad \cdots\cdots (2.7)$$

여기서 $$\sigma_{Y_t} = \sqrt{\frac{\Sigma (Y_t - \overline{Y}_t)^2}{n-1}} \quad \cdots\cdots (2.8)$$

$$\sigma_{Y_{t-1}} = \sqrt{\frac{\Sigma (Y_{t-1} - \overline{Y}_{t-1})^2}{n-1}} \quad \cdots\cdots (2.9)$$

이다.

〈表 2.1〉의 資料 계산하면

$$r_{Y_t, Y_{t-1}} = \frac{-27}{(12)(11.6)} = -0.19 \text{ (時差 1)}$$

$$r_{Y_t, Y_{t-2}} = 0.22 \text{ (時差 2)}$$ 이다.

系列 Y_t가 定常的(Stationary), 즉 $\overline{Y}_t = \overline{Y}_{t-1}$ 이라면 式 (2.7)을 單純化할 수 있고 時系列의 自己相關關係가 더 直接的으로 적용된다. 대개의 경우 Y_{t-1}가 觀測值보다 적은 한 時差를 갖는다는 사실은 無視되는 작은 差가 있게 된다.

특히 系列이 定常的(趨勢가 없는)이기 때문이다. 따라서 式(2.8)과 式(2.9)의 값은 같은 값을 갖는다. 이런 單純化된 假定을 利用하면 式(2.7)은 다음과 같다.

$$r_{Y_t, Y_{t-1}} = \frac{\sum_{t=1}^{n-1} (Y_t - \overline{Y}_t)(Y_{t-1} - \overline{Y}_t)}{\alpha_{Y1}^2} \quad \cdots\cdots\cdots (2.10)$$

$$= \frac{\sum_{t=1}^{n-1} (Y_t - \overline{Y}_t)(Y_{t-1} - \overline{Y}_t)}{\sum_{t=1}^{n} (Y_t - \overline{Y}_t)^2} \quad \cdots\cdots\cdots\cdots (2.11)$$

定常性의 假定은 Y_{t-1}와 Y_{t-2} 間의 자기상관계수가 Y_t와 Y_{t-1} 間의 自己相關係數와 같고 平均이 같으므로 $Y_t\, Y_{t-1}$의 合을 平均하면 $Y_{t-1}\, Y_{t-2}$와 같게 된다.

따라서 式(2.11)이 一般式이므로 한기간 時系列의 모든 時差에 利用할 수 있다.

式(2.11)은

$$\gamma_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \overline{Y}_t)(Y_{t-1} - \overline{Y}_t)}{\sum_{t=1}^{n} (Y_t - \overline{Y}_t)^2} \quad \cdots\cdots\cdots\cdots\cdots (2.12)$$

인데 여기서 γ_1은 時差 1의 自己相關係數이므로 시차 2, 3, 4 … n에 대한 自己相關關係는 γ_k로 定義한다. 즉 一般式

$$\gamma_k = \frac{\sum_{t=1}^{n=k} (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2} \quad \cdots\cdots\cdots\cdots\cdots (2.13)$$

을 利用하여 計算한다.

式(2.13)에 Y_{t+k}의 合은 式(2.12)을 다시 定理한 것이다. 이것은 定常性을 假定하기 때문에 그 結果들에 아무런 영향을 주지 않는다.

式(2.13)은 回歸의 相關係數로써 정확하게 說明되어질 수 있다. 즉 式(2.13)의 自乘은 Y_t와 Y_{t+k} 間의 AR線分이 $\overline{Y}_t$ 平均線分보다 더 나은 것인지를 意味하는데 설명된 변동이 총변동에 비율로 $0 < \gamma_k^2 < 1$인 값을 갖는다.

式(2.13)을 〈表 2.1 〉에 資料를 이용하여 어떻게 계산되어지는가를 알아보자.

$$\gamma_1 = \frac{(13-10)(8-10) + (8-10)(15-10) + (15-10)(4-10) + \cdots\cdots + (14-10)(12-10))}{(13-10)^2 + (8-10)^2 + (15-10)^2 + (4-10)^2 + \cdots\cdots + (14-10)^2 + (12-10)^2}$$

$$= \frac{(3)(-2) + (-2)(5) + (5)(-6) + (5)(-6) + (-6)(2) + \cdots\cdots\cdots + (4)(2)}{3^2 + (-2)^2 + 5^2 + (-6)^2 \cdots\cdots\cdots + 4^2 + 2^2}$$

$$= \frac{-17}{144} = -0.188$$ 이 된다.

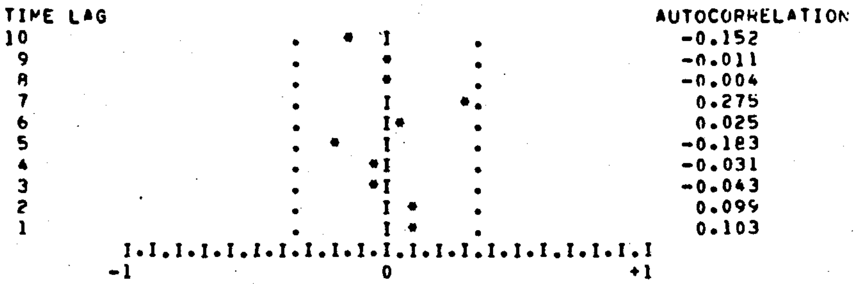
式(2.13)은 式(2.7)보다 자기상관을 계산하기 더 효율적이다. 더군다나, 非定常性이 資料에 存在할때 非定常性의 判別을 용이하게 하는 성질이 있다. 式(2.13)을 利用하여 r_2, r_3 도 계산할 수 있다.

3.2 系列 任意性 判別

資料가 任意的인지 아닌지를 判別하기 위해 自己相關을 利用한다. 만일 임의의 自己相關關係가 " 0 "으로부터 有意的 差가 있다면 몇개의 時差의 자기상관계수를 살펴보아야 한다.

〈그림 2.1 〉에서와 같이 任意性이 存在하는지 안하는지를 判別함에 있어 첫단계로 自己相關係數를 圖示하는 것이 편리하다.

自己相關係數는 平均 " 0 "와 標準誤差 " $1/\sqrt{n}$ "인 正規曲線에 근사한 標準分布를 갖는다. 따라서 모든 標本에 근거를 둔 自己相關係數의 95 %는 平均±(1.96)×(標準誤差)에 의해 計算된 범위내에 있어야 한다.



〈그림 2.1 〉

즉 계산된 自己相關係數가

$$-1.96 \times (\text{標準誤差}) \leq r_k \leq 1.96 \times (\text{標準誤差})$$ 범위내에 있다면 그 系列을 任意的이라고 判別할 수 있다.

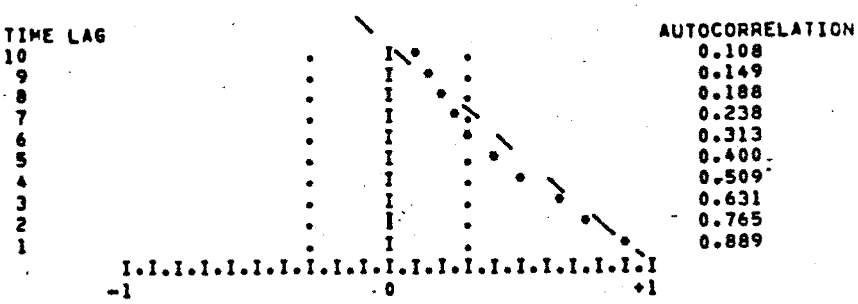
3.3 定常性 存在의 判別

定常性이란 資料가 連續的으로 增加하거나 減少하는 現狀이 없는것이 말한다, 즉 자료가 반드시 時間축에 따라 水平的이어야 한다. 다시 말하면, 資料가 상수평균을 中心으로 變動하고 時間에 독립이어야 한다.

系列에서 資料의 定常性의 特徵을 自己相關係數로 살펴봄으로써 쉽게 判別할 수가 있다. 定常資料의 自己相關이 時差 2 또는 時差 3 이후에 " 0 " 에 근접하는 반면에 非定常系列은 몇 時差後에 " 0 "과 有意的差가 있다.

이를 表示하면 非定常資料의 自己相關이 增加한 時差數에 따라 그 傾向이 오른쪽에서 왼쪽으로 (대각선방향)진행한다.

〈그림 2.2 〉는 非定常系列에 대한 自己相關의 PLOT를 나타낸 것이다.



〈그림 2 . 2 〉

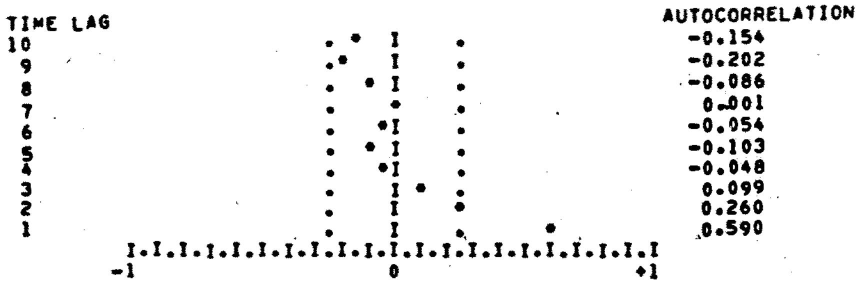
連續된 값들 間에 서로 높은 相關關係가 있으면 資料에 趨勢가 存在함을 意味한다.

만일 系列에 任意性이나 다른 變動이 存在하지 않는다면 많은 非定常系列에 自己相關이 직선으로 나타나며 系列에 다른 變動이 存在하면 自己相關이 직선 주위에서 미세한 변동을 하게 된다.

〈그림 2.2 〉는 2 번째 또는 3 번째값 이후에 자기상관이 " 0 "에 근사

하지 않고 趨勢를 나타내기 때문에 非定常系列임을 알 수 있다.

또한 〈그림 2.3〉은 自己相關이 定常的이지만 非任意系列임을 나타내고 있다. 즉 時差 그 이후에 자기상관이 "0"과 有意的 差가 없음을 알 수 있다. 1과 2期間의 時差에 대한 有意的 자기상관이 있음은 추세이외의 또 다른 형태의 유형이 存在함을 意味한다.



〈그림 2 . 3〉

3.4 非定常性 除去

어떤 종류의 趨勢로 인하여 자기상관형태에 영향을 주는 허위자기상관이 산출되는 경향이 있다.

따라서 時系列分析을 하기전에 우선 資料가 갖고 있는 非定常性을 除去해야 되고 除去하는 方法으로 階差(Differencing)를 통상적으로 이용한다.

예로서, 線形과 非任意性을 갖고 있는 單純系列 2,4,6,8,……20인 연속값들에게 1次 階差를 취하면, 4-2,6-4,8-6,…, 20-18이 된다.

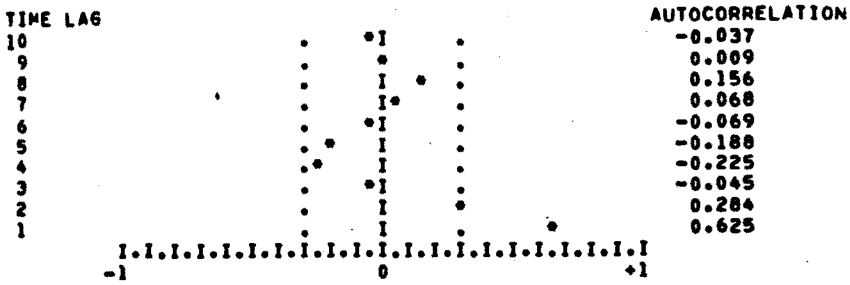
따라서 새로이 생긴 系列이 2,2,…,2가 된다. 이 系列은 분명히 定常的이다.

그러므로 定常性을 갖게 하기 위해선 새로운 系列이 생기는데 이것은 連續期間間의 階差

$Z_t = X_{t+1} - X_t$ ·· (2.14)가 된다.

새로운 계열 Z_t 는 n-1개(즉 원계열보다 1개 적은)의 값을 갖고, 만일에 原資料에서 趨勢가 線形(첫번째 差數에)이라면 새로운 系列은 正常化된 것이다.

〈그림 2.2 〉에서 使用된 資料에 1次 階差를 하면 〈그림 2.4 〉와 같은 自己相關의 系列을 갖게 된다.



〈그림2 . 4 〉

〈그림 2.4 〉에서 첫번째와 두번째 自己相關係數는 " 0 "과 有意的 差는 있지만 그외의 것은 " 0 "과 有意的差가 없는데 이것은 1差階差系列에서 資料가 定常化된 것을 뜻한다. 만일에 1差階差된 資料의 自己相關이 두번째 또는 세번째 時差(lag)이후에 " 0 "에 근접하지 않으면 그것은 아직도 定常化되지 않았음을 의미한다. 따라서 1차 계차된 자료를 계속하여 1차계차를 취해야 한다.

즉 $W_t = Z_{t+1} - Z_t$ ·· (2.15)

W_t 는 n-2개의 값을 갖는 2次階差된 系列이다. 式(2.14)을 式 (2. 15)에 代入하면

$W_t = (X_{t+2} - X_{t+1}) - (X_{t+1} - X_t) = X_{t+2} - 2X_{t+1} + X_t$ …… (2.16)

이 된다.

계차의 계산과정을 (表2.2)에 나타내었다.

〈表2 . 2〉

時　間(1)	時　系　列(2)	1次階差(3)	2次階差(4)
1	2.44	2.86	0.81
2	5.30	3.67	1.24
3	8.97	4.91	0.79
4	13.88	5.70	1.71
5	19.58	7.41	1.55
6	26.99	8.96	0.95
7	35.95	9.91	-0.07
8	45.86	9.84	1.82
9	55.70	11.66	0.61
10	67.36	12.27	0.23
11	79.63	12.5	-
12	92.13	-	-

이 表를 보면 두번째 수준에서 非定常性이 포함되어 있음을 안다. 表의 제 3열은 1次階次인데 式(2.14)을 이용하면

$Z_1 = X_2 - X_1 \;=\; 5.3 - 2.44 = 2.86$

$Z_2 = X_3 - X_2 \;=\; 8.97 - 5.3 \;=\; 3.67$

$\vdots$

$Z_{11} = X_{12} - X_{11} = 92.13 - 79.63 = 12.5$ 이다.

Z_{12} 은 計算할 수 없기 때문에 1次階差된 계열은 n- 1개의 觀測値만 갖게 된다.

제 4 열은 2次階差로써 式(2.15)을 이용한다. 즉

$$W_1 = Z_2 - Z_1 = 3.67 - 2.86 = 0.81$$

$$W_2 = Z_3 - Z_2 = 4.91 - 3.67 = 1.24$$

$$\vdots \qquad\qquad \vdots \qquad \vdots$$

$$W_{10} = Z_{11} - Z_2 = 12.5 - 12.27 = 0.23$$

이 새로운 系列은 W_{11}, W_{10} 을 계산할 수없기 때문에 두개의 값을 잃은 n-2개의 값을 갖는다.

또한 2차계차는 式(2.16)을 利用하여 구할 수도 있다. 즉

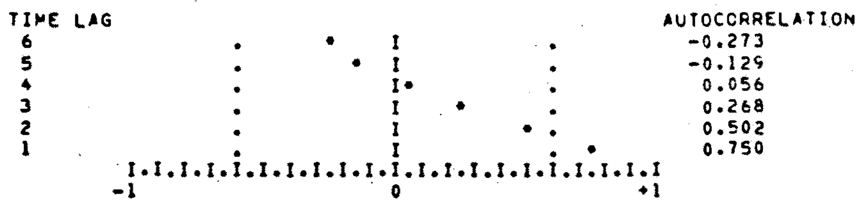
$$W_1 = X_3 - 2X_2 + X_1 = 8.97 - 2(5.30) + 2.44 = 0.81$$

$$W_2 = X_4 - 2X_3 + X_2 = 13.88 - 2(8.97) + 5.30 = 1.24$$

$$\vdots \qquad\qquad \vdots \qquad \vdots$$

$$W_{10} = X_{12} - 2X_{11} + X_{10} = 92.13 - 2(79.63) + 67.36 = 0.23$$

만일에 〈表 2.2 〉의 時系列 資料에 관하여 아는 것이 아무것도 없다면 이 分析에서의 1단계는 제 2열의 原時系列資料에 대한 自己相關을 計算한다. 이 自己相關係數를 〈그림 2.5 〉에 나타내었다.



〈그림 2.5 〉

〈그림 2.6〉에서 보면 自己相關에서 趨勢가 있음을 나타내기 때문에 이 資料가 定常的이 아님을 알 수 있다.

따라서 2 단계는 제 3 열의 階差된 資料를 利用하여 自己相關係數를 〈그림 2.6〉에 나타내었다.

〈그림 2.6〉

위 그림의 자기상관은 〈그림 2.5〉의 자기상관과는 유의적으로 많은 差가 없다. 즉 自己相關이 "0"에 급격히 접근하지 않고 있는데 이것은 1次階差된 系列이 아직도 非定常的임을 의미하는 것이다. 따라서 다른 계차(1차계차된 계열을 다시 1차계차 또는 2차계차)를 취하여 〈表 2.2〉의 제 4 열에 나타난 2차계차된 資料의 自己相關을 구한다. 2차階差된 系列의 自己相關은 〈그림 2.7〉에 도시하였는데 이 수준에서 定常性이 나타나고 있다. 定常性이 되기 위해선 時差 2 또는 3 내에 자기상관이 "0"에 접근할 때까지 계차를 계속해야 하는 기계적인 일로 축소할 수 있다.

〈그림 2.7〉

그런데 첫번째 水準 또는 두번째 水準에서 實質資料가 非定常性을 내포하기 때문에 2차階差 이상에 계차를 계속적으로 취할 필요는 없다.

