

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA TOÁN - CƠ - TIN HỌC

—o0o—



BÁO CÁO

HỌC PHẦN PHÂN TÍCH CHUỖI THỜI GIAN

Chủ đề

Phân tích giá quả bơ

Thành viên: Nguyễn Văn An  
Nguyễn Bá Mạnh  
Giảng viên: Đỗ Thị Minh Tâm  
ThS. Hoàng Thị Cành

# LỜI MỞ ĐẦU

Trong bối cảnh nền kinh tế toàn cầu ngày càng phát triển, việc nghiên cứu và phân tích các yếu tố ảnh hưởng đến thị trường hàng hóa đã trở thành một nhu cầu cấp thiết. Đặc biệt, sản phẩm quả bơ với vai trò là một loại thực phẩm phổ biến, không chỉ được tiêu thụ rộng rãi mà còn mang giá trị dinh dưỡng cao, đã thu hút sự chú ý của nhiều nhà nghiên cứu và nhà kinh doanh.

Báo cáo này được thực hiện nhằm phân tích xu hướng giá của quả bơ trên thị trường Hoa Kỳ trong giai đoạn từ năm 2015 đến năm 2018. Thông qua việc áp dụng các phương pháp phân tích dữ liệu và mô hình chuỗi thời gian, chúng tôi mong muốn làm rõ các yếu tố ảnh hưởng đến giá bơ, từ đó đưa ra các nhận định, dự đoán và đề xuất phù hợp.

Trong quá trình thực hiện dự án, mặc dù chúng em đã cố gắng hết sức, nhưng không tránh khỏi những thiếu sót và hạn chế. Chúng em trân trọng mọi ý kiến đóng góp từ giảng viên, nhằm làm cho dự án của chúng em trở nên hoàn thiện hơn.

Báo cáo gồm 3 phần sau:

Phần 1: Lời mở đầu.

Phần 2: Tổng quan về đề tài.

Phần 3: Tài liệu tham khảo.

# Mục lục

<b>LỜI MỞ ĐẦU</b>	<b>1</b>
<b>TỔNG QUAN VỀ ĐỀ TÀI</b>	<b>5</b>
<b>1 Cơ sở nghiên cứu giá bơ</b>	<b>6</b>
1.1 Tổng quan về dữ liệu . . . . .	6
1.1.1 Nguồn dữ liệu . . . . .	6
1.1.2 Các biến chính trong bộ dữ liệu . . . . .	6
1.1.3 Phạm vi dữ liệu . . . . .	7
1.1.4 Đặc điểm nổi bật của dữ liệu . . . . .	7
1.2 Phân tích dữ liệu . . . . .	8
1.2.1 So sánh 2 loại bơ . . . . .	8
1.2.2 Mối quan hệ giữa giá và tổng khối lượng . . . . .	9
1.2.3 Các kiểu mẫu . . . . .	10
<b>2 Phân tích chuỗi thời gian</b>	<b>22</b>
2.1 Phân tích dữ liệu theo năm tháng . . . . .	22
2.2 Chuỗi thời gian cố định . . . . .	25
<b>3 Dự đoán chuỗi thời gian</b>	<b>28</b>
3.1 Phương pháp ARIMA . . . . .	29
3.1.1 Thử nghiệm mô hình . . . . .	29
3.1.2 Mô hình tốt nhất . . . . .	29
3.1.3 Đánh giá mô hình . . . . .	29
3.2 ARIMA với loại bơ thường . . . . .	31
3.2.1 Mô hình ARIMA được lựa chọn . . . . .	31
3.2.2 Hiệu quả của mô hình . . . . .	31
3.2.3 Dự báo giá bơ . . . . .	32
3.3 ARIMA với loại bơ hữu cơ . . . . .	32
3.3.1 Mô hình được lựa chọn . . . . .	32
3.3.2 Hiệu quả mô hình . . . . .	33
3.3.3 Kết quả dự báo . . . . .	33
3.3.4 Hạn chế . . . . .	33

**TÀI LIỆU THAM KHẢO**

**35**

# Danh sách hình vẽ

1.1	Giá bơ mỗi loại . . . . .	8
1.2	Giá bơ theo thời gian . . . . .	9
1.3	Giá trung bình của bơ theo thời gian . . . . .	10
1.4	Phân phối giá 2 loại bơ trung bình theo năm . . . . .	11
1.5	Giá trung bình của 2 loại bơ theo tháng . . . . .	12
1.6	Biến động giá bơ . . . . .	13
1.7	Biến động giá bơ theo độ lệch chuẩn . . . . .	14
1.8	Chu kỳ mùa vụ cả 2 loại bơ . . . . .	15
1.9	Biểu đồ thay đổi giá bơ thông dụng . . . . .	16
1.10	Thay đổi giá bơ hữu cơ . . . . .	17
1.11	Giá bơ theo mùa . . . . .	18
1.12	Mối quan hệ giữa khối lượng và giá . . . . .	19
1.13	Sản lượng thị trường của bơ thường . . . . .	20
1.14	Sản lượng thị trường của bơ hữu cơ . . . . .	21
2.1	Giá trung bình của 2 loại bơ theo năm tháng . . . . .	22
2.2	Biểu đồ Polar cho 2 loại bơ . . . . .	23
2.3	Biểu đồ chuỗi phụ theo tháng . . . . .	24
2.4	Biểu đồ tương quan . . . . .	26
2.5	Biểu đồ tự tương quan hàng tuần . . . . .	27
3.1	Điểm khác biệt của bơ loại thường . . . . .	28
3.2	Phần dư của ARIMA . . . . .	30
3.3	Dự đoán với bơ loại thường . . . . .	31
3.4	Dự đoán với bơ loại hữu cơ . . . . .	32

# TỔNG QUAN VỀ ĐỀ TÀI

Bơ Hass đã trở thành một phần không thể thiếu trong chế độ ăn của nhiều người, đặc biệt là giới trẻ. Sự phổ biến của bơ không chỉ gắn liền với giá trị dinh dưỡng mà còn mang tính biểu tượng cho lối sống lành mạnh và hiện đại. Tuy nhiên, cùng với sự yêu thích ngày càng tăng, giá bơ cũng biến động đáng kể, đặc biệt ở những năm gần đây.

Ở Mỹ, việc tiêu thụ bơ đã có ảnh hưởng lớn đến các xu hướng kinh tế, từ giá cả đến cung ứng. Câu chuyện hài hước nhưng đầy ẩn ý về "thế hệ trẻ không mua được nhà vì chi tiêu quá nhiều cho món bánh mì bơ" đã làm nổi bật tầm quan trọng của việc hiểu rõ hơn về giá và lượng cung cấp của bơ trên thị trường. Từ đó, việc phân tích các mẫu biến động giá bơ có thể mang lại những giá trị hữu ích trong việc dự đoán xu hướng tiêu dùng và đưa ra các chiến lược kinh doanh phù hợp.

Mục tiêu của bài phân tích này là khám phá các yếu tố ảnh hưởng đến giá bơ và nhận diện các mẫu biến động theo thời gian và khu vực. Cụ thể: phân tích sự khác biệt về giá giữa bơ thông thường và bơ hữu cơ; khám phá các mẫu biến động giá theo mùa; xác định mối quan hệ giữa lượng cung và giá; đưa ra những nhận định giúp người tiêu dùng và nhà bán lẻ đưa ra quyết định phù hợp hơn về thời điểm và địa điểm mua/bán bơ.

Dữ liệu được sử dụng trong bài phân tích này được cung cấp bởi Hass Avocado Board, tập trung vào giá và lượng bơ được bán tại các khu vực trên toàn nước Mỹ từ năm 2015 đến năm 2018. Phạm vi phân tích bao gồm: Phân loại bơ, thời gian, khu vực.

Phân tích này không chỉ giúp người tiêu dùng nhận biết thời điểm và khu vực mua bơ với giá rẻ nhất mà còn hỗ trợ các doanh nghiệp tối ưu hóa chiến lược kinh doanh, quản lý chuỗi cung ứng và dự đoán nhu cầu thị trường. Với việc nhận diện các mẫu biến động theo mùa và khu vực, bài viết này còn có thể trở thành nền tảng cho các nghiên cứu tương lai về chuỗi cung ứng nông sản và hành vi tiêu dùng.

# Chương 1

## Cơ sở nghiên cứu giá bơ

### 1.1 Tổng quan về dữ liệu

#### 1.1.1 Nguồn dữ liệu

Dữ liệu sử dụng trong bài phân tích được cung cấp bởi Hass Avocado Board, một tổ chức cung cấp dữ liệu quét bán lẻ trực tiếp từ các máy tính tiền tại các kênh phân phối như siêu thị, câu lạc bộ, cửa hàng tiện lợi, hiệu thuốc và các kênh bán lẻ quân đội; chuyên theo dõi và báo cáo dữ liệu về lượng tiêu thụ và giá bơ Hass. Dữ liệu được tải về vào tháng 5 năm 2018 và bao gồm thông tin chi tiết về giá trung bình, khối lượng tiêu thụ, và các mã sản phẩm PLU (Product Lookup Unit) cho từng loại bơ.

#### 1.1.2 Các biến chính trong bộ dữ liệu

Bộ dữ liệu bao gồm nhiều trường thông tin quan trọng, trong đó các biến chính có ý nghĩa cụ thể như sau:

- **Date (Ngày):** Ngày quan sát dữ liệu, được lưu dưới dạng tuần (weekly data). Biến này giúp theo dõi xu hướng giá và lượng qua các thời điểm.
- **AveragePrice (Giá trung bình):** Giá trung bình của một quả bơ (tính theo đơn vị). Đây là biến chính dùng để phân tích và so sánh giá giữa các loại bơ, khu vực, và thời gian.
- **Type (Loại bơ):** Gồm hai loại chính:
  - *Conventional (Bơ thông thường):* Loại bơ phổ biến nhất.
  - *Organic (Bơ hữu cơ):* Loại bơ được trồng và canh tác theo phương pháp hữu cơ, thường có giá cao hơn.
- **Year (Năm):** Năm diễn ra quan sát, giúp xác định các xu hướng dài hạn hoặc sự khác biệt theo từng năm.
- **Region (Khu vực):** Tên thành phố hoặc khu vực ghi nhận dữ liệu. Dữ liệu bao gồm 54 khu vực, từ

các thành phố lớn như New York, Los Angeles, đến các vùng ít phổ biến hơn.

- **Total Volume (Tổng lượng tiêu thụ):** Số lượng bơ được bán ra (tính theo đơn vị). Biến này đại diện cho cung trên thị trường.
- **PLU Codes (Mã sản phẩm):**
  - 4046: Số lượng bơ loại PLU 4046 được bán.
  - 4225: Số lượng bơ loại PLU 4225 được bán.
  - 4770: Số lượng bơ loại PLU 4770 được bán.
- **Total Bags (Tổng túi):** Số lượng bơ được bán theo túi.
- **Small Bags (Túi nhỏ):** Số lượng túi nhỏ được bán.
- **Large Bags (Túi lớn):** Số lượng túi lớn được bán.
- **XLarge Bags (Túi rất lớn):** Số lượng túi rất lớn được bán.

### 1.1.3 Phạm vi dữ liệu

Thời gian: Dữ liệu bao gồm các quan sát từ năm 2015 đến năm 2018, giúp phân tích các xu hướng dài hạn cũng như các biến động ngắn hạn theo mùa.

Không gian: Dữ liệu được thu thập từ 54 khu vực trên toàn nước Mỹ, bao gồm các thị trường lớn như New York, Chicago, và Los Angeles, cũng như các khu vực ít được biết đến hơn như Albany hoặc Boise.

### 1.1.4 Đặc điểm nổi bật của dữ liệu

Dữ liệu rất chi tiết, bao gồm cả loại bơ, thời gian, và khu vực, cho phép thực hiện các phân tích phức tạp như so sánh giữa các loại bơ hoặc dự đoán theo khu vực.

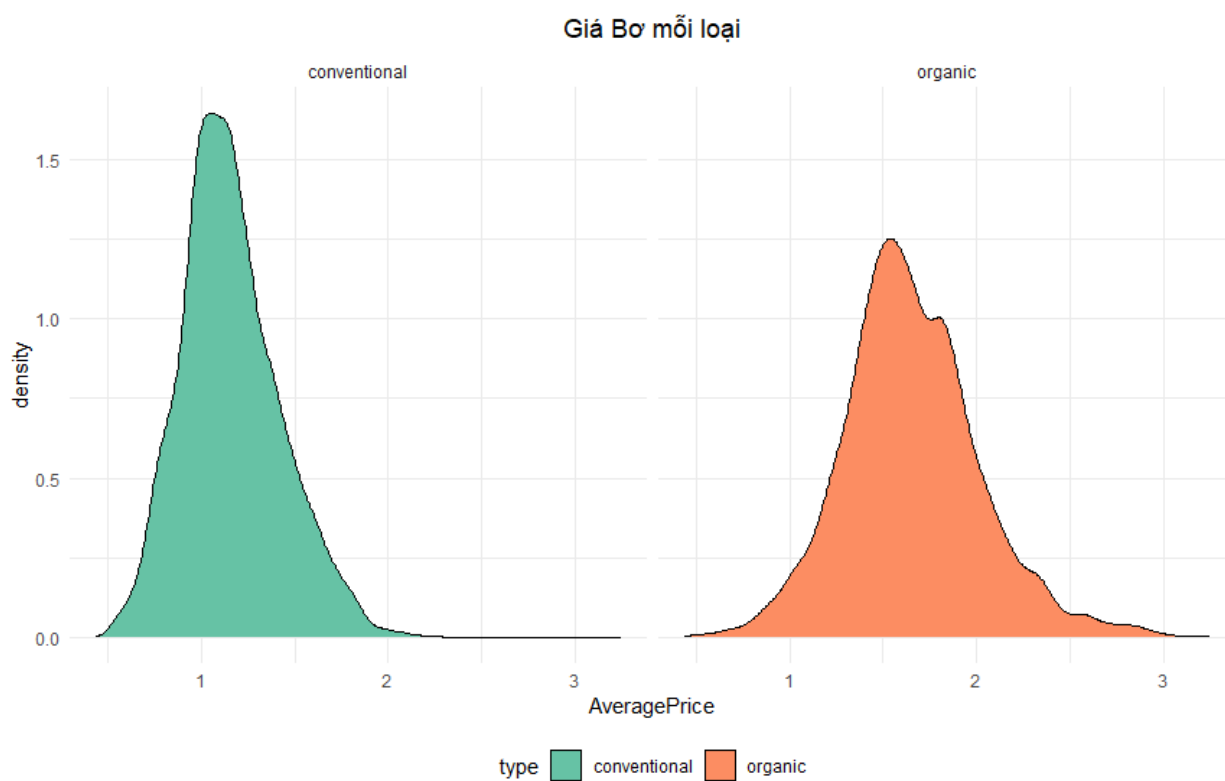
Dữ liệu phân chia hai loại bơ (thông thường và hữu cơ), giúp dễ dàng nhận diện sự khác biệt về giá và lượng tiêu thụ.

