ARIMA automodeler

MODEL STRUCTURE IDENTIFIER

RECURSIVE DIFFERENCER/ INTEGRATER

PARAMETER ESTIMATOR

N-STEP AHEAD PREDICTOR

Stationary Series

Non-stationary and/or Seasonal series

Forecasts for Original Series

Differenced Series

Estimated Parameters

Estimated Parameters

Forecasts for Original Series

Forecasts for Differenced Series

y(t)

y(t)

(t+n)

(t+n)

Modeling process:

* Thành phần nhận dạng mô hình(Model identification) : thành phần này ước lượng và xấp xỉ mối liên hệ trong dữ liệu thu thập được. Thành phần này nhận dạng được các mẫu một cách tự động bằng cách phân tích hệ số tự tương quan và hệ số tự tương quan bán phần theo các độ trễ khác nhau. Do đó sẽ tránh được việc người dự đoán phải dựa vào các đồ thị dữ liệu để xác định mô hình.
* Thành phần lấy hiệu/tổng hợp (Recursive differencer/integrator): dùng để chuyển đổi các chuỗi dữ liệu không tĩnh hoặc có tính mùa sang chuỗi dữ liệu tỉnh trước khi áp dụng các giải thuật để ước lượng các tham số cho mô hình. Bởi vì mô hình ARIMA chỉ phù hợp với chuỗi dữ liệu tĩnh, các thành phần bị loại bỏ khỏi chuỗi dữ liệu trong quá trình chuyển đổi phải được tích hợp lại vào dữ liệu khi dự đoán.
* Thành phần ước lượng tham số (Parameter estimator): ước lượng tham số của mô hình bằng giải thuật tối thiểu sai số đệ quy. Giải thuật này sử dụng đệ quy để giảm chi phí tính toán và kĩ thuật tối thiểu lỗi để tối thiểu bình phương các ước lượng lỗi. Giải thuật này có hai giả định là: mô hình phải tuyến tính với các tham số và chuỗi lỗi phải có phân phối tự nhiên với giá trị trung bình là 0 và phương sai không đổi.
* Thành phần dự đoán (N-step ahead predictor): tạo giá trị dự đoán cho các giai đoạn tiếp theo dựa vào mô hình đã ước lượng được.

1. Thành phần nhận dạng mô hình
   1. Phân tích hệ số tự tương quan
      1. Hệ số tự tương quan
      2. Hệ số tự tương quan riêng phần
      3. Tính chất mô hình trong ACFs và PACFs

Mô hình ARIMA sử dụng tính chất của ACFs và PACFs để kết hợp một quá trình với một mô hình thích hợp: mô hình tự hồi quy AR(p), mô hình trung bình di động MA(q), mô hình kết hợp tự hồi quy và trug bình di động ARMA(p,q).

* Mô hình tự hồi quy AR(p): quá trình có ACFs với các độ trễ giảm chậm về không theo hình sin hoặc theo hàm mũ, trong khi đó PACFs giảm nhanh về không sau một vài độ trễ.
* Mô hình trung bình di động MA(q): quá trình có tính chất ngược lại với mô hình AR(p), ACFs giảm nhanh về không sau một vài độ trễ, trong khi đó PACFs với các độ trễ giảm chậm về không theo hình sin hoặc theo hàm mũ.
* Mô hình kết hợp tự hồi quy và trug bình di động ARMA(p,q): quá trình có ACFs và PACFs giảm chậm về không.
  + 1. Kĩ thuật nhận dạng mô hình

Dựa vào tính chất của ACFs và PACFs trong mỗi mô hình ARIMA, ta có thể xây dựng phương pháp để tự động nhận dạng ra mô hình:

* Nhận dạng các hệ số tương quan có ý nghĩa trong ACFs và PACFs sử dụng khoảng tin cậy thống kê.
* Trích xuất mô hình từ những hệ số tương quan đó
  + - 1. Nhận dạng hệ số tương quan có ý nghĩa