3.5 季節性 判別

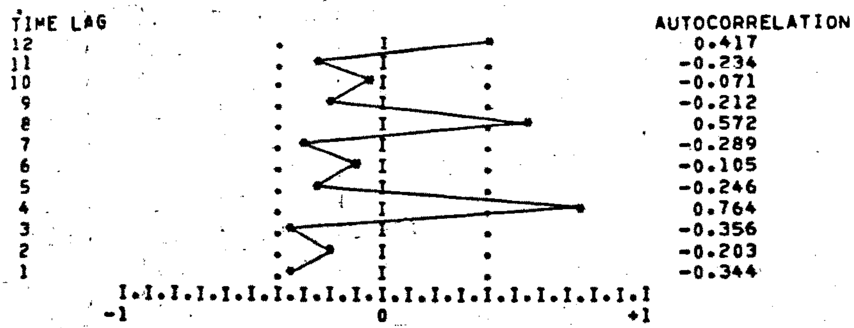
季節性이란 일정한 時間間隔에 따라 반복하는 하나의 형태 (월 = 12, 24, 분기 = 4, 8)로 定義한다.

예를들면 난방용 유류 판매량은 겨울에 높고, 여름에는 낮다. 이것은 12개월 주기의 계절형태가 있음을 뜻하는 것이다. 만일에 이 形態가 一次的인 形態라고 한다면 12개월 時差의 自己相關係數가 系節性의 存在를 意味하는 높은 " + "의 값을 갖게 된다.

또한 " 0 "과 有意的 差가 없다면 1년 주기의 各月들이 첫번째 年度에서 다음 년도에 생기는 形態와 일치하지 않는 無相關 (任意的)이 된다. 이런 資料는 季節性이 없다.

定常資料에 대해 定常性은 " 0 "와 有意的 差가 있는 2 또는 3 以上 時差의 自己相關係數를 判別하여 구할 수 있다. " 0 "과 有意的 差가 있는 自己相關이 資料에 存在할때 어떤 形態의 有形이 存在함을 意味하는 것이다.

季節性을 判別하기 위해 이런 주기의 높은 自己相關이 있는지 보아야 한다. 〈그림 2 . 8 〉에 나타나 있는 自己相關係數는 기간 4 (분기별 자료) 마다 높은 자기 상관이 보이고 있어 계절유형이 存在함을 알 수 있다.



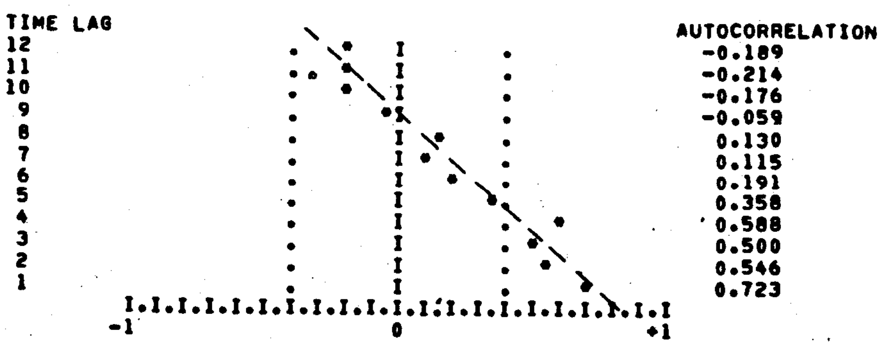
〈그림 2 . 8 〉

즉, $r_4 = 0.764$ 는 "0"과 有意的 差가 있음을 볼 수 있고 또 $r_8 = 0.572$ 도 有意的이며 $r_{12} = 0.417$ 은 신뢰한계선 상에 있다. 이런 季節性의 週期的 特性 ($r_4 > r_8 > r_{12}$)은 세기간의 相關 모두가 "0"과 유의적 差가 있다는 사실을 알 수 있다.

實際的으로 自己相關係數 $r_4 = 0.764$ 는 기간 4 의 간격을 갖는 季節性이 있음을 判別하기에 충분하다.

더불어 r_8 와 r_{12} 의 값도 크다는 사실로 이를 확인할 수 있다. 季節性이 단지 形態만 있다면 自己相關의 그림으로 쉽게 볼 수 있고 또는 단순히 다른 時差의 自己相關을 살펴봄으로써 알 수가 있다. 季節性도 系列에 趨勢와 같은 다른 形態가 混合되어 있을 때 判別하기가 容易한 것이 아니다. 資料에 非定常性이 있으면 類似한 自己相關이 나타나기 때문에 趨勢性이 더 강해져서 季節性이 적어지게 된다. 이 문제는 資料가 定常性일 때만 季節性을 決定하여 해결할 수 있다. 이것은 규칙처럼 자료에서 季節性을 결정하기 전에 추세의 나타남을 階差方法으로 資料를 變形하여 定常化하여야 함을 의미하는 것이다.

〈그림 2 . 9〉는 非定常資料系列에 대한 自己相關 그림이다.



〈그림2 . 9〉

위 그림에서 自己相關의 움직임이 趨勢가 있음을 알 수 있다. 즉 4번째 期間 時差에 대응하는 自己相關은 $r_4 = 0.588$이다. 이 값은 2번째 期間時差의 값보다 크며 마찬가지로 r_8도 바로 이전의 값보다 크다. 이것은 季節性의 存在를 意味하지만 분명하게는 判別되지 않는다. 趨勢를 찾아낸 다음에 그 系列을 반드시 階差한 後 階差된 系列의 자기상관을 計算하여야 한다. 그 결과가 앞의 〈그림 2 . 8 〉에 나타난 plot이다.

"0"과 有意的 差가 있는 季節要因을 제외하고는 自己相關이 有意的 差가 있는 것이 하나도 없기 때문에 階差된 系列은 定常的이다. 〈그림 2 . 8 〉은 週期가 4인 季節形態가 명백히 나타나고 있어 自己相關分析에 있어서 두가지 質問에 대한 해결을 전부 할 수 있기 때문에 平滑化 模型을 決定하고 季節性 주기를 利用하고자 그 資料를 圖示할 필요는 없는 것이다. 趨勢와 季節性을 갖는 資料에 모든 自己相關들은 〈그림 2 . 8 〉과 〈그림 2 . 9 〉에서 보여주는 것과 같이 나타나지 않는다. 趨勢가 季節性과 비교할 때 더 강하면 原資料의 自己相關은 한 線上에 있게 된다. 다른 極端的인 경우에 季節性이 나타나지 않고 趨勢만 나타나게 된다. 즉 자기상관분석에서 첫단계인 非定常性을 반드시 제거하면 다른 形態를 說明할 수 있다. 따라서 다음 세가지 경우가 있을 수 있다.

1) 自己相關들이 〈그림 2 . 1 〉의 形態로 움직일 것이다. ("0"과 有意的 差가 없음)
2) 自己相關들이 〈그림 2 . 3 〉에서와 같은 어떤 形態가 나타나게 될 것이다. (단, 첫번째, 두번째, 세번째 自己相關만이 有意性이 있음)
3) 自己相關들이 〈그림 2 . 8 〉에 자기상관과 類似해 질 것이다. 이것은 季節有形이 나타난 것이다.

4. 有意性 檢定 (χ^2 －檢定)

몇개의 自己相關係數가 “ 0 ”과 有意的 差가 있는지 없는지를 決定할 수 있는 程度를 나타내는 하나의 檢定統計量 (Q統計量)을 Box & Pierce (1970)가 展開시켰다.

이 檢定統計量은 自己相關係數가 χ^2 －分布에 따름을 根據로 두고 있다. 만일에 檢定統計量으로 計算된 값이 χ^2 分布表의 統計量의 값보다 적다면 檢定하기 위해 使用된 自己相關은 “ 0 ”과 有意的 差가 없다는 것이다. 이것은 自己相關을 발생시키는 資料가 任意的인 것을 意味한다.

만일에 計算된 χ^2 －檢定量 (즉 Q－統計量)이 χ^2 分布表로 부터 찾은 값보다 크다면 自己相關은 “ 0 ”과 有意的 差가 있는데 이것은 어떤 模型이 存在함을 暗示하는 것이다.

Q－統計量은

$$\chi^2 = n \cdot \sum_{k=1}^{m} r_k^2$$ 인데

여기서 m 은 計算된 最大時差이다.

〈그림 2 . 6 〉의 自己相關에 대한 χ^2－값 ($n = 12$)은

$$\chi^2 = 12 \cdot [(0.751)^2 + (0.475)^2 + (0.213)^2 + (-0.01)^2 + (-0.215)^2 + (-0.346)^2] = 12 \cdot (0.98) = 11.76$$ 이며

마찬가지로 〈그림 2 . 7 〉의 χ^2 값은

$$\chi^2 = 12 [(-0.202)^2 + (-0.217)^2 + (0.155)^2 + (0.151)^2 + (-0.322)^2] = 12 \cdot (0.238) = 2.86$$ 이다.

〈그림 2 . 7 〉에 대한 χ^2 값이 보다 작다고 하는 〈그림 2 . 7 〉의 自己相關이 〈그림 2 . 6 〉의 自己相關에 비하여 任意的인 것에 보다 접근된 것임을 意味하는 것이다.

χ^2 一檢定은 自己相關 및 資料에서의 任意性의 크기(범위)를 決定하는 첫번째 근사치로써 利用할 수 있다.

또한 이 χ^2 一檢定量은 誤差의 自己相關에도 應用된다. 만일에 殘差의 χ^2의 計算된 값이 χ^2 分布表의 값보다 작다면(주어진 自由度에 비해) 自己相關이 "0"과 有意的 差가 없다는 假設은 그대로 維持가 되는데 다만 任意誤差가 남아있기 때문에 殘差는 任意的으로 分布를 하며 使用된 模型은 적절한 模型임을 暗示하는 것이다.

5. 偏自己相關關係

偏自己相關關係는 X에 대하여 다른 時差의 효과가 계속될 때 X_t와 X_{t-q} 간에 얼마나 關聯이 있는가를 測定하는데 利用된다.

時系列分析에서 偏自己相關關係의 利用목적은 豫測을 위해 적합한 ARMA 模型을 判別하는데 도움을 주기 위한 것이다.

偏自己相關은 AR(m)모형의 마지막 項으로 정의한다.

즉, $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, …… $\hat{\phi}_{m-1}$, $\hat{\phi}_n$는 式(2.18)~式(2.22)에서 볼 수 있는 것과 같은 任意의 AR模型의 m개 偏自己相關이다.

$$X_t = \hat{\phi}_1 X_{t-1} + e_t \quad \cdots\cdots (2.18)$$

$$X_t = \phi_1 X_{t-1} + \hat{\phi}_2 X_{t-2} + e_t \quad \cdots\cdots (2.19)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \hat{\phi}_2 X_{t-3} + e_t \quad \cdots\cdots (2.20)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots\cdots + \hat{\phi}_{m-1} X_{t-m+1} + e_t \quad \cdots\cdots (2.21)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots\cdots + \phi_{m-1} X_{t-m+1} + \hat{\phi}_m X_{t-m} + e_t \quad \cdots (2.22)$$

물론 이들의 값을 定하기 위해 $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3 \cdots \hat{\phi}_{n-1}, \hat{\phi}_m$ 에 대해 方程式을 풀 수 있다. 그런데 이 計算은 극히 많은 時間이 요구되어 진다. 따라서 自己相關係數에 근거를 둔 $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \cdots\cdots \hat{\phi}_{m-1}, \hat{\phi}_m$ 의 推定値를 구하는 것이 좀 더 바람직하다. 이런 推定은 다음과 같은 方法에 의해서 구할 수 있다.

만일에 式(2.18)의 양변에 X_{t-1} 을 곱하면 그 結果는

$$X_{t-1} \cdot X_t = \phi_1 X_{t-1} X_{t-1} + X_{t-1} e_t \quad \cdots\cdots (2.23)$$

인데 이 式(2.23)에다 期待値를 취하면

$$E(X_{t-1} X_t) = \phi_1 E(X_{t-1} X_{t-1}) + E(X_{t-1} e_t)$$

$$r_1 = \phi_1 r_0 \quad \cdots\cdots (2.24)$$

가 되는데 이것은 定義에 의해서

$E(X_{t-1} X_t) = r_1$, $E(X_{t-1} X_{t-1}) = r_0$, $E(X_{t-1} e_t) = 0$ 이기 때문이다.

만일에 式(2.24)의 양변을 r_0 로 나누면

$$r_1 / r_0 = \frac{\phi_1 r_0}{r_0}$$

$$\rho_1 = \phi_1 \quad \cdots\cdots (2.25)$$

이 되는데

$\rho_k = (r_k / r_0)$ 이기 때문이다.

따라서 $\hat{\phi}_1 = \hat{\rho}_1 = r_1$ 이다.

즉 時差1에 대한 偏自己相關은 ρ_1 이거나 또는 이의 標本推定値 r_1 이다. 대개 式(2.18)～(2.22)의 양변에 r_{t-k} 를 곱하고 期待値를 취한 다음 r_0 로 나누면 하나의 聯立方程式(Yule-Walker 方程式이라고 함)이 생기는데 이것으로 $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3 \cdots\cdots \hat{\phi}_{m-1}, \hat{\phi}_m$ 을 풀 수가 있다. 따라서 이 값들은 m 時差에 이르기까지의 偏自己相關의 推定値로써 利用할 수가 있다.

偏自己相關이 무엇이고 이것을 어떻게 求하는가를 理解한 다음의 관심은 적합한 ARMA模型을 判別하는데 어떻게 이것을 利用할 수 있는가에 있다. 만일에 주어진 系列에서 발생되는 基本的인 過程이 AR(1) 模型이라면 $\hat{\phi}_1$만 "0"과 有意的 差가 있지만 반면에 $\hat{\phi}_2$, $\hat{\phi}_3$ ……$\hat{\phi}_{m-1}$, $\hat{\phi}_m$은 統計的으로 有意하지 않다는 것으로 理解되어진다. 만일에 實質的인 발생과정이 AR(2)이라면 $\hat{\phi}_1$과 $\hat{\phi}_2$만 有意的이며 나머지 推定된 값들은 有意하지 않게 된다. 高次數의 AR模型에 대해서도 같은 의미로 말할 수 있다.

바꿔말하면 $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ ……$\hat{\phi}_{m-1}$, $\hat{\phi}_m$이 計算되어진 方法때문에 이것들은 생성된 資料에 진정한 AR模型의 次數에 이르기까지 단지 "0"과 有意的 差가 있다는 것이다. 그러면 模型判別에서 만일에 두개의 有意的인 偏自己相關만 存在한다고 하면 그 생성과정은 제2차 模型이며 豫測模型의 次數는 AR(2)가 될 것이다. 또한 P개의 有意的 偏自己相關이 存在한다면 그 次數 AR(P)가 된다. 따라서 判別目的에 대하여 만일 模型이 AR模型(自己相關係數가 指數函數型으로 "0"에 수렴)이라면 偏自己相關은 模型의 次數를 決定하는데 利用한다. 그 次數는 有意的인 偏自己相關의 數와 같게 된다.

만일에 생성과정이 AR보다는 오히려 MA模型이라면 偏自己相關은 AR模型에 적합되도록 設定되었기 때문에 MA模型의 次數를 決定하는데 利用되지는 않는다. 그러나 AR模型에 대한 自己相關의 값과 같은 의미로 움직이도록 時差에서 다음 時差까지 종속성을 도입하게 되면 偏自己相關이 指數型으로 "0"에 근접하게 된다. 判別目的에 대하여 偏自己相關이 時差P 이후에 任意값으로 떨어지는 것이 나타나지는 않지만 "0"에 指數型으로 수렴할 때는 실질 생성과정은 MA模型이라고 假定할 수 있다.

단지 P次數의 偏自己相關이 "0"과 有意的 差가 있다고 할 때 그 模型이 AR(P)임을 가정하고 偏自己相關이 指數型으로 "0"에 수렴할 때 그 模型을 MA模型으로 假定한다.

6. 時系列 分析節次

自己相關係數를 利用하여 機械的인 方法으로 時系列 分析을 할 수 있음을 알았다. 그 過程은 다음의 단계와 같다.

(1) 原系列의 自己相關係數를 求한다.

만일에 自己相關係數들이 두번째 또는 세번째 값 이후에 급격히 "0"에 접근하면 그 資料의 原形態가 定常的임을 意味한다. 임의의 有形 (단계(3)을 참조)을 살펴보아야 한다. 만일 "0"에 접근하지 않으면 그것은 非定常性임을 의미하게 된다((2)을 참조).

(2) 自己相關이 非定常性임을 意味할 때 原系列에 대해 1次階差를 취하이 자기상관을 計算한다.

만일 이 自己相關係數들이 定常性임을 나타내면 나머지 有形에 대하여 살펴본다(단계(3)을 참조).

그런데 그 自己相關이 여전히 非定常的이라면 다시 1次階差를 취하고 이들의 自己相關係數를 검토한다. 대부분 實際的인 目的에 따라 2次階差 정도로 資料를 定常系列로 變換시키게 된다.

(3) 만일 資料가 定常性이라면 이들의 自己相關係數를 살펴본다. 3보다 큰 次數를 갖는 系列의 自己相關係數가 有意的이라면 이것은 가장 큰 自己相關係數의 時差에 대응하는 주기를 갖는 季節形態를 意味하는 것이다. 또한 첫번 세개의 自己相關係數가 有意的이라면 다른 非季節類型을 제시하

는 것이다.

끝으로 만일 "0"과 有意的 差가 있는 自己相關係數가 하나도 없다면 아무런 形態가 存在하지 않으며 나머지는 任意性임을 제시하는 것이다.

自己相關의 PLOT는 原系列의 PLOT와는 완전히 틀린다는 것을 상기하여야 한다. 原系列의 PLOT는 有形의 形態를 判別하는데 도움을 주는 視覺的인 補助道具이다. 自己相關關係는 原系列에서 存在하는 有形의 結果이다. 自己相關關係의 PLOT는 原系列에 관한 것과 이들의 特徵을 밝혀주는 것이다.

Computer의 광범위한 有用性에 따라서 自己相關關係를 計算한다는 것은 이를 利用하고자 하는데 있어서는 귀찮은 일은 아니다. 이것은 時系列分析을 하는데 相對的으로 쉽게 해주고 있다.

끝으로 주의할 점은 다음과 같다.