Dữ liệu được thu thập từ các kênh bán lẻ thực tế, đảm bảo tính chính xác và độ tin cậy.



## 1.2 Phân tích dữ liệu

### 1.2.1 So sánh 2 loại bơ



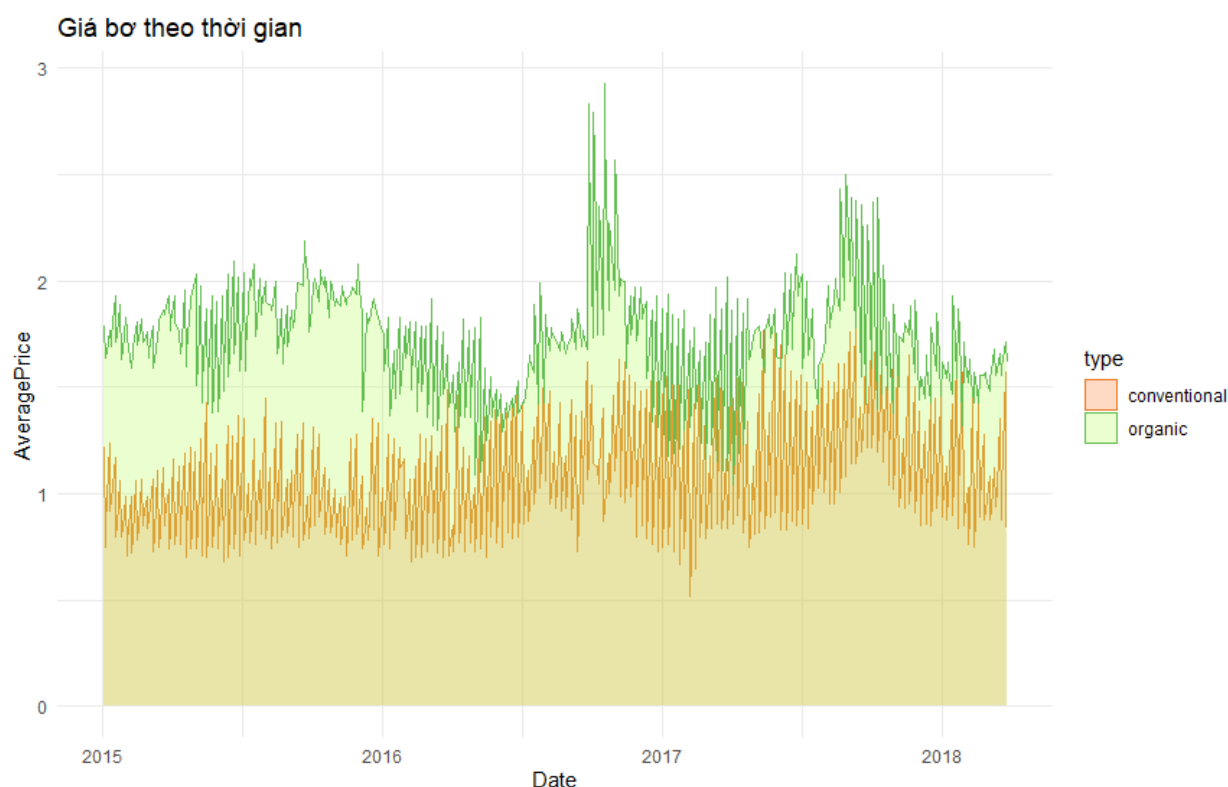
Hình 1.1: Giá bơ mỗi loại

Bơ Conventional: Giá ổn định (khoảng \$1.00), chiếm 97.19% thị phần với khối lượng trung bình 1,653,212.90 đơn vị.

Bơ conventional thống trị thị trường nhờ hiệu quả kinh tế và khả năng tiếp cận cao.

Bơ Organic: Giá dao động (\$1.00 - \$3.00), chỉ chiếm 2.81% thị phần với khối lượng 47,811.21 đơn vị.

Bơ organic có giá trị cao, phù hợp với xu hướng tiêu dùng bền vững, dù thị phần còn nhỏ.



Hình 1.2: Giá bơ theo thời gian

Bơ organic luôn có giá trung bình cao hơn bơ conventional, dao động từ \$1.5 - \$3. Trong khi đó, giá bơ conventional ổn định hơn, nằm trong khoảng \$1 - \$1.5.

Xu hướng chung:

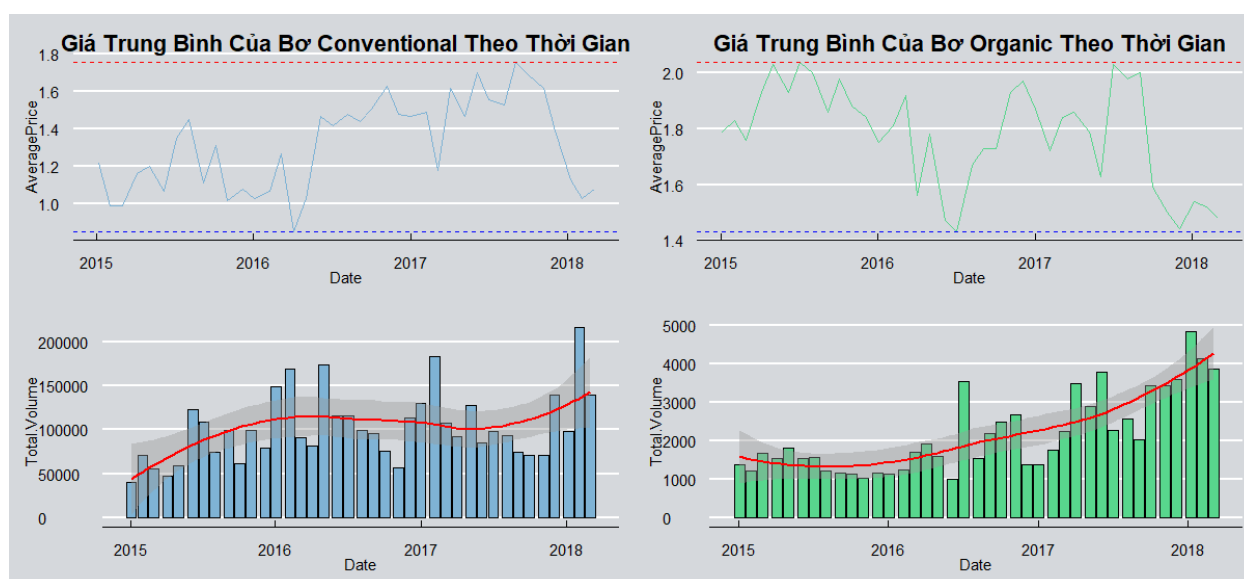
- Bơ organic thể hiện sự biến động mạnh hơn, phản ánh tính nhạy cảm với thị trường hoặc nhu cầu khách hàng cao cấp.
- Bơ conventional giữ giá ổn định, phù hợp với nhu cầu tiêu dùng phổ thông.

Giai đoạn 2016-2017 có nhiều biến động mạnh hơn, đặc biệt với giá bơ organic, cho thấy tác động từ thị trường hoặc yếu tố mùa vụ.

Khách hàng chọn bơ organic phải chấp nhận rủi ro giá cao và biến động, trong khi bơ conventional phù hợp với sự ổn định và hiệu quả kinh tế.

### 1.2.2 Mối quan hệ giữa giá và tổng khối lượng

Thông thường, có một mối quan hệ nghịch đảo giữa nguồn cung và giá cả. Khi có sự sản xuất dư thừa bơ, điều này sẽ tác động tiêu cực đến giá bơ trên thị trường. Hãy xem liệu điều này có đúng với cả bơ conventional và organic hay không.



Hình 1.3: Giá trung bình của bơ theo thời gian

Bơ Conventional: Vào cuối năm 2017, chúng ta có thể thấy giá bơ giảm mạnh, đồng thời lượng bơ trên thị trường tăng lên đáng kể.

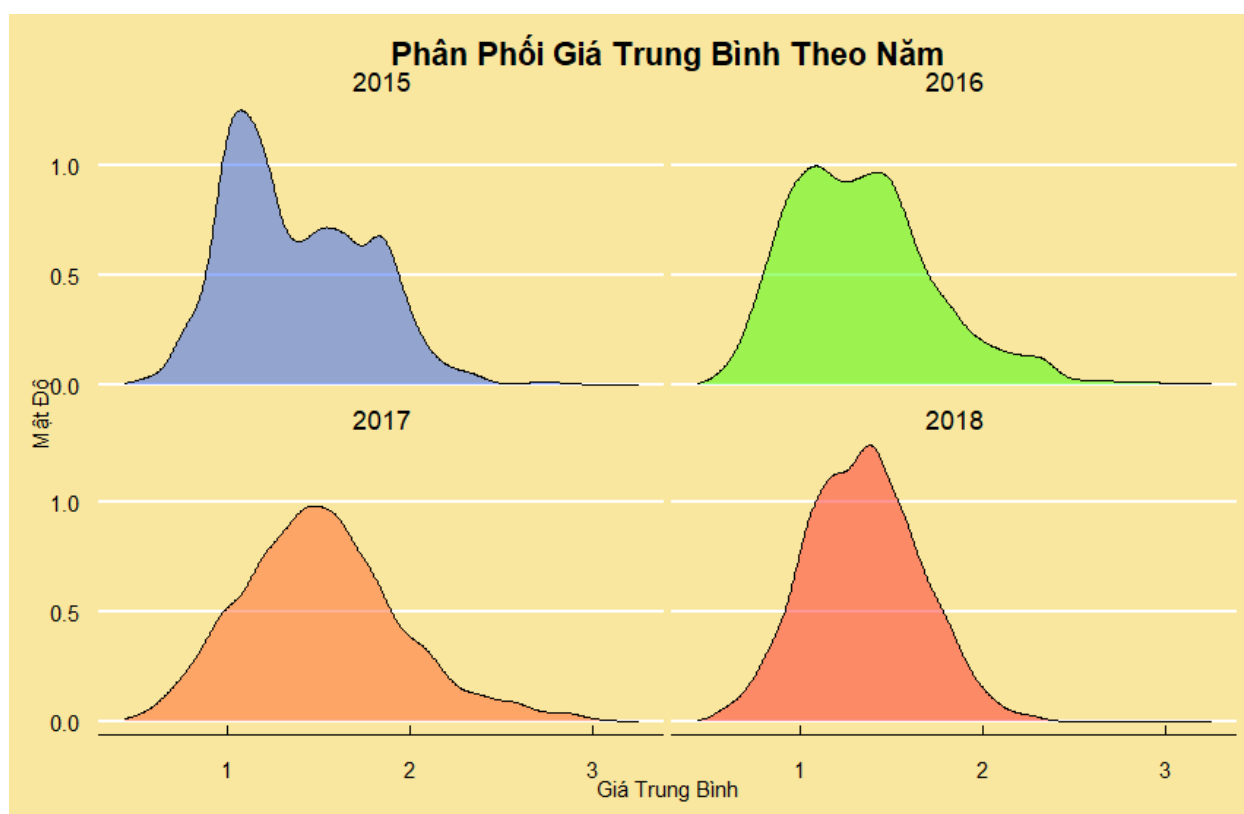
Bơ Organic: Tương tự, đối với bơ organic, vào cuối năm 2017 cũng xảy ra hiện tượng giá giảm mạnh và một sự gia tăng lớn về khối lượng.

Đỉnh khối lượng: Lưu ý rằng mỗi khi khối lượng đạt đỉnh, đó là dấu hiệu cho một đợt giảm giá bơ sắp tới.