Trong các quá trình, các giá trị gần nhau sẽ liên kết với nhau hơn những giá trị xa nhau. Do đó, trong các quá trình tĩnh hệ số tự tương quan và tự tương quan riêng phần giảm về không tại những độ trễ lớn. Barlett đã đưa ra công thức để xấp xỉ độ lêch chuẩn của ACF và PACFs:



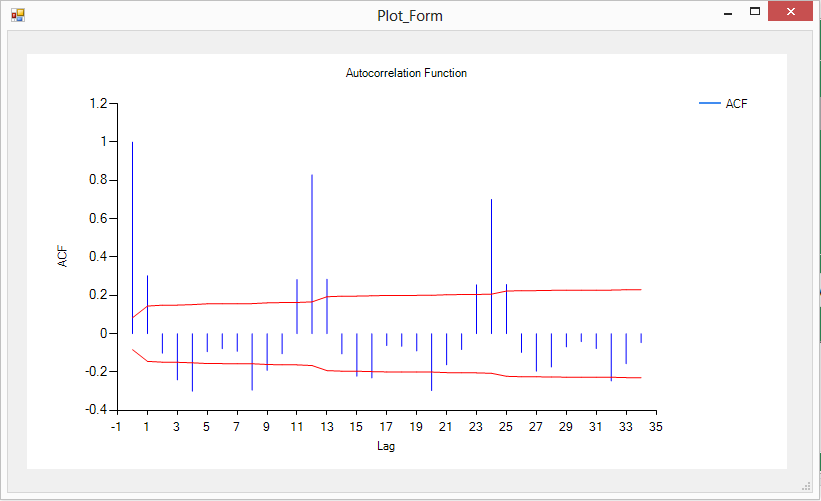


Ta có thể tự động chọn những hệ số tương quan có ý nghĩa bằng cách chọn những giá trị vượt qua khoản tin cậy.

* + - 1. Tự động xác định chu kì mùa

Đối với những quá trình có tình mùa, các giá trị quan sát không chỉ phụ thuộc vào những giá trị lân cận mà còn phụ thuộc vào những giá trị cách nó một bội số của chiều dài mùa. Do vậy để xác định chu kì mùa, ta cần:

* Xác định các hệ số tương quan tối ưu: các hệ số này cao hơn các hệ số lân cận và là giá trị cao nhất cục bộ.
* Tính khoảng cách giữa các hệ số tương quan tối ưu: sự sai khác về độ trễ giữa các hệ số liên tiếp
* Nhận dạng khoảng cách được lặp lại nhiều nhất, nếu tần suất của khoảng cách này lớn thì đó chính là chu kì mùa.



* 1. Tự động phân tích hệ số tự tương quan

Hầu hết các ACFs và PACFs đều có mô hình thuộc một trong ba dạng sau:

* Giảm từ từ trong các quá trình không tĩnh
* Giảm theo hình sin hoặc hàm mũ trong quá trình tĩnh
* Giảm nhanh trong quá trình tĩnh
  + 1. Kiểm tra độ trễ

Nếu các hệ số tương quan chuẩn, không bị nhiễu và đơn điệu giảm thì ta có thể xác định loại mô hình dựa vào vị trí của độ trễ có ý nghĩa cuối cùng.

If( hight frequencies are monotonically decreasing)

{

If ( hight frequencies stay within MIN\_LAGS) then

Pattern = abrupt cut-off

Else if (hight frequencies extend beyond MAX\_LAGS) then

Pattern = slow decay

Else Pattern = exponential decay

}

Tuy nhiên, các hệ số nhiễu thường xuyên xuất hiện trong ACFs và PACFs vì chịu ảnh hưởng của sai số mẫu hoặc các hệ số lớn lân cận.