原系列의 有形이 非季節的일 때 定常性은 時差가 2 또는 3이상의 모든 自己相關關係가 "0"과 有意的 差가 없을 때를 意味하는 것이다. 그런데 原系列이 季節的일 때는 이 규칙은 언제나 有意하지는 못한다. 季節有形은 時差가 2 또는 3이상의 自己相關關係가 "0"이 아니어야 한다는 原因이 된다. 그러한 경우에 자기상관이 어떻게 빨리 "0"에 수렴되는가를 검토하는 것보다 오히려 PLOT상에 "0"을 中心으로 自己相關이 어떻게 分布되어지는 가를 살펴보는 것이 오히려 더 낳다.

χ^2 — 檢定에 대한 것도 같다. 下限값은 대개보다 任意系列임을 意味하지만 항상 그렇지는 않은 것이다. 混合된 趨勢, 季節性, 任意性이라는 것들은 이의 값들을 왜곡시킬 수가 있다.

Ⅲ. ARIMA(Autoregressive integrated Moving Average)模型

1. AR(Auto-Regressive)模型

AR模型의 一般的模型은 다음과 같은 形式이다.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots\cdots + \phi_p X_{t-p} + e_t \quad \cdots\cdots\cdots \quad (3.\ 1)$$

式(3. 1)은 獨立變數가 從屬變數의 過去값과 誤差項으로 이루어져 있기 때문에 一般回歸方程式

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots\cdots + b_k X_k \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.\ 2)$$

와는 다르다.

式(3. 2)도 AR(自己回歸)이지만 回歸母數는 線形最少自乘法을 통해 推定되어지는 반면에 式(3. 1)의 AR母數는 非線形最少自乘法을 使用하여 구할 수가 있다. 또한 式(3. 2)의 分散도 最少自乘法으로 計算되고 式(3. 1)의 分散은 非線形最少自乘法으로 獨立變數들이 서로 相關된다는 사실을 說明할 수 있다. 이 點이 주요한 差異點이다.

式(3. 1)의 一般的인 AR(p)模型은 p次數에 의존하는 몇가지 形態를 취할 수 있다.

p = 1일때 1次 AR模型 또는 AR(1)이다.

一般模型은 AR(p)로 나타낸다. AR模型을 使用하기전에 이模型이 次數 p는 반드시 判別되어야 한다. 포함되어야 하는 項의 數로 분류되는 p의 근사치는 自己相關係數로 검토하여 求할 수 있다. AR(p)類型의 性質은 數理的인 形態를 살펴보는 것이 보다 理解하기가 좋다.

예를들면 AR(1)模型은

$$X_t = \phi_1 X_{t-1} + e_t \quad \cdots\cdots (3.3)$$

와 같이 나타낸다.

$X_{t-1} = \phi_1 X_{t-2} + e_{t-1}$ 를 式(3. 3)에 代入하면 式(3. 3)은

$$X_t = \phi^2 X_{t-2} + \phi_1 e_{t-1} + e_t \quad \cdots\cdots (3.4)$$

가 된다.

마찬가지로 $X_{t-2} = \phi_1 X_{t-3} + e_{t-2}$ 를 式(3. 4)에 代入하면

$$\begin{aligned} X_t &= \phi_1^2(\phi_1 X_{t-3} + e_{t-2}) + \phi_1 e_{t-1} + e_t \\ &= \phi_1^3 X_{t-3} + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t \quad \cdots\cdots (3.5) \end{aligned}$$

이다.

最初의 觀測値에 X_{t-i} 의 代入을 계속하면

$$X_t = \phi_1^{n-1} X_{t-n+1} + \phi_1^{n-2} e_{t-n+2} + \cdots + \phi_1^3 e_{t-3} + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t \quad \cdots\cdots (3.6)$$

로 나타내어진다.

AR(1)模型이 이 指數形으로 減少하므로 過去時系列값이 加重된 單純指數平滑法의 式

$$\begin{aligned} F_{t+1} = \alpha X_1 &+ (1-\alpha) X_{t-1} + (1-\alpha)^2 X_{t-2} + (1-\alpha)^3 X_{t-3} \\ &+ \alpha(1-\alpha)^4 X_{t-4} + \alpha(1-\alpha)^5 X_{t-5} + \cdots\cdots + \alpha(1-\alpha)^{N-1} X_{t-(N-1)} \end{aligned}$$

과 類似하다.

AR(2)模型은 다음과 같다.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t \quad \cdots\cdots (3.7)$$

式(3. 7)을 過去誤差에 의해서 표현될 수 있다.

즉 連續期間의 誤差의 交叉(cross)곱을 포함한 誤差를 제외하고는 加

重值는 式(3. 6)과 비슷하다.

예를들면 부분적인 代替는

$$X_t = (\phi_1^2 + \phi_2)\phi_1 X_{t-3} + (\phi_1^2 + \phi_2)\phi_2 X_{t-4} + \phi_1^2\phi_2 e_{t-2} + \phi_1 e_{t-1} + e_t \quad \cdots\cdots (3.8)$$

로 나타내어진다.

式(3. 1)에서 나타난 一般AR(p) 模型은 매우 적응성이 높고 變動하기 쉬운 時系列方法으로 생각된다. 이것은 p를 判別함으로써 모든 時系列의 形態를 單純하게 취급할 수 있는 模型의 一般的인 水準을 다시 나타내는 것이다.

2. MA(Moving Average) 模型

p가 거의 작을때 AR(p) 模型을 갖고 分離할 수 없는 몇가지 時系列 有形이 存在할 수도 있다. 그런데 다행스럽게도 "MA(移動平均)"라고 불리는 一般的 模型의 또 다른 形으로 이러한 gap을 채울 수가 있다. 1938年에 Wold는 任意의 離散時系列을 AR模型 또는 MA模型으로 表現할 수 있음을 입증하였다. MA模型은 X_t의 實測過去值 p의 線形函數로서 X_t를 表現하는 AR模型에 대응하여 過去誤差의 線形結合에 근거를 둔 X_t의 豫測值를 나타내고 있다. 一般的 MA模型은

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots\cdots - \theta_q e_{t-q} \quad \cdots\cdots (3.9)$$

式(3. 9)을 移動平均(Moving Aoerage)模型이라 부르고 MA(q)模型으로 나타낸다.

MA(1)模型은

$$X_t = e_t - \theta_1 e_{t-1} \quad \cdots\cdots (3.10)$$

이다. 또는 $X_t = e_t - \theta_1(X_{t-1} + \theta_1 e_{t-2})$ 이다.

왜냐하면 $X_{t-1} = e_{t-1} - \theta_1 e_{t-2}$ 이거나

$e_{t-1} = X_{t-1} + \theta_1 e_{t-2}$ 이기 때문에

$$X_t = e_t - \theta_1 X_{t-1} - \theta_1^2 e_{t-2} \quad \cdots\cdots (3.11)$$

이 된다.

e_{t-2}의 값을 式(3、11)에 代入하면

$$X_{t-2} = e_{t-2} - \theta_1 e_{t-3} \quad \cdots\cdots (3.12)$$

또는 $e_{t-2} = X_{t-2} + \theta_1 e_{t-3}$ 로 주어진다.

式(3. 12)을 式(3. 11)에 代入하면

$$X_t = e_t - \theta_1 X_{t-1} - \theta_1^2 X_{t-2} - \theta_1^3 e_{t-3} \quad \cdots\cdots (3.13)$$

이 代入을 계속하면 式(3. 13)은

$X_t = e_t - \theta_1 X_{t-1} - \theta_1^2 - \theta_1^3 X_{t-3} \cdots\cdots \theta_1^{n-1} X_{t-(n-1)} - \theta_1^n X_{t-n}$ 이 된다.

분명히 MA(1)模型은 單純指平滑模型과 일치한다.

MA(2)模型은

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \quad \cdots\cdots (3.14)$$

로써 이式은 式(3. 13)과 비슷하게 表現할 수 있고 $X_t X_{t-1}$의 교차 항을 포함하게 된다.

MA(3)模型은

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} \quad \cdots\cdots (3.15)$$

와 같이 쓰며 반면에 一般的인 MA(q)模型은 式(3. 9)에서 나타난 것과 같은 模型이다.

3. 結合 ARMA(Mixed Autoregressive Moving Average) 模型

AR(p)模型과 MA(q)模型을 追加하게 되면 같은 方程式 안에 AR과 MA을 結合시킬 수가 있다. 이러한 類의 가장 一般的인 경우는 ARMA(p, q)模型이라고 불리운다. AR模型과 MA模型은 이러한 부류의 특수한 경우이다.

예를들면 AR(2)模型은 ARMA(2. 0)模型으로 表現할 수 있고 MA(1) 模型은 ARMA(0. 1)과 같이 나타낼수가 있다. 가장 一般的인 ARMA模型은 次數가 p와 q이다.

式(3. 1)과 式(3. 9)을 結合하면 다음과 같이 된다.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \cdots\cdots \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \cdots\cdots \theta_q e_{t-q} \quad (3.16)$$

특히 ARMA(1. 1)은

$$X_t = \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \quad (3.17)$$

인데 이式(3. 17)

$$X_t = -\theta_1 X_{t-1} + \phi_1^2 X_{t-2} + \phi_1 \theta_1 X_{t-2} + e_t + \phi e_{t-1} - \theta_1^2 e_{t-2} - \phi_1 \theta_1 e_{t-2} \quad (3.18)$$

과 대응한다.

式(3. 18)은 AR과 MA을 結合한다. 즉 式(3. 17)에

$X_{t-1} = \phi_1 X_{t-2} + e_{t-1} - \theta_1 e_{t-2}$ 와

$e_{t-1} = X_{t-1} - \phi_1 X_{t-2} + \theta_1 e_{t-2}$ 을 代入하여 구한다.

ARMA(2, 1)模型은

$X_t = \phi_1 X_{t-1} - \phi_2 X_{t-2} + e_t - \theta_1 e_{t-1}$ 이다.

따라서 式(3.18)은 몇개의 項을 포함한 過去値와 過去誤差 및 AR과 MA 母數 둘 다 結合한 것을 利用한 ARMA模型들임을 알 수 있다. 이 포괄적인 특징 때문에 이들의 結合된 模型은 AR(1)과 MA(1)을 각기 利用하는 것보다는 효율적이다.

4. 季節模型(Seasonal Model)

季節的時系列은 한 週期에 의한 類型에 부가함에 每1번째 期間에 일어나는 系列에서 오랜동안 반복된 形이기 때문에 ARMA模型을 豫測하는데 季節的 時系列은 또다른 차원의 어려움을 부가시키고 있다. 季節類型을 豫測하기 위하여 반드시 季節的母數가 포함되어야 한다. 非季節的 模型에 있어서와 같이 自己回歸形態가 되거나 移動平均形態가 될 수 있다.(또한 結合模型을 검토할 수 있지만 대개 실제에 있어서 그것이 不必要하다) 季節的 ARMA 模型을 表現하고 理解하기 위해서는 후향연산자(B)의 개념이 반드시 검토되어야 한다.

이 연산자는 數學的 意味는 없지만 模型을 나타내는데 보다 容易하도록 하고자 使用하는 것이다. 후향연산자(Backshift Operator)는 $B^m X_t$로써 定義하는데

예를들면, $BX_t = X_{t-1}$ 또는 $Be_t = e_{t-1}$

$$B^2 X_t = X_{t-2} \quad \text{또는} \quad B^2 e_t = e_{t-2}$$

$$\vdots$$

$$B^m X_t = X_{t-m} \quad \text{또는} \quad B^m e_t = e_{t-m} \quad \text{이다.}$$

ARMA模型은 후향연산자에 의해서 表現할 수가 있다. 예를들면 AR(1) 模型은

$$X_t = \phi_1 X_{t-1} + e_t \quad \cdots\cdots (3.19)$$

또는 $X_t - \phi_1 X_{t-1} = e_t$ 이다.

그런데 $X_{t-1} = BX_t$ 이기 때문에 式(3.19)은

$X_t - \phi_1 BX_t = e_t$ 또는 $(1-\phi_1 B) X_t = e_t$ 와 같이 表現할 수 있다.

같은 方法으로 AR (2) 模型에 대해서는 $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = e_t$ 이다.

$X_{t-1} = BX_t$, $X_{t-2} = B^2 X_t$ 이기 때문에 이式은

$X_t - \phi_1 BX_t - \phi_2 B^2 X_t = e_t$

또한 $(1-\phi_1 B - \phi_2 B^2) X_t = e_t \quad \cdots\cdots (3.20)$

와 같이 表現할 수 있다.

AR (2)模型을 갖고 1次階差를 使用하였다고 假定하자.

즉 $(X_t - X_{t-1}) = \phi_1 (X_{t-1} - X_{t-2}) + \phi_2 (X_{t-2} - X_{t-3}) + e_t$ 이다.

$X_{t-1} = BX_t$, $X_{t-2} = B^2 X_t = BX_{t-1}$, $X_{t-3} = B^3 X_t = BX_{t-2}$ 이기 때문에 $(1-B) X_{t-1} = \phi_1 (1-B) + \phi_2 (1-B) X_{t-2} + e_t \cdots\cdots (3.21)$ 이다.

또한 $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$ 이므로

式(3,21)은

$$(1-B) X_t = \phi_1 (1-B) BX_t + \phi_2 (1-B) B^2 X_t = e_t$$

또는 $(1-B)X_t-\phi_1(1-B)BX_t-\phi_2(1-B)B^2X_t = e_t$ ....(3.22)

이 된다.

만일 $(1-B)X_t$ 를 공통인자로 使用하였다면 式(3,22)는

$$X_t(1-B)(1-\phi_1B-\phi_2B^2) = e_t$$

또는 $(1-B)(1-\phi_1B-\phi_2B^2)X_t = e_t$ .... (3.23)

이 된다.

式(3.23)은 1次 階差한 系列$(1-B)$에 관하여 AR(2)模型 (즉 $1-\phi_1B-\phi_2B^2$)와 동일하다. 이러한 表現 形態는 ARMA 模型이 乘法的인 것이기 때문에 可能하다. 같은 方法으로 MA模型이나 ARMA 模型도 후향연산자(B)로 表現할 수 있다.

MA(1) 模型은

$X_t = e_t - \theta_1e_{t-1}$ 인데

$e_{t-1} = Be_t$ 이므로

이것은 $X_t = e_t - \theta_1Be_t$ 이거나

$X_t = (1-\theta_1B)e_t$ .... (3.24)

로 表現할 수 있다. 마찬가지로 MA(2) 模型은

$X_t = (1-\theta_1B-\theta_2B^2)e_t$ .... (3.25)

이고, 1次 階差 系列에 대한 MA(2) 模型은

$(1-B)X_t = (1-\theta_1B-\theta_2B^2)et$ .... (3.26)

이다.

ARMA(1,1)模型은

$(1-\phi_1B)X_t = (1-\theta_1B)e_t$ .... (3.27)

인데 1次 階差 ARMA(1,1)模型은

$$(1-B)(1-\phi_1 B)X_t = (1-\theta_1 B)e_t \quad \cdots\cdots\cdots\cdots \quad (3.28)$$

이며 2次階差 ARMA(1,2) 模型은

$$(1-2B+B^2)(1-\phi_1 B)X_t = (1-\theta_1 B-\theta_2 B^2)e_t \cdots\cdots \quad (3.29)$$

이다.

이 후향연산자의 利用은 우선 보기에 서투른 것 같이 보이지만 몇가지 資料의 관념에 대한 것이 밝혀진다면 이것을 취급하는 것이 쉽다. 母數없는 연산자는 B의 높은 指數와 같은 水準인 階差를 意味한다. 그런데 母數를 알게되면 그것은 ARMA模型으로 나타난다. 그 模型의 次數와 形態는 B의 指數와 母數의 有形과 判別할 수 있다. (만일 母數가 Φ_i 이라면 AR模型이 다시 表現되며 Θ_i 이면 MA模型으로 表現되어 진다) 일단 季節的 系列를 갖는다고 假定하자. 乘法模型을 假定하면 ARMA 模型은 다음의 두가지 部分으로 構成되어 적합시키고자 할 것이다.

1) 規則的 非季節的 部分

2) 몇개의 季節的 母數

判別에 의해서 初期에는 系列의 季節性을 무시할 수 있고, 전과 같이 ARMA模型을 選擇할 수 있다. 季節的 有形의 判別에 의해서 非季節的 部分을 무시하고 (季節的 模型을 선택함으로써) AR 模型이나 MA 模型에 의하여 決定되는지 안되는지를 결정하게 된다. 좀더 상세히 說明하면 非季節的 部分은 階差되지 않은 (式(3.27) 참조) ARMA(1,1) 이라고 假定하자. 만일 時系列이 月別로 季節的 有形을 보인다면 模型은

$(1-\phi_1 B)(1-\phi_{12} B^{12})X_t = (1-\theta_1 B)e_t$: 季節性이 AR에 있다면

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.30)$$

이거나

$(1-\phi_1 B) X_t = (1-\theta_1 B)(1-\theta_{12} B^{12}) e_t$: 季節性이 MA에 있다면

$$\cdots\cdots \quad (3.31)$$

이 된다. 여기서 $(1-\phi_{12} B^{12}) X_t = X_t - \phi_{12} B^{12} X_t$

따라서 $X_t = \phi_{12} X_{t-12}$ 이다 $\cdots\cdots$ (3.32)

유사하게 $(1-\theta_{12} B^{12}) e_t = e_t - \theta_{12} B^{12} e_t$

$$= e_t - \theta_{12} e_{t-12} \quad \cdots\cdots \quad (3.33)$$

이다.

즉 式(3.30)이나 式(3.31)은 季節性을 고려한 12個月前의 X_t 또는 e_t 를 使用한 季節母數를 포함하게 된다.

1次 階差의 水準을 갖는 MA(2) 季節模型(式(3.29) 참조)은 다음 두가지 形態中의 하나를 취할 수 있다.