### 1.2.3 Các kiểu mẫu

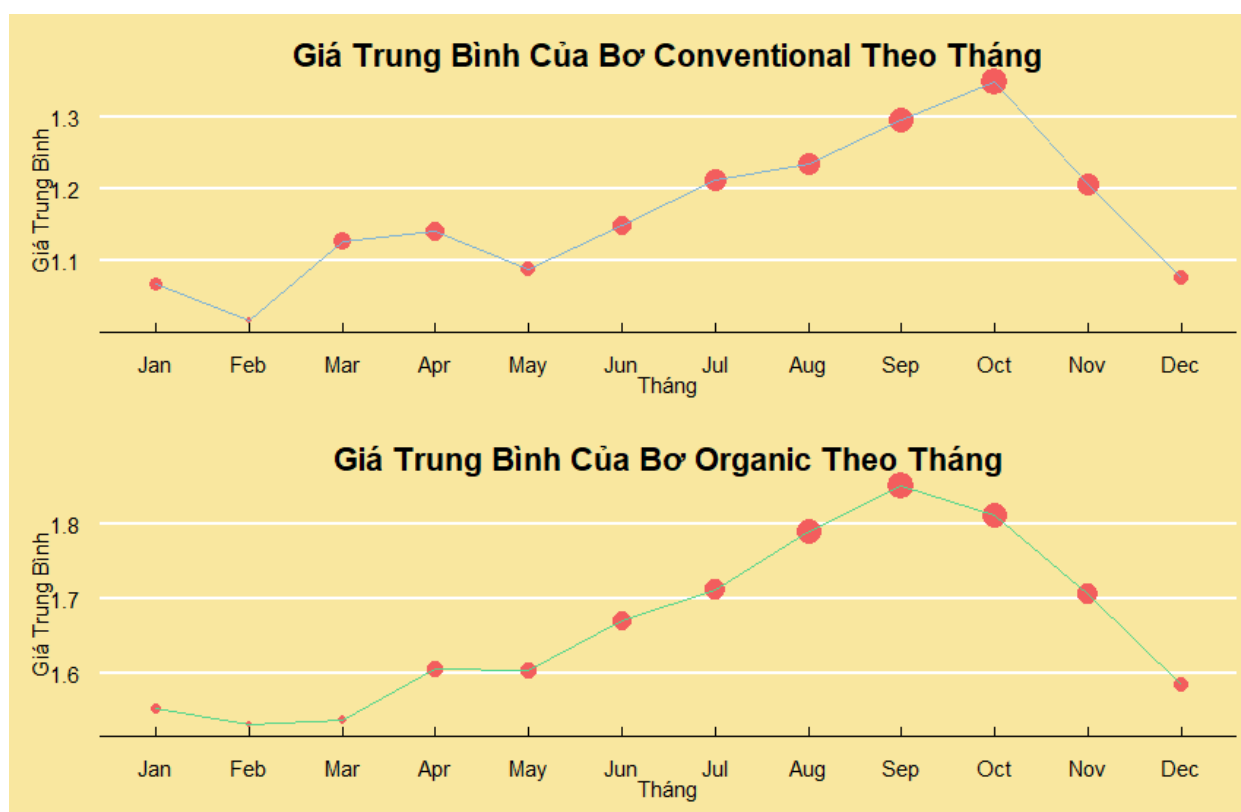
#### a, Mẫu theo mùa

Trong phần này, chúng ta sẽ cố gắng tìm hiểu xem có bất kỳ mô hình theo mùa đáng kể nào lặp lại hay không. Điều này có nghĩa là có những xu hướng lặp lại nào mà giá bơ có xu hướng tăng hay không. Ví dụ, vào tháng 5 hàng năm, giá bơ thường tăng vì một lý do cụ thể nào đó.



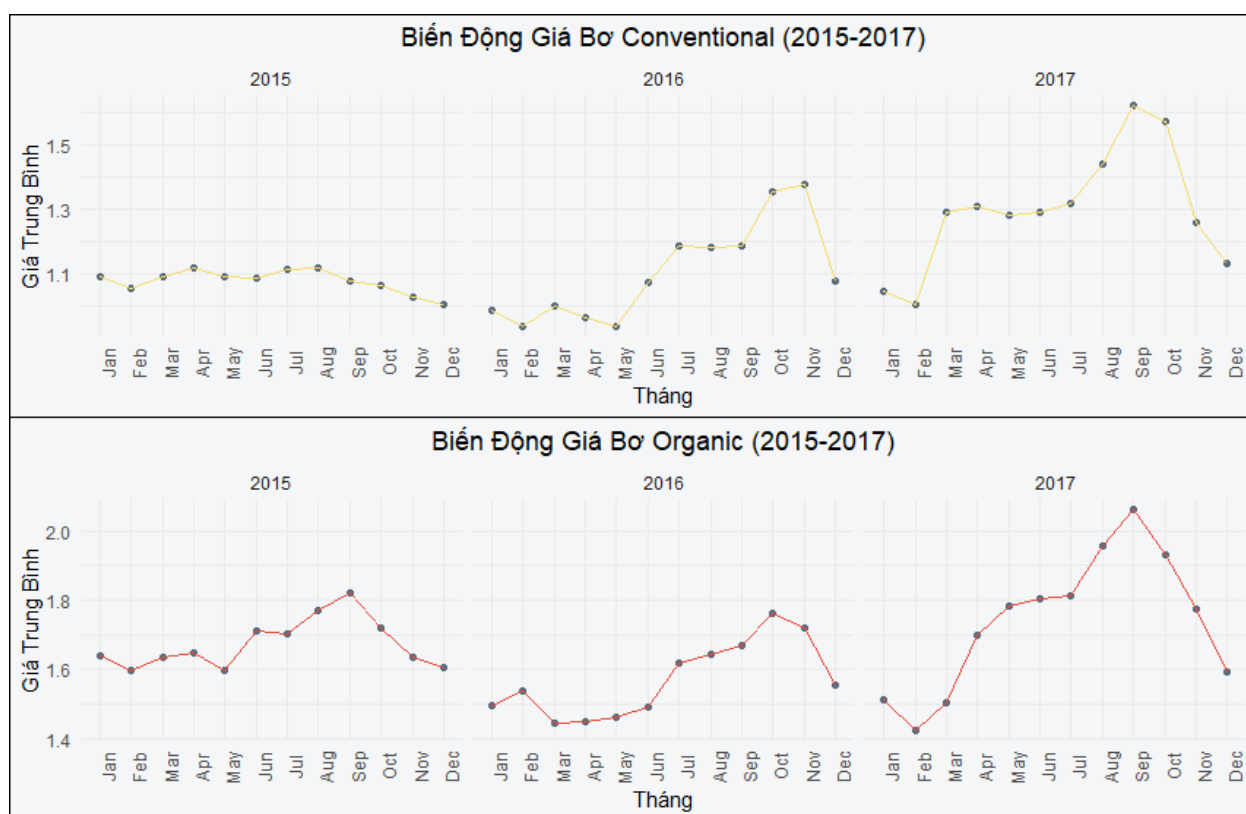
Hình 1.4: Phân phối giá 2 loại bơ trung bình theo năm

Có vẻ như hầu hết giá trong năm 2015 đối với bơ conventional đều nằm trong khoảng \$1.00. Trong khi đó, vào các năm 2016 và 2017, mật độ giá cao hơn một chút.



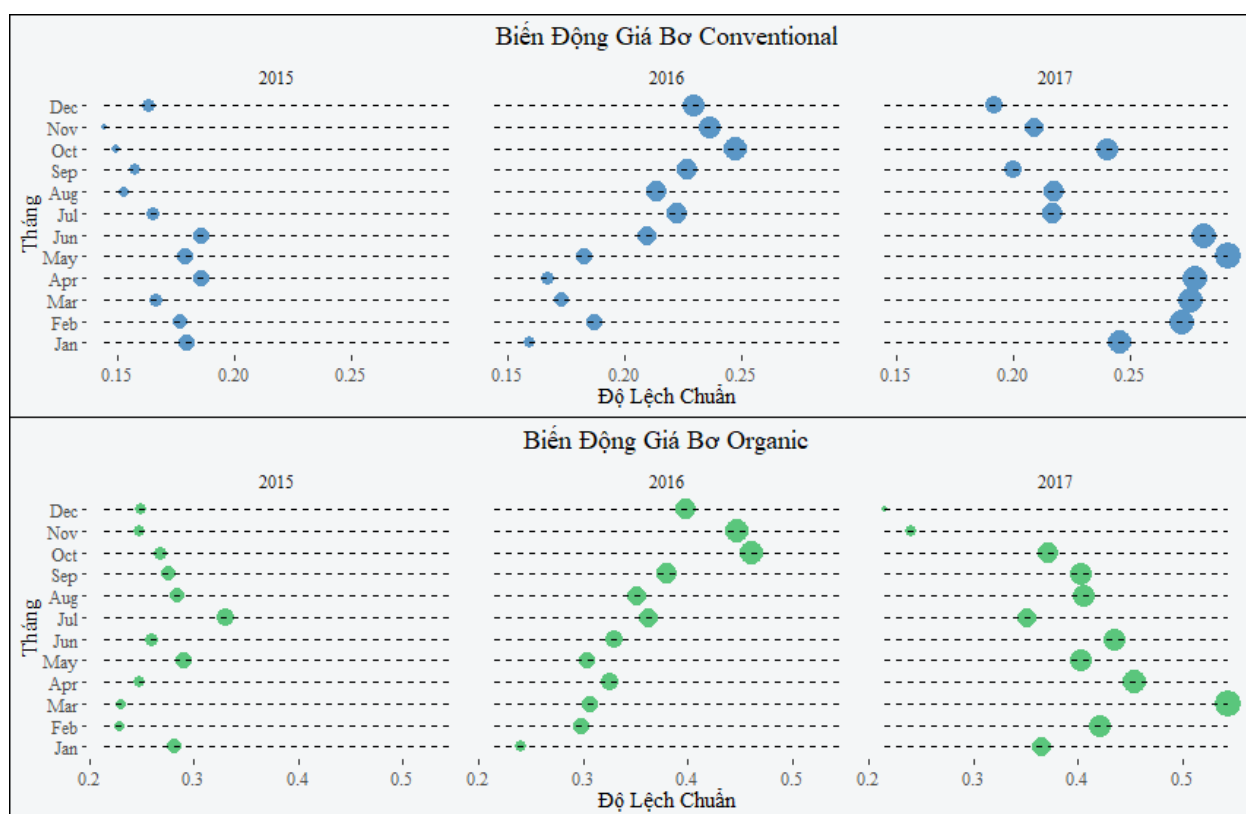
Hình 1.5: Giá trung bình của 2 loại bơ theo tháng

Có vẻ như hầu hết các đỉnh giá của cả bơ conventional và organic đều xuất hiện trong khoảng tháng 9 và tháng 10.



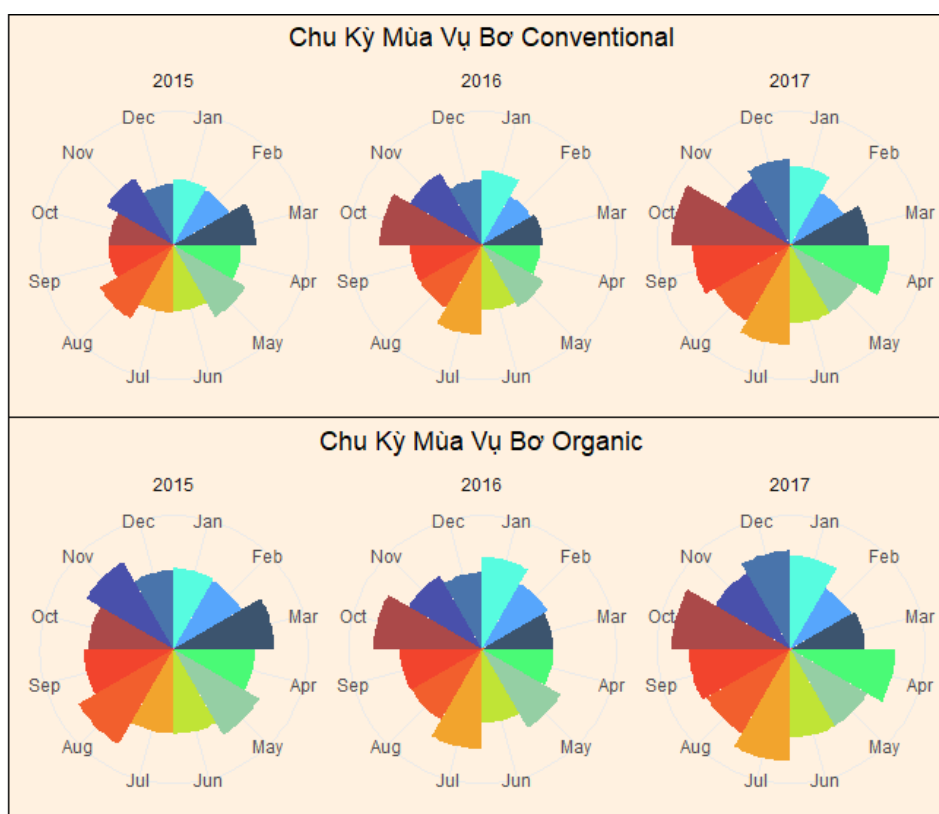
Hình 1.6: Biến động giá bơ

Vào cuối năm giá bơ giảm mạnh. Lý do nào có thể khiến giá giảm vào cuối năm như vậy?



