

UNIwersytet im. Adama Mickiewicza w Poznaniu
Wydział Matematyki i Informatyki

Piotr Chlebicki

Kierunek: matematyka

Specjalność: statystyka matematyczna i analiza danych

Numer albumu: 456516

**Wektorowe uogólnione modele
liniowe oraz ich wybrane
zastosowania**

**Vector generalized linear models and their selected
applications**

Praca magisterska
napisana pod kierunkiem
prof. UAM dra hab. Waldemara Wołyńskiego

POZNAŃ 2024

Spis treści

Streszczenie	5
Abstract	7
Rozdział 1. Wprowadzenie i notacja	9
1.1. Notacja używana w pracy	9
1.2. Wprowadzenie	10
Rozdział 2. Wektorowy uogólniony model liniowy	11
2.1. Postać ogólna	11
2.1.1. Macierze ograniczeń i macierz wektorowego modelu liniowego	12
2.2. Estymacja współczynników regresji	13
2.2.1. Algorytm iteracyjnie ważonych najmniejszych kwadratów	13
2.2.2. Reszty Pearsona	16
2.2.3. Macierz projekcji	17
2.3. Testy w modelu	17
2.3.1. Testy istotności współczynników modelu	17
2.3.2. Analiza dewiancji	19
2.4. Obecnie istniejące implementacje	19

Rozdział 3. Przykłady zastosowań	21
3.1. Model liniowy z warunkową heteroskedastycznością	21
3.1.1. Opis modelu	21
3.1.2. Symulacja numeryczna	25
3.2. Model Dixona-Colesa	28
3.2.1. Opis modelu i sposobu porównania modeli	31
3.2.2. Wyniki	33
3.3. Inne popularne godne uwagi przykłady	35
Rozdział 4. Estymacja wielkości populacji	37
4.1. Wprowadzenie i motywacja	37
4.2. Metoda wieloźródłowa	38
4.2.1. Modele podstawowe	39
4.2.2. Podział metod estymacji częstościowej	40
4.2.3. Porównanie podejść w metodzie wieloźródłowej	45
4.3. Przykład wykorzystania na rzeczywistych danych	48
Podsumowanie	51
Bibliografia	53
Dodatek A. Wyniki symulacji z rozdziału 3	57
Dodatek B. Analiza dewiancji z rozdziału 3	69
Dodatek C. Statystyki opisowe modeli z rozdziału 4	71

Streszczenie

Klasa uogólnionych modeli liniowych jest popularnym narzędziem analizy danych ze względu na swoją prostotę jak i ilość użytecznych modeli statystycznych, które można przedstawić w tejże klasie. Mniej popularną oraz nowszą klasą modeli statystycznych są wektorowe uogólnione modele liniowe, o której można myśleć jako o rozszerzeniu uogólnionych modeli liniowych na przypadek wielo-parametryczny, wielowymiarowy oraz na przypadek, gdy nie operujemy rodziną wykładniczą. W pracy przedstawiony został zarys teorii wektorowych uogólnionych modeli liniowych w kontekście wnioskowania częstościowego wraz z pewnymi uwagami praktycznymi odnośnie implementacji oraz wybrane zastosowania w testowaniu hipotez, modelowaniu wielowymiarowych procesów dyskretnych oraz estymacji wielkości populacji metodą *capture-recapture*.

Praca została zrealizowana w ramach grantu Narodowego Centrum Nauki OPUS 20 nr 2020/39/B/HS4/00941 pt. "Statystyka cudzoziemców bez spisu powszechnego – jakość, integracja danych i estymacja"
(kierownik dr Maciej Beręsewicz)

Abstract

The class of generalized linear models is a popular data analysis tool due to its simplicity and the number of usefull statistical models that can be presented in this class. A less popular and newer class of statistical models are the vector generalized linear models, which can be thought of as an extension of the generalized linear models to the multi-parameter, multi-dimensional cases and to the case when we parametric assumptions on the data cannot be encompassed within an exponential family. The paper presents an outline of the theory of the vector generalized linear models in the context of frequentist inference, along with some practical notes of computational nature and selected applications in hypothesis testing, modeling of multidimensional discrete processes and population size estimation using the *capture-recapture* method.

This thesis was supported by the National Science Center grant OPUS 20 2020/39/B/HS4/00941 "Towards census-like statistics for foreign-born populations – quality, data integration and estimation"
(principal investigator dr Maciej Beręsewicz)

Wprowadzenie i notacja

1.1. Notacja używana w pracy

$(\mathbf{A} \mid \mathbf{B})$ Macierz utworzona przez dołączenie kolumn macierzy \mathbf{B} do macierzy \mathbf{A} tzn. dla:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \vdots & \vdots & \dots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j'} \\ b_{21} & b_{22} & \dots & b_{2j'} \\ \vdots & \vdots & \dots & \vdots \\ b_{i1} & b_{i2} & \dots & b_{ij'} \end{pmatrix}$$

rezultatem operacji $(\mathbf{A} \mid \mathbf{B})$ będzie macierz:

$$(\mathbf{A} \mid \mathbf{B}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & b_{11} & b_{12} & \dots & b_{1j'} \\ a_{21} & a_{22} & \dots & a_{2j} & b_{21} & b_{22} & \dots & b_{2j'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & b_{i1} & b_{i2} & \dots & b_{ij'} \end{pmatrix}$$

$\mathcal{I}_A(x)$ Funkcja indykatorowa równa 0, gdy: $x \notin A$ i 1, gdy $x \in A$.

$\text{Dg}(\cdot)$ Operacja tworzenia macierzy diagonalnej z elementów argumentu.

\odot Iloczyn Hadamarda wektorów, dla iloczynu wektora \mathbf{Y} i macierzy \mathbf{X} przyjmujemy umownie, że $\mathbf{Y} \odot \mathbf{X} := \text{Dg}(\mathbf{Y})\mathbf{X} =: \mathbf{X} \odot \mathbf{Y}$.

$\text{vect}(\cdot)$ Wektoryzacja macierzy.

$\psi^{(m)}$ Funkcja poligamma m -tego rzędu, zdefiniowana jako $\psi^{(m)} = \frac{d^{m+1}}{dz^{m+1}} \ln \Gamma(z)$, przyjmijmy skrótowo $\psi := \psi^{(0)}$.

\otimes Iloczyn Kroneckera.

\mathbf{I}_n Macierz jednostkowa rozmiaru n .

$\binom{A}{a_1, a_2, \dots, a_k}$ Symbol wielomianowy, gdzie $A = \sum_i a_i$.

$\mathbf{x}_{(i)}, \mathbf{x}_i$ Jeżeli \mathbf{X} jest macierzą, to przez \mathbf{x}_i rozumiemy jako kolumnowy wektor stanowiący i -tą kolumnę macierzy \mathbf{X} a przez $\mathbf{x}_{(i)}$ wierszowy wektor stanowiący i -ty wiersz macierzy \mathbf{X} . Jako pewne nadużycie notacji przyjmujemy, że $\mathbf{X}_{(i)}$ (odpowiednio \mathbf{X}_i) jest cechą której realizacja to $\mathbf{x}_{(i)}$ (odpowiednio \mathbf{x}_i).

$\mathbf{X}_n = o_{\mathbb{P}}(a_n) \iff \frac{1}{a_n} \mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{0}$, probabilistyczna notacja o .

$\chi^2(k)$ Rozkład chi kwadrat z k stopniami swobody.
 \mathbb{E}, \mathbb{V} Operatory wartości oczekiwanej i wariancji. Operatory warunkowe oznaczone są analogicznie.
 $\mathcal{N}(\mu, \sigma^2), \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ Odpowiednio: rozkład normalny o wartości oczekiwanej μ i wariancji σ^2 , wielowymiarowy rozkład normalny o wartości oczekiwanej $\boldsymbol{\mu}$ i macierzy wariancji-kowariancji $\boldsymbol{\Sigma}$.
 $\mathcal{P}(\lambda)$ Rozkład Poissona z parametrem λ .
 $\mathcal{U}(a, b)$ Ciągły rozkład jednostajny na przedziale (a, b) .
 $b(p)$ Rozkład Bernoullego, gdy $X \sim b(p)$ to $\mathbb{P}[X = x] = p^x(1 - p)^{1-x}$.
 n, N Odpowiednio rozmiar próby oraz rozmiar skończonej populacji.
 $R \xrightarrow{d} Q, R \xrightarrow{\mathbb{P}} Q, R \xrightarrow{\text{a.s.}} Q$ zbieżność odpowiednio: względem rozkładu, miary probabilistycznej i prawie na pewno.