1) $(1-B)(1-\phi_{12} B^{12}) X_t = (1-\theta_1 B - \theta_2 B^2) e_t$

$(1-B)$ ↓ 계차 1의 수준　$(1-\phi_{12} B^{12})$ ↓ AR의 계절모수　$(1-\theta_1 B - \theta_2 B^2)$ ↓ MA(2) 비계절성

2) $(1-B) X_t = (1-\theta_1 B - \theta_2 B^2)(1-\theta_{12} B^{12}) e_t$

$(1-B)$ ↓ 계차 1의 수준　$(1-\theta_1 B - \theta_2 B^2)$ ↓ MA(2) 비계절성　$(1-\theta_{12} B^{12})$ ↓ MA의 계절모수

5. ARIMA (Autoregressive Integrated Moving Average) 模型

Box-Jenkins (박스-젠킨스)는 ARMA模型과 階差方法을 結合하여 다음의 式과 같은 ARIMA模型 즉 ARIMA (p',d,q)模型을 만들었다.

$$\phi_q(B)(1-B)^d X_t = \theta_q(B) e_t \quad \cdots\cdots \quad (3.34)$$

$$\begin{cases} \phi_q(B) = 1-\phi_1 B-\phi_2 B^2-\cdots\cdots-\phi_q B^p \\ \theta_q(B) = 1-\theta_1 B-\theta_2 B^2-\cdots\cdots-\theta_q B^q \\ \text{B : 후향연산자} \\ \text{d : 階差의 次數} \end{cases}$$

예를들어, ARIMA (1,1,1)模型을 式으로 나타내면

$$\phi_1(B)(1-B)^1 X_t = \theta_1(B)e_t \quad \cdots\cdots \quad (3.35)$$

또는 $X_t = X_{t-1} + \phi_1 X_{t-1} - \phi_1 X_{t-2} + e_t - \theta_1 e_{t-1}$ 이다.

ARIMA模型에 季節類型이 存在할 경우 季節模型을 結合하여 乘法 ARIMA (p,d,q)(P.D,Q)模型으로 表現된다.

즉, 乘法(multiplicative) ARIMA(p,d,q)(P,D,Q)$_s$ 模型의 一般式은

$$\phi_p(B)\Phi_p(B)^s(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)e_t \quad \cdots\cdots(3.36)$$

여기서
$$\begin{cases} \phi_p(B) = 1-\phi_1 B - \cdots\cdots - \phi_p B^p \\ \theta_q(B) = 1-\theta_1 B - \cdots\cdots - \theta_q B^q \\ \Phi_p(B^s) = 1-\Phi_1 B^{1\cdot s} - \cdots\cdots - \Phi_p B^{p\cdot s} \\ \Theta_Q(B^s) = 1-\Theta_1 B^{1\cdot s} - \cdots\cdots - \Theta_Q B^{p\cdot s} \end{cases}$$

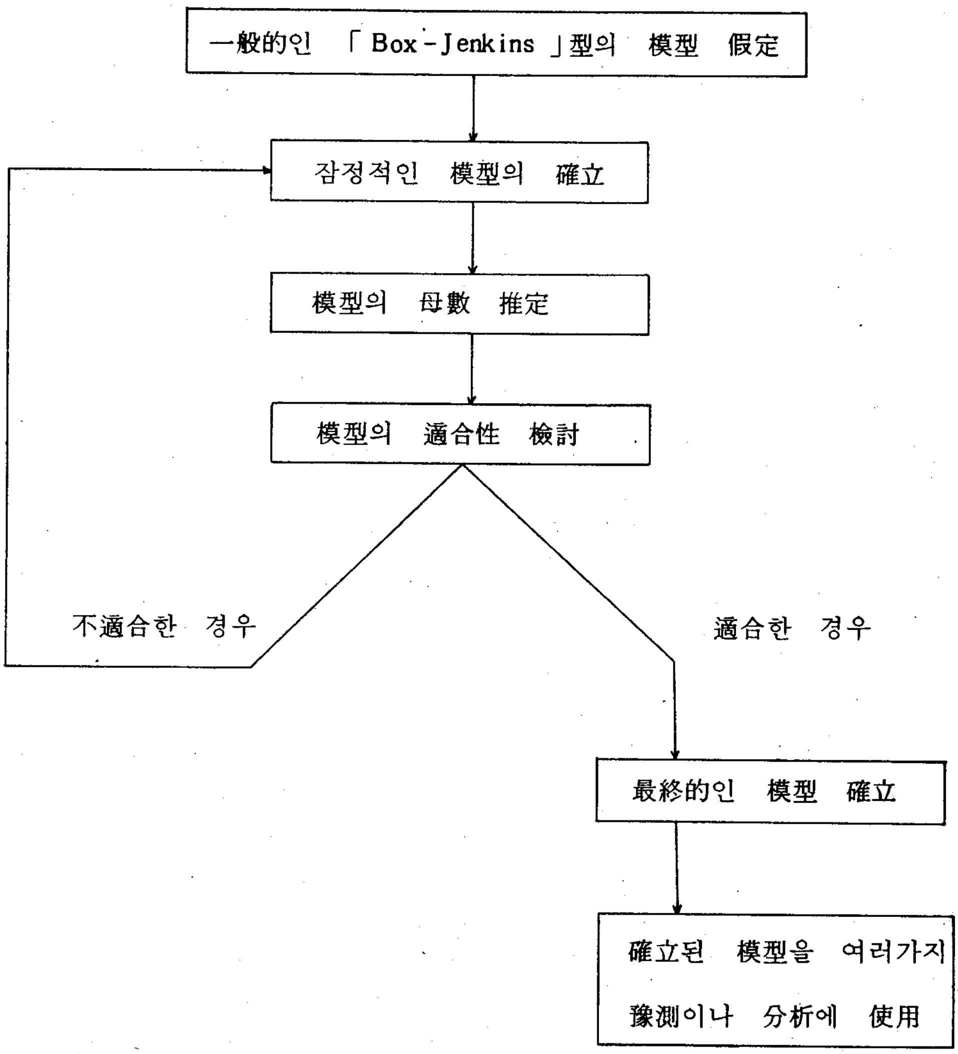
Ⅳ. ARIMA 模型 選定方法

ARIMA 模型은 Box & Jenkins 에 의해 광범위하게 硏究되었으며 이들의 이름이 자주 時系列分析, 豫測 및 管理 (control) 에 應用된 ARMA 過程과 같이 使用되어진다. Yule(1926) 에 의해 처음 AR 模型이 소개되었으며 Walker (1931) 에 의해 一般化된 반면에 MA 模型은 Slutzky(1937) 에 의해 처음 使用되었다. 그런데 이 分野에서 가장 가치있는 것으로 간주되는 것은 ARIMA 模型의 理論的 基礎를 준비한 Wold(1938) 의 硏究이다. Wold 의 硏究의 實績은 ARMA 模型이 두가지 방향에서 발전되었는데 그것은 유효한 判別과 推定節次 (AR, MA, ARMA, ARIMA 模型에 대하여) 그리고 季節的인 時系列을 포함하여 擴張한다.

Box & Jenkins(1970)는 單一變量 時系列 ARMA模型을 利用하고 이해하는데 必要하고 적합한 정보 (내용) 을 잘 理解할 수 있는 方法을 효과적으로 잘 종합하였다. Box - Jenkins 접근 方法의 기본은 〈그림 4.1〉에 요약하였는데 세가지 過程

1. 判別 (Identification)
2. 推定 및 檢定 (Estimation 및 Testing)
3. 豫測 (Forecast)

으로 構成되어 있다.



〈그림 4.1〉

1. 判別 (Identification)

判別 目的은

$$X_t = \phi_2 X_{t-1} + \phi_2 X_{t-2} + \cdots\cdots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

$$\cdots\cdots\cdots\cdots\cdots \theta_q e_{t-q} \cdots\cdots\cdots\cdots \quad (4.1)$$

와 같이 ARMA (p ,q) 模型의 一般的인 形態로부터 特定한 ARMA 模型을 選擇하려는 것이다. 적당한 p 와 q 를 선택하기 위해서는 系列에서 계산된 自己相關과 偏自己相關係數를 살펴보아야 한다.

1.1) 偏自己相關係數

式(4.1)에서 $q = 0$ 라 假定하고 계속해서 p = 0 ,1 ,2 ,3 …… p 라 하자. 그러면 式(4.1)은

$$X_t = e_t \cdots\cdots\cdots\cdots \quad (4.2)$$

$$X_t = \phi_1 X_{t-1} + e_t \cdots\cdots\cdots\cdots \quad (4.3)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t \cdots\cdots\cdots\cdots \quad (4.4)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + e_t \cdots\cdots\cdots\cdots \quad (4.5)$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots\cdots + \phi_p X_{t-p} + e_t \cdots\cdots\cdots \quad (4.6)$$

이 된다.

만일에 式(4.3)의 實際次數가 p = 0 이라면 ϕ_1 은 " 0 "과 有意的差가 없게 된다. 그래서 ϕ_1 을 계산할 수 있다면 이 模型은 AR(0)임을 判別할 수 있다. 그러나 實際次數가 p = 1 이라면 式(4.3)에서 계산된 ϕ_1 의 값은 " 0 "과 有意的 差가 있게 된다. 만일 式(4.4)이 使用되면 實際次數가 p = 1 일 때 ϕ_2 는 統計的으로 有意的이 아니다. 비슷한 理由로 實際次數가 p = 2 이라면 ϕ_1 은 " 0 "과 有意的 差가 있으며 또한 ϕ_2 도 有意的이지만 ϕ_3 는 그렇지가 않다. 왜냐하면 實際 AR 模型은 단지 2 次模型이기 때문이다.

一般的으로 AR(p) 模型의 p번째 母數는 AR模型이 적어도 p次이거나 또는 p次보다 높을때 " 0 "과 有意的 差가 있다고만 하게 된다.

대개 AR模型의 實際次數는 未知이지만 式(4.3)을 利用하여 ϕ_1을 계산할 수가 있고 式(4.4)을 利用하여 ϕ_2를 式(4.5)에서 利用하여 ϕ_3 …… 등등을 계산할 수 있다. 만일 ϕ_1이 "0"과 有意的 差가 없다면 그 模型은 AR(0) 이다. 또 ϕ_1이 有意的이고 ϕ_2가 有意的이 아니라면 그 模型은 AR(1)이며, 마찬가지로 ϕ_p가 有意的이고 ϕ_{p+1}이 有意的이 아니라면 그 模型은 AR(p)이다.

式(4.3), (4.4), (4.5), (4.6)에서 ϕ_1, ϕ_2, ϕ_3 …… ϕ_p의 값을 計算하기 위해서는 많은 시간을 소비하여야 한다. 따라서 $k=1, 2, 3$ …… p에 대해

式 $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_2 \rho_{k-3} + \cdots\cdots + \phi_p \rho_{k-p}$

여기서 $\rho_k = \dfrac{r_k}{r_0}$ …………………………………………………………… (4.7)

을 利用하여 推定值를 구할 수 있다.

$k=1$일때 式(4.7)은 $\rho_0 = 1$(ρ_1은 時差 1의 자기상관계수)이기 때문에

$$\rho_1 = \phi_1 \rho_0, \quad \phi_1 = \rho_1 \qquad (4.8)$$

이 된다.

$k=2$일때 式(4.7)은

$$\rho_1 = \phi_1 + \phi_2 \rho_1, \quad \rho_2 = \phi_1 \rho_1 + \phi_2 \qquad (4.9)$$

이 되고

ϕ_2에 대해 式(4.9)을 풀면

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} \quad \cdots\cdots (4.10)$$

이 된다.

式(4.10)에서 $r_1 = \rho_1$ 으로 대체(즉 표본추정치가 ρ_1과 ρ_2에 대해 使用됨)하면

$$\phi_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$ 과 같다.

만일 $k=3$이라면 式(4.7)은

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_3 \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 \quad \cdots\cdots (4.11) \\ \rho_3 &= \phi_1\rho_2 + \phi_2\rho_1 + \phi_3 \end{aligned}$$

이 된다.

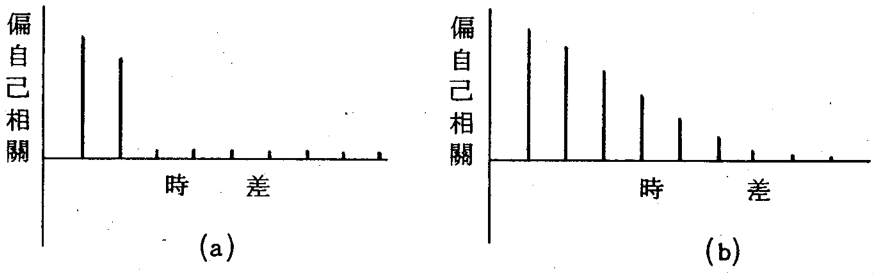
$r_1 = \rho_1$, $r_2 = \rho_2$, $r_3 = \rho_3$로 대입하고 式(4.11)을 풀면 AR模型의 次數를 決定하는데 推定値로써 使用할 수 있는 ϕ_3에 대한 값을 얻을 수가 있다.

마찬가지로 $k=p$라 놓으면 일단의 時系列 資料의 p개 偏自己相關을 利用하여 ϕ_p를 計算할 수 있다. AR模型의 次數를 判別한다는 것은 이들 偏自己相關을 검토함으로써 할 수 있다. 次數는 단순히 "0"과 有意的 差가 있는 偏自己相關의 數와 같아진다. p時差를 통해서 얻은 偏自己相關은 有意的인 반면에 나머지 項은 "0"에 가까워진다. 이 切斷値 p는 AR模型의 次數가 된다.

移動平均 (Moving Average: MA) 模型은 式(4.3) ,(4.4),(4.6)과 같이 다를 수가 없다. 그러한 경우 ϕ_1, ϕ_2, ϕ_3 …… ϕ_p 를 推定하고자 한다는 것은 系列에 관한 적합하지 못한 模型에 適合시키려고 하는 것이다. MA 模型에서 相異한 X_t 값은 서로 從屬的이므로 MA 系列에 AR 模型을 適合시키기 위해서는 무한 數의 ϕ_1 項이 必要하다.

MA 系列에 대해서 偏自己相關母數의 값은 初期에는 크며, 그 크기는 時差가 증가함에 따라 감소한다. 그러므로 MA 模型의 偏自己相關은 AR模型에서와 같이 P時差 以後에 切斷值를 갖지 않지만 계속적으로 천천히 " 0 " 에 근사한다. 따라서 偏自己相關은 自己相關길이의 時差로서 큰값에서 작은 값으로 指數形 감소가 계속된다. 偏自己相關에 대한 몇가지 典型的인 形은 〈그림 4 .2 〉와 같다.

〈그림 4 .2 〉



〈그림 4.2 〉의 a) 는 2개의 偏自己相關이 " 0 "과 有意的差가 있다. 나머지는 有意的差가 없다. 이것은 그 時差以後에 " 0 "에 근사하기 때문에 AR 模型임을 제시하고 b) 는 偏自己相關이 指數形으로 " 0 "에 감소하기 때문에 MA 模型임을 제시한다.

1.2) 自己相關係數

AR模型과 MA模型의 自己相關函數와 偏自己相關函數는 매우 다르게 행동한다.

AR模型의 自己相關係數는 〈그림 4.2〉 b)의 形態(指數型으로 減少하는)로써 나타나며 MA模型의 偏自己相關係數는 〈그림 4.2〉 a)의 有形(q時差以後 "0"에 근접하는)을 나타내고 있다. (이것은 바로 AR 模型과 MA模型에 대한 偏自己相關係數의 形態의 逆을 나타낸 것임에 有意해야 한다).

自己相關係數는, 時差가 서로 다른, 같은 變數間의 관련 程度를 나타내는 것이다. 그러한 聯關性이 存在하는 AR模型에서 이의 크기는 模型의 次數를 決定한다.

예를들면 AR(1)模型에서

$X_t = \phi_1 X_{t-1} + e_t$ ·· (4.12)

$X_t = r_1 X_{t-1}$ ·· (4.13)

式(4.13)은 $r_1 = \hat{\phi}_1$을 式(4.12)에 대입한 結果이다. 또 式(4.13)은 X_t와 X_{t-1}(時差 1)가 r_1이나 또는 이의 推定値 r_1을 통해서 關聯되어 있는 狀態이다. 또한 이것은 다음과 같은 경우와 같다.

$X_{t-1} = \rho_1 X_{t-2}$ ·· (4.14)

式(4.14)을 式(4.13)에 代入하면

$X_t = \rho_1(\rho_1 X_{t-2}) = \rho_1^2 X_{t-2}$ ·· (4.15)

이다.

따라서 X_t와 X_{t-2}는 ρ_1^2을 통해 연관되어 있다. 마찬가지로

$X_t = \rho_1^s X_{t-s}$ ·· (4.16)

이다.

AR(1)과의 自己相關은 X_t 와 X_{t-s} 間의 時差가 증가함에 따라 指數的으로 점차 소멸되어질 것이다. 비록 自己相關이 점차 소멸되어 진다고 하더라도 AR(P)模型 (P > 1)에 대해서 自己相關係數의 形態는 비슷하다. 그것은 AR(2) 模型을 얻기 위해서

$\phi_1 = \dfrac{r_1(1-r_2)}{1-r_1^2}$, $\phi_2 = \dfrac{r_2 - r_1^2}{1-r_1^2}$ 을 代入함으로써 求할 수가 있다.

MR 模型에 대해서 連續的인 X_t 값 間의 聯關은 MR 模型의 次數에 제한된다.

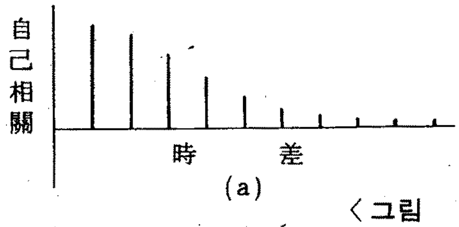
예를들면 MA(1) 模型에서

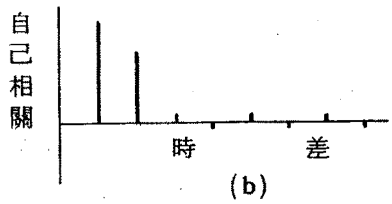
$$X_t = e_t - \theta_1 e_{t-1} \cdots\cdots (4.17)$$

式(4.17)은 X_t 가 나머지 항들에 의하지 않고 단지 e_{t-1} 에 의해서만 X_{t-1} 에 聯關되어 있다. 따라서 1보다 더 긴 時差의 自己相關係數는 存在하지 않는다. ("0"과 有意的 差가 없다). 그러므로 時差 1以後의 自己相關係數에서 반드시 절단되어야 한다.