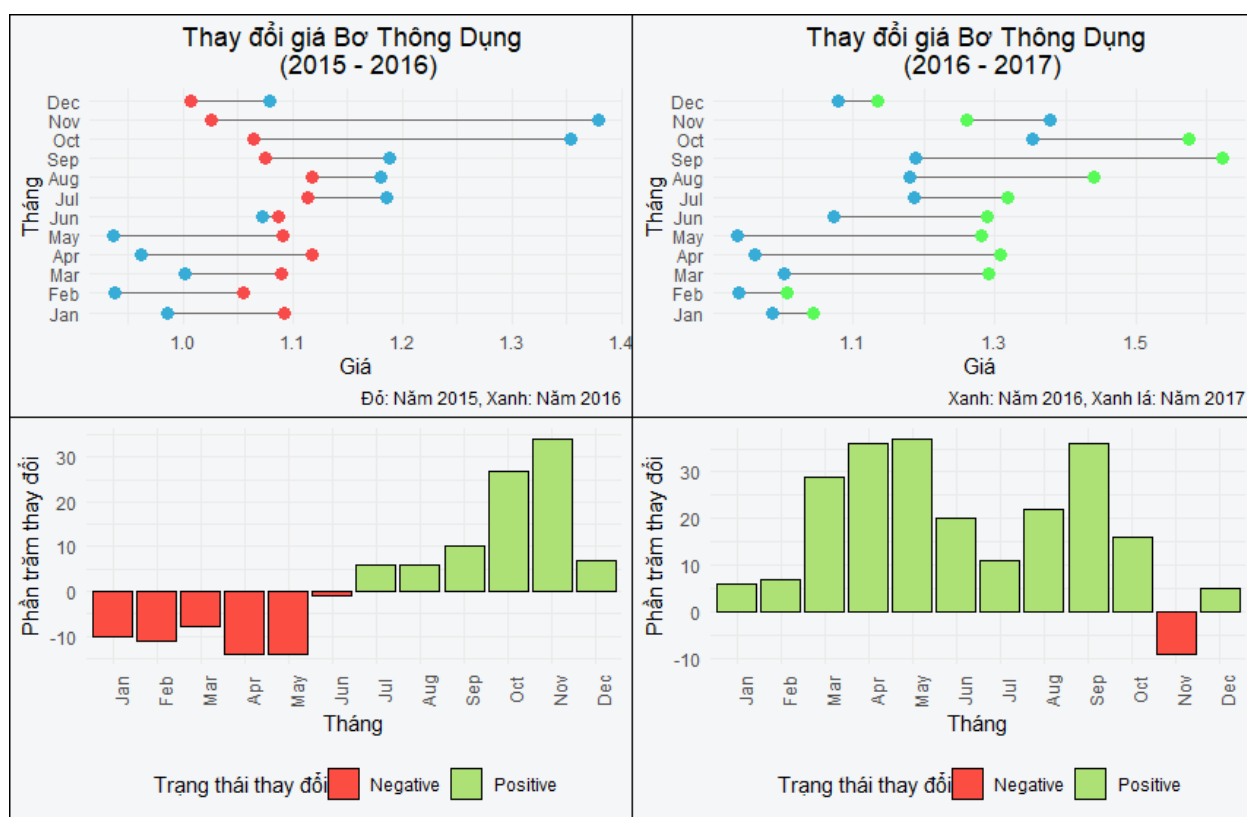
Hình 1.7: Biến động giá bơ theo độ lệch chuẩn

Độ lệch chuẩn chỉ đơn giản là căn bậc hai của phương sai. Chúng ta có thể thấy rằng trong năm 2017, thị trường bơ đã trải qua sự biến động cao nhất đối với cả bơ conventional và organic.

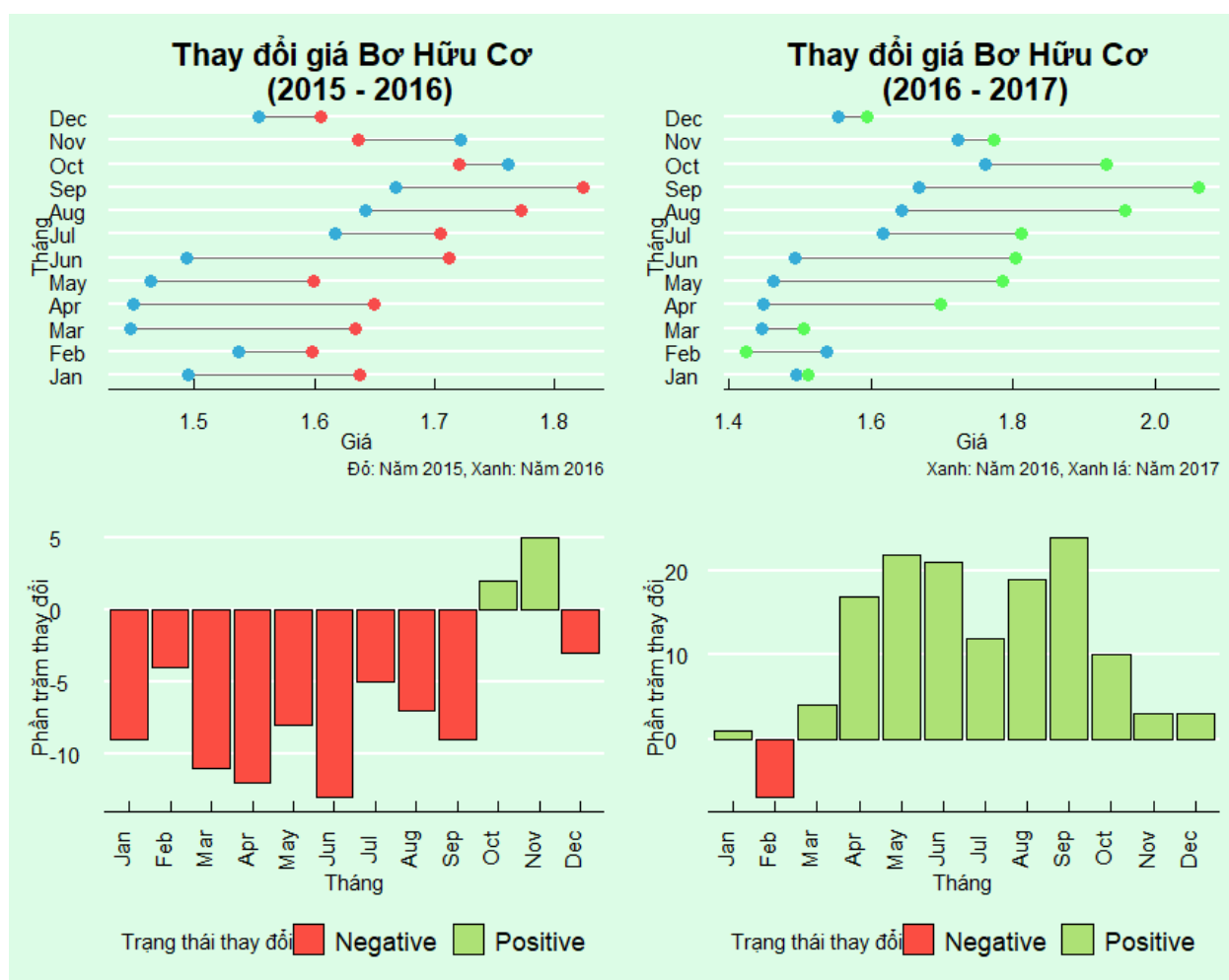


Hình 1.8: Chu kỳ mùa vụ của cả 2 loại bơ





Hình 1.9: Biểu đồ thay đổi giá bơ thông dụng



Hình 1.10: Thay đổi giá bơ hữu cơ

Sự khác biệt phần trăm đối với bơ conventional: Ở đây, chúng ta muốn xem liệu có sự gia tăng phần trăm qua các năm về giá trung bình mỗi tháng của từng loại bơ hay không. Từ năm 2015 đến 2016, giá bơ conventional bị chậm lại trong sáu tháng đầu năm, sau đó tăng mạnh trong sáu tháng cuối năm. Từ năm 2016 đến 2017, giá bơ conventional đã thể hiện rất tốt trong năm 2017 so với năm 2016.

Sự khác biệt phần trăm đối với bơ organic: Đối với bơ organic thì câu chuyện hoàn toàn khác, ít nhất là từ năm 2015 đến 2016. Năm 2016, bơ organic có hiệu suất kém hơn gần như toàn bộ các tháng so với năm 2015. Sau đó, nó thể hiện tốt hơn trong năm 2017, gần như ở tất cả các tháng so với năm 2016.

Năm 2017 là một năm tốt cho bơ: Dựa trên xu hướng này, chúng ta có thể thấy rằng năm 2017 là một năm tốt cho bơ. Liệu dự báo của chúng ta (sẽ thực hiện ở phần sau của dự án) có dự đoán rằng xu hướng này sẽ tiếp tục vào năm 2018 không?

Các công thức cần nắm được:

- Tỷ lệ phần trăm tăng trưởng:

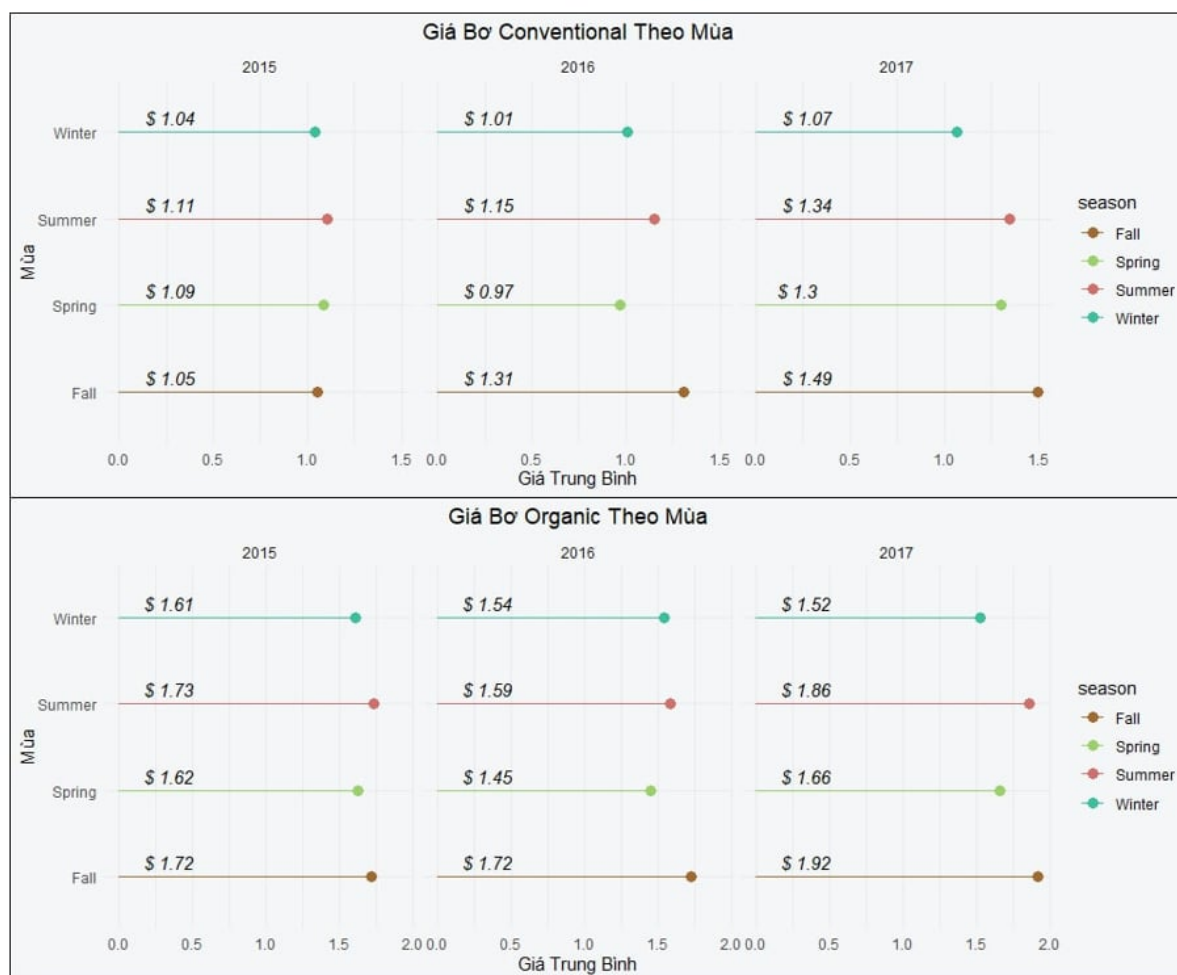
$$\text{Tỷ lệ phần trăm tăng trưởng} = \frac{\text{Giá trị Kết thúc} - \text{Giá trị Bắt đầu}}{\text{Giá trị Bắt đầu}}$$

- Độ lệch chuẩn:

$$\text{Độ lệch chuẩn} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

- Trong đó:

- $x$ : Giá trị thực tế trong tập dữ liệu.
- $\bar{x}$ : Giá trị trung bình của tập dữ liệu.
- $n$ : Số lượng giá trị trong tập dữ liệu.



Hình 1.11: Giá bơ theo mùa

Bơ Organic luôn có giá trung bình cao hơn đáng kể so với bơ Conventional trong cùng mùa.

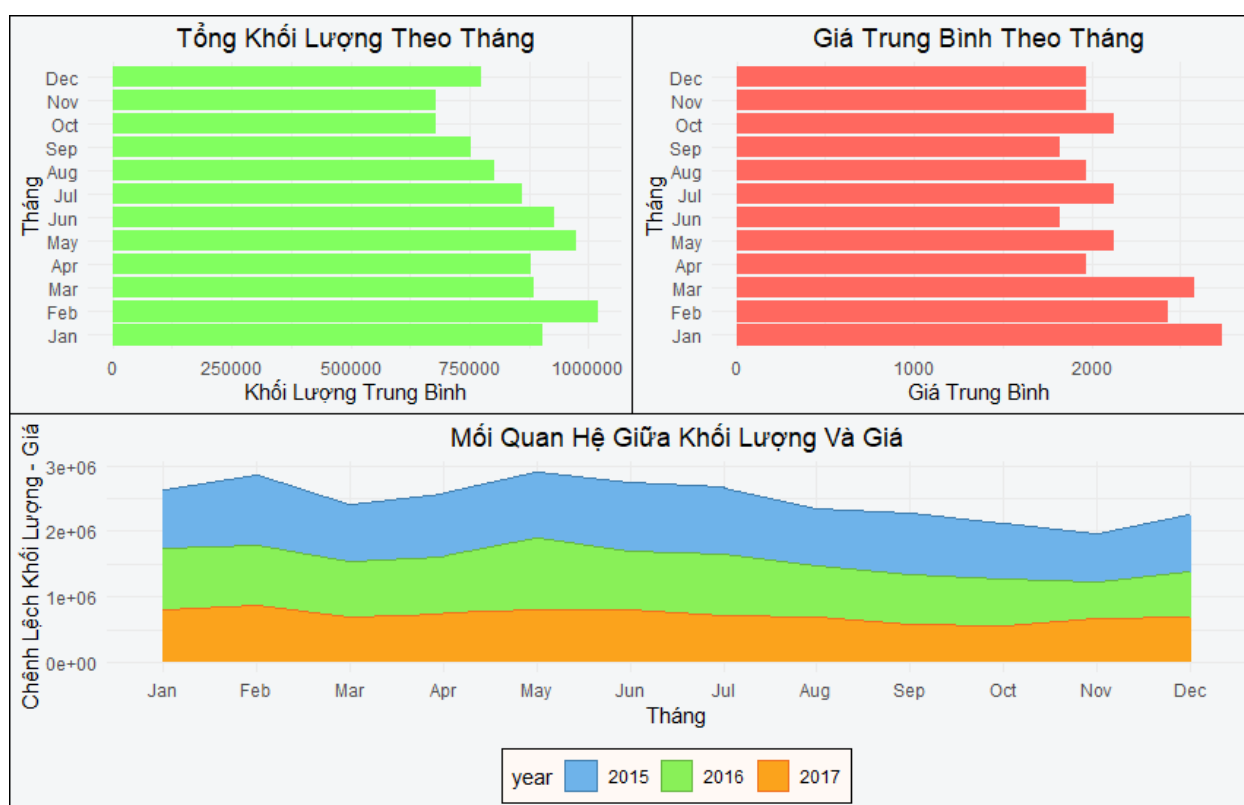
Sự chênh lệch lớn nhất xuất hiện vào mùa thu, khi giá bơ Organic vượt xa giá bơ Conventional.