1.2. Wprowadzenie

Celem tej pracy jest przedstawienie teorii i wybranych zastosowań klasy wektorowych uogólnionych modeli liniowych. Jest to klasa modeli statystycznych, która uogólnia klasę modeli uogólnionych modeli liniowych i posiada bardzo szeroką gamę zastosowań.

Praca zorganizowana jest w następujący sposób:

- W drugim rozdziale w sposób szczegółowy opisana została ogólna postać wektorowego uogólnionego modelu liniowego wraz z najpopularniejszym algorytmem optymalizacyjnym jak i najważniejsze statystyki używane do oceny modelu.
- Rozdział trzeci przedstawia dwa przykładowe zastosowania omawianej teorii, które wykraczają poza granicę uogólnionych modeli liniowych.
- Celem rozdziału czwartego jest przedstawienie teorii estymacji wielkości zamkniętych populacji metodą *capture-recapture*, która może zostać przedstawiona w całości jako specjalny przypadek wektorowego uogólnionego modelu liniowego.
- Praca kończy się krótkim podsumowaniem.

Wszystkie obliczenia w pracy wykonane zostały w języku R (R Core Team 2023) a kod dostępny jest na publicznym repozytorium GitHub <https://github.com/Kertoo/MastersThesis>.

Wiele technicznych terminów związanych z teorią nie ma jeszcze polskich odpowiedników, (nawet sam termin wektorowe uogólnione modele liniowe (*ang. vector generalized linear models*) nie jest używany w żadnych pracach naukowych, o których wiadomo autorowi) dlatego obok tłumaczenia na język polski w nawiasie zapisany będzie anglojęzyczny odpowiednik wyrażenia, gdy pojawia się po raz pierwszy.

Wektorowy uogólniony model liniowy

2.1. Postać ogólna

Niech $\mathbf{Y} = (Y_1 \dots Y_Q)^T$ będzie wektorem cech zależnych, a macierz \mathbf{y} realizacją tego wektora losowego, ponadto niech \mathbf{X} będzie macierzą zmiennych niezależnych rozmiaru $n \times p$. W problemach regresji badany jest związek pomiędzy rozkładem wektora losowego \mathbf{Y} a wierszami macierzy \mathbf{X} będącymi realizacjami wektora \mathbf{X}_j i nazywanymi zmiennymi niezależnymi. Rozważane modele regresji będą charakteryzowane przez następujące elementy:

- Regularnego warunkowego rozkładu zmiennej zależnej $\mathbb{P}(\mathbf{Y}|\mathbf{X}_{(k)}, \mathbf{B})$ z M elementowym wektorem parametrów rozkładu $\boldsymbol{\theta}$ zdeterminowanym przez wektor zmiennych niezależnych $\mathbf{X}_{(k)}$ i macierz współczynników regresji $\mathbf{B} = (\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2 \ \dots \ \boldsymbol{\beta}_M) = (\boldsymbol{\beta}_{(1)}^T \ \boldsymbol{\beta}_{(2)}^T \ \dots \ \boldsymbol{\beta}_{(p)}^T)^T$ gdzie $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_M$ są wektorami kolumnowymi a $\boldsymbol{\beta}_{(1)}, \boldsymbol{\beta}_{(2)}, \dots, \boldsymbol{\beta}_{(p)}$ to wektory wierszowe.
- M wektorów predyktorów liniowych¹ $\boldsymbol{\eta}_k = \mathbf{X}\boldsymbol{\beta}_k$ dla $k = 1, \dots, M$.
- M funkcji łączących g_1, \dots, g_M przekształcających predyktory liniowe na parametry rozkładu.

Na wektorowe uogólnione modele liniowe (dalej VGLM) omawiane w tej pracy można patrzeć jak na 'wielowymiarowe' (w sensie wielowymiarowości zmiennych zależnych \mathbf{Y}) uogólnione modele liniowe (dalej GLM) lub GLM-y z większą liczbą parametrów rozkładu prawdopodobieństwa.

Powyższe podsumowanie nie wyczerpuje tematu modeli VGLM, ponieważ w przeciwieństwie do GLM'ów nie wymagane jest by rozkład warunkowy $\mathbf{Y}|\mathbf{X}_{(k)}$ należał do rodziny wykładniczej, jak np. w regresji kwantylowej.

Praca nie wyczerpie także tematu zastosowań VGLM'ów, ze względów objętościowych nie byłoby to możliwe. Część z nieomówionych w tej pracy zastosowań jest opisane w książce Thomas W. Yee 2015, która zawiera także dokładniejszy przegląd teorii VGLM'ów oraz na podstawie której napisany jest ten wprowadzający rozdział.

¹ W powiązonym zagadnieniu modeli addytywnych, które także jest opisane w książce Thomas W. Yee 2015 ale wykracza poza zakres pracy przyjmuje się, że $\eta_{ik} = \beta_1 + \sum_{t=2}^p f_t(x_{it}), i = 1, \dots, n$ dla pewnych $f_2, \dots, f_p \in C^\infty(\mathbb{R})$.

2.1.1. Macierze ograniczeń i macierz wektorowego modelu liniowego

Istotną rolę w teorii VGLM'ów pełnią tak zwane macierze ograniczeń (ang. *constraint matrices*) pozwalające na różnorodną specyfikację zależności pomiędzy poszczególnymi elementami macierzy \mathbf{B} . Decydują one o postaci macierzy wektorowego modelu liniowego \mathbf{X}_{VLM} zastępującego w teorii VGLM'ów macierz \mathbf{X} w tej teorii oznaczaną najczęściej przez \mathbf{X}_{LM} i nazywaną macierzą modelu liniowego. Podobnie jak w GLM wektor współczynników regresji $\boldsymbol{\beta} = \text{vect}(\mathbf{B})$ i macierz predyktorów liniowych $\boldsymbol{\eta}$ przy znajomości macierzy \mathbf{X}_{VLM} połączone są przez równanie:

$$\mathbf{X}_{\text{LM}}\mathbf{B} = \begin{pmatrix} \boldsymbol{\eta}_{(1)} \\ \boldsymbol{\eta}_{(2)} \\ \vdots \\ \boldsymbol{\eta}_{(n)} \end{pmatrix}, \quad \text{vect}(\boldsymbol{\eta}) = \mathbf{X}_{\text{VLM}}\boldsymbol{\beta}, \quad (2.1)$$

gdzie przy wektoryzacji \mathbf{B} pomijamy "elementy nieistotne" tzn. takie, o których z góry zakładamy, że są równe 0 i nie są estymowane a macierz \mathbf{X}_{VLM} tworzona jest za pomocą równania:

$$\mathbf{X}_{\text{VLM}} = \left(\mathbf{X}_{\text{LM},1} \otimes \mathbf{H}_1 \mid \mathbf{X}_{\text{LM},2} \otimes \mathbf{H}_2 \mid \dots \mid \mathbf{X}_{\text{LM},p} \otimes \mathbf{H}_p \right), \quad (2.2)$$

gdzie $\mathbf{X}_{\text{LM},k}$ to k -ta kolumna macierzy \mathbf{X}_{LM} . Macierz ta ma rozmiar $nM \times \sum_{k=1}^p \text{ncol}(\mathbf{H}_k)$. Taka konstrukcja macierzy wektorowego modelu liniowego jest tylko konwencją i poprawność obliczeń zostanie zachowana jeżeli zastąpimy $\mathbf{X}_{\text{LM},t} \otimes \mathbf{H}_t$ przez $\mathbf{H}_t \otimes \mathbf{X}_{\text{LM},t}$ (zmienić się mogą tylko kolejności elementów w macierzach/wektorach). W szczególnych przypadkach, np. gdy macierze $\mathbf{H}_1, \dots, \mathbf{H}_p$ określają indeksy sumowania w modelu postaci $\boldsymbol{\eta}_1 = \sum_t \mathbf{X}_{\text{LM},t} \beta_{t,1}, \dots, \boldsymbol{\eta}_M = \sum_t \mathbf{X}_{\text{LM},t} \beta_{t,M}$ (tzn. określają obecność lub brak zależności predyktorów od konkretnych zmiennych), macierze te można doprowadzić do postaci blokowo-diagonalnej tzn:

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_M \end{pmatrix} \quad (2.3)$$

dla odpowiednio dobranych macierzy $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M$, która także ma swoje zalety (tzn. w pewnych przypadkach doprowadza do prostszych obliczeń).

Od macierzy $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_p$ wymaga się aby:

- Każda z nich miała M wierszy i pomiędzy 1 a M kolumn (nie muszą być tego samego rozmiaru).
- Każda z nich była znana oraz stała (nie estymujemy macierzy \mathbf{H}_k).
- Każda z nich była pełnego rzędu, tzn. żeby rząd każdej \mathbf{H}_k był równy ilości jej kolumn co zapewnia istnienie macierzy $(\mathbf{H}_k^T \mathbf{H}_k)^{-1}$.