만일 過程이 MA(2)模型이라면 X_t 는 e_{t-1} 과 e_{t-2} 에 의해서만 聯關이 있다. 그러므로 처음 두개의 自己相關이 "0"과 有意的 差가 있고 나머지는 "0"으로 떨어질 것이다. 一般的으로 過程이 MA(q) 模型이라면 첫번째 q 自己 相關이 "0"과 有意的差가 있을 것이지만 그것은 그렇지 않다.

〈그림 4.3〉은 自己相關係數의 전형적인 形態의 例示이다.





〈그림 4.3〉

〈그림 4.3〉에서 a)는 自己相關이 指數型으로 "0"에 감소하므로 AR 模型임을 제시한다.

b)는 단지 처음 2개의 自己相關만이 "0"과 有意的 差가 있기 때문에 MA(2)模型임을 제시한다.

1.3. 自己相關係數와 偏自己相關係數 모두를 使用하는 경우

〈그림 4.2〉와 〈그림 4.3〉에서의 정보는 서로 補完的이다. 예를들면 〈그림 4.2(b)〉는 그 過程이 MA模型임을 암시하고 있으며 〈그림 4.3(b)〉는 이의 次數가 2임을 判別해 주고 있다. 또한 〈그림 4.3(a)〉는 AR模型을 암시하며 〈그림 4.2(a)〉는 이의 次數가 2임을 判別해 준다.

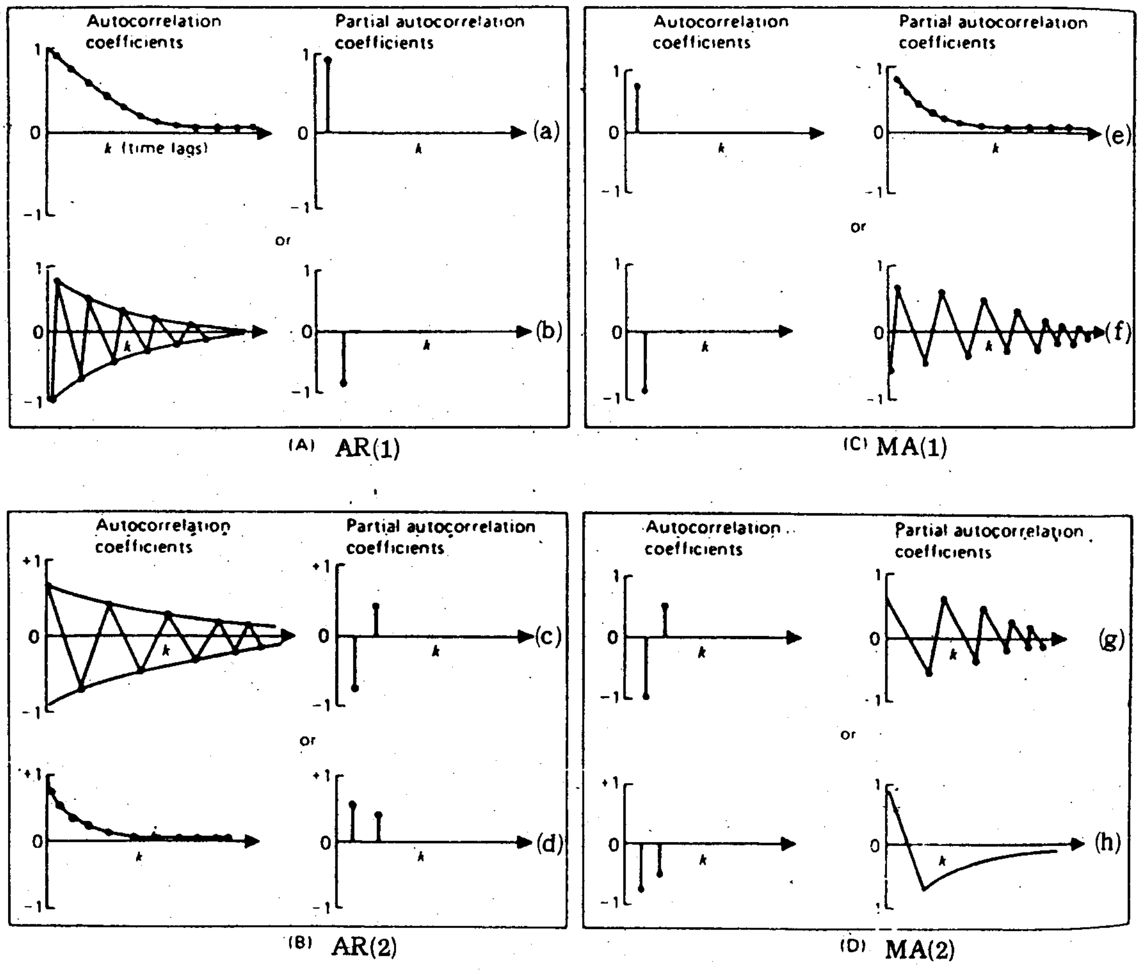
그러므로 純粹한 AR模型이나 MA模型에 대해서 自己相關이거나 偏自己相關이 "0"으로 점점 근사할때 自己相關인 경우는 AR模型〈그림 4.3(a)〉을 가리키며, 偏自己相關인 경우는 MA模型〈그림 4.2(b)〉을 가리킨다. 따라서 값이 "0"으로 절단(떨어)질때 時差를 살펴봄으로써 模型의 次數를 判別하게 된다.

〈그림 4.2〉와 〈그림 4.3〉은 自己相關과 偏自己相關係數의 形態의 가능한 경우만 나타낸 것이다.

그들이 類似한 類型을 가질지라도 많은 變動이 存在한다. 예를들면, 만일 ϕ_1이 陰數이면 式(4, 14)의 ρ_1은 "−"가 되지만 式(4, 15)의 ρ_2는 "+"가 될 것이다. 그러므로 自己相關의 부호는 바뀌고 "0"으로 사라지면서 有意的인 偏自己相關은 "−"가 될 것이다. 또한 模型이 AR(2)이고, ϕ_1이 "+"이고 ϕ_2가 "−"이라면 自己相關과 偏自己相關이 有意的으로 相異한 類型에서 움직일 것이다.

그러나 단지 두개의 偏自己相關은 " 0 "과 有意的 差가 있을 것이지만 몇가지 變化되는 모양에서 自己相關은 " 0 "으로 사라질 것이다.

〈그림 4 . 4〉는 AR(1), AR(2), MA(1), MA(2)模型의 自己相關係數와 偏自己相關係數의 기대되는 形態를 보이고 있다. 높은 次數의 模型은 비슷한 類型으로 움직일 것이지만 自己相關函數나 偏自己相關函數中에 두개 이상의 有意的인 係數를 갖게 될 것이다.

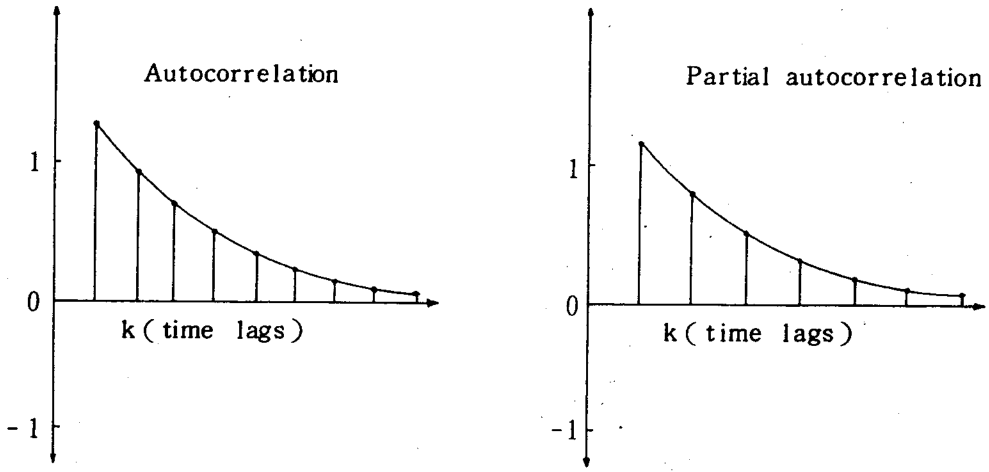


〈그림 4 . 4〉

1.4 ARMA 模型

ARMA 模型은 自己相關과 偏自己相關係數의 形態에서와 ARMA 模型의 公式에 의해서 모두 AR과 MA概念의 特性이 結合되어 있다. ARMA 模型에 대해 대응된 自己相關과 偏自己相關係數를 얻기 위하여 반드시 AR 및 MA 自己相關과 偏自己相關係數를 완전히 습하여야 한다.

따라서 ARMA (1.1) 模型에 대해서 〈그림 4.4 (a)〉와 〈그림 4.4 (e)〉는 〈그림 4.5〉에서 보여주는 類型을 얻기 위해 結合할 수 있다.

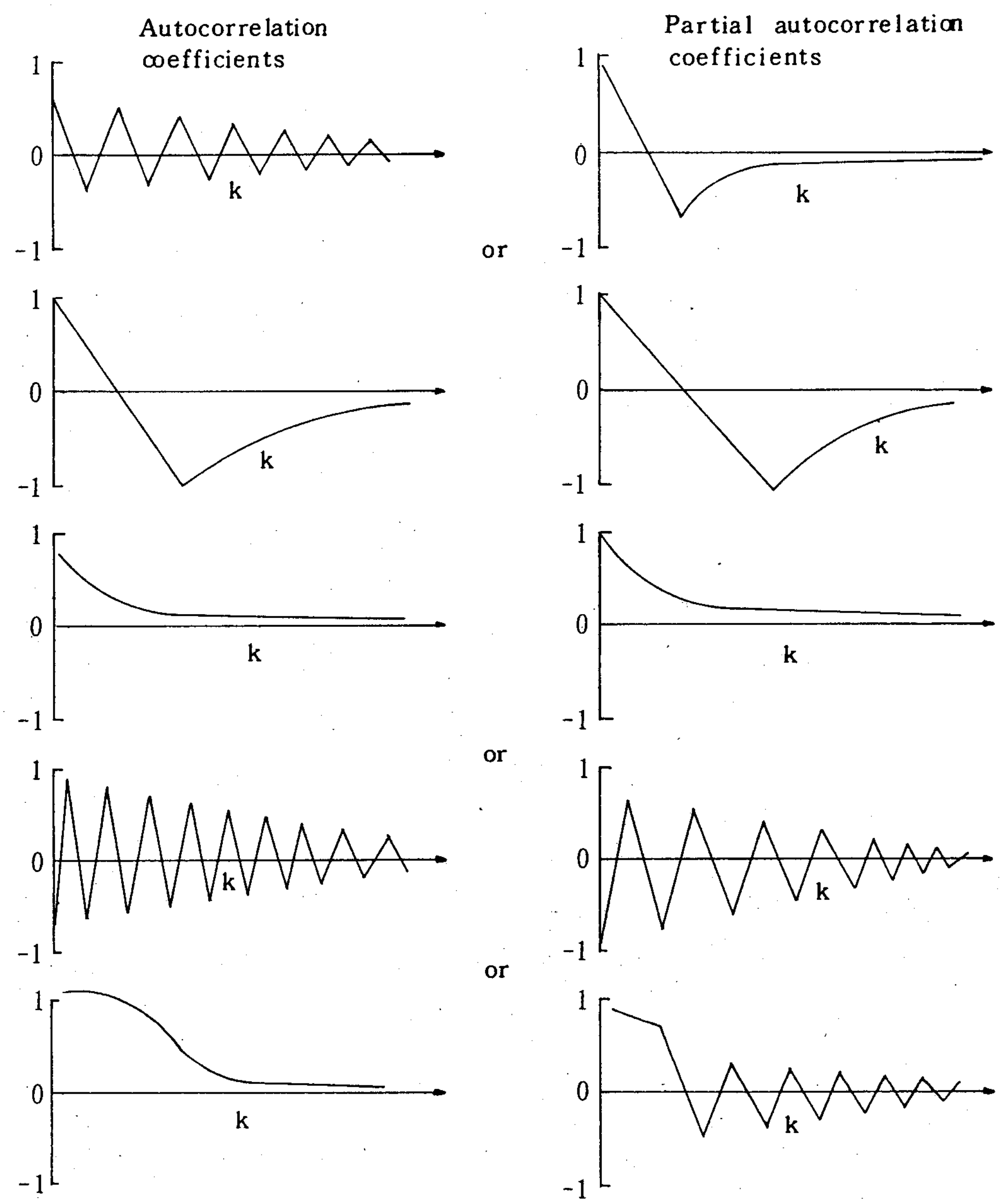


〈그림 4.5〉 ARMA (1.1) 模型의 自己相關과 偏自己相關

〈그림 4.5〉는 〈그림 4.4 (e)〉의 自己相關의 모양과 〈그림 4.4 (a)〉의 偏自己相關의 모양을 갖는다. " 0 "과 有意的 差가 있는 單一自己相關과 偏自己相關은 〈그림 4.5〉의 一部分에 포함된다. 따라서 實質過程이 結合 ARMA模型일 때 自己相關 및 偏自己相關 모두가 " 0 "으로 사라지는 時間에만 存在할 것이다.

〈그림 4.6〉은 ARMA(1.1)模型에 대해서 얻어진 또다른 가능한 有形

을 가리킨다. 이것은 單純히 〈그림 4.4〉의 AR(1)과 MA(1)模型의 가능한 結合이다.



〈그림 4.6〉 **結合 ARMA(1.1) 模型인 自己相關과 偏自己相關**

1.5 暫定 ARMA(p.q) 模型 判別

Box-Jenkins 方法論에서 주요한 過程은 時系列의 自己相關係數와 偏自己相關係數를 검토하여 적합한 ARMA (p.q) 模型을 判別하려는 것이다. 비록 이런 過程이 직접적이라 할지라도 기계적인 수준에서는 제거될 수 없다. 이것은 人間의 判別이 必要하고 공교롭게도 그것은 다음의 두가지 理由에 의해서 單一模型의 정확한 判別이 항상 나타나지는 않는다.

① 自己相關이나 偏自己相關은 特定한 模型을 분명히 암시하지 않는다는 점.

② 自己相關이나 偏自己相關은 하나의 模型보다는 더 많은 模型을 가리킨다는 점.

실제에 있어서 系列의 任意性때문에 自己相關과 偏自己相關은 항상 확실한 類型에 따라서 움직이지 않는다. 앞에서 지적한 自己相關과 偏自己相關의 形態는 理論的이며 noise (잡음)가 存在할 때 自己相關이나 偏自己相關은 그들의 理論值보다 더 높거나 낮을 수도 있다.

任意性에 의한 變動量은 처음부터 알려지지 않기 때문에 自己相關이나 偏自己相關의 形은 "暫定的"인 ARMA (p.q)模型을 推論하는데 使用되어 진다. 自己相關과 偏自己相關은 平均이 "0" 標準偏差 "$\frac{1}{\sqrt{n}}$"인 正規分布를 한다고 알려져 있다. 이런 假定下에서 信賴區間을 設定할 수 있고 주어진 自己相關 또는 偏自己相關이 "0"과 有意的 差가 있을 것이라는 기회를 決定하기 위해 使用된다. 95% 信賴區間은 有意的이 되기 위해서는 약 $\frac{2}{\sqrt{n}}$보다 더 큰 自己相關 또는 偏自己相關을 요구한다.

이것은 自己相關과 偏自己相關의 實質的인 形態를 定하기 위해 개략적인 規則으로써 利用될 수 있다.

그런데 模型의 判別을 하기전에 먼저 3가지 단계의 節次에 관하여 개략적으로 살펴보는 것이 有用하다.

(1) 系列의 定常化

만일 時系列이 定常的이 아니라면 허위의 自己相關이 算出될 수 있기 때문에 모형판별에 방해되는 結果가 될 것이다. 따라서 系列이 定常的이 아니면 階差(differencing)의 적당한 水準을 취함으로써 定常系列化한다.

(2) 自己相關 및 偏自己相關의 檢討(PLOT 形態에서 선택)

指數型으로 "0"에 떨어지는 相關關係를 반드시 判別하여야 한다. 만일 사라지는 것이 自己相關係數間에 일어난다면 그것은 AR模型을 암시하며 偏自己相關係數間에 일어난다면 그것을 MA模型을 나타낸다. 또한 指數型으로 양쪽 모두 떨어진다면 그것은 結合ARMA 模型임을 암시한다.

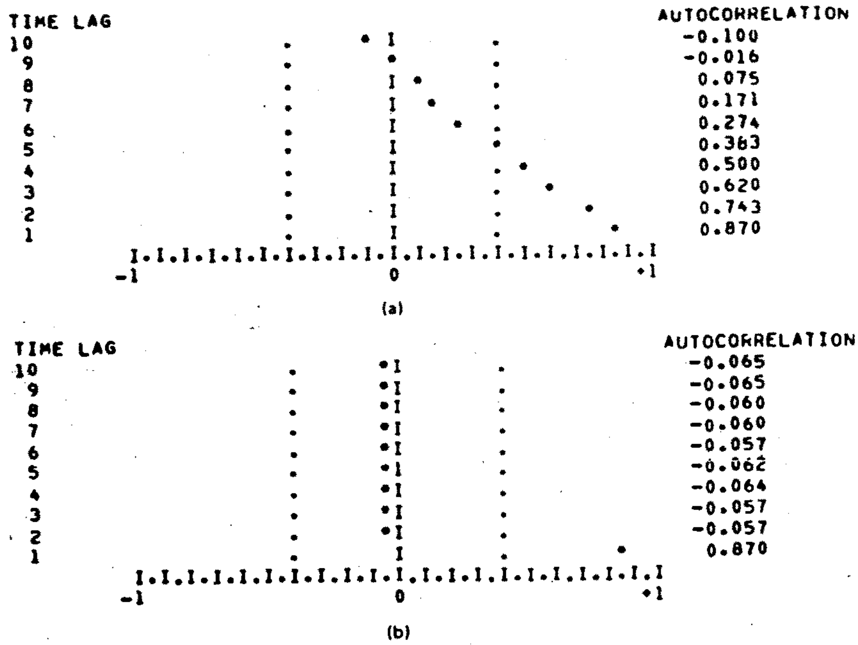
(3) AR 또는 MA模型의 次數를 決定하기 위한 나머지 相關關係의 檢討("0"으로 떨어지지 않는 것)

이 決定은 "0"과 有意的 差가 있는 自己相關이나 偏自己相關의 數를 세어 봄으로써 한다. 結合ARMA 模型에 대해서 AR次數는 偏自己相關에서 決定되어지며 MA次數는 自己相關에서 決定한다.