Mùa xuân là thời điểm có giá trung bình thấp nhất cho cả hai loại bơ, có thể do nguồn cung dồi dào hoặc nhu cầu giảm.

### b, Mẫu theo khu vực

Chúng ta sẽ phân tích hai yếu tố chính là khối lượng bơ và giá của các loại bơ khác nhau ở mỗi khu vực. Chúng ta sẽ thấy mối tương quan nghịch đảo giữa cung và cầu và cách nó ảnh hưởng đến giá. Vì chúng ta có 54 khu vực, chúng ta sẽ chỉ sử dụng một vài khu vực để ghi chú mối tương quan nghịch đảo này.

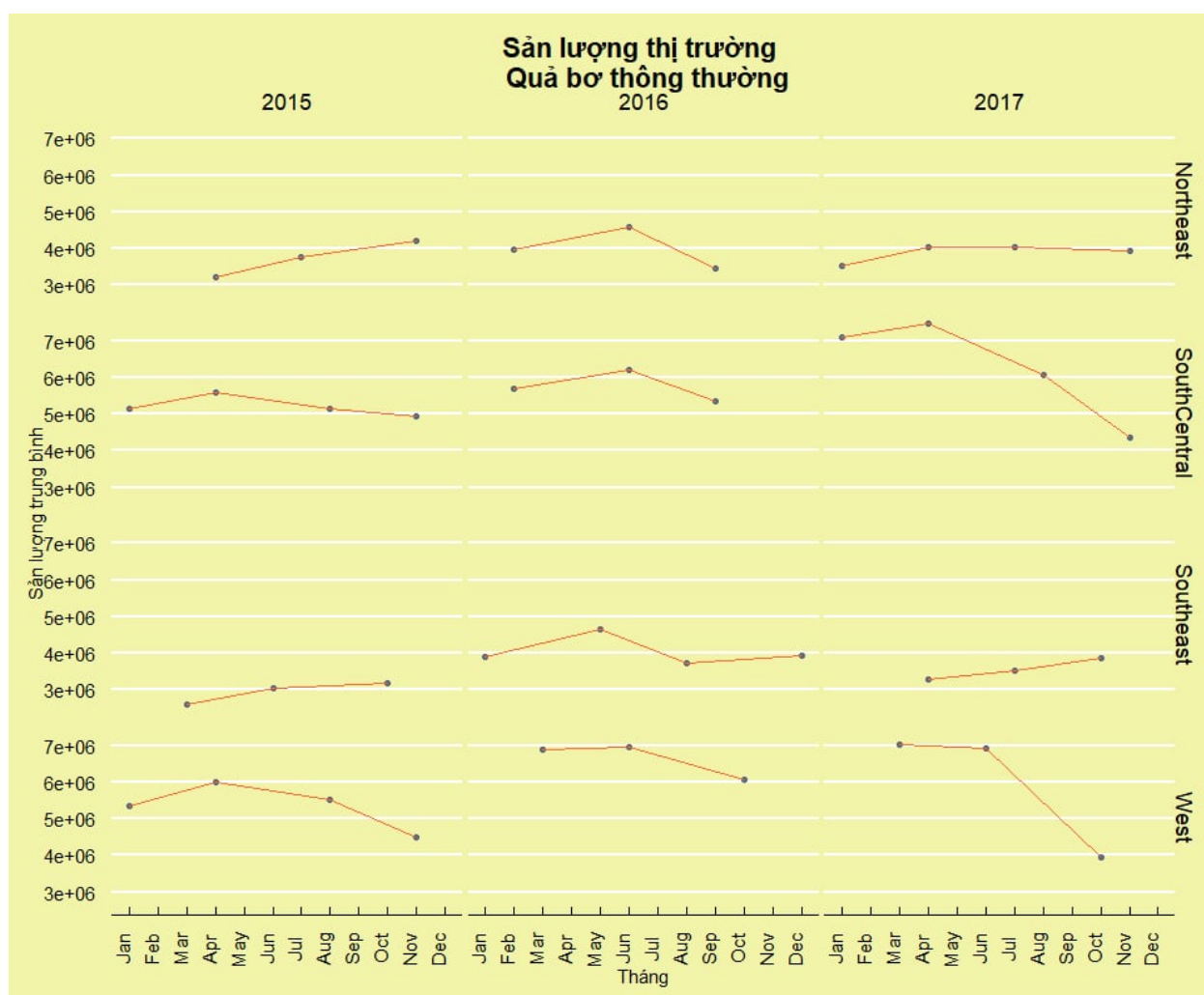
Chênh lệch Khối lượng - Giá = Tổng Khối lượng Bơ Sản xuất – Giá Bơ Trung bình



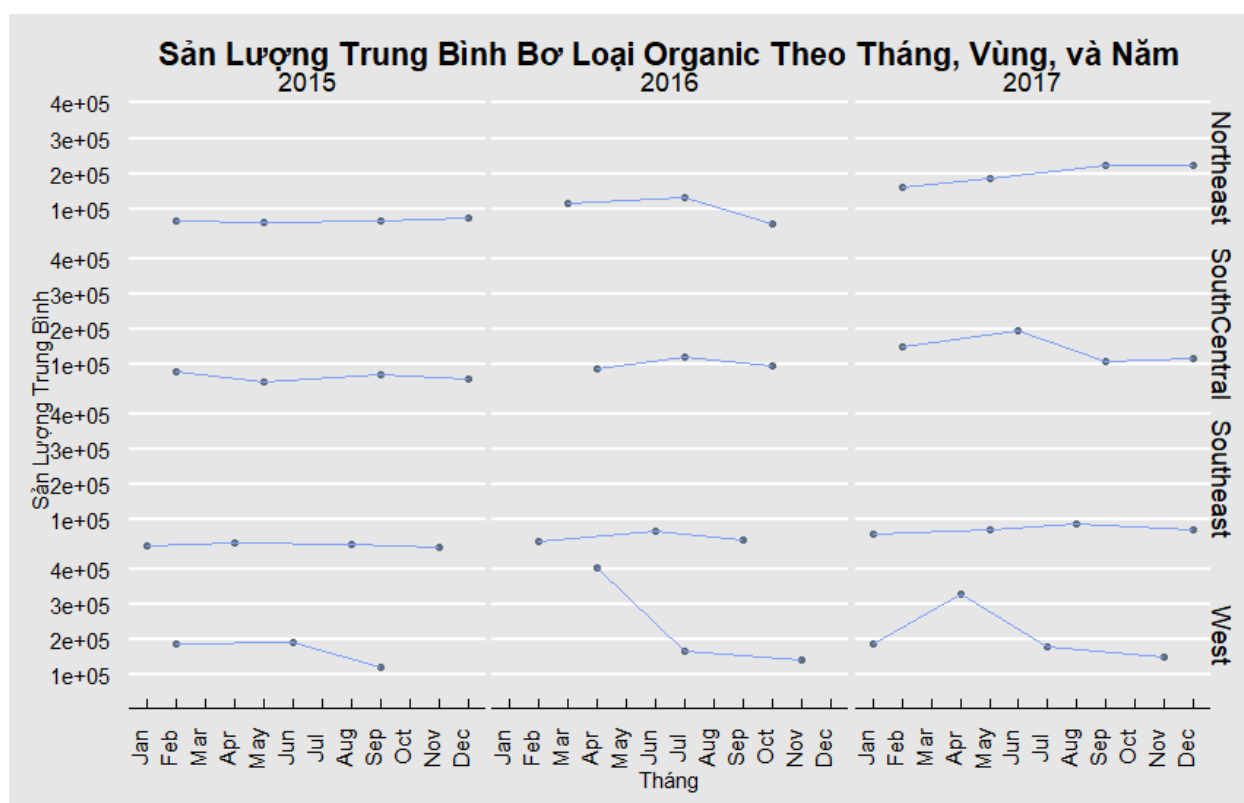
Hình 1.12: Mối quan hệ giữa khối lượng và giá

Tháng 2 có khối lượng trung bình cao nhất so với 2 tháng 1 và 3, cho thấy sự tăng mạnh nhu cầu vào đầu năm. Nhưng ngược lại giá trung bình thì tháng 2 lại hệt hơn so với 2 tháng đó.

Sự chênh lệch giữa giá và khối lượng cho thấy ảnh hưởng của cung-cầu, khi khối lượng tăng, giá có xu hướng giảm.



Hình 1.13: Sản lượng thị trường của bơ thường



Hình 1.14: Sản lượng thị trường của bơ hữu cơ

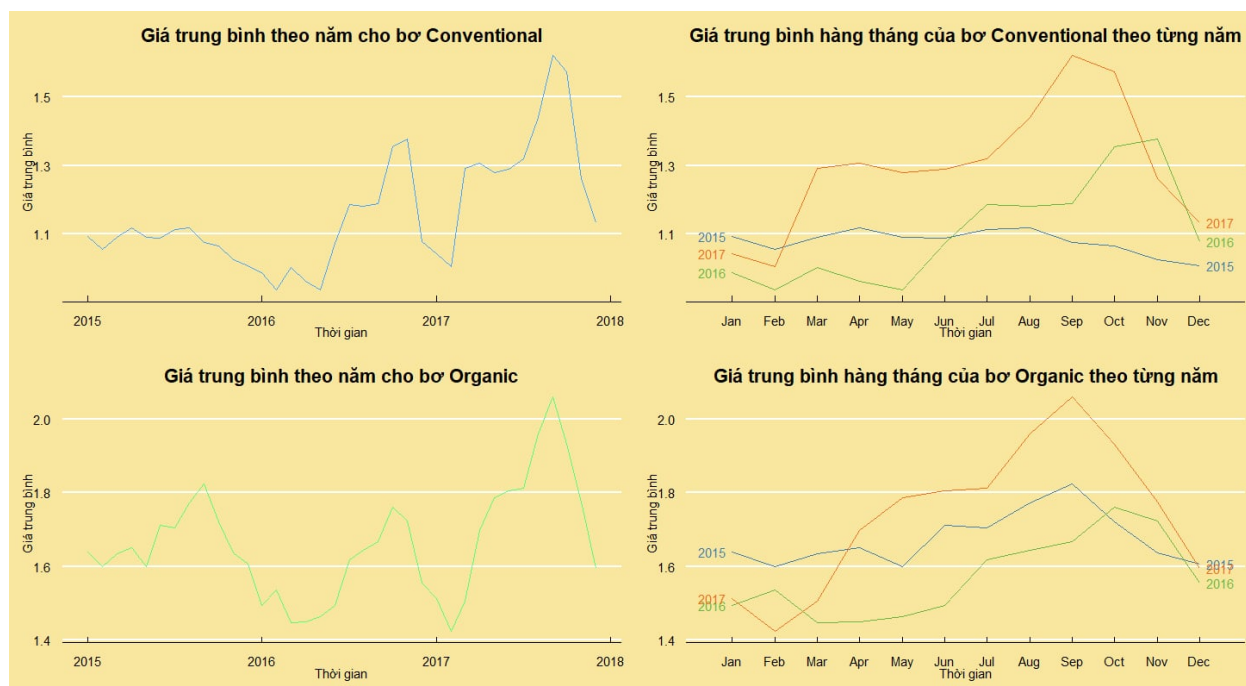
Sản lượng ở West không ổn định và xu hướng giảm ở cả 2 loại bơ vào giữa năm đến cuối năm