2.2. Estymacja współczynników regresji

Najczęściej wykorzystywaną metodą estymacji wektorowych uogólnionych modeli liniowych, podobnie jak uogólnionych modeli liniowych, jest metoda największej wiarygodności. Różnica w tym przypadku polega na tym, że z wyjątkiem modelu o trywialnych macierzach ograniczeń (tzn. $\mathbf{H}_1 = \dots = \mathbf{H}_p = \mathbf{I}_M$) lub sytuacji gdy wszystkie poza jedną macierzą ograniczeń są jedno-kolumnowe, algorytm Newtona i podobne uniwersalne algorytmy optymalizacyjne jak algorytm Nelder-Meada czy BFGS często nie dają zadowalających rezultatów.

Obecnie standardowym algorytmem optymalizacyjnym w kontekście modeli GLM jest algorytm iteracyjnie ważonych najmniejszych kwadratów (IRLS). Algorytm ten uogólnimy na przypadek wektorowych uogólnionych modeli liniowych. O ile w przypadku GLM'ów najczęściej użycie algorytmu IRLS tylko przyspiesza zbieżność i można by używać zwykłej metody (co nie jest niespotykane) o tyle w klasie VGLM w ogólności tylko ten algorytm cechuje się porządnymi własnościami w zadowalająco "ogólnym" przypadku.

2.2.1. Algorytm iteracyjnie ważonych najmniejszych kwadratów

Rozważmy (jednowymiarowy) model liniowy w którym obserwacje nie są niezależne:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2.4)$$

dla pewniej znanej, dodatnio określonej macierzy $\boldsymbol{\Sigma}$. Istnieje wtedy odwracalna macierz \mathbf{U} taka, że $\mathbf{U}^T \mathbf{U} = \boldsymbol{\Sigma}^*$ wtedy pomnożenie (2.4) lewostronnie przez \mathbf{U}^{-1} sprowadza (2.4) do:

$$\mathbf{U}^{-1} \mathbf{y} = \mathbf{U}^{-1} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}', \quad \boldsymbol{\epsilon}' \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n). \quad (2.5)$$

Jest to zwykły model liniowy i znamy postać estymatora najmniejszych kwadratów $\hat{\boldsymbol{\beta}}$ (a także największej wiarygodności):

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left((\mathbf{U}^{-1} \mathbf{X})^T \mathbf{U}^{-1} \mathbf{X} \right)^{-1} (\mathbf{U}^{-1} \mathbf{X})^T \mathbf{U}^{-1} \mathbf{y} \\ &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} \end{aligned}$$

który w kontekście modelu (2.4) nazywany jest uogólnionym estymatorem najmniejszych kwadratów. Własność ta jest zachowana gdy \mathbf{Y} jest macierzą a ϵ ma rozkład macierzowy normalny.

Algorytm IRLS (dla GLM'ów) polega na iteracyjnym rozwiązywaniu zadania najmniejszych kwadratów dla zmiennej zależnej \mathbf{Z} zdefiniowanej w każdej operacji poprzez:

$$Z_k = \eta_k + \frac{\partial \ell_k}{\partial \eta_k} \left(-\frac{\partial^2 \ell_k}{\partial \eta_k^2} \right)^{-1},$$

gdzie $\eta_k = \mathbf{x}_{(k)}\boldsymbol{\beta}$ ($\boldsymbol{\beta}$ z obecnej iteracji) a ℓ_k to logarytm funkcji wiarygodności k -tej jednostki.

Uogólnienie tej idei na wektorowe uogólnione modele liniowe prowadzi do następującego algorytmu:

Algorytm: IRLS

- 1: Zainicjuj z wartościami początkowymi $\text{converged} \leftarrow \text{FALSE}$, $\text{iter} \leftarrow 1$, $\boldsymbol{\beta} \leftarrow \text{start}$
- 2: **while** NOT converged and $\text{iter} \leq \text{maxiter}$ **do**
- 3: $\mathbf{W}^- \leftarrow \mathbf{W}$, $\boldsymbol{\beta}^- \leftarrow \boldsymbol{\beta}$, $\ell^- \leftarrow \ell(\boldsymbol{\beta})$
- 4: Dla $k = 1, \dots, n$ używając wartości $\boldsymbol{\eta}_{(k)}$ czyli wierszy macierzy $\boldsymbol{\eta} = (\boldsymbol{\eta}_1 \dots \boldsymbol{\eta}_M)$ z obecnej iteracji skonstruuj:

$$\mathbf{W}_k \leftarrow \mathbb{E} \left[-\frac{\partial^2 \ell_{(k)}}{\partial \boldsymbol{\eta}_{(k)}^T \partial \boldsymbol{\eta}_{(k)}} \right]$$

- 5: Uaktualnij pseudo-residua (wierszowo) przez:

$$\mathbf{Z}_{(k)} \leftarrow \boldsymbol{\eta}_{(k)} + \mathbf{W}_k^{-1} \frac{\partial \ell_k}{\partial \boldsymbol{\eta}_{(k)}}$$

- 6: Skonstruuj macierz \mathbf{W} używając odpowiedniej konwencji na przykład:

$$\mathbf{W} \leftarrow \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{W}_n \end{pmatrix}$$

dla konwencji (2.2), lub dla konwencji (2.3):

$$\mathbf{W} \leftarrow \left(\text{Dg}(W_{ij1}, W_{ij2}, \dots, W_{ijn}) \right)_{ij} =$$

$$\begin{pmatrix} \text{Dg}(W_{111}, \dots, W_{11n}) & \text{Dg}(W_{121}, \dots, W_{12n}) & \dots & \text{Dg}(W_{1p1}, \dots, W_{1pn}) \\ \text{Dg}(W_{211}, \dots, W_{21n}) & \text{Dg}(W_{221}, \dots, W_{22n}) & \dots & \text{Dg}(W_{2p1}, \dots, W_{2pn}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Dg}(W_{p11}, \dots, W_{p1n}) & \text{Dg}(W_{p21}, \dots, W_{p2n}) & \dots & \text{Dg}(W_{pp1}, \dots, W_{ppn}) \end{pmatrix}$$

gdzie W_{ijk} to element w i -tym wierszu kolumnie j -tej kolumny dla macierzy wag k -tej jednostki.

```

7:   $\beta \leftarrow (\mathbf{X}_{\text{VLM}}^T \mathbf{W} \mathbf{X}_{\text{VLM}})^{-1} \mathbf{X}_{\text{VLM}}^T \mathbf{W} \text{vect}(\mathbf{Z})$ 
8:   $\eta \leftarrow \mathbf{X}_{\text{VLM}} \beta$ 
9:  converged  $\leftarrow (\ell(\beta) - \ell^- < \ell^- \cdot \varepsilon)$  or  $(\|\beta - \beta^-\|_\infty < \varepsilon)$ 
10: if NOT converged then
11:   iter  $\leftarrow$  iter + 1
12: else
13:   if  $(\ell(\beta) - \ell^- < 0)$  then
14:     $\beta \leftarrow \beta^-$ ,  $\mathbf{W} \leftarrow \mathbf{W}^-$ 
15:   end if
16: end if
17: end while
18: Zwróć  $\beta, \mathbf{W}, \ell(\beta)$ 

```

minimalizującym na każdym kroku formę kwadratową:

$$\sum_{i=1}^n (\mathbf{Z}_{(i)} - \eta_{(i)})^T \mathbf{W}_i (\mathbf{Z}_{(i)} - \eta_{(i)}). \quad (2.6)$$

Różnicą względem GLM'ów jest to, że używamy macierzy informacji uaktualniając macierze \mathbf{W}_k zamiast wartości skalarnych.

Możliwe jest korzystanie z hesjanu tzn. z $\frac{\partial^2 \ell_k}{\partial \eta_{(k)}^T \partial \eta_{(k)}}$ prowadzi to nawet do szybszej (kwadratowej) zbieżności algorytmu. Jest to jednak niewskazane, ponieważ konieczne do zbieżności (choć nie dostateczne) jest to, żeby macierze \mathbf{W} były dodatnio określone co w praktyce może nie być spełnione gdy używamy hesjanu (nazywanego zazwyczaj w kontekście VGLM'ów zaobserwowaną macierzą informacji) zamiast macierzy informacji.

Zauważmy też, że postać macierzy \mathbf{W} wynikająca z konwencji (2.2) nie jest zbyt dobrze przystosowana do nowoczesnych metod programowania statystycznego. W praktyce więc konieczna jest implementacja specyficznych metod przechowywania informacji o macierzy \mathbf{W} .