이러한 3단계를 實行함으로써 定常系列에 대한 ARMA(p.q)模型을 判別할 수 있다. 이 過程을 〈表 4.1〉의 系列을 利用하여 說明하여 보자. 이 系列을 가장 잘 說明하는 ARMA模型을 決定하려는 첫번째 단계에서 自己相關 및 偏自己相關의 PLOT 〈그림 4.7〉를 檢討한다.

〈表 4.1〉

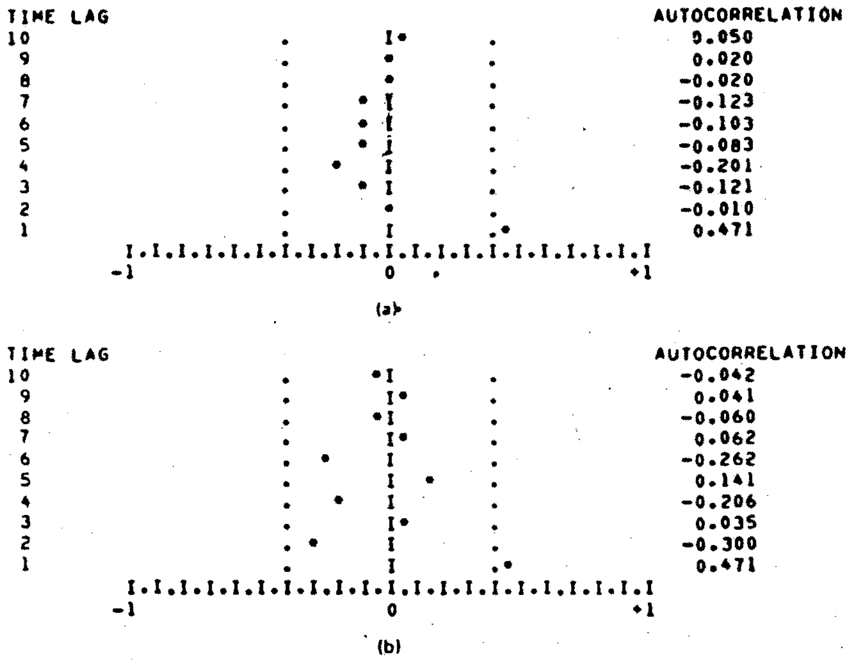
기 간	관 측 치	기 간	관 측 치	기 간	관 측 치
1	22,665	9	107,930	17	184,705
2	32,861	10	116,722	18	194,733
3	43,623	11	124,937	19	204,726
4	54,606	12	133,885	20	215,358
5	65,140	13	144,446	21	225,484
6	76,421	14	154,768	22	234,735
7	87,345	15	165,617	23	246,186
8	98,158	16	175,654	24	258,504



〈그림 4.7〉 原系列의 自己相關(a)와 偏自己相關(b)

自己相關이 “ 0 ”으로 빨리 떨어지지 않는다는 사실은 (2차 또는 3차 時差以後) 原系列이 非定常的임을 암시한다.

다음 단계는 1次 階差를 취하는 것이며 따라서 〈그림 4.8〉에서 보여지는 것처럼 自己相關과 偏自己相關을 다시 計算한다.



〈그림 4.8〉 1次 階差된 系列의 自己相關 (a) 偏自己相關 (b)

지금 自己相關은 定常性 (相關係數들이 “ 0 ”으로 아주 빨리 떨어지고 있음) 임을 암시하고 있으며 自己相關과 偏自己相關의 消滅現狀을 볼 수 있다.

〈 그림 4.8(b)〉 에서 볼 수 있는 것처럼 偏自己相關에 대해서는 이러한 消滅現狀이 일어나며 부호가 바뀐 狀態에서 “ 0 ”으로 떨어지고 (指數型같

이) 있으며 〈그림 4.4 (f)〉의 有形과 많이 類似하다. 이 類型은 MA 模型임을 意味한다.

結果的으로 다음 단계는 다만 自己相關이 " 0 "과 有意的 差가 있다는 것이 분명해지는 곳의 自己相關係數를 살펴보는 것이다. 「" • "선의 바깥쪽 : 즉 $\pm 2/\sqrt{n}$」 結論은 MA 模型의 次數는 1이고 過程이 MA(1) 또는 동일한 ARMA (0, 1)이라는 것이다.

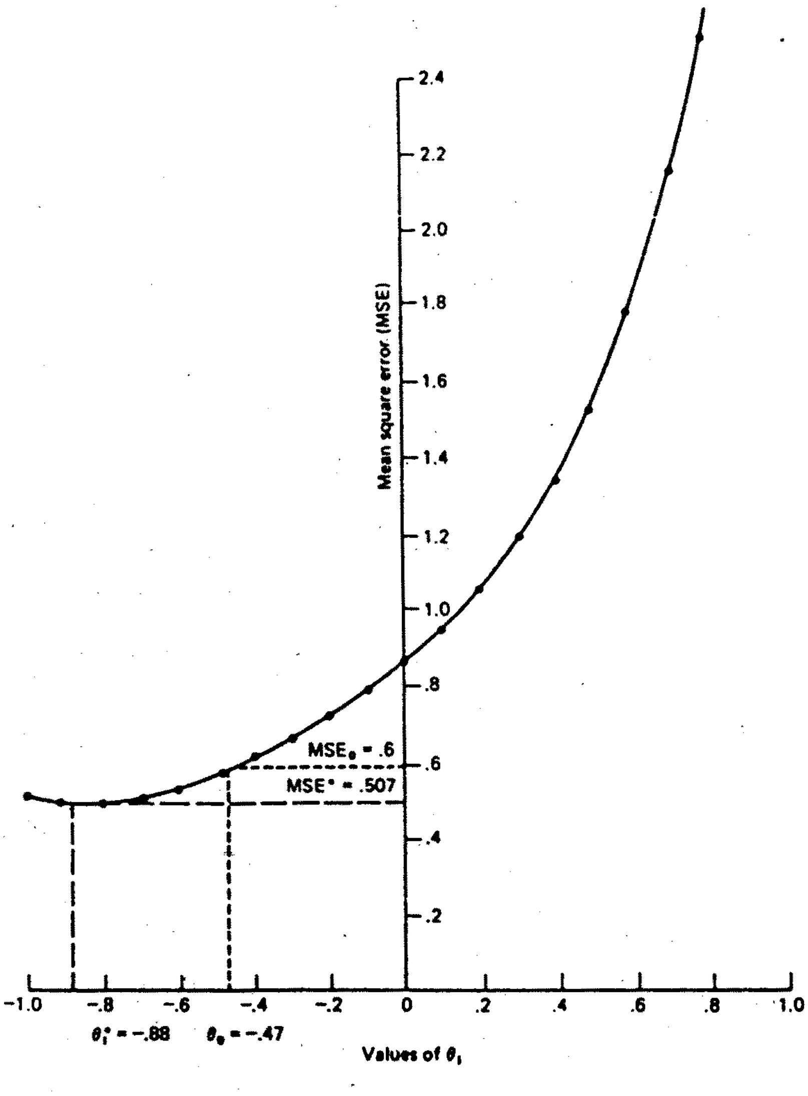
2. 母數의 推定

〈表 4.1〉의 系列에 1次 階差에 관한 ARMA (0, 1)을 적합시키기 위한 決定이 일단 정해지면 반드시 MA(1)模型의 母數의 값이 推定되어야 한다. 그것은 式(4.18)에서 θ_1의 값이 MSE (Mean Square Error : 自乘誤差平均)가 最少化되는 것을 求하게 된다.

$$(1-B)X_t = e_t - \theta_1 e_{t-1} \quad \cdots\cdots \quad (4.18)$$

따라서 $-1 < \theta_1 < 1$임이 알려져 있고 이 범위안에서 θ_1의 相異한 값을 알려고 하는 것인데 MSE를 計算하여 MSE를 最少로 하는 θ_1^* 의 값을 선택하게 된다. 〈그림 4.10〉은 θ_1의 相異한 값에 대한 MSE를 보여주고 있다.

θ_1^* 의 最適値는 -0.88로써 이것은 MSE가 0.507일때 주어진 값이다. 여러번의 施行과 誤謬에 의해서 最量 θ_1을 求함에 있어서 時間消費와 不必要한 計算上의 작업을 피하기 위해 自己相關을 포함하는 函數에 근거를 둔 初期推定值에서 출발할 수가 있다.



〈그림 **4.10**〉

이러한 特定 예제에 대하여 初期推定値는 0.6의 MSE에 대응하는 θ_0 $= -0.47$이다. 이 출발점에서 " -1 "와 " $+1$ "間의 전체범위를 살펴보는 것보다는 -0.88의 最適値를 움직이는 것이 더 쉽다.

最量 θ_1 을 求하기 위해 單純한 施行과 誤謬過程을 利用하는 대신에 Gauss-Newton 制約最適化 접근방법에 근거를 둔 方法을 應用하는 것보다 유효적이다. ARMA 模型을 應用함에 있어서의 주요한 어려움은 반복적인 施行과 誤謬過程이 포함되는 관계로 模型의 次數가 1보다 더 증가할 때는 注目해야 할 計算事項이 요구되어 진다.

〈表 4.1〉에서의 系列에 대하여 마지막 模型은

$$(1-B)X_t = e_t - 0.88\,e_{t-1} \qquad \cdots\cdots (4.19)$$

이 되며 다음 단계는 式(4.19)의 模型이 적합한 模型인지 아닌지를 檢討하는 것이다.

3. 模型의 妥當性 檢證

일단 式(4.19)이 推定되었다면 이 系列에서 說明하는 근사성은 式(4.18)의 殘差를 살펴봄으로써 決定하게 된다.

즉, $e_t = (1-B)X_t + 0.88\,e_{t-1}$

이 殘差들에 대한 自己相關은 어떠한 有形으로 正義되지 않는다.(즉 모든 自己相關이 點線 범위내에 있다(그림 4.11), 더욱이 自己相關의 χ^2 값은 2.06인 반면에 95% 信賴水準에서 自由度 22에 대한 χ^2 分布表의 대응값이 12.3이다.

$\chi_t^2 < \chi_0^2$ 이므로 式(4.19)에서 보여진 MA(1) 模型은 〈表 4.1〉의 時系列의 未來값을 豫測하는 적절한 模型이라고 結論지울 수 있다.

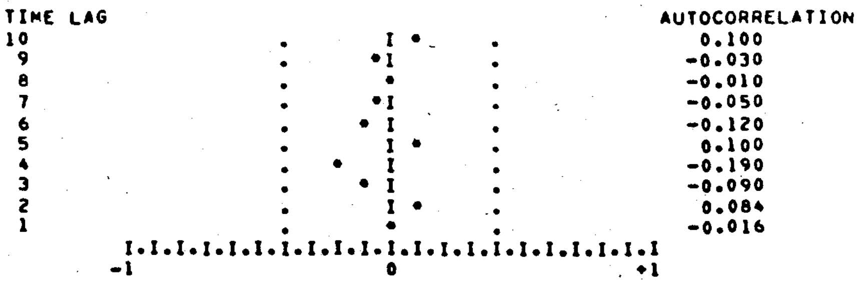
判別過程으로 되돌아가서 보다 좋은 模型을 判別할려고 할 必要가 있다. 1次 階差를 취한 後 判別過程에서 잘못 해석된 相異한 有形이 만들어

졌다고 假定하다.

MA (1)을 선택하는 대신에 MA (2)가 선택되었다. 이것은 原系列에 관한 ARMA (1,1)을 適合시킴으로써 잘못된 것이 아니다. 두번째 MA母數의 推定値는 “ 0 ”에 아주 근사한 결과가 되는데 그것은 MA(2) 模型이 式 (4.19)의 AM (1) 模型과 거의 일치한다는 것을 意味한다. 여기서 단지 문제가 되는 것은 필요한 것보다 더 많은 母數를 利用하였다. 만일 두번째 MA 母數가 適合되기 위해 아무것도 첨가할 수 없다면 이것은 보다 적은 自由度를 갖게 되며 많은 計算期間이 필요하지 않게 된다. 마찬가지로 1次 階差한 系列에 대한 ARMA (1,1) 模型 (ARIMA (1,1,1))을 利用하기로 하였다.

AR 母數의 값이 “ 0 ”에 매우 근사하기 때문에 이 模型은 MA (1)模型보다는 더 좋지 못하다. 한개 이상의 模型이 任意的인 分布를 하고 그 分布의 MSE가 근사적으로 같은 殘差가 주어졌을때 모수들이 적은 模型을 利用 (Parsimony의 原理)하는 것이 바람직하다.

推定 및 檢證은 Box－Jenkins 方法의 2번째 단계로써 완결된 것이다. 마지막 過程은 時系列의 未來값을 豫測하기 위하여 推定하고 檢證된 模型을 利用한다.



〈그림 4.11〉

4. 豫 測

ARMA模型의 豫測의 有用性을 例示하고자 〈表 4.1〉의 系列에 대하여 判別된 MA (1) 模型을 利用하고자 한다.

式(4.19)에서 求한 模型은

$(1-B)X_t = e_t - 0.88\, e_{t-1}$ 이다.

24個의 資料點들이 存在하기 때문에 期間 24의 結果가 利用 可能하다고 假定하면 期間 25, 26, 27, 28에 대한 豫測을 준비하는 過程을 살펴보자.

期間 25에 대한 豫測은

$$(1-B)X_{25} = e_{25} - 0.88e_{24} \quad \cdots\cdots \quad (4.21)$$

을 이용한다.

$$(1-B)X_{25} = X_{25} - X_{24} \quad \cdots\cdots \quad (4.22)$$

이기 때문에 式(4.22)을 (4.21)에 代入하면

$$X_{25} - X_{25} = e_{25} - 0.88\, e_{24}$$

또는 $X_{25} = X_{24} + e_{25} - 0.88\, e_{24}$ ······ (4.23)

이다.

그런데 X_t가 平均 $\bar{X}$ 주위의 偏差形으로 表現되어 있다. 즉 $X_t = X_t^c - \bar{X}$ 이다.

그러므로 $X_{25} = X_{25}^c - \bar{X}$ 이다. ······ (4.24)

式(4.24)을 式(4.23)에 代入하면

$$\begin{aligned} X_{25}^c &= \bar{X} + X_{24} + e_{25} - 0.88\, e_{24} \\ &= 10.254 + 258.504 - 0.88(-0.776) + e_{25} \\ &= 269.44 + e_{25} \end{aligned}$$

로 주어진다.

여기서 $\bar{X}$ 의 값은 〈表 4.1 〉의 系列을 平均한 것이고, X_{24}는 系列의 가장 최근의 實際값이고, e_{24} 는 지난 過去의 모든 값에 관한 模型을 適用하여 알게 된다.

期間 26 에 대하여

$$X^c_{26} = \bar{X} + X_{25} + e_{26} - 0.88\ e_{25} \quad \cdots\cdots \quad (4.26)$$

그런데 X_{25} 에 대한 實際값이나 e_{25} 도 期間 24 에서 알 수는 없다. 이것은 MA (1) 模型의 豫測値가 한 期間前보다 더 以前의 期間에 제한되었음을 假定한다. 그것은 $e_{25}=0$ 이고 式 (4.26)에 대한 實際値로서 式 (4.25)에서 X_{25} 의 推定値를 使用한다고 假定하는 것이 可能하다.

이것을 Bootstrapping 이라 부르며 結果的으로 요구되는 未來의 期間에 대한 豫測値를 얻기 위해 이것이 허용되는 것이다. 예를 들면 이 경우에

$$X^c_{26} = 10{,}254 + 269{,}44 + e_{26} = 279.674 + e_{26}$$

$$X^c_{27} = 10{,}254 + 279{,}674 + e_{27} = 289.948 + e_{27}$$

$$X^c_{38} = 10{,}254 + 289{,}948 + e_{27} = 300.202 + e_{27}$$

같은 方法으로 다른 模型과 相異한 階差의 程度를 使用하여 豫測할 수 있다. 그러나 다음의 경우를 주의해야 한다.

1) 만일 系列이 平均의 偏差形으로 表示되어 있다면 平均을 더하라.

2) 系列에 利用된 階差의 水準을 고려하라.

後者의 생각은 AR 模型이 應用될 때 1 次階差 以上이 利用될 때 약간 어렵게 된다.

標準的인 Computer Program은 自動的으로 (1)과 (2) 모두를 行할 수 있고 요구한 豫測値의 數를 利用者에게 줄 수 있다. 그런데 豫測에 포함된

것을 例示하기 위하여 2次 階差된 系列에 대해 判別된 AR (2)模型의 例를 說明하기로 한다.

이 模型은

$$(1-B)^2 X_t = 0.5(1-B)^2 X_{t-1} - 0.3(1-B)^2 X_{t-2} + e_t \quad \cdots\cdots (4.27)$$

인 것으로 求한다.

여기서 $(1-B)^2 X_t$ 는 X_t 의 2次 階差로 正義하도록 使用된 것인데 다음과 같이 單純化한다.

$$\begin{aligned}(1-B)^2 X_t &= (1-B)X_t - (1-B)X_{t-1} \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - X_{t-1} - X_{t-1} + X_{t-2} \\ &= X_t - 2X_{t-1} + X_{t-2} \quad \cdots\cdots (4.28)\end{aligned}$$

같은 方法으로

$$(1-B)^2 X_{t-1} = X_{t-1} - 2X_{t-2} + X_{t-3} \quad \cdots\cdots (4.29)$$

$$(1-B)^2 X_{t-2} = X_{t-2} - 2X_{t-3} + X_{t-4} \quad \cdots\cdots (4.30)$$

式(4.27)의 左邊에 式(4.28)을 代入하고

式(4.27)의 우변에 式(4.29)와 (4.30)을 代入하면

$$X_t - 2X_{t-1} + X_{t-2} = 0.5(X_{t-1} - 2X_{t-2} + X_{t-3}) - 0.3(X_{t-2} - 2X_{t-3} + X_{t-4}) + e_t$$

로 주어진다.