## Chương 2

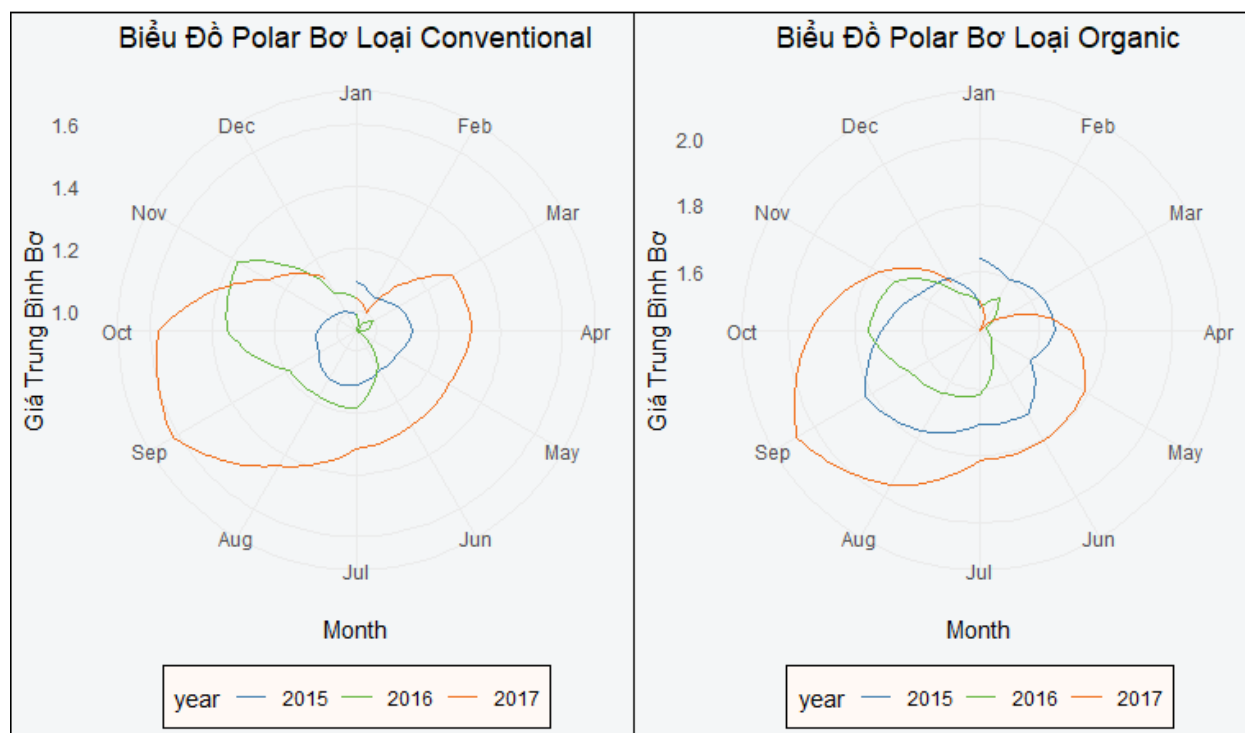
# Phân tích chuỗi thời gian

### 2.1 Phân tích dữ liệu theo năm tháng

Sử dụng thư viện Autoplot cho Chuỗi thời gian. Thư viện autoplot cho phép chúng ta thấy các mẫu giữa các năm khác nhau.



Hình 2.1: Giá trung bình của 2 loại bở theo năm tháng



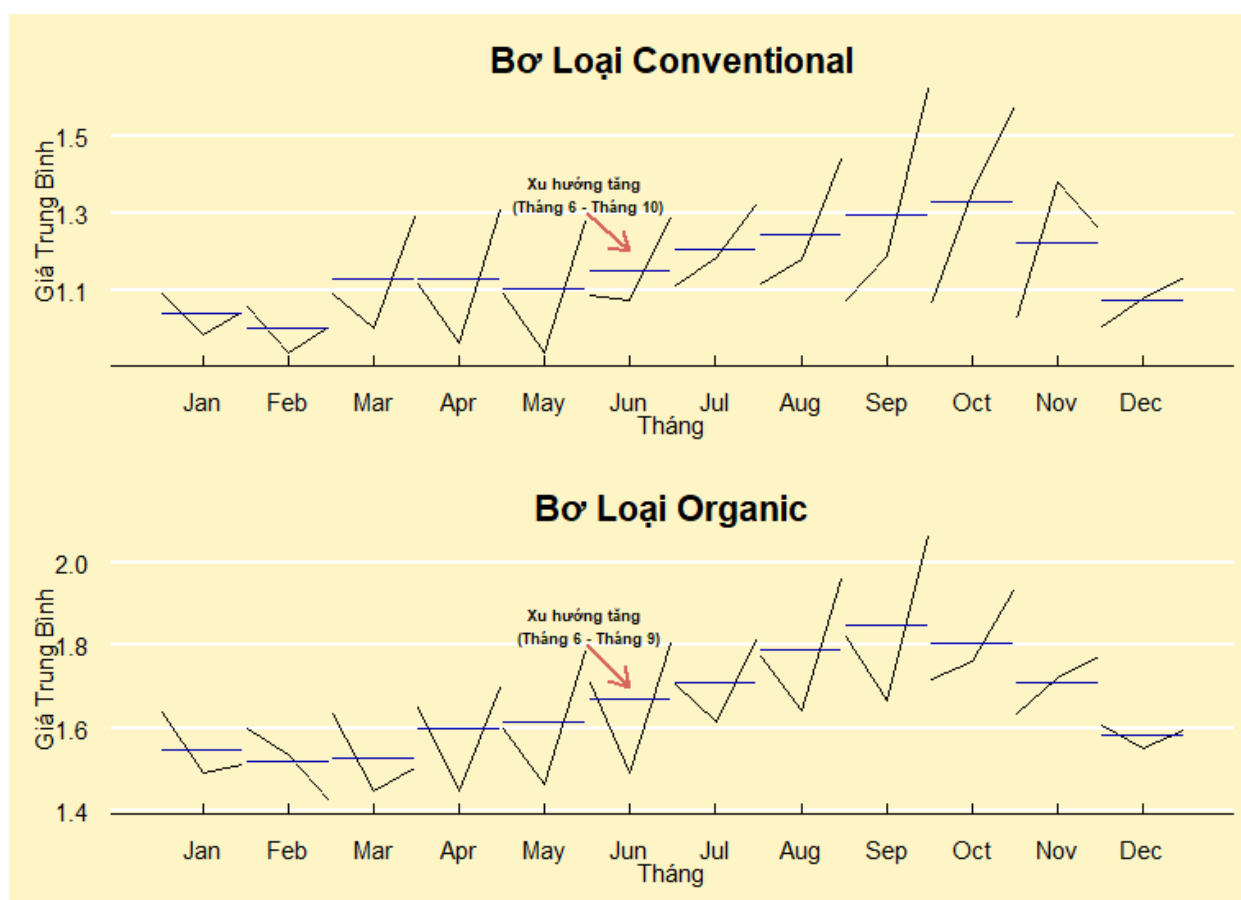
Hình 2.2: Biểu đồ Polar cho 2 loại bơ

Giá năm 2017 là cao nhất nói chung.

Giá bơ tăng đối với cả loại hữu cơ và thông thường.

Giá thấp nhất xảy ra vào năm 2015





Hình 2.3: Biểu đồ chuỗi phụ theo tháng

Đây là một dạng khác để phân tích các xu hướng khác nhau theo mùa.

Đường màu xanh: Là giá trị trung bình của tháng bơ trong một tháng cụ thể.

Đường màu đen: Đây là cách để xem các biến động trong suốt mỗi tháng.

Nhận xét:

Xu hướng tăng: Xu hướng tăng đáng kể nhất xảy ra giữa tháng 6 - tháng 9 đối với bơ hữu cơ và tháng 6 - tháng 10 đối với các loại bơ thông thường. Cả hai loại bơ đều có xu hướng tương tự.

Những tháng đắt nhất để mua bơ: Giống như trong phân tích trước, thông qua hình ảnh trực quan này, chúng tôi xác nhận rằng những tháng có giá trung bình cao nhất là tháng 9 và tháng 10, tuy nhiên, tháng 8 cũng khá đắt.

Biểu đồ này cho chúng ta biết điều gì? Đối với mỗi tháng, chúng ta có ba đỉnh. Điều đó biểu thị đỉnh cao nhất trong mỗi tháng của mỗi năm. Vì chúng ta đang sử dụng giá của năm 2015, 2016 và 2017 nên chúng ta có thể thấy giá của mỗi năm diễn biến như thế nào trong cả mười hai tháng. Điều này sẽ giúp chúng ta xem liệu có bất kỳ mô hình theo mùa chính nào không. Ngoài ra, chúng ta có thể thấy mức thay đổi trung bình

diễn ra như thế nào, cho phép chúng ta biết liệu có bất kỳ xu hướng chính nào trong mỗi tháng hay không.

## 2.2 Chuỗi thời gian cố định

Biểu đồ tự tương quan cho chúng ta biết điều gì?

Trước hết, chúng ta phải hiểu khái niệm độ trễ. Vậy độ trễ là gì? Hãy coi độ trễ là các khoảng thời gian. Trong trường hợp này, chúng ta coi độ trễ là các khoảng thời gian hàng tháng. Bây giờ mục tiêu chính của tự tương quan là gì? Chúng ta muốn xem liệu có mối quan hệ tuyến tính nào từ độ trễ đầu tiên (tháng) hay không. Vì vậy, chúng ta muốn xem liệu có một số mô hình nhất định nào đó trái ngược với tháng đầu tiên hay không. Tương quan ở độ trễ bằng 0 luôn là 1 vì nó tương quan với chính nó. Ở độ trễ một, chúng ta thấy rằng tương quan gần bằng 1, điều này cho chúng ta biết rằng tương quan với tháng đầu tiên (độ trễ 0) là tương tự, do đó xu hướng có tương quan cao với tháng 1 (tháng 1).

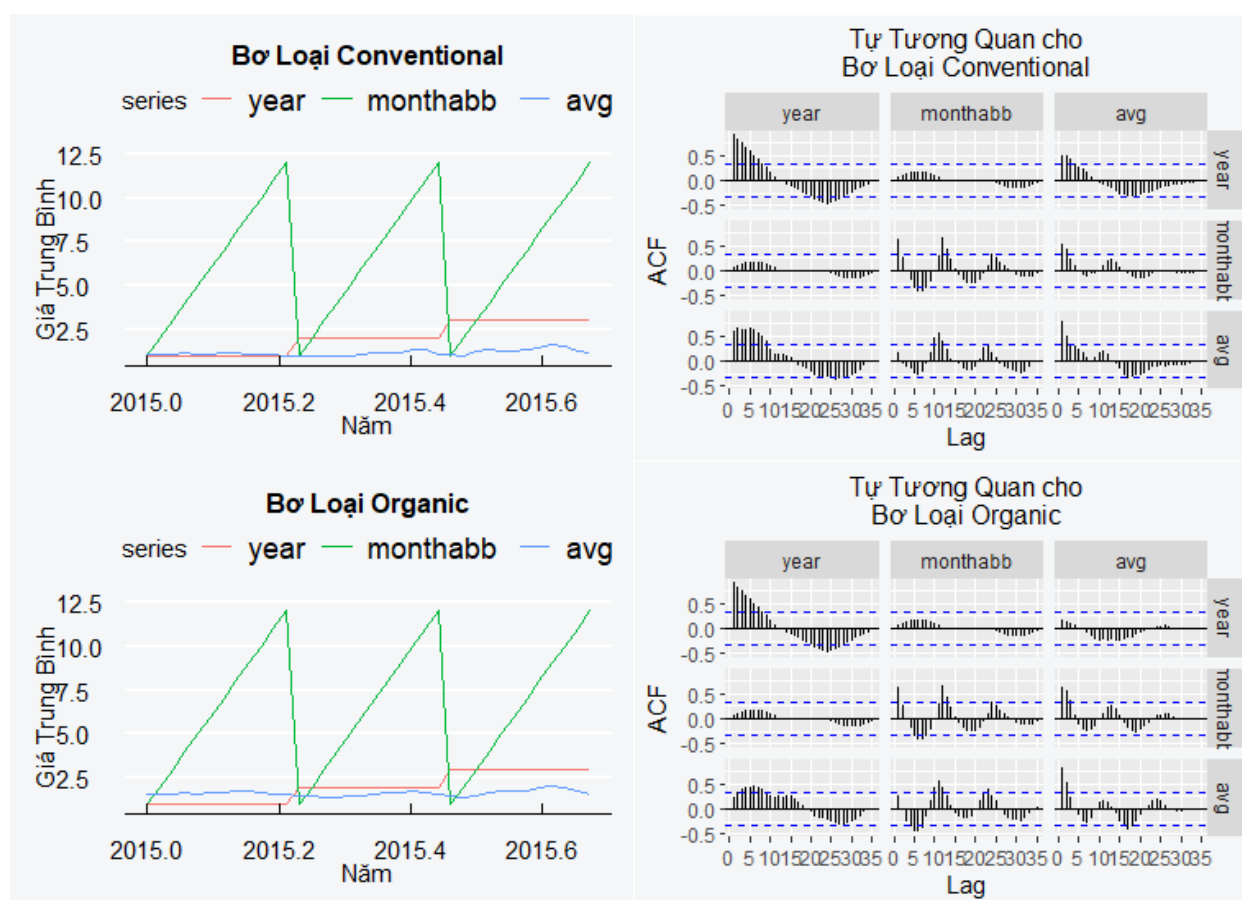
Chúng ta có thể kết luận gì về tự tương quan?

Có một tự tương quan cao cho đến độ trễ thứ ba (tháng 3), điều này cho chúng ta biết rằng có một mối quan hệ tuyến tính cao giữa những tuần này với tháng đầu tiên. Tuy nhiên, chúng ta không thấy mối quan hệ tuyến tính nào giữa các tháng khác so với tháng đầu tiên. Vì vậy, chúng ta có thể kết luận rằng không có mô hình nhất quán nào cho thấy bất kỳ mối quan hệ tuyến tính nào với giá hiện thị trong tuần đầu tiên đối với cả bơ thông thường và bơ hữu cơ.