Warunek zbieżności $\hat{\beta}$ względem normy $\|\cdot\|_\infty$ i relatywnej zbieżności ℓ można zastąpić innymi warunkami zbieżności algorytmu.

Uwagi odnośnie implementacji algorytmu

W praktyce często można spotkać się z sytuacją, w której nie znaleźliśmy jeszcze minimum ale krok algorytmu "przestrzelił" tzn. gdy ℓ^- jest daleko

od minimum globalnego i $\ell(\beta) < \ell^-$ dlatego w praktyce rekomendowany jest uaktualnienie:

$$\beta \leftarrow \beta^- + \Xi(\beta - \beta^-),$$

dla pewnego Ξ i w sytuacji gdy $\ell(\beta) < \ell^-$ zamiast kończyć pracę algorytmu szukamy najmniejszego $h \in \mathbb{N}$ o własności $\ell(\beta) > \ell^-$ dla uaktualnienia:

$$\beta \leftarrow \beta^- + \frac{\Xi}{2^h}(\beta - \beta^-),$$

takie warunkowe zmniejszanie wartości Ξ nazywa się połowieniem kroku (ang. *half-stepping*).

Możliwe jest też uwzględnienie przesunięcia (ang. *offset*) tzn. sytuacji gdy znamy macierz \mathbf{o} taką, że "prawdziwy" wektor predyktorów liniowych w modelu wyraża się przez:

$$\eta = \mathbf{X}_{VLM}\beta + \mathbf{o},$$

wtedy oprócz odpowiedniego dostosowania aktualizacji η_k pseudo-residua aktualizowane są przez:

$$\mathbf{Z}_{(k)} \leftarrow \eta_{(k)} + \mathbf{W}_k^{-1} \frac{\partial \ell_k}{\partial \eta_{(k)}} - \mathbf{o}_{(k)}.$$

Algorytm IRLS można także stosować do maksymalizacji dowolnej funkcji celu, nie tylko do funkcji wiarygodności, tak długo jak wartość oczekiwana względem zmiennej zależnej od hesjanu, lub sam hesjan tej funkcji jest ujemnie określony. Przedstawiony został w postaci maksymalizacji logarytmu funkcji wiarygodności bo mimo wszechstronności najczęściej stosowany jest właśnie do tego zadania.

2.2.2. Reszty Pearsona

Zauważmy, że wielkość (2.6) można przedstawić, jako:

$$\begin{aligned} \sum_{i=1}^n \left(\mathbf{W}_i^{-1} \frac{\partial \ell_i}{\partial \eta_{(i)}} \right)^T \mathbf{W}_i \left(\mathbf{W}_i^{-1} \frac{\partial \ell_i}{\partial \eta_{(i)}} \right) &= \sum_{i=1}^n \left(\frac{\partial \ell_i}{\partial \eta_{(i)}} \right)^T \mathbf{W}_i^{-1} \frac{\partial \ell_i}{\partial \eta_{(i)}} \\ &= \sum_{i=1}^n \left(\mathbf{U}_i^{-1} \frac{\partial \ell_i}{\partial \eta_{(i)}} \right)^T \mathbf{U}_i^{-1} \frac{\partial \ell_i}{\partial \eta_{(i)}}, \end{aligned}$$

gdzie $\mathbf{U}_i \mathbf{U}_i^T = \mathbf{W}_i \implies (\mathbf{U}_i^{-1})^T \mathbf{U}_i^{-1} = \mathbf{W}_i^{-1}$, dlatego macierz (lub wektor, gdy $M = 1$):

$$\mathbf{r} = \left(\left(\mathbf{U}_1^{-1} \frac{\partial \ell_1}{\partial \eta_{(1)}} \right) \quad \left(\mathbf{U}_2^{-1} \frac{\partial \ell_2}{\partial \eta_{(2)}} \right) \quad \cdots \quad \left(\mathbf{U}_n^{-1} \frac{\partial \ell_n}{\partial \eta_{(n)}} \right) \right)^T,$$

definiuje się jako uogólnienie reszt Pearsona dla VGLM'ów. Definicja ma sens tylko gdy $n > p$ i pokrywa się ze standardowymi definicjami dla GLM'ów.

Wyprowadzenie. *Zauważmy, że w przypadku gdy $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, przyjmując standardowo $g_1 = Id, g_2 = \ln$ i $\boldsymbol{\eta}_1 = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\eta}_2 = \eta_2 = \ln(\sigma)$ mamy:*

$$\begin{aligned} U_k^{-1} &= \begin{pmatrix} \exp(\eta_2) & 0 \\ 0 & \frac{1}{\sqrt{2}} \exp(\eta_2) \end{pmatrix} \\ \frac{\partial \ell_k}{\partial \boldsymbol{\eta}_{(k)}} &= \begin{pmatrix} -\exp(-2\eta_2) (y_k - \eta_{1k}) \\ -\exp(-\eta_2) + \exp(-3\eta_2) (y_k - \eta_{1k})^2 \end{pmatrix} \\ \Rightarrow U_k^{-1} \frac{\partial \ell_k}{\partial \boldsymbol{\eta}_{(k)}} &= \begin{pmatrix} -\exp(-\eta_2) (y_k - \eta_{1k}) \\ -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \exp(-2\eta_2) (y_k - \eta_{1k})^2 \end{pmatrix} \end{aligned}$$

Więc w modelu regresji liniowej przedstawionym jako VGLM pierwsza kolumna \mathbf{r} jest wektorem reszt Pearsona w standardowym modelu regresji liniowej zdefiniowanych jako:

$$\mathbf{r}_1 = \frac{1}{\hat{\sigma}} \begin{pmatrix} y_1 - \mathbf{x}_{(1)} \hat{\boldsymbol{\beta}} \\ y_2 - \mathbf{x}_{(2)} \hat{\boldsymbol{\beta}} \\ \vdots \\ y_n - \mathbf{x}_{(n)} \hat{\boldsymbol{\beta}} \end{pmatrix}, \quad \hat{\sigma} = \sqrt{\frac{1}{n-p} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})},$$

których kwadraty $\mathbf{r}_1^T \mathbf{r}_1$ sumują się do $n - p$.

2.2.3. Macierz projekcji

Dla modeli VGLM definicja macierzy projekcji \mathcal{H} , uzasadniona przez algorytm IRLS, to macierz \mathcal{H} spełniająca równość:

$$\mathbf{Z} = \mathcal{H}\boldsymbol{\eta},$$

używając wartości wyjściowych $\mathbf{W}, \mathbf{Z}, \boldsymbol{\eta}$ z ostatniej iteracji algorytmu IRLS. Wyraża się ona bezpośrednio przez wzór:

$$\mathcal{H} = \mathbf{U} \mathbf{X}_{\text{VLM}} (\mathbf{X}_{\text{VLM}}^T \mathbf{W} \mathbf{X}_{\text{VLM}})^{-1} \mathbf{X}_{\text{VLM}}^T \mathbf{U}^T, \quad (2.7)$$

gdzie $\mathbf{U}^T \mathbf{U} = \mathbf{W}$ jest rozkładem Choleskiego \mathbf{W} , macierz jest tego samego rozmiaru co \mathbf{W} tzn. $nM \times nM$.

2.3. Testy w modelu

2.3.1. Testy istotności współczynników modelu

Testowanie istotności współczynników $\boldsymbol{\beta}$ regresji dla VGLM jest analogiczne do testowania istotności we wszystkich modelach w których minimalizowana

jest funkcja celu, a oszacowania $\hat{\beta}$ są asymptotycznie normalne i efektywne.

Dla uproszczenia podamy postaci statystyk przy funkcji celu będącą logarytmem funkcji wiarygodności ℓ . Popularne metody testowania hipotezy:

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0, \end{cases} \quad (2.8)$$

to:

$$\begin{cases} \left((\hat{\beta} - \beta_0)^T \left(\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \beta^T \partial \beta} \right] \right)^{-1} \right) \Big|_{\beta=\beta_0} (\hat{\beta} - \beta_0) & \text{test Walda} \\ \left(\left(\frac{\partial \ell}{\partial \beta} \right) \left(\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \beta^T \partial \beta} \right] \right)^{-1} \left(\frac{\partial \ell}{\partial \beta} \right)^T \right) \Big|_{\beta=\beta_0} & \text{score test.}^2 \end{cases}$$

Te statystyki mają rozkład asymptotyczny $\chi^2(p)$. Możliwe jest także wykorzystanie testu ilorazu wiarygodności który zostanie omówiony w następnej podsekcji.

W teście Walda oraz score teście za wariancję można przyjąć także jej oszacowanie w postaci odwrotność macierzy informacji w punkcie $\hat{\beta}$.