項들을 合하면

$$X_t = 2.5X_{t-1} - 2.3X_{t-2} + 1.1X_{t-3} - 0.3X_{t-4} + e_t$$

이다.

이 例에서 平均은 0.727 이고 마지막 네번째 資料값은

$X_{24} = 140$, $X_{23} = 114$, $X_{22} = 133$, $X_{21} = 158$ 이다.

위의 式을 利用하고 平均을 더하면 期間 25에 대한 豫測値로써

$$X^c_{25} = 0.727 + 2.5\,(\,140\,) - 2.3\,(\,114\,) + (\,1.1\,)\,(\,133\,) - 0.3\,(\,158\,)$$

$= 187.427$ 이다.

Bootstrap 豫測値는 $X_{25} = 187.427$ 로 假定함으로써 期間 26에 대한 값을 얻을 수 있는데,

$$X_{26} = 0.727 + 2.5\,(\,187.427\,) - 2.3\,(\,140\,) + 1.1\,(\,114\,) - 0.3\,(\,133\,)$$

$= 232.795$ 이다.

같은 方法으로 期間 25와 26에 대한 豫測값을 實際값이라고 假定하면 X_{27} 에 대한 豫測値로써 271.433 등을 얻을 수가 있다.

요약하면 Box − Jenkins 方法論은 3가지 過程으로 構成되어 있다. 過程(Ⅰ)에 있어서는 적당한 ARMA (p . q) 模型이 判別되었다.

過程(Ⅱ)에 있어서는 模型의 母數가 推定되고 模型의 殘差가 任意的인지 아닌지를 보기 위해 模型을 檢定한다.

그것은 殘差에 대한 自己相關을 計算하고, 그들의 " 0 " 平均 주위에서 任意的으로 分布를 하는지 않는지를 決定하여 行할 수가 있다.

過程(Ⅲ)에 있어서는 적당한 模型이 時系列의 未來값을 豫測하기 위하여 利用한다.

〈附錄〉

I. ARIMA模型選定을 위한 SAS Package 利用方法

SAS Package에서 BOX – JENKINS 시계열 분석을 수행하는 Procedure는 SAS／ETS에 PROC ARIMA이며 일변량 시계열 모형, 외적영향 모형, 입력전이 함수모형등에 관한 것이 함께 내재되어 있다.

ARIMA는 PROC문(statement)에 의해 불려지고, IDENTIFY, ESTIMATE, FORECAST문등에 의해 수행된다. 이 문들은 다음과 같은 순서에 의해 수행된다.

```
PROC ARIMA options;
  IDENTIFY VAR = variable options;
  ESTIMATE options;
  FORECAST LEAD = number options;
  BY variables;
```

1.1) PROC ARIMA Statement.

이 문은 시계열자료를 지정해 주는 것으로 다룰 수 있는 시계열의 크기와 사용자가 이용할 options은 다음과 같다.

1.1.1) 시계열의 크기

최소 30개에서 최대 2,000개의 관측치가 있어야 한다.

1.1.2) options 설명

DATA = SAS dataset : 분석하려는 DATA NAME을 지정한다. 만일 PROC 문과 IDENTIFY문에 DATA NAME이 다를 경우에는 IDENTIFY

문에 지정된 DATA NAME으로 분석한다. 또한 DATA NAME 이 PROC 또는 IDENTIFY문 양쪽 모두 지정되지 않았을 때는 가장 최근에 지정한 DATA NAME으로 분석한다.

OUT = SAS dataset : 시계열 예측치에 DATA NAME을 지정한다. 만일 OUT =에 지정한 NAME이 PROC와 FORECAST 문에 다르게 지정했다면 FORECAST문에 지정한 NAME으로 이용한다.

1.2) IDENTIFY Statement

이 문은 모형을 식별하는데 도움을 줄 통계량을 계산한다. 즉 원시계열 또는 계차된 시계열의 자기상관함수, 역자기상관함수, 편자기상관함수 및 Chi-square 값을 계산하고 plot한다.

1.2.1) Options 설명

VAR = variable (d_1, d_2, ………, d_k) ;

: 시계열에 지정된 변수이름.

괄호속에 나열된 수들은 시계열 시차에 의한 계차(differencing)를 요구할때 괄호속에 수를 기입한다.

예를 들어 시계열에 연속계차 1차, 계절계차 S(=12)차를 할 때 X (1,12)로 표시한다. 여기서 X는 시계열에 지정된 변수이름.

DATA = SAS dataset : 분석하려는 DATA NAME을 지정한다. 만일 DATA NAME이 없으면 PROC문에 지정된 DATA NAME으로 수행하고, PROC문에서 NAME이 없으면 가장 최근에 지정된 DATA NAME으로 수행한다.

NLAG = number : 자기상관값, plot 를 프린트하고자 할때 고려될 시차수, 시차수를 지정하지 않으면 24 값을 갖는다.

NOPRINT : IDENTIFY문에서 산출된 통계량들과 plot 를 프린트 시키지 않을 때 지정한다.

1.3) ESTIMATE Statement

이 문은 IDENTIFY문에서 지정된 계열의 모형 모수를 추정하고, 모형의 적합성을 검토한다.

1.3.1) Options설명

P = lag : 모형의 자기회귀 부분에 차수를 지정한다.

모형에 자기회귀 모수가 적합하지 않을 때 차수를 지정하지 않는다.

예를 들어 P = 2를 지정하면 연속자기회귀 모수 2차 모형으로 ARIMA(2 , d , 0)이다.

P = (lag, lag, …… lag)

……(lag, lag, …… lag) : 계열의 계절성이 있을 때 이용한다.

예를 들어 P = (1,2)(12)는 연속 자기회수모수 2차, 계절자기회귀모수 12차인 모형으로 ARIMA (2, d, 0) (1, D, 0) 이다.

Q = lag : 모형의 이동평균 부분에 차수를 지정한다.

차수가 지정되지 않으면 모형에 이동평균모수가 적합하지 않음을 의미한다.

예를 들어 Q = 1은 연속이동평균 모수 1차인 모형으로 ARIMA (0 , d , 1)이다.

Q = (lag, lag, … lag)

…(lag, lag, … lag) : 계열의 계절성이 있을 때 이용한다.

예를 들어 Q = (1 , 2) (12)는 연속이동평균모수 2 차, 계절이동평균모수 12 차인 모형으로 ARIMA (0 , d , 2) (0 , D , 1)이다.

자기회귀모수와 이동평균모수가 혼합되어 있는 ARIMA (1 ; d , 1) (1 , D , 1) 모형을 표시할 경우 P = (1) (12) Q = (1) (12) 로 나타낸다.

NOCONSTANT : 모형에 상수항을 포함시키지 않을 경우 선택한다.

NOPRINT : ESTIMATE 문에서 산출된 값들을 프린트하지 않는다.

PRINTALL : ESTIMATE 문에서 산출된 과정과 결과를 모두 프린트 한다.

PLOT : 잔차들의 표본자기상관, 표본역자기상관, 표본편자기상관함수를 PLOT 한다.

1.4) FORECAST Statement

이 문은 ESTIMATE 문에서 산출된 모수 추정치를 이용하여 시계열의 예측값을 계산한다.

예측값은 또한 SAS dataset 의 Output 가 될 수 있다.

1.4.1) Options 설명

LEAD = number 시계열 연장 (예측) 기간의 수

OUT = SAS dataset : 예측치를 포함한 다른 값들의 산출결과 (output)에 NAME 을 지정한다. NAME 이 지정되지 않으면 PROC 문에서 지정된 OUT = NAME 이 이용된다.

ID = variable : Out Put 의 NAME에 관측치로 식별되어진 변수 이름을 지정한다.

INTERVAL = interval 관측치간에 시간 간격을 지정한다. 예를들어, YEAR, QTR, MONTH, WEEK, DAY 등이다.

NOPRINT : 예측치와 이에 관련된 값들이 프린트 안된다.

PRINTALL : 예측치와 이에 관련된 모든 값들이 프린트 된다.

1.5) BY Statement

이 문은 ARIMA에서 BY변수(variables)에 의해 지정한 자료에 따라 Dataset를 처리하는데 이용한다.

1.6) 산출결과

ARIMA과정에서 산출된 통계량을 나열해 보면 다음과 같다.

① 변수의 이름, 변수의 평균, 표준편차, 계차 후 관측치수

② 자기상관값과 PLOT

③ 역자기상관값과 PLOT

④ 편자기상관값과 PLOT

⑤ χ^2 —값(chi-square-value)

⑥ 모수들의 추정치, 표준오차, t—율, 시차(lags)

⑦ 분산추정치, 표준오차 추정치, 오차들의 수

⑧ 추정치들간의 상관관계

⑨ 오차들의 자기상관, 역자기상관, 편자기상관 값들과 PLOT

⑩ 예측값과 신뢰한계

1.7) 예제연구

산업생산지수 (1970 년 1 월 ~ 1983 년 12 월)의 자료를 이용하여 ARIMA 모형을 판별하는 과정을 SAS Package program으로 작성해 보자.

이 자료에 연속계차 1 차, 계절계차 1 차 (12) 를 한후 연속자기회귀모수 1 차, 연속이동평균모수 1 차, 계절이동평균모수 1 차를 적용할때 모형은 ARIMA (1 , 1 , 1) (0 , 1 , 1)$_{12}$ 이다. 이 모형에 대한 Program은 다음과 같다.

```
DATA SERIESA ;
 T = _N_  ;
 INPUT X : 6.1QQ;
 CARDS ;
154   147   172   177   186   182   183   179   177   190   187   197
176   181   202   207   220   216   207   209   211   201   212   219
201   187   215   223   243   244   235   235   237   259   263   276
265   240   286   292   318   305   310   331   331   361   358   366
349   360   395   403   418   429   422   389   399   399   400   437
423   393   455   455   463   471   471   478   492   522   517   556
519   525   594   593   618   641   651   645   625   666   664   683
639   609   691   721   740   779   742   767   756   796   792   862
794   738   886   836   946   941   898   946   921   964   986   1076
943   954   1063  1040  1096  1051  1017  990   1035  982   1007  1034
949   910   1019  1006  1030  1002  1002  990   957   1035  1036  1068
1012  927   1103  1138  1170  1134  1140  1165  1137  1196  1165  1193
1048  1052  1151  1134  1205  1194  1181  1155  1208  1201  1254  1291
1213  1132  1289  1322  1374  1405  1382  1400  1387  1432  1444  1475
PROC ARIMA DATA=SERIESA ;
 IDENTIFY VAR=X (1,12)  NLAG=24 ;
 ESTIMATE P=1  Q=(1) (12) NOCONSTANT PLOT ;
 FORECAST LEAD=24  OUT=FOR  ID=T  PRINTALL ;
```

이 Program에 의한 산출결과는 다음과 같다.

```
                              SAS                                10:51 FR
                         ARIMA PROCEDURE

   ①  NAME OF VARIABLE          =  X
       PERIODS OF DIFFERENCING=  1,12.

       MEAN OF WORKING SERIES=  0.104516
       STANDARD DEVIATION     =  3.49978
       NUMBER OF OBSERVATIONS=       155

                      ② AUTOCORRELATIONS

LAG COVARIANCE CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1         STD
  0    12.2484    1.00000  |                    |********************|           0
  1   -6.20594   -0.50667  |          **********|   .                |   0.0803219
  2    3.36573    0.27479  |                 .  |*****               |   0.0993134
  3  -0.643549   -0.05254  |                 . *|  .                 |    0.103625
  4     0.8182    0.06680  |                 .  |*  .                |    0.103793
  5   0.656129    0.05357  |                 .  |*  .                |    0.104075
  6  -0.331806   -0.02709  |                 . *|   .                |    0.104253
  7    1.32368    0.10807  |                 .  |** .                |    0.104298
  8   -2.33261   -0.19044  |                ****|   .                |    0.105018
  9    1.36719    0.11162  |                 .  |** .                |    0.107223
 10   -1.80248   -0.14716  |                .***|   .                |     0.10797
 11    3.46378    0.28279  |                 .  |******              |    0.109255
 12   -6.04757   -0.49374  |          **********|    .               |    0.113981
 13    2.48723    0.20307  |                .   |****.               |    0.125943
 14   -1.52922   -0.12485  |                .  **|    .              |    0.129021
 15   -1.21489   -0.09919  |                .  **|    .              |    0.129793
 16    1.27013    0.10370  |                .   |**  .               |    0.130286
 17   -2.42156   -0.19770  |               .****|    .               |    0.130313
 18    1.64698    0.13446  |                .   |***  .              |    0.132732
 19   -2.49698   -0.20386  |               .****|    .               |    0.133607
 20    1.84918    0.15097  |                .   |***  .              |    0.135599
 21  -0.120354   -0.00983  |                .   |    .               |     0.13663
 22   -1.28228   -0.10469  |                .  **|    .              |    0.135684
 23    1.38296    0.11291  |                .   |**  .               |      0.1372
 24   -1.31864   -0.10766  |               .  **|     .              |    0.137799
                           '.' MARKS TWO STANDARD ERRORS
```

③ INVERSE AUTOCORRELATIONS

```
LAG CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1     0.33179  |                 .  |*******             |
  2    -0.00415  |                 .  |  .                 |
  3    -0.07814  |                 .**|  .                 |
  4    -0.05163  |                 . *|  .                 |
  5    -0.10098  |                 .**|  .                 |
  6    -0.21166  |                ****|  .                 |
  7    -0.02183  |                 .  |  .                 |
  8     0.05369  |                 .  |* .                 |
  9    -0.02765  |                 . *|  .                 |
 10    -0.06742  |                 . *|  .                 |
 11     0.04496  |                 .  |* .                 |
 12     0.44740  |                 .  |*********           |
 13     0.16738  |                 .  |***                 |
 14     0.09978  |                 .  |**.                 |
 15     0.07753  |                 .  |**.                 |
 16     0.03787  |                 .  |* .                 |
 17    -0.01948  |                 .  |  .                 |
 18    -0.09712  |                 .**|  .                 |
 19    -0.03214  |                 . *|  .                 |
 20    -0.02043  |                 .  |  .                 |
 21    -0.01478  |                 .  |  .                 |
 22     0.02604  |                 .  |* .                 |
 23     0.02132  |                 .  |  .                 |
 24     0.13135  |                 .  |***                 |
```

④ PARTIAL AUTOCORRELATIONS

```
LAG CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1    -0.50667  |          **********|  .                 |
  2     0.02431  |                 .  |  .                 |
  3     0.12872  |                 .  |***                 |
  4     0.11186  |                 .  |**.                 |
  5     0.13568  |                 .  |***                 |
  6     0.03622  |                 .  |* .                 |
  7     0.09224  |                 .  |**.                 |
  8    -0.17050  |                 ***|  .                 |
  9    -0.13445  |                 ***|  .                 |
 10    -0.16814  |                 ***|  .                 |
 11     0.26175  |                 .  |*****               |
 12    -0.32143  |              ******|  .                 |
 13    -0.29748  |              ******|  .                 |
 14    -0.06311  |                 . *|  .                 |
 15    -0.11819  |                 .**|  .                 |
 16     0.03069  |                 .  |* .                 |
 17    -0.05921  |                 . *|  .                 |
 18     0.07209  |                 .  |* .                 |
 19     0.04453  |                 .  |* .                 |
 20    -0.06089  |                 . *|  .                 |
 21     0.10083  |                 .  |**.                 |
 22    -0.18012  |                ****|  .                 |
 23     0.18467  |                 .  |****                |
 24    -0.23733  |               *****|  .                 |
```

AUTOCORRELATION CHECK FOR WHITE NOISE

⑤

TO LAG	CHI SQUARE	DF	PROB	AUTOCORRELATIONS					
6	54.32	6	0.000	-0.507	0.275	-0.053	0.067	0.054	-0.027
12	122.96	12	0.000	0.108	-0.190	0.112	-0.147	0.283	-0.494
18	146.41	18	0.000	0.203	-0.125	-0.099	0.104	-0.193	0.134
24	164.48	24	0.000	-0.204	0.151	-0.010	-0.105	0.113	-0.108

ARIMA: LEAST SQUARES ESTIMATION

⑥

PARAMETER	ESTIMATE	STD ERROR	T RATIO	LAG
MA1,1	-0.0373899	0.183872	-0.20	1
MA2,1	0.60052	0.0712113	8.43	12
AR1,1	-0.47056	0.163333	-2.88	1

⑦

VARIANCE ESTIMATE = 6.33745
STD ERROR ESTIMATE = 2.51743
NUMBER OF RESIDUALS= 155

⑧

CORRELATIONS OF THE ESTIMATES

	MA1,1	MA2,1	AR1,1
MA1,1	1.000	0.060	0.896
MA2,1	0.060	1.000	0.126
AR1,1	0.896	0.126	1.000

⑨ AUTOCORRELATION CHECK OF RESIDUALS

TO LAG	CHI SQUARE	DF	PROB	AUTOCORRELATIONS					
6	6.79	3	0.079	0.001	0.028	0.056	0.060	0.114	0.142
12	15.21	9	0.085	0.067	-0.122	-0.003	-0.059	0.113	-0.121
18	22.41	15	0.097	0.030	-0.122	-0.144	-0.023	-0.062	0.017
24	30.69	21	0.079	-0.094	0.072	-0.037	-0.147	0.064	0.061
30	42.65	27	0.028	0.147	0.031	0.023	-0.052	0.192	0.002