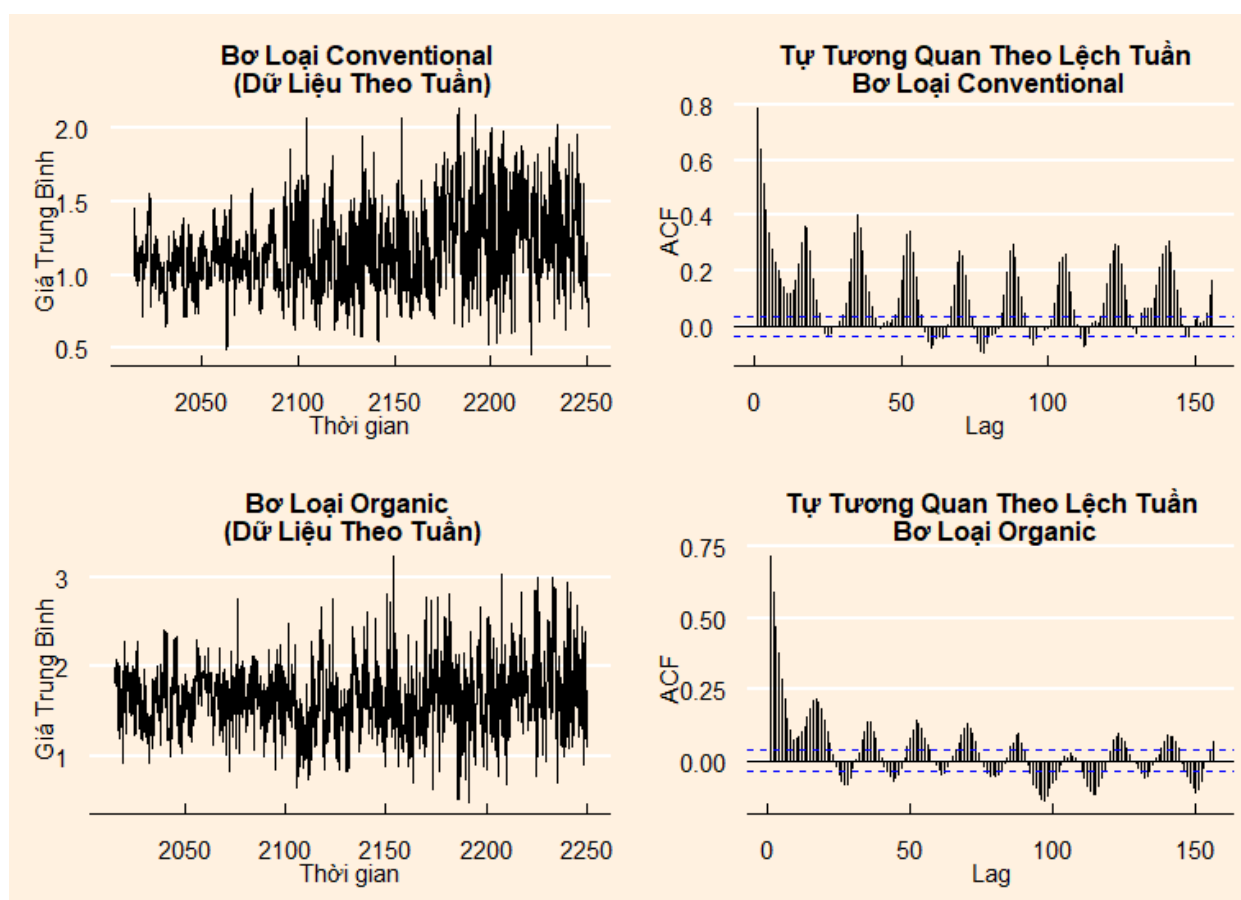
Công thức tự tương quan:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.1)$$

Tự tương quan giúp chúng ta xác định xem có những mẫu nhất định trong dữ liệu của mình hay không. Vậy mục đích chính của việc sử dụng Tự tương quan là gì và làm thế nào nó có thể giúp chúng ta xác định những mẫu nhất định trong tập dữ liệu chuỗi thời gian của mình? Trước hết, từ "Tự động" gợi ý cho chúng ta rằng chúng ta đang có kế hoạch sử dụng tương quan nội bộ. Tương quan nội bộ có nghĩa là gì? Về cơ bản, chúng ta đang lấy một ngày cụ thể (trong ví dụ bên dưới, chúng ta sử dụng tháng 1) và chúng ta muốn xem liệu giá bơ có một số loại tương quan với giá tháng 1 hay không. Trong hình ảnh trực quan đầu tiên, độ trễ là biểu diễn của các tháng. Vì vậy, chúng ta có thể thấy rằng tương quan càng cao khi tháng càng gần tháng 1/ Điều này có nghĩa là giá tháng 2 tương tự như giá bơ của tháng 1. Khi các tháng trôi qua, tương quan trái ngược với tháng 1 trở nên thấp hơn, cho chúng ta biết rằng không có mẫu cụ thể nào đối với tháng 1. Biến động giá trong những tháng đó ngoại trừ tháng 2 và tháng 3 không có giá tương quan cao so với giá tháng 1.



Hình 2.4: Biểu đồ tương quan

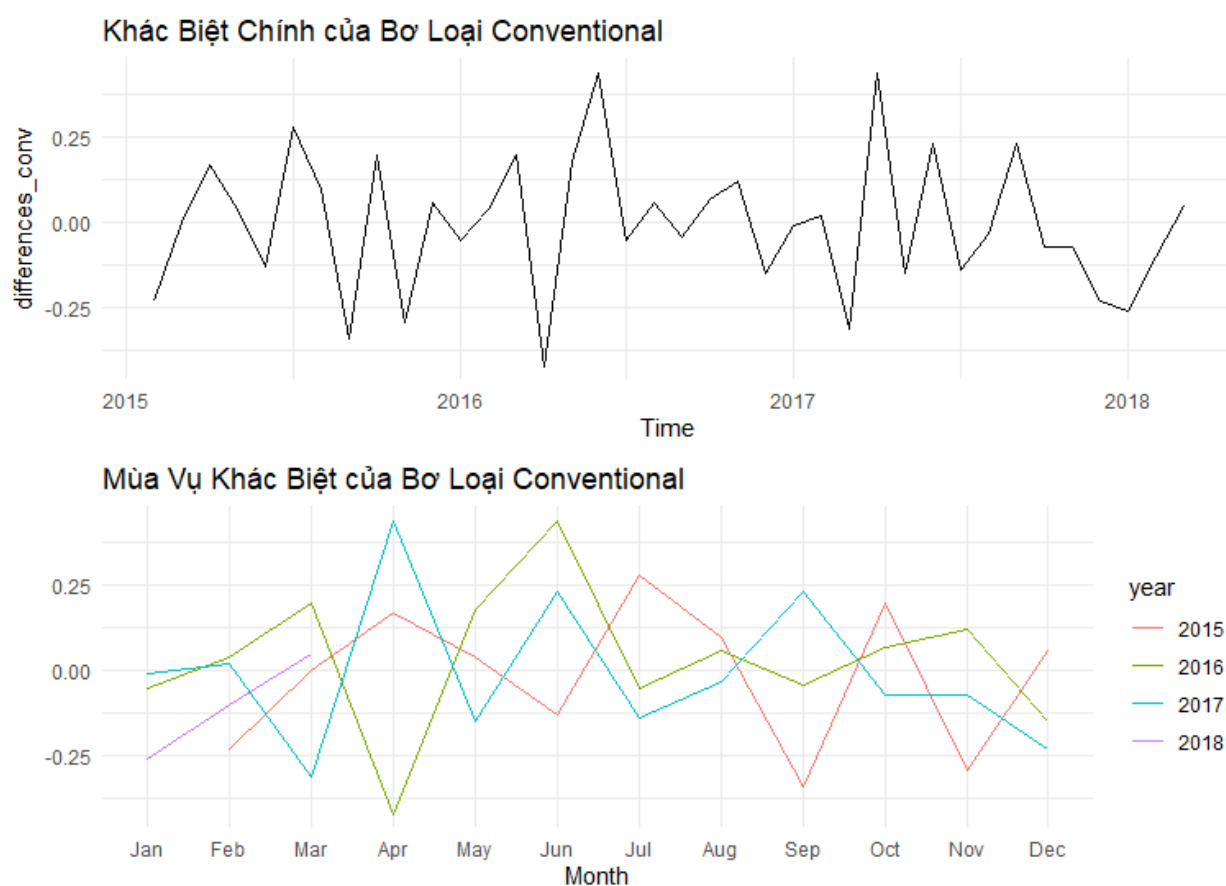


Hình 2.5: Biểu đồ tự tương quan hàng tuần

## Chương 3

# Dự đoán chuỗi thời gian

Trong phần này, chúng ta sẽ sử dụng 2 mô hình để đưa ra những dự đoán chính xác nhằm xem các mô hình sắp tới sẽ như thế nào đối với giá bơ.



Hình 3.1: Điểm khác biệt của bơ loại thường

Cả hai biểu đồ nhấn mạnh rằng giá bơ có sự biến động không chỉ theo thời gian mà còn theo mùa. Thời điểm tháng 3-4 hàng năm và giai đoạn năm 2016-2017 là những thời điểm cần chú ý, có thể do các yếu tố cung cầu hoặc sự kiện thị trường đặc biệt

## 3.1 Phương pháp ARIMA

### 3.1.1 Thử nghiệm mô hình

Để dự báo giá bơ, các mô hình ARIMA với các thông số khác nhau đã được thử nghiệm. Các mô hình được đánh giá dựa trên tiêu chí AIC (Akaike Information Criterion) nhằm chọn ra mô hình có độ phù hợp cao nhất với dữ liệu.

Các mô hình được thử nghiệm bao gồm ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0) với các biến thể theo mùa như (0,1,0)[12], (1,1,0)[12].

Mô hình có giá trị AIC thấp nhất là ARIMA(0,1,1)(1,1,0)[12], được chọn làm mô hình tốt nhất.

### 3.1.2 Mô hình tốt nhất

#### ARIMA(0,1,1)(1,1,0)[12]

Cấu trúc mô hình:

ARIMA(0,1,1): Bao gồm phần sai khác bậc 1 ( $d=1$ ) để dữ liệu trở nên tĩnh và một thành phần MA(1) để bắt nhiễu trong chuỗi thời gian.

(1,1,0)[12]: Thành phần theo mùa với chu kỳ 12 tháng, đại diện cho yếu tố mùa vụ trong dữ liệu.

Thông số mô hình:

MA(1) coefficient: -0.5852 ( $\pm 0.1547$ ).

Seasonal AR(1) coefficient: -0.5753 ( $\pm 0.1753$ ).

sigma<sup>2</sup>: 0.05354, cho thấy mức độ nhiễu thấp trong mô hình

### 3.1.3 Đánh giá mô hình

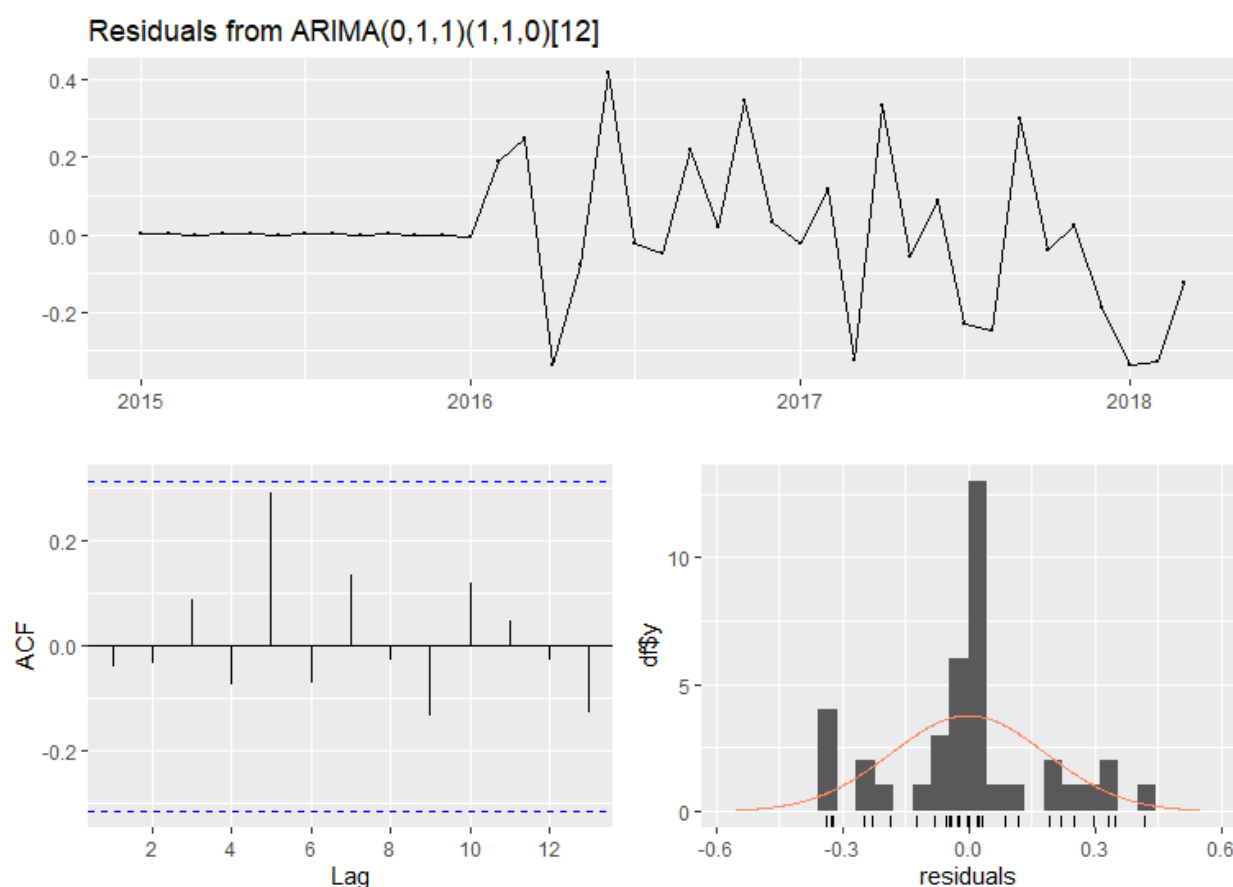
Sai số mô hình:

RMSE (Root Mean Square Error): 0.1815  $\rightarrow$  Sai số bình phương trung bình thấp.