* + 1. Kiểm tra tỉ lệ thay đổi trung bình

1. Thành phần lấy hiệu/tổng hợp

Trong thực tế, các quá trình rất phức tạp và thường mang tính xu hướng hoặc tính mùa. Do đó, các quá trình cần được biến đổi thành các quá trình tĩnh trước khi đưa vào thành phần nhận dạng mô hình. Lấy hiệu là kĩ thuật được sử dụng phổ biến nhất.

* 1. Khái niệm cơ bản về lấy hiệu

Lấy hiệu là kĩ thuật biến đổi nhằm làm tĩnh giá trị trung bình của quá trình không tĩnh và mang tính nùa. Ở đây có hai loại lấy hiệu tùy thuộc vào khoảng lấy hiệu

* Lấy hiệu thông thường: kĩ thuật này dùng cho quá trình không tĩnh và không có tính mùa. Khoảng lấy hiệu là một, tức là những giá trị liên tiếp nhau bị trừ để tạo ra quá trình mới: . Ví dụ với quá trình không tĩnh sau: 70.3, 100.5, 103.2, 160.7, 189.5, 220.2, 250.4

100.5 130.2 160.7 189.5 220.2 250.4

- 70.3 100.5 130.2 160.7 189.5 220.2

30.2 29.7 30.5 28.8 30.7 30.2

Sau khi lấy hiệu trên quá trình, ta thu được một quá trình tĩnh mới dao động quanh giá trị trung bình 30.

* Lấy hiệu theo mùa: kĩ thuật này dùng cho quá trình có tính mùa. Khoảng lấy hiệu là một giá trị lớn hơn một S, chiều dài của mùa. Mỗi giá trị trong quá trình chuyển đổi chính là sự sai khác giữa các giá trị cách nhau S giai đoạn trong quá trình cũ: . Ví dụ với quá trình có tình mùa sau: 3.2, 6.3, 9.5, 12.1, 50.0, 6.3, 9.7, 12.3, 15.1, 53.1, 9.4, 12.5, 15.5, 18.2, 56.4

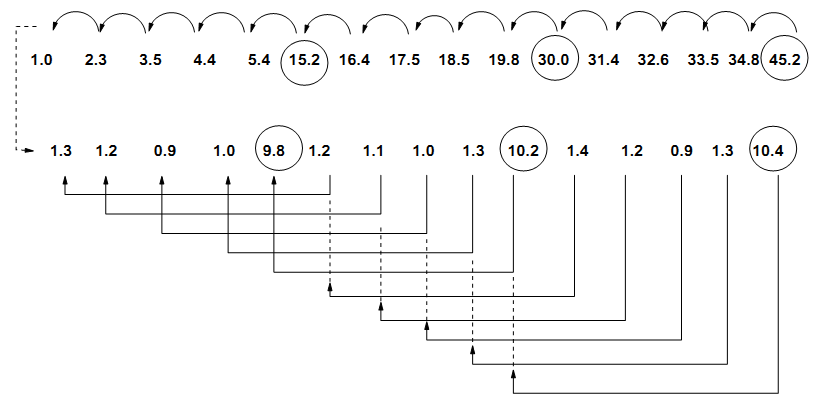
6.3 9.7 12.3 15.1 53.1 9.4 12.5 15.5 18.2 56.4

- 3.2 6.3 9.5 12.1 50.0 6.3 9.7 12.3 15.1 53.1

3.1 3.4 2.8 3.0 3.1 3.1 2.8 3.2 3.1 3.3

Sau khi lấy hiệu với khoảng lấy hiệu là 5, chiều dài của mùa thì ta được một quá trình tĩnh dao động xung quanh giá trị trung bình 3.

Với một số quá trình chỉ với một lần lấy hiệu ta có thể lấy được quá trình tĩnh. Nhưng đối với một số quá trình, ta phải lấy liệu nhiều lần và kết hợp cả hai phương pháp mới có được quá trình tĩnh.



Lấy hiệu thông thường

Lấy hiệu theo mùa S=5

1. Thành phần ước lượng tham số
2. Thành phần dự đoán