⑨ AUTOCORRELATION PLOT OF RESIDUALS

```
LAG COVARIANCE CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1       STD
  0    6.33745    1.00000  |                    |********************|           0
  1 0.00575367    0.00091  |                  . |  .                 |   0.0303219
  2   0.174935    0.02760  |                  . |* .                 |    0.030322
  3   0.415107    0.06550  |                  . |* .                 |   0.0303332
  4    0.37851    0.05973  |                  . |* .                 |   0.0307268
  5   0.719668    0.11356  |                  . |**.                 |   0.0310113
  6   0.902394    0.14239  |                  . |***                 |   0.0320319
  7   0.422328    0.06664  |                  . |* .                 |   0.0335113
  8  -0.773642   -0.12207  |                  .**|  .                |   0.0339532
  9 -0.0525666   -0.00829  |                  . |  .                 |   0.0350007
 10  -0.375776   -0.05929  |                  . *|  .                |    0.035096
 11   0.713425    0.11257  |                  . |**.                 |   0.0353621
 12  -0.765523   -0.12079  |                  .**|  .                |   0.0363145
 13      0.193    0.03045  |                  . |* .                 |   0.0373984
 14  -0.775239   -0.12233  |                  .**|  .                |   0.0374663
 15  -0.915638   -0.14443  |                 .***|   .               |   0.0385637
 16   -0.14629   -0.02303  |                 .   |   .               |   0.0390715
 17  -0.395476   -0.06240  |                 .  *|   .               |   0.0391097
 18     0.1084    0.01710  |                 .   |   .               |    0.039338
 19  -0.593069   -0.09358  |                 . **|   .               |   0.0394039
 20   0.458239    0.07231  |                 .   |*  .               |   0.0310317
 21  -0.235364   -0.03714  |                 .  *|   .               |   0.0314015
 22  -0.945036   -0.14912  |                 .***|   .               |   0.0314988
 23   0.408724    0.06449  |                 .   |*  .               |   0.0330535
 24   0.385505    0.06083  |                 .   |*  .               |   0.0333415
                           '.' MARKS TWO STANDARD ERRORS
```

⑨ INVERSE AUTOCORRELATIONS

```
LAG CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1    -0.01063  |                  .  |  .                 |
  2     0.02371  |                  .  |* .                 |
  3    -0.08321  |                  .**|  .                 |
  4    -0.08345  |                  .**|  .                 |
  5    -0.08780  |                  .**|  .                 |
  6    -0.14381  |                  ***|  .                 |
  7     0.01780  |                  .  |  .                 |
  8     0.09496  |                  .  |**.                 |
  9    -0.02694  |                  . *|  .                 |
 10     0.06627  |                  .  |* .                 |
 11    -0.17414  |                  ***|  .                 |
 12     0.10463  |                  .  |**.                 |
 13    -0.05479  |                  . *|  .                 |
 14     0.13755  |                  .  |***                 |
 15     0.10295  |                  .  |**.                 |
 16     0.00495  |                  .  |  .                 |
 17     0.04587  |                  .  |* .                 |
 18    -0.03765  |                  . *|  .                 |
 19     0.00195  |                  .  |  .                 |
 20    -0.08739  |                  .**|  .                 |
 21    -0.03740  |                  . *|  .                 |
 22     0.15196  |                  .  |***                 |
 23    -0.06092  |                  . *|  .                 |
 24    -0.00795  |                  .  |  .                 |
```

⑨ PARTIAL AUTOCORRELATIONS

```
LAG CORRELATION -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1    0.00091  |                   .  |  .                 |
  2    0.02760  |                   .  |* .                 |
  3    0.06550  |                   .  |* .                 |
  4    0.05927  |                   .  |* .                 |
  5    0.11123  |                   .  |**.                 |
  6    0.13962  |                   .  |***                 |
  7    0.06156  |                   .  |* .                 |
  8   -0.14682  |                   ***|  .                 |
  9   -0.05073  |                   . *|  .                 |
 10   -0.09977  |                   .**|  .                 |
 11    0.09233  |                   .  |**.                 |
 12   -0.13690  |                   ***|  .                 |
 13    0.05634  |                   .  |* .                 |
 14   -0.09261  |                   .**|  .                 |
 15   -0.10623  |                   .**|  .                 |
 16   -0.03958  |                   . *|  .                 |
 17   -0.05195  |                   . *|  .                 |
 18    0.04632  |                   .  |* .                 |
 19   -0.02232  |                   .  |  .                 |
 20    0.12134  |                   .  |**.                 |
 21    0.04981  |                   .  |* .                 |
 22   -0.18294  |                  ****|  .                 |
 23    0.07309  |                   .  |* .                 |
 24    0.00953  |                   .  |  .                 |
```

SAS

```
MODEL FOR VARIABLE X
NO MEAN TERM IN THIS MODEL.
PERIODS OF DIFFERENCING= 1,12.

 AUTOREGRESSIVE FACTORS
                   FACTOR 1
1+ 0.470568**(1)

 MOVING AVERAGE FACTORS
                   FACTOR 1
1+.0373899B**(1)
                   FACTOR 2
1-0.600528**(12)
```

FORECASTS FOR VARIABLE X

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%	ACTUAL	RESIDUAL
14	18.1000	2.5174	13.1659	23.0341	18.1000	0.0000
15	20.0353	2.5174	15.1013	24.9694	20.2000	0.1647
16	20.8944	2.5174	15.9603	25.8284	20.7000	-0.1944
17	21.5927	2.5174	16.6587	26.5268	22.0000	0.4073
18	21.4270	2.5174	16.4929	26.3611	21.6000	0.1730
19	21.7065	2.5174	16.7724	26.6405	20.7000	-1.0065
20	20.7329	2.5174	15.7989	25.6670	20.8000	0.0671
21	20.3672	2.5174	15.4332	25.3013	21.1000	0.7328
22	22.1921	2.5174	17.2581	27.1262	20.1000	-2.0921
23	20.8041	2.5174	15.8700	25.7381	21.2000	0.3959
24	21.5560	2.5174	16.6220	26.4901	21.9000	0.3440
25	19.9540	2.5174	15.0200	24.8881	20.1000	0.1460
26	20.4643	2.5174	15.5302	25.3983	18.7000	-1.7643
27	21.5292	2.5174	16.5951	26.4633	21.5000	-0.0292
28	21.7825	2.5174	16.8485	26.7166	22.3000	0.5175
29	23.2380	2.5174	18.3039	28.1720	24.3000	1.0620
30	23.4973	2.5174	18.5632	28.4313	24.4000	0.9027
31	23.8990	2.5174	18.9649	28.8331	23.5000	-0.3990
32	23.5674	2.5174	18.6333	28.5015	23.5000	-0.0674
33	23.4030	2.5174	18.4689	28.3370	23.7000	0.2970
34	23.9981	2.5174	19.0640	28.9321	25.9000	1.9019
35	25.3745	2.5174	20.4405	30.3086	26.3000	0.9255
36	27.1485	2.5174	22.2145	32.0826	27.6000	0.4515
37	25.4392	2.5174	20.5051	30.3732	26.5000	1.0608
38	25.8665	2.5174	20.9324	30.8005	24.0000	-1.8665
39	27.3050	2.5174	22.3709	32.2390	28.6000	1.2950
40	28.2913	2.5174	23.3573	33.2254	29.2000	0.9087
41	30.6787	2.5174	25.7446	35.6128	31.8000	1.1213
42	31.0936	2.5174	26.1596	36.0277	30.5000	-0.5936
43	30.4559	2.5174	25.5219	35.3900	31.0000	0.5441
44	30.4110	2.5174	25.4769	35.3451	33.1000	2.6890
45	32.2355	2.5174	27.3015	37.1696	33.1000	0.8645
46	34.2776	2.5174	29.3436	39.2117	35.6000	1.3224
47	35.5932	2.5174	30.6592	40.5273	35.8000	0.2068
48	37.1452	2.5174	32.2112	42.0793	36.6000	-0.5452
49	35.0677	2.5174	30.1336	40.0018	34.9000	-0.1677
50	33.7731	2.5174	28.8390	38.7072	36.0000	2.2269
51	38.2535	2.5174	33.3194	43.1875	39.5000	1.2465
52	40.0895	2.5174	35.1554	45.0235	40.3000	0.2105
53	42.1200	2.5174	37.1859	47.0541	41.8000	-0.3200
54	41.3370	2.5174	36.4029	46.2710	42.9000	1.5630
55	42.0157	2.5174	37.0816	46.9498	42.2000	0.1843
56	43.2445	2.5174	38.3105	48.1786	38.9000	-4.3445
57	40.6591	2.5174	35.7250	45.5931	39.9000	-0.7591
58	41.2858	2.5174	36.3517	46.2198	39.9000	-1.3858
59	40.7948	2.5174	35.8607	45.7288	40.0000	-0.7948
60	40.9048	2.5174	35.9708	45.8389	43.7000	2.7952
61	40.8528	2.5174	35.9188	45.7869	42.3000	1.4472
62	41.9794	2.5174	37.0454	46.9135	39.3000	-2.6794
63	43.8305	2.5174	38.8965	48.7646	45.5000	1.6695
64	44.9375	2.5174	40.0034	49.8716	45.5000	0.5625
65	47.5849	2.5174	42.6509	52.5190	48.3000	0.7151
66	47.8836	2.5174	42.9495	52.8176	47.1000	-0.7836
67	47.3072	2.5174	42.3732	52.2413	47.1000	-0.2072
68	46.0677	2.5174	41.1336	51.0018	47.8000	1.7323

FORECASTS FOR VARIABLE X

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%	ACTUAL	RESIDUAL
69	47.5599	2.5174	42.6259	52.4940	49.2000	1.6401
70	49.9232	2.5174	44.9892	54.8573	52.2000	2.2768
71	51.4818	2.5174	46.5478	56.4159	51.7000	0.2182
72	54.0298	2.5174	49.0957	58.9639	55.6000	1.5702
73	53.2328	2.5174	48.2987	58.1668	51.9000	-1.3328
74	51.5090	2.5174	46.5749	56.4431	52.5000	0.9910
75	56.1007	2.5174	51.1666	61.0347	59.4000	3.2993
76	58.8187	2.5174	53.8846	63.7527	59.3000	0.4813
77	61.7230	2.5174	56.7889	66.6571	61.8000	0.0770
78	61.1985	2.5174	56.2645	66.1326	64.1000	2.9015
79	62.7036	2.5174	57.7695	67.6376	65.1000	2.3964
80	64.3834	2.5174	59.4494	69.3175	64.5000	0.1166
81	65.4923	2.5174	60.5582	70.4264	62.3000	-3.1923
82	65.4706	2.5174	60.5365	70.4046	66.6000	1.1294
83	65.3484	2.5174	60.4143	70.2824	66.4000	1.0516
84	69.2503	2.5174	64.3163	74.1844	68.3000	-0.9503
85	66.2707	2.5174	61.3366	71.2048	63.9000	-2.3707
86	64.1756	2.5174	59.2415	69.1096	60.9000	-3.2756
87	67.3680	2.5174	62.4339	72.3020	69.1000	1.7320
88	68.0899	2.5174	63.1559	73.0240	72.1000	4.0101
89	73.2342	2.5174	68.3001	78.1682	74.0000	0.7658
90	74.8669	2.5174	69.9328	79.8009	77.9000	3.0331
91	76.7563	2.5174	71.8222	81.6903	74.2000	-2.5563
92	75.5922	2.5174	70.6582	80.5263	76.7000	1.1078
93	74.9971	2.5174	70.0630	79.9312	75.6000	0.6029
94	78.7984	2.5174	73.8643	83.7324	79.6000	0.8016
95	78.9143	2.5174	73.9802	83.8483	79.2000	0.2857
96	81.7519	2.5174	76.8178	86.6859	86.2000	4.4481
97	81.0114	2.5174	76.0774	85.9455	79.4000	-1.6114
98	79.4894	2.5174	74.5553	84.4234	73.8000	-5.6894
99	82.0442	2.5174	77.1101	86.9782	88.6000	6.5558
100	86.2924	2.5174	81.3583	91.2265	88.6000	2.3076
101	91.4480	2.5174	86.5140	96.3821	94.6000	3.1520
102	94.8499	2.5174	89.9158	99.7840	94.1000	-0.7499
103	93.9094	2.5174	88.9753	98.8435	89.8000	-4.1094
104	91.8208	2.5174	86.8868	96.7549	94.6000	2.7792
105	92.1347	2.5174	87.2006	97.0688	92.1000	-0.0347
106	96.2626	2.5174	91.3285	101.1966	96.4000	0.1374
107	95.6744	2.5174	90.7403	100.6084	98.6000	2.9256
108	101.8083	2.5174	96.8743	106.7424	102.6000	0.7917
109	98.1091	2.5174	93.1751	103.0432	94.8000	-3.3091
110	92.7162	2.5174	87.7821	97.6502	95.4000	2.6838
111	103.3384	2.5174	98.4044	108.2725	106.3000	2.9616
112	106.7130	2.5174	101.7789	111.6470	104.0000	-2.7130
113	109.0362	2.5174	104.1021	113.9703	109.6000	0.5638
114	109.6889	2.5174	104.7548	114.6229	105.1000	-4.5889
115	104.9953	2.5174	100.0612	109.9293	101.7000	-3.2953
116	104.3766	2.5174	99.4426	109.3107	99.0000	-5.3766
117	99.7866	2.5174	94.8525	104.7207	103.5000	3.7134
118	104.5632	2.5174	99.6291	109.4972	98.2000	-6.3632
119	102.9195	2.5174	97.9854	107.8535	100.7000	-2.2195
120	103.9347	2.5174	99.0007	108.8688	103.4000	-0.5347
121	97.9614	2.5174	93.0273	102.8955	94.9000	-3.0614
122	94.4535	2.5174	89.5194	99.3875	91.0000	-3.4535
123	102.2849	2.5174	97.3509	107.2190	101.9000	-0.3849
124	101.1483	2.5174	96.2142	106.0824	100.6000	-0.5483

FORECASTS FOR VARIABLE X

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%	ACTUAL	RESIDUAL
125	105.4313	2.5174	100.4972	110.3653	103.0000	-2.4313
126	102.6579	2.5174	97.7239	107.5920	100.2000	-2.4579
127	97.9901	2.5174	93.0560	102.9241	100.2000	2.2099
128	99.2855	2.5174	94.3514	104.2195	99.0000	-0.2855
129	100.6742	2.5174	95.7402	105.6083	95.7000	-4.9742
130	97.6222	2.5174	92.6881	102.5563	103.5000	5.8778
131	101.5312	2.5174	96.5971	106.4652	103.6000	2.0688
132	107.8777	2.5174	102.9436	112.8117	106.8000	-1.0777
133	99.8749	2.5174	94.9408	104.8089	101.2000	1.3251
134	98.1275	2.5174	93.1935	103.0616	92.7000	-5.4275
135	105.8703	2.5174	100.9363	110.8044	110.3000	4.4297
136	106.3508	2.5174	101.4167	111.2848	113.8000	7.4492
137	115.6922	2.5174	110.7581	120.6262	117.0000	1.3078
138	115.4031	2.5174	110.4690	120.3371	118.4000	2.9969
139	115.2638	2.5174	110.3297	120.1978	114.0000	-1.2638
140	114.9450	2.5174	110.0110	119.8791	116.5000	1.5550
141	114.5106	2.5174	109.5765	119.4447	113.7000	-0.8106
142	117.8164	2.5174	112.8823	122.7504	119.6000	1.7836
143	119.2864	2.5174	114.3523	124.2205	116.5000	-2.7864
144	121.7023	2.5174	116.7682	126.6364	119.3000	-2.4023
145	113.0268	2.5174	108.0928	117.9609	104.8000	-8.2268
146	103.4100	2.5174	98.4759	108.3440	105.2000	1.7900
147	116.1407	2.5174	111.2067	121.0748	115.1000	-1.0407
148	117.6115	2.5174	112.6775	122.5456	118.4000	0.7885
149	120.7710	2.5174	115.8369	125.7050	120.5000	-0.2710
150	120.5784	2.5174	115.6443	125.5125	119.4000	-1.1784
151	116.8240	2.5174	111.8899	121.7580	118.1000	1.2760
152	118.2836	2.5174	113.3495	123.2176	115.5000	-2.7836
153	115.4477	2.5174	110.5136	120.3817	120.8000	5.3523
154	122.0357	2.5174	117.1016	126.9697	120.1000	-1.9357
155	121.6666	2.5174	116.7325	126.6006	125.4000	3.7334
156	125.8921	2.5174	120.9580	130.8261	129.1000	3.2079
157	119.2908	2.5174	114.3567	124.2248	121.3000	2.0092
158	117.7322	2.5174	112.7981	122.6662	113.2000	-4.5322
159	127.5151	2.5174	122.5810	132.4491	128.9000	1.3849
160	129.0724	2.5174	124.1384	134.0065	132.2000	3.1276
161	134.5620	2.5174	129.6279	139.4960	137.4000	2.8380
162	135.6611	2.5174	130.7271	140.5952	140.5000	4.8389
163	136.6648	2.5174	131.7307	141.5988	138.2000	1.5352
164	137.7709	2.5174	132.8368	142.7050	140.0000	2.2291
165	140.1612	2.5174	135.2271	145.0953	138.7000	-1.4612
166	142.0933	2.5174	137.1592	147.0274	143.2000	1.1067
167	143.8959	2.5174	138.9619	148.8300	144.4000	0.5041
168	148.0379	2.5174	143.1038	152.9720	147.5000	-0.5379

--------FORECAST BEGINS--------

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%
169	138.6836	2.5174	133.7495	143.6177
170	133.7384	2.6937	128.0668	139.4100
171	147.2240	3.4839	140.3957	154.0523
172	149.6567	3.8759	142.0600	157.2534
173	153.4903	4.2787	145.1042	161.8763
174	154.2637	4.6261	145.1968	163.3306
175	152.0279	4.9581	142.3102	161.7456
176	152.4246	5.2655	142.1049	162.7444
177	152.6124	5.5573	141.7203	163.5044
178	155.7805	5.8339	144.3463	167.2147
179	157.2797	6.0983	145.3272	169.2321
180	160.5506	6.3516	148.1018	172.9994
181	151.6658	6.9340	138.0755	165.2562
182	146.7528	7.3201	132.4058	161.0998
183	160.2232	7.7518	145.0300	175.4165
184	162.6631	8.1311	146.7265	178.5998
185	166.4933	8.5067	149.8205	183.1662
186	167.2683	8.8604	149.9023	184.6344
187	165.0318	9.2033	146.9936	183.0699
188	165.4289	9.5325	146.7454	184.1122
189	165.6164	9.8514	146.3081	184.9247
190	168.7846	10.1600	148.8715	188.6973
191	170.2838	10.4598	149.7835	190.7841
192	173.5547	10.7508	152.4837	194.6253