0.1213  $\rightarrow$  Sai số tuyệt đối trung bình nhỏ, chứng minh dự báo chính xác.

MAPE (Mean Absolute Percentage Error): 9.33%  $\rightarrow$  Sai số phần trăm nhỏ hơn 10%, rất tốt cho các dự báo thực tế.

Phần dư (Residuals): Phần dư không có tự tương quan mạnh ( $ACF1 = -0.0403$ ) và kiểm định Ljung-Box ( $p\text{-value} = 0.4139$ ) chỉ ra rằng phần dư là ngẫu nhiên và không chứa thông tin chưa được mô hình hóa.



Hình 3.2: Phần dư của ARIMA

Phần dư không chỉ được sử dụng để hiểu sự khác biệt giữa mô hình dự báo của chúng ta và giá trị thực tế, mà còn cho chúng ta biết về khả năng xảy ra các biến động "bất thường" trong một khoảng thời gian cụ thể. Hãy cùng khám phá các biểu đồ dưới đây cho cả hai loại bơ, conventional và organic. Dưới đây là những phát hiện của tôi:

Bơ Conventional: Có hai đỉnh bất thường trong biểu đồ bơ conventional. Điều này có thể cho thấy các biến động giá bất thường, có khả năng xảy ra do một sự kiện đặc biệt nào đó. Bơ Organic: Chúng ta cũng thấy một số đỉnh bất thường trên biểu đồ bơ organic, nhưng không lớn bằng so với bơ conventional. Tuy nhiên, vẫn có những đỉnh bất thường cần được đánh giá thêm.

Hiểu về mô hình ARIMA:

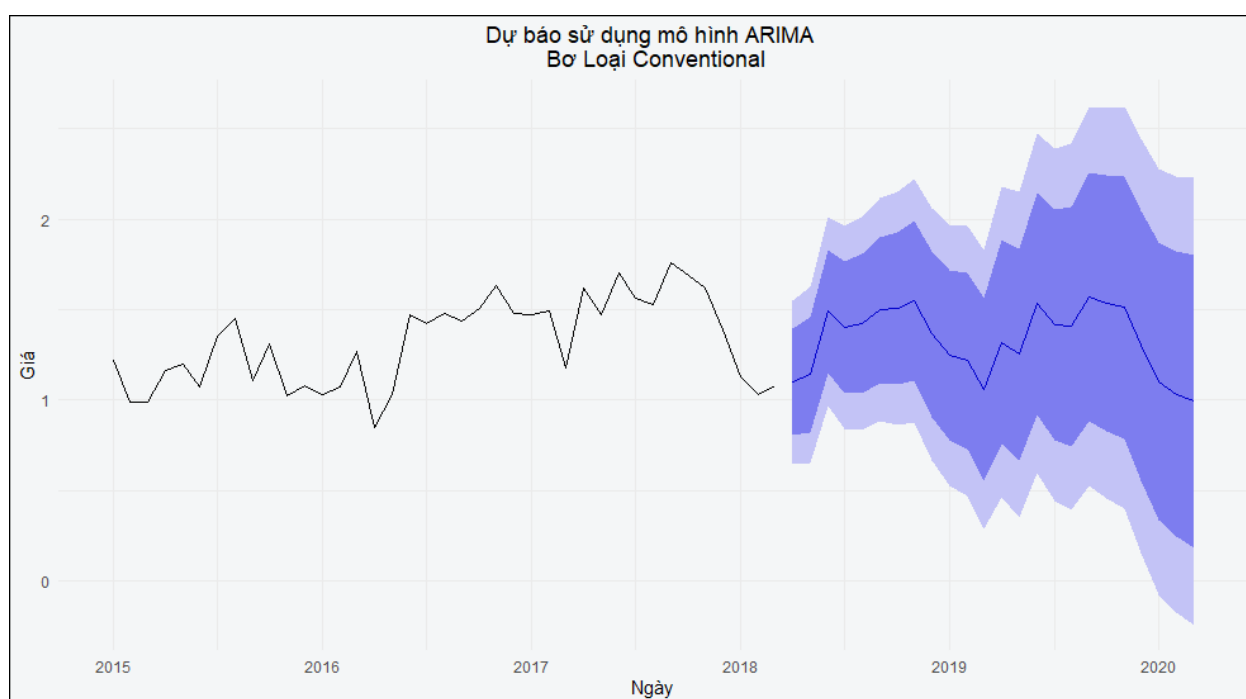
Tự hồi quy (AR - Auto Regressive): Điều này có nghĩa là các giá trị ở thời điểm trước đó có thể ảnh hưởng đến các giá trị hiện tại và tương lai. Mô hình ARIMA sử dụng các quan sát có độ trễ để dự báo các quan sát tương lai. Một trọng số được thêm vào các quan sát trước đó, và trọng số này có thể thay đổi tùy thuộc vào mức độ gần đây của các quan sát trong quá khứ. Quan sát càng gần thời điểm hiện tại thì trọng số càng lớn.

Tích phân (I - Integrated): Nếu có các xu hướng nhất quán trong sự biến động của giá trong quá khứ,

thời chuỗi thời gian có khả năng không tính (non-stationary), nghĩa là tính mùa vụ vẫn tồn tại trong sự biến động của giá. Phần "Integrated" trong mô hình ARIMA giúp loại bỏ tính mùa vụ trong tập dữ liệu, trong trường hợp có các mô hình lặp lại ổn định. Bậc sai phân trong mô hình ARIMA giúp xử lý vấn đề xu hướng và tính mùa vụ.

Trung bình trượt (MA - Moving Average): Phần trung bình trượt giúp loại bỏ ảnh hưởng của các biến động ngẫu nhiên trong giá bơ. Nếu có một sự kiện bất thường dẫn đến việc giá bơ tăng đột biến, trung bình trượt sẽ giúp làm "mượt" chuỗi dữ liệu và giúp mô hình chuỗi thời gian không bị ảnh hưởng bởi các biến động này.

## 3.2 ARIMA với loại bơ thường



Hình 3.3: Dự đoán với bơ loại thường

### 3.2.1 Mô hình ARIMA được lựa chọn

Mô hình  $ARIMA(0,1,1)(1,1,0)[12]$  là mô hình phù hợp nhất để dự báo giá bơ. Mô hình được lựa chọn dựa trên các tiêu chí như AIC, AICc và BIC, đồng thời các kiểm tra dư (Residual) chứng minh rằng mô hình phù hợp với dữ liệu.

### 3.2.2 Hiệu quả của mô hình

Sai số bình phương trung bình (RMSE) là 0.1815 và sai số tuyệt đối trung bình (MAE) là 0.1213, cho thấy mô hình có khả năng dự báo tương đối chính xác.

Sai số trung bình phần trăm (MAPE) là 9.33%, nghĩa là dự báo có sai số tương đối thấp so với giá trị



thực tế.

Kiểm tra Ljung-Box cho thấy phần dư là ngẫu nhiên ( $p\text{-value} = 0.4139$ ), chứng minh rằng mô hình không bỏ sót yếu tố quan trọng nào.

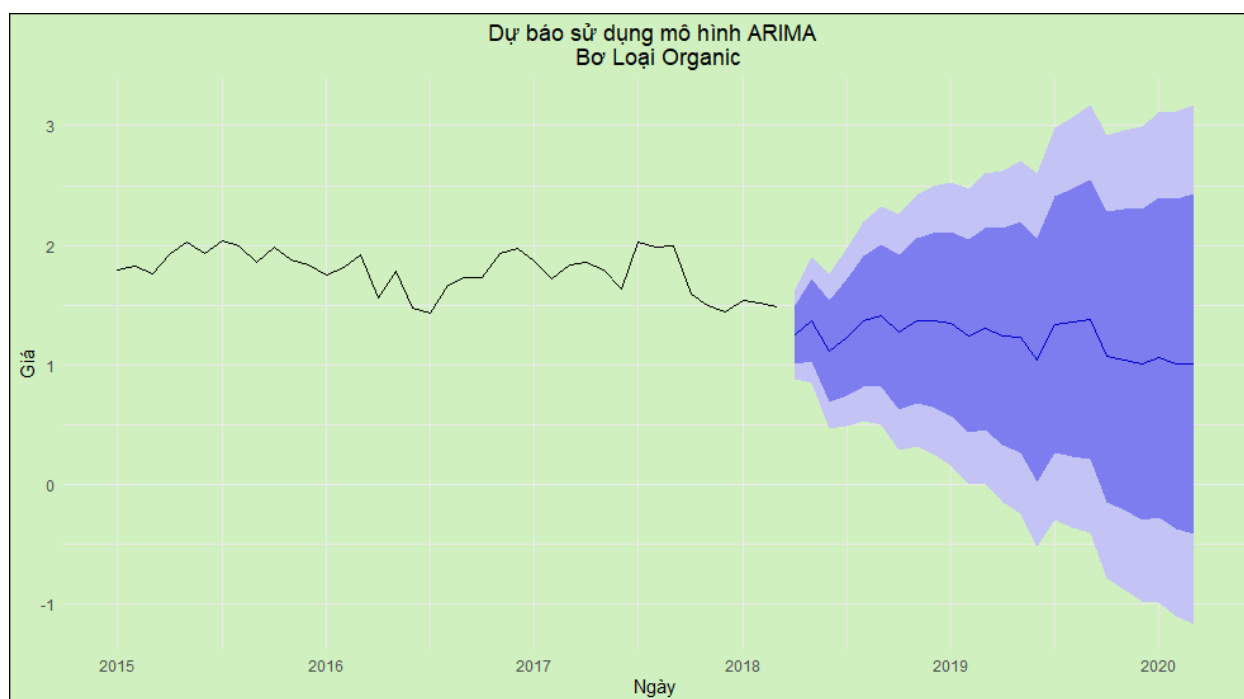
### 3.2.3 Dự báo giá bơ

Dự báo cho thấy giá bơ có xu hướng tăng nhẹ trong các tháng mùa hè (tháng 6-8) và giảm dần vào cuối năm.

Ví dụ, giá bơ dự báo trong tháng 6/2018 là 1.489 USD, tăng so với các tháng đầu năm. Vào tháng 12/2019, giá giảm xuống còn 1.300 USD, phù hợp với xu hướng giảm giá vào cuối năm.

Dải khoảng tin cậy mở rộng theo thời gian, cho thấy sự không chắc chắn gia tăng khi dự báo xa hơn.

## 3.3 ARIMA với loại bơ hữu cơ



Hình 3.4: Dự đoán với bơ loại hữu cơ

### 3.3.1 Mô hình được lựa chọn

Mô hình  $ARIMA(0,1,0)(1,1,0)[12]$  là lựa chọn phù hợp để dự báo giá bơ hữu cơ, xử lý tốt tính mùa vụ và xu hướng dài hạn trong dữ liệu.

### 3.3.2 Hiệu quả mô hình

Mô hình đạt được sai số trung bình phần trăm thấp ( $MAPE = 5.74\%$ ), cho thấy độ chính xác cao trong việc dự báo giá bơ hữu cơ.

Phần dư không có sự tự tương quan mạnh ( $ACF1 = -0.1943$ ), chứng minh rằng mô hình đã nắm bắt tốt các đặc tính của dữ liệu.

### 3.3.3 Kết quả dự báo

Dự báo cho thấy giá bơ hữu cơ dao động trong khoảng 1.00 - 1.41 USD từ tháng 4/2018 đến tháng 3/2020.

Giá tăng nhẹ trong các tháng giữa năm (tháng 6 - 8) và giảm dần vào cuối năm.

### 3.3.4 Hạn chế

Khoảng tin cậy rộng ở các giá trị dự báo xa cho thấy sự không chắc chắn tăng cao khi thời gian dự báo xa hơn.

# KẾT LUẬN

Dựa trên phân tích đã thực hiện, chúng tôi rút ra một số quan sát sau:

Bơ hữu cơ đắt hơn: Như dự đoán, giá bơ hữu cơ cao hơn nhiều so với bơ thông thường.

Các mô hình tương tự theo loại: Mặc dù có một số khác biệt giữa hai loại bơ, nhưng hầu hết các mô hình (xu hướng, mùa vụ) đều tương tự giữa hai loại bơ này.

Năm 2017 - Năm thành công của bơ: Năm 2017 được xem là năm tốt nhất cho giá bơ. Một trong những yếu tố có thể góp phần vào điều này là sự phát triển ổn định của nền kinh tế nói chung, nhưng cũng có thể có các yếu tố tiềm ẩn khác ảnh hưởng đến giá bơ trên thị trường.

Biến động cao: Mặc dù năm 2017 có giá cao nhất, nhưng cũng là năm có biến động giá cao nhất. Điều này có nghĩa là giá bơ trong năm này có sự dao động lớn hơn so với hai năm trước đó.

Nên mua bơ trước mùa thu! Chúng tôi nhận thấy một mô hình nhất quán là giá bơ có xu hướng tăng lên khi mùa thu đến. Điều này áp dụng cho cả bơ thông thường và bơ hữu cơ.

Xu hướng giảm trong dài hạn: Dựa trên mô hình ARIMA, dự báo cho thấy giá của cả hai loại bơ có xu hướng giảm trong dài hạn. Mặc dù có thể có áp lực tăng giá trong ngắn hạn, nhưng mô hình dự báo rằng giá bơ sẽ giảm sau khoảng 2 năm kể từ hiện tại.

# TÀI LIỆU THAM KHẢO

[1] Janio Martinez Bachmann.(2018,04.09). Price of Avocados || Pattern Recognition Analysis.  
<https://www.kaggle.com/code/janiobachmann/price-of-avocados-pattern-recognition-analysis/notebook#notebook-container>