Test Walda można za pomocą metody δ uogólnić na dowolną nieliniową (ciągle różniczkowalną) funkcję c zmiennej β dla hipotezy:

$$\begin{cases} H_0 : c(\beta) = c(\beta_0) \\ H_1 : c(\beta) \neq c(\beta_0), \end{cases} \quad (2.9)$$

jako:

$$\left(c(\hat{\beta}) - c(\beta_0) \right)^T \left(\frac{\partial c}{\partial \beta} \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \beta^T \partial \beta} \right] \frac{\partial c^T}{\partial \beta} \right)^{-1} \Big|_{\beta=\beta_0} \left(c(\hat{\beta}) - c(\beta_0) \right)^T.$$

Niestety w ogólności liniowy test Walda w (2.8) i test nieliniowy dla (2.9) mogą sugerować sprzeczne konkluzje nawet, gdy c jest bijekcją co spowodowane jest nie-liniowością operatora wariancji.

Score test jest prostszy pod względem obliczeniowym i w wielu praktycznych przykładach wykazuje szybszą zbieżność do rozkładu asymptotycznego niż test Walda jak zademonstrowano na przykład w artykule Zimmer, Draxler, and Debelak 2022. Naturalnie uogólnić można też przedstawione testy dla zmodyfikowanej hipotezy (2.8) dla kontrastów.

² Nazywany czasami w ekonometrii testem współczynników Lagrange'a, lub testem Rao.

2.3.2. Analiza dewiancji

Podobnie jak w przypadku testów istotności współczynników bezpośrednio do VGLM'ów uogólnia się testy dewiancji. Jeżeli mamy wektor cech niezależnych $(X_1 \ X_2 \ \dots \ X_r)$ i chcemy przetestować, które ze składowych X_1, \dots, X_r wektora wpływają istotnie na jakość modelu (przy ustalonych macierzach ograniczeń) można przetestować układ hipotez:

$$\begin{cases} H_0 & : \ell_{-k} = \ell \\ H_1 & : \ell_{-k} < \ell \end{cases} \quad \text{Dla: } k = 1, \dots, r, \quad (2.10)$$

gdzie ℓ jest logarytmem funkcji wiarygodności w modelu ze wszystkimi zmiennymi X_1, \dots, X_r a ℓ_{-k} jest logarytmem funkcji wiarygodności w modelu po usunięciu (tylko) zmiennej x_k . Najczęściej w celu weryfikacji hipotezy (2.10) wykorzystuje się statystykę testową:

$$T = 2 \left(\hat{\ell} - \hat{\ell}_{-k} \right) \xrightarrow{d} \chi^2(b), \quad (2.11)$$

gdzie b to liczba kolumn macierzy \mathbf{X}_{VLM} usuniętych w przy usuwaniu zmiennej x_k czyli $\text{ncol}(\mathbf{H}_k)$ (w przypadku gdy $Q = 1 = M$ wartość b to 1). Nazwa analiza dewiancji wywodzi się z równości:

$$T = 2 \left(\hat{\ell}^* - \hat{\ell}_{-k} \right) - 2 \left(\hat{\ell}^* - \hat{\ell} \right),$$

gdzie ℓ^* jest logarytmem funkcji wiarygodności w modelu nasyconym tzn w modelu z identyczną liczbą obserwacji i zmiennych niezależnych, wielkość $2 \left(\hat{\ell}^* - \hat{\ell} \right)$ zwana jest dewiancją modelu.

2.4. Obecnie istniejące implementacje

Większość modeli, które zostaną zaprezentowane w pracy zaimplementowane jest w pakiecie języka do obliczeń statystycznych R, VGAM (T. W. Yee 2023) i dostępnym na platformie CRAN.

Najważniejszą funkcją w pakiecie (w kontekście wektorowych uogólnionych modeli liniowych) jest `vglm` działająca analogicznie do funkcji `glm` z pakietu `stats`. Wywołania tej funkcji mają zazwyczaj postać:

```
object_vglm <- vglm(
  formula = cbind(Y_1, Y_2, Y_3) ~ X_1 + X_2 + X_3,
  family  = poissonff(),
  data    = data.frame(),
  weights = c(),
  control = vglm.control(
```

```

        epsilon = 1e-7,
        maxit    = 30,
        half.stepsizing = TRUE
    ),
    constraints = list(
        "(Intercept)" = matrix(),
        "X_1"          = matrix(),
        "X_2"          = matrix(),
        "X_3"          = matrix()
    )
)

```

Jeżeli podany zostanie rozkład jednowymiarowy a zmienna zależna jest wielowymiarowa dopasowany zostanie model, w którym Y_1, Y_2, Y_3 są niezależne warunkowo przy X_1, X_2, X_3 .

Podobnie jak w przypadku `stats::glm` procedury takie jak konstrukcja macierzy (2.7) analiza dewiancji (2.10) czy testy istotności (2.8) za pomocą testu Walda czy testy ilorazu wiarygodności mają zaimplementowane odpowiednie metody dla klasy `vglm` typu `S4`. Nazwy w liście `constraints` zawierającej macierz $\mathbf{H}_1, \dots, \mathbf{H}_p$ muszą mieć te same nazwy co nazwy w wektorze `coef(object_vglm)`.

W książce Thomas W. Yee 2015 opisane oprócz różnych aspektów modeli VGLM i modeli addytywnych są także aspekty dotyczące wykorzystania pakietu oraz implementacji metod. Z pakietu tego korzystać będziemy w każdym przykładzie na rzeczywistych danych dlatego poświęcone zostało miejsce na opis jego funkcjonalności.

Przykłady zastosowań

3.1. Model liniowy z warunkową heteroskedastycznością

3.1.1. Opis modelu

Pierwszym nasuwającym się przykładem wektorowego uogólnionego modelu liniowego jest model regresji liniowej z warunkową heteroskedastycznością, to znaczy model z jednowymiarową zmienną zależną Y , który można zapisać jako:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \exp(\mathbf{X}_2\boldsymbol{\beta}_2)\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n). \quad (3.1)$$

Istnieją znane i sprawdzone metody radzenia sobie z heteroskedastycznością reszt regresji w modelu liniowym np. estymatory "kanapkowe" (ang. *sandwich estimators*) opisane dokładnie w pracy Zeileis 2004. Będziemy chcieli porównać, czy znajomość postaci heteroskedastyczności w modelu może polepszyć jakość testów istotności dla parametrów związanych z wartością oczekiwaną $\hat{\boldsymbol{\beta}}_1$.

Estymatory odporne macierzy wariancji w modelu liniowym

W tym celu przypomnimy postać estymatorów macierzy $\mathbb{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]$ w zwykłym modelu regresji liniowej:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}), \quad (3.2)$$

z założeniem, że \mathbf{X} i $\boldsymbol{\epsilon}$ są niezależne. Ponieważ:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y},$$

to także:

$$\begin{aligned} \mathbb{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbb{V}[\mathbf{Y}] \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right)^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned} \quad (3.3)$$

Żeby znaleźć estymator $\mathbb{V}[\widehat{\beta}|\mathbf{X}]$ należy zatem znaleźć estymator dla Σ .

Jeżeli przyjmiemy założenie, że macierz Σ jest diagonalna popularnymi estymatorami są:

$$\text{HC0: } \hat{\Sigma} = \text{Dg}(\epsilon \odot \epsilon),$$

$$\text{HC1: } \hat{\Sigma} = \frac{1}{n-p} \text{Dg}(\epsilon \odot \epsilon),$$

$$\text{HC2: } \hat{\Sigma} = \text{Dg}\left(\frac{1}{1-\mathbf{h}} \odot \epsilon \odot \epsilon\right),$$

$$\text{HC3: } \hat{\Sigma} = \text{Dg}\left(\frac{1}{(1-\mathbf{h})^2} \odot \epsilon \odot \epsilon\right),$$

$$\text{HC4: } \hat{\Sigma} = \text{Dg}\left(\left(\frac{1}{(1-h_i)^{\max\{h_i/\bar{h}, 4\}}}\right)_{i=1,\dots,n} \odot \epsilon \odot \epsilon\right),$$

$$\text{HC4m: } \hat{\Sigma} = \text{Dg}\left(\left(\frac{1}{(1-h_i)^{\min\{1, \frac{n}{p}h_i\} + \min\{1.5, \frac{n}{p}h_i\}}}\right)_{i=1,\dots,n} \odot \epsilon \odot \epsilon\right),$$

$$\text{HC5: } \hat{\Sigma} = \text{Dg}\left(\left(\frac{1}{(1-h_i)^{\frac{1}{2} \min\{\frac{n}{p}h_i, \max\{4, \frac{7n}{10p}\|\mathbf{h}\|_\infty\}\}}}\right)_{i=1,\dots,n} \odot \epsilon \odot \epsilon\right),$$

gdzie h_i to elementy diagonalne macierzy $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ a \mathbf{h} to wektor złożony z tych elementów, operacje odejmowania (dodawania) wielkości skalarnych od wektorów rozumiemy jako $a - (+)\mathbf{x} := (a \ a \ \dots \ a)^T - (+)\mathbf{x}$. Definiującą własnością estymatorów HC[0-5] jest to, że po podstawieniu $\hat{\Sigma}$ do $\mathbb{V}[\widehat{\beta}|\mathbf{X}]$ estymatory te są zgodne.

Estymator HC0, zaproponowany w pracy Eicker 1963, był chronologicznie pierwszym estymatorem "heteroskedastycznie-zgodnym" dla macierzy Σ . Powodem wprowadzenia estymatorów HC[1-5] jak i nadal trwających badań w tym temacie jest poprawienie własności estymatora HC0, w szczególności test dla istotności składowych wektora $\hat{\beta}$ bazujący na estymatorze HC0 jest zbyt liberalny w skończonych próbach.

Przy założeniu modelu postaci VGLM (3.1) estymatory największej wiarygodności dla $\hat{\beta}_1, \hat{\beta}_2$ w tym modelu (uwaga w ogólności estymator dla $\hat{\beta}_1$ nie jest estymatorem najmniejszych kwadratów) także są zgodne.

Estymatory odporne macierzy wariancji dla wektorowych uogólnionych modeli liniowych

W przypadku gdy macierze ograniczeń określają tylko to czy dany parametr rozkładu zależy (liniowo), bądź nie od konkretnej zmiennej niezależnej to estymatory "kanapkowe" można rozszerzyć na wektorowe uogólnione modele

liniowe w taki sam sposób w jaki konstruuje się je dla uogólnionych modeli liniowych i ogólnych estymatorów które zdefiniowane są przez minimalizację funkcji celu Ψ . Na podstawie teorii z artykułu Zeileis 2006 dla estymatora $\boldsymbol{\theta}$ zdefiniowanego, jako:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \Psi(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}),$$

oraz korzystając z twierdzenia 6.10 z książki White 1994 przy minimalnych założeniach o regularności rozkładu $\mathbf{Y}|\mathbf{X}$, spełnionych dla modelu (3.1), mamy:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_p \left(\mathbf{0}, \mathbb{E} \left[-\frac{\partial^2 \Psi}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right]^{-1} \underbrace{\mathbb{V} \left[\frac{\partial \Psi}{\partial \boldsymbol{\theta}} \right]}_{:= \boldsymbol{\Sigma}} \mathbb{E} \left[-\frac{\partial^2 \Psi}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right]^{-1} \right),$$

łatwo sprawdzić, że taka definicja pokrywa się z (3.3) dla zwykłego modelu liniowego.

Estymatory HC[0-5] w tym kontekście przyjmują postać:

$$\begin{aligned} \text{HC0: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC1: } \hat{\boldsymbol{\Sigma}} &= \frac{1}{n-p} \text{Dg} \left(\frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC2: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\frac{1}{1-\mathbf{h}} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC3: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\frac{1}{(1-\mathbf{h})^2} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC4: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\left(\frac{1}{(1-h_i)^{\max\{\frac{h_i}{h}, 4\}}} \right)_{i=1, \dots, nM} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC4m: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\left(\frac{1}{(1-h_i)^{\min\{1, \frac{n}{p}h_i\} + \min\{1.5, \frac{n}{p}h_i\}}} \right)_{i=1, \dots, nM} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \\ \text{HC5: } \hat{\boldsymbol{\Sigma}} &= \text{Dg} \left(\left(\frac{1}{(1-h_i)^{\frac{1}{2} \min\{\frac{n}{p}h_i, \max\{4, \frac{7n}{10p}\|\mathbf{h}\|_\infty\}\}}} \right)_{i=1, \dots, nM} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \odot \frac{\partial \Psi}{\partial \boldsymbol{\eta}'} \right), \end{aligned}$$

gdzie \mathbf{h} to wektor utworzony z elementów diagonalnych macierzy \mathcal{H} z (2.7) oraz $\boldsymbol{\eta}' = \text{vect}(\boldsymbol{\eta})$.

Wyprowadzenie. Dla modelu (3.1) przy macierzy \mathbf{X} jest w konwencji (2.3), logarytm funkcji wiarygodności to:

$$\ell = -\frac{n}{2} \ln(2\pi) - \sum_{k=1}^n \ln \sigma_k - \frac{1}{2} \sum_{k=1}^n \left(\frac{y_k - \mu_k}{\sigma_k} \right)^2$$

$$= -\frac{n}{2} \ln(2\pi) - \sum_{k=1}^n \mathbf{x}_{2(k)} \boldsymbol{\beta}_2 - \frac{1}{2} \sum_{k=1}^n \left(\frac{y_k - \mathbf{x}_{1(k)} \boldsymbol{\beta}_1}{\exp(\mathbf{x}_{2(k)} \boldsymbol{\beta}_2)} \right)^2,$$

gdzie $\mathbf{x}_{1(k)}$, $\mathbf{x}_{2(k)}$ to k -ty wiersz macierzy \mathbf{X}_1 , \mathbf{X}_2 . Gradient funkcji ℓ ma postać:

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\beta}_1^T} &= \begin{pmatrix} (y_1 - \mathbf{x}_{1(1)} \boldsymbol{\beta}_1) \exp(-2\mathbf{x}_{2(1)} \boldsymbol{\beta}_2) \\ (y_2 - \mathbf{x}_{1(2)} \boldsymbol{\beta}_1) \exp(-2\mathbf{x}_{2(2)} \boldsymbol{\beta}_2) \\ \vdots \\ (y_n - \mathbf{x}_{1(n)} \boldsymbol{\beta}_1) \exp(-2\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) \end{pmatrix} \mathbf{X}_1^T \\ &= \sum_{k=1}^n (y_k - \mathbf{x}_{1(k)} \boldsymbol{\beta}_1) \exp(-2\mathbf{x}_{2(k)} \boldsymbol{\beta}_2) \mathbf{x}_{1(k)}^T = \sum_{k=1}^n \frac{y_k - \mu_k}{\sigma_k^2} \mathbf{x}_{1(k)}^T \\ \frac{\partial \ell}{\partial \boldsymbol{\beta}_2^T} &= \begin{pmatrix} (y_1 - \mathbf{x}_{1(1)} \boldsymbol{\beta}_1)^2 \exp(-3\mathbf{x}_{2(1)} \boldsymbol{\beta}_2) - \exp(-\mathbf{x}_{2(1)} \boldsymbol{\beta}_2) \\ (y_2 - \mathbf{x}_{1(2)} \boldsymbol{\beta}_1)^2 \exp(-3\mathbf{x}_{2(2)} \boldsymbol{\beta}_2) - \exp(-\mathbf{x}_{2(2)} \boldsymbol{\beta}_2) \\ \vdots \\ (y_n - \mathbf{x}_{1(n)} \boldsymbol{\beta}_1)^2 \exp(-3\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) - \exp(-\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) \end{pmatrix} \\ &\quad \begin{pmatrix} \exp(\mathbf{x}_{2(1)} \boldsymbol{\beta}_2) & 0 & \dots & 0 \\ 0 & \exp(\mathbf{x}_{2(2)} \boldsymbol{\beta}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \exp(\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) \end{pmatrix} \mathbf{X}_2^T \\ &= \begin{pmatrix} (y_1 - \mathbf{x}_{1(1)} \boldsymbol{\beta}_1)^2 \exp(-2\mathbf{x}_{2(1)} \boldsymbol{\beta}_2) - 1 \\ (y_2 - \mathbf{x}_{1(2)} \boldsymbol{\beta}_1)^2 \exp(-2\mathbf{x}_{2(2)} \boldsymbol{\beta}_2) - 1 \\ \vdots \\ (y_n - \mathbf{x}_{1(n)} \boldsymbol{\beta}_1)^2 \exp(-2\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) - 1 \end{pmatrix} \mathbf{X}_2^T \\ &= \sum_{k=1}^n \left(\left(\frac{y_k - \mathbf{x}_{1(k)} \boldsymbol{\beta}_1}{\exp(\mathbf{x}_{2(k)} \boldsymbol{\beta}_2)} \right)^2 - 1 \right) \mathbf{x}_{2(k)}^T = \sum_{k=1}^n \left(\left(\frac{y_k - \mu_k}{\sigma_k} \right)^2 - 1 \right) \mathbf{x}_{2(k)}^T, \end{aligned}$$

z macierzą informacji:

$$\begin{aligned} \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \right] &= Dg \left(\exp(-2\mathbf{x}_{2(1)} \boldsymbol{\beta}_2), \exp(-2\mathbf{x}_{2(2)} \boldsymbol{\beta}_2), \dots, \exp(-2\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) \right) \\ \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} \right] &= 2Dg \left(\exp(-2\mathbf{x}_{2(1)} \boldsymbol{\beta}_2), \exp(-2\mathbf{x}_{2(2)} \boldsymbol{\beta}_2), \dots, \exp(-2\mathbf{x}_{2(n)} \boldsymbol{\beta}_2) \right) \\ \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \right] &= \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} \right] = \mathbf{0}. \end{aligned}$$

3.1.2. Symulacja numeryczna

Opis symulacji

Aby porównać testy istotności dla współczynników regresji bazujące na modelu liniowym z estymatorami HC[0-5] i analogicznych testów w modelu (3.1) przeprowadzona została symulacja numeryczna w 4 warunkach, gdy hipoteza zerowa jest prawdziwa/fałszywa i gdy heteroskedastyczność jest/nie jest wyjaśniona w pełni przez zmienną objaśniającą.

W symulacji wykorzystane zostaną zmienne $X_1 \sim b\left(\frac{1}{2}\right)$, $X_2 = \sin(X'_2)$, gdzie $X'_2 \sim \mathcal{U}(0, 1)$ generowane niezależnie dla każdej jednostki i niezależnie między sobą. W modelu ze zmiennymi niezależnymi X_2 , X_1X_2 interesować nas będzie testowanie efektu zmiennej X_1X_2 .

Efekty zmiennych niezależnych na zmiennej zależnej Y wyspecyfikowane będą w następujący sposób:

D1 $Y = -1 + X_2 + X_1X_2 + \exp\left(\frac{1}{2} + X_1\right)\varepsilon$, gdzie $\varepsilon \sim \mathcal{N}(0, 1)$,

D2 $Y = -1 + X_2 + X_1X_2 + \exp\left(\frac{1}{2} + X_1 + Z\right)\varepsilon$, gdzie $\varepsilon \sim \mathcal{N}(0, 1)$, $Z \sim \mathcal{N}(1, 1)$,

D3 $Y = -1 + X_2 + \exp\left(\frac{1}{2} + X_1\right)\varepsilon$, gdzie $\varepsilon \sim \mathcal{N}(0, 1)$,

D4 $Y = -1 + X_2 + \exp\left(\frac{1}{2} + X_1 + Z\right)\varepsilon$, gdzie $\varepsilon \sim \mathcal{N}(0, 1)$, $Z \sim \mathcal{N}(1, 1)$,

z ε i Z generowanymi niezależnie i niezależnie od wszystkich innych zmiennych. W symulacji z efektami D3, D4 hipoteza zerowa (braku efektu zmiennej X_1X_2 przy znajomości X_1 i X_2) jest prawdziwa a w przypadku D1, D2 jest fałszywa.

Dodatkowo macierze ograniczeń w rozważanym modelu VGLM to:

- I_2 dla wyrazu wolnego,
- $\begin{pmatrix} 0 & 1 \end{pmatrix}^T$ dla zmiennej X_1 ,
- $\begin{pmatrix} 1 & 0 \end{pmatrix}^T$ dla zmiennej X_2 ,
- $\begin{pmatrix} 1 & 0 \end{pmatrix}^T$ dla zmiennej X_1X_2 .

Wyniki symulacji

W każdej specyfikacji efektu sprawdzony zostanie test w próbach o rozmiarach $n = 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300$ z 10000 powtórzeniami dla każdej kombinacji efektu i n .

Na ilustracjach 3.1 i 3.1.2 znajdują się wyniki symulacji porównującej testy w wektorowy model liniowy (3.1) i testy w modelu z estymatorem najmniejszych kwadratów $\hat{\beta}_1$ dla β_1 i estymatorem HC0 dla Σ . Pełne wyniki znajdują się w materiałach dodatkowych A.

Dodatkowo w tabeli 3.1.2 porównana została moc testów Breuscha-Pagan dla homoskedastyczności (wybrany jako przykładowy popularny test w modelu liniowym) i test za pomocą testowania istotności współczynnika dla zmiennej X_1 .



Rysunek 3.1. Empiryczna moc (na górze) oraz empiryczny błąd pierwszego rodzaju (na dole) testu w zależności od rozmiaru próby i modelu wraz z przedziałami ufności.

$n \setminus$ Efekty	Test współczynników VGLM				Test Breuscha-Pagana			
	D1	D2	D3	D4	D1	D2	D3	D4
25	0.92	0.75	0.92	0.75	0.30	0.11	0.30	0.11
50	1.00	0.85	1.00	0.85	0.70	0.13	0.70	0.13
75	1.00	0.90	1.00	0.90	0.93	0.16	0.93	0.16
100	1.00	0.93	1.00	0.93	0.99	0.20	0.99	0.20
125	1.00	0.95	1.00	0.95	1.00	0.24	1.00	0.24
150	1.00	0.96	1.00	0.96	1.00	0.27	1.00	0.27
175	1.00	0.97	1.00	0.97	1.00	0.30	1.00	0.30
200	1.00	0.97	1.00	0.97	1.00	0.33	1.00	0.33
225	1.00	0.98	1.00	0.98	1.00	0.36	1.00	0.36
250	1.00	0.98	1.00	0.98	1.00	0.39	1.00	0.39
275	1.00	0.99	1.00	0.99	1.00	0.43	1.00	0.43
300	1.00	0.99	1.00	0.99	1.00	0.44	1.00	0.44

Tabela 3.1. Test współczynników jak i test Breuscha-Pagana bardzo szybko osiągają moc bliską 100% przy efektach $D1$ i $D3$.

Można w niej zauważyć wyraźną przewagę testu opartego na współczynniku X_1 , co jest zgodne z intuicją, że wiedza o postaci heteroskedastyczności powinna wpływać na poprawę procedury testowej. We wszystkich testach rozważany był (nominalny) poziom istotności $\alpha = 5\%$.

Na podstawie testu proporcji dla każdej pary testu bazującego na macierzy HC i modeli (3.1), (3.2) istotne różnice występują w:

- Teście bez poprawki na heteroskedastyczność przy $n = 25, 50, 75, 100$ dla $D1$, $n = 50, 75$ dla $D2$, w których to przypadkach test w modelu (3.1) ma wyższą moc i przy $n = 25$ dla $D3$, $n = 25, 50, 75$ dla $D4$ gdzie ten sam test ma większy empiryczny błąd pierwszego rodzaju.
- Teście $HC0$ przy $n = 25$ w warunkach $D2$ gdzie test w modelu (3.1) cechuje się wyższą mocą.
- Teście $HC1$ przy $n = 25$ w warunkach $D2/D3/D4$ gdzie test w modelu (3.1) cechuje się wyższą mocą i wyższym empiryczny poziom błędu 1-szego rodzaju, który dla $D3$ jest wyższy dla obu testów od 0.05 a w przypadku $D4$ nie różniącym się istotnie od 0.05.
- Teście $HC2$ przy $n = 25$ w warunkach $D2, D4$ gdzie test w modelu (3.1) cechuje się wyższą mocą i wyższym choć mniejszym niż 0.05 empiryczny poziom błędu 1-szego rodzaju.
- Teście $HC3$ przy $n = 25$ we wszystkich warunkach eksperymentalnych gdzie test bazujący na (3.1) ma większą moc i wyższy empiryczny poziom błędu 1-szego rodzaju, który w $D3$ jest dla obu testów wyższy niż 0.05 a dla $D4$ niższy niż 0.05.

- W pozostałych warunkach eksperymentalnych nie zaobserwowano istotnych różnic pomiędzy testami bazującymi na modelach (3.1), (3.2) z odpowiednimi typami macierzy odpornych na heteroskedastyczność.

Występują więc istotne różnice pomiędzy testami dla modeli (3.2), (3.1), głównie w niewielkich próbach. Testy bazujące na modelu (3.1) są z reguły bardziej liberalne, różnice jednak pomimo istotności są zazwyczaj niewielkie. Intuicyjnym wyjaśnieniem tego faktu jest to, że estymatory $\hat{\beta}_1, \hat{\beta}_2$ w modelu (3.1) są niezależne przez co poprawna specyfikacja modelu nie prowadzi do znacznego polepszenia jakości testu. Pełne wyniki można znaleźć w materiałach dodatkowych A.

3.2. Model Dixona-Colesa

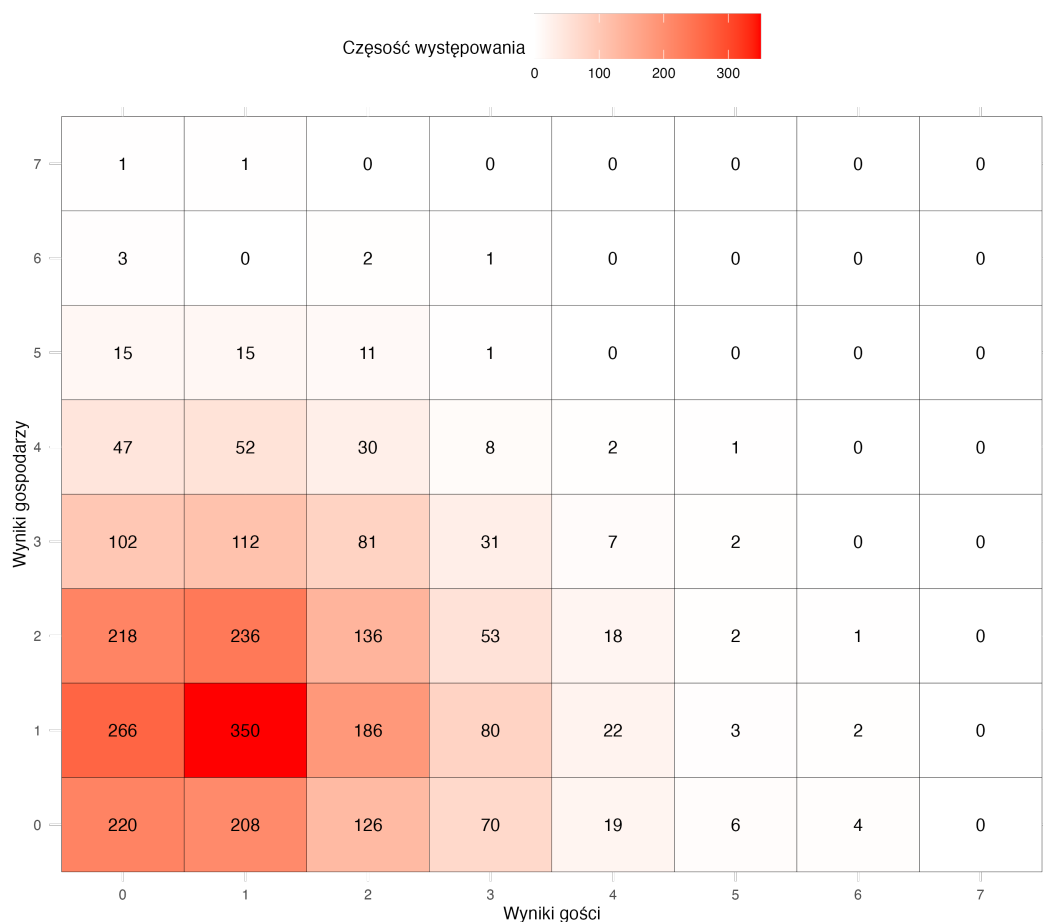
Ciekawym przykładem modeli VGLM są modele statystyczne opisujące wyniki konkurencji sportowych pomiędzy zawodnikami/drużynami jak modele Bradley'a-Terry'ego z pracy Bradley and Terry 1952, działający na systemie porównań podobnym do rankingu szachowego ELO. Modeli takich najczęściej nie da się przedstawić jako GLM.

Wykorzystamy w tym celu dane meczowe z polskiej ekstraklasy z sezonów od 2013/2014 do 2022/2023 (do dnia 6 lutego 2023). Zgromadzone zostały wyniki bramkowe spotkań w ekstraklasie oraz następujące statystyki pomeczowe:

- posiadanie piłki przez obie drużyny,
- ilość sytuacji 'krytycznych' (z wysokim prawdopodobieństwem zdobycia bramki) dla obu drużyn,
- liczba strzałów i niecelnych strzałów obu drużyn,
- liczba rzutów różnych obu drużyn,
- liczba zagrań 'spalonych' obu drużyn,
- liczba strzałów obronionych przez bramkarzy obu drużyn,
- liczba fauli, żółtych i czerwonych kartek otrzymanych przez graczy obu drużyn,
- posiadanie piłki przez obie drużyny,

na podstawie danych zebranych ze strony www.wyniki.pl. Kod do programu zbierającego dane jak i same dane dostępne są na repozytorium GitHub https://github.com/LukaszChrostowski/Football_Results_Predictions.

W pracy Maher 1982 w podobnym kontekście wykorzystany został dwuwymiarowy model Poissona, w którym założono, że składowe wektora $(X \ Y)$ reprezentującym liczbę bramek zdobytych odpowiednio przez drużynę gospodarzy i gości są niezależne. Rozkład tego wektora przedstawiony jest na ilustracji 3.2. Pomimo niespełnionych założeń model ten uznany został za adekwatny do zastosowania w zagadnieniach interesujących autorów.



Rysunek 3.2. Ilość zaobserwowanych wyników bramkowych w sezonach od 2013/2014 do 2022/2023.

Założenie takie wydawało się dość nie kontrowersyjne, ponieważ dobrze znanym zjawiskiem w literaturze dotyczącej modelowania wyników w piłce nożnej jest to, że test χ^2 dla niezależności zmiennych X i Y wskazuje zazwyczaj, że nie ma istotnych różnic w zaobserwowanym rozkładzie $(X \ Y)$ od rozkładu oczekiwanego przy niezależności X i Y . Przykładowo dla naszych danych wynik testu χ^2 z p-wartością otrzymaną przy 10^5 powtórzeniach symulacji Monte-Carlo to p-wartość w przybliżeniu 0.6218 z wartością statystyki testowej 34.519.

Pomimo tego faktu model z pracy Maher 1982 nie jest już stosowany w praktyce. Jednym z powodów jest założenie niezależności, od którego "uwolniono się" w modelu (stosowanym do dzisiaj) z pracy Dixon and Coles 2002,

w którym wprowadzona została poprawka w postaci rozkładu:

$$\mathbb{P}[(X, Y) = (x, y)] = \frac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot \frac{\lambda_2^y e^{-\lambda_2}}{y!} \cdot \tau(x, y; \lambda_1, \lambda_2, \rho), \quad (3.4)$$

gdzie $\lambda_1, \lambda_2, \rho$ są parametrami rozkładu oraz:

$$\tau(x, y; \lambda_1, \lambda_2, \rho) = \begin{cases} 1 - \lambda_1 \lambda_2 \rho & \text{gdy: } x = y = 0, \\ 1 + \lambda_1 \rho & \text{gdy: } x = 0, y = 1, \\ 1 + \lambda_2 \rho & \text{gdy: } x = 1, y = 0, \\ 1 - \rho & \text{gdy: } x = y = 1, \\ 1 & \text{w przeciwnym wypadku.} \end{cases} \quad (3.5)$$

Bezpośrednią motywacją dla autorów pracy była poprawa dwuwymiarowych modeli Poissona, które zdaniem autorów nie były zadowalające w badanym kontekście czyli strategii obstawiania wyników jak i dobierania odpowiednich wypłat przez bukmacherów.

Wprowadzenie. *Brzegowy rozkład zmiennej Y to $\mathcal{P}(\lambda_2)$, ponieważ:*

$$\begin{aligned} \mathbb{P}[Y = y] &= \sum_{k=0}^{\infty} \mathbb{P}[(X, Y) = (x, y)] \\ &= e^{-(\lambda_1 + \lambda_2)} \begin{cases} (1 - \lambda_1 \lambda_2 \rho) + \lambda_1(1 + \lambda_2 \rho) + \sum_{x=2}^{\infty} \frac{\lambda_1^x}{x!} & \text{gdy: } y = 0, \\ \lambda_2(1 + \lambda_1 \rho) + \lambda_1 \lambda_2(1 - \rho) + \lambda_2 \sum_{x=2}^{\infty} \frac{\lambda_1^x}{x!} & \text{gdy: } y = 1, \\ \frac{\lambda_2^y}{y!} \sum_{x=0}^{\infty} \frac{\lambda_1^x}{x!} & \text{gdy: } y > 1, \end{cases} \\ &= e^{-(\lambda_1 + \lambda_2)} \begin{cases} -\lambda_1 \lambda_2 \rho + \lambda_1 \lambda_2 \rho + e^{\lambda_1} & \text{gdy: } y = 0, \\ \lambda_1 \lambda_2 \rho - \lambda_1 \lambda_2 \rho + \lambda_2 e^{\lambda_1} & \text{gdy: } y = 1, \\ \frac{\lambda_2^y}{y!} e^{\lambda_1} & \text{gdy: } y > 1, \end{cases} \\ &= \frac{\lambda_2^y}{y!} e^{-\lambda_2}. \end{aligned} \quad (3.6)$$

Argument z (3.6) można poprzez symetrię zastosować do zmiennej X i analