Các mô hình hiện tại cho các vấn đề có kết quả hợp thành và biến giải thích dựa vào việc biến đổi dữ liệu hợp thành từ $S\_D$ đến $R\_{D-1}$. Có hai phương pháp biến đổi được khuyến nghị cho dữ liệu hợp thành trong ngữ cảnh hồi quy. Phương pháp biến đổi đầu tiên là biến đổi ALR (Aitchison, 1986), trong đó cho $z \in S\_D$,

\[

alr(z)\_j = ln\left ( \frac{z\_j}{z\_D} \right ),j=1,...,D-1

]\  
phương pháp biến đổi thứ 2 là biến đổi ilr(Egozcue *et al.*, [2003](https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13465#biom13465-bib-0016); Hron *et al.*, [2012](https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13465#biom13465-bib-0025); Filzmoser *et al.*, [2018](https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13465#biom13465-bib-0020)), ở đó:  
\[

ilr(z)\_j = \sqrt{\frac{D-j}{D-j+1}}\left ( \frac{z\_j}{(\prod \_{k=j+1}^{D}z\_k)^{\frac{1}{D-j}}} \right ), j=1,...,D-1

\]

Cả hai phương pháp biến đổi này đều cho phép các hợp thành biến đổi được sử dụng như các biến đổi và kết quả trong một mô hình hồi quy tuyến tính tiêu chuẩn mà không cần phải hạn chế các hệ số hồi quy (Hron et al., 2012). Trong khi Van den Boogaart và Tolosana-Delgado (2013) lưu ý rằng hai phương pháp biến đổi có thể được sử dụng thay thế cho nhau trong việc tạo ra các mô hình tương tự, mỗi phương pháp đều có ưu và nhược điểm riêng của nó.

Phương pháp biến đổi ILR là đồng dạng, nghĩa là nó bảo tồn góc và khoảng cách, nhưng khó hiểu hơn (Egozcue et al., 2003; Van den Boogaart và Tolosana-Delgado, 2013; Filzmoser et al., 2018). Mô hình ILR được trình bày bởi Chen et al. (2017) giả định rằng đối với một kết quả $y \in S^D$ và biến giải thích $x \in S^{D\_S}$ , rằng:

\[

E[ilr(y)\_k|x]=\beta \_{0k} + \sum\_{j=1}^{D\_S-1}\beta \_{jk}ilr(x)\_j, k=1,...,D\_r-1.

\]

Ở đây, $\beta \_{11}$ có thể được hiểu là hiệu ứng của việc tăng giá trị tương đối của x1 lên 1 so với phần còn lại của $x$, giữ nguyên tỷ lệ giữa các thành phần khác của $x$ không đổi, đối với sự thay đổi của giá trị tương đối của $y\_1$ so với phần còn lại của y ; các hệ số hồi quy khác không có ý nghĩa giải thích (Hron et al., 2012; Chen et al., 2017). Để thu được các hiệu ứng của sự thay đổi tương đối của mỗi phần của $x$ lên $y$, người ta phải sử dụng phép hoán vị $z^l=(z\_l,z\_1,…,z\_{l-1},z\_{l+1},…,z\_D)$ và ước lượng $D\_r.D\_s$ các mô hình riêng biệt trong đó:  
\[

E[ilr(y^{l\_1})\_k|x]=\beta \_{0k}^{l\_1,l\_2} + \sum\_{j=1}^{D\_S-1}\beta \_{jk}^{(l\_1,l\_2)}ilr(x^{l\_2})\_j, k=1,...,D\_r-1,l\_1=1,...,D\_r,l\_2=1,..,D\_s

\]

Các hệ số quan trọng sau đó sẽ là $}\beta \_{11}^{(l\_1,l\_2)}$ cho mỗi kết hợp của $l\_1$ và $l\_2$ (Chen et al., 2017; Filzmoser et al., 2018). Do việc ước lượng tham số được thực hiện bằng cách sử dụng ước lượng hợp lý cực đại tiêu chuẩn cho các mô hình hồi quy tuyến tính, quy trình này không tốn kém về mặt tính toán. Tuy nhiên, việc sử dụng nhiều phiên bản của một mô hình để thu được một tập hợp các hệ số không thể được hiểu cùng nhau là không mong muốn.

Có hai điểm tiêu cực bổ sung. Thứ nhất, phép biến đổi ILR không cho phép có giá trị 0 trong các hợp thành. Do đó nếu x hoặc y là biến phân loại hoặc là các hợp thành có giá trị 0, khung việc này không thể sử dụng, mặc dù biến phân loại vẫn nằm trong không gian đơn hình. Thứ hai, các hệ số quan trọng chỉ có thể được hiểu trong ngữ cảnh của sự thay đổi tương đối của mỗi phần của dữ liệu hợp thành đến giá trị trung bình hình học. Điều này không cho phép giải thích đơn giản các hệ số trong ngữ cảnh của tác động trực tiếp của việc thay đổi giá trị của x trong không gian đơn hình đến giá trị kỳ vọng của y trong không gian đơn hình (Morais et al., 2018). Việc thiếu một giải thích đơn giản cho các hệ số trong (2) đã buộc các nhà thực hành phải dựa vào các kỹ thuật đồ họa để hiển thị bề mặt phản ứng ước lượng của y như một hàm của x (Nguyen et al., 2018). Một người có thể cố gắng khắc phục những vấn đề này bằng cách sử dụng phép biến đổi ALR vì nó chỉ cần chạy một mô hình và các hệ số hồi quy dễ hiểu hơn (Billheimer et al., 2001) trong ngữ cảnh của các biến đổi của hợp thành. Tuy nhiên, từ quan điểm của một nhà thực hành, các biến đổi có thể khó hiểu hơn so với sự thay đổi trực tiếp (khác biệt) trong phản ứng do thay đổi trong biến giải thích và chỉ có thể dễ dàng hình dung cho các hợp thành có ba hoặc ít hơn các thành phần (Billheimer et al., 2001). Ngoài ra, các giải thích của ALR trên không gian đơn hình phụ thuộc vào các giả định phân phối (như logistic-normal) cho kết quả hợp thành (Billheimer et al., 2001). Hơn nữa, ALR không bảo tồn đồng đẳng và không cho phép có giá trị 0 và 1. Cuối cùng, cả hai phép biến đổi log đều dựa vào giả định về phân phối chuẩn cho kết quả hợp thành đã được biến đổi, cũng như tính tuyến tính chỉ trong không gian Euclide đã được biến đổi.

Alenazi (2019) tiếp cận khác biệt đối với hồi quy hợp thành, khi chỉ biến đổi biến giải thích hợp thành x. Trong khi Alenazi (2019) quan tâm hơn đến độ chính xác dự đoán hơn là giải thích và sử dụng một phép biến đổi phức tạp dựa trên thành phần chính, người ta có thể sử dụng bất kỳ phép biến đổi nào t (ví dụ, các phép biến đổi ALR hoặc ILR). Mô hình hồi quy được giả định là đặc tả (pseudo-ML) đa thức nhân tạo (Papke và Wooldridge, 1996; Mullahy, 2015; Murteira và Ramalho, 2016):

\[

E[y\_k|x] = \frac{exp\left ( \beta\_{0k} +\sum\_{j=1}^{D\_s-1}\beta\_{jk}t(x)\_j \right )}{1+\sum\_{k=1}^{D\_r-1}\left [ exp\left ( \beta\_{0k}+ \sum\_{j=1}^{D\_s-1}\beta\_{jk}t(x)\_j \right ) \right ]}

K=1,…,D\_r-1

E[y\_{D\_r}|x] = \frac{1}{1+\sum\_{k=1}^{D\_r-1}\left [ exp\left ( \beta\_{0k}+ \sum\_{j=1}^{D\_s-1}\beta\_{jk}t(x)\_j \right ) \right ]}

\]

Murteira và Ramalho (2016) bàn luận cả hai phương pháp hồi quy cực đại gần đúng và cực đại (QML và ML) để ước lượng các hệ số. Tuy nhiên, Alenazi (2019) sử dụng một phương pháp QML cho phép các giá trị 0 trong $y$ (Papke và Wooldridge, 1996; Mullahy, 2015; Murteira và Ramalho, 2016) và không đưa ra bất kỳ giả định nào về phân phối của $y$.

**3 PHẢN HỒI TUYẾN TÍNH TRỰC TIẾP CỦA CÁC BIẾN TỔNG HỢP TRÊN SIMPLEX**

Đối với hồi quy dữ liệu hợp thành trên dữ liệu hợp thành mà không cần biến đổi và cho phép có các giá trị 0 trong cả y và x, chúng ta mô hình trực tiếp kỳ vọng của y như một hàm tuyến tính của x như là:  
\[

E[y|x] = \sum\_{j=1}^{D\_s}x\_jb\_j

\]

ở đó các $b\_j$ là các chiều vector của $D\_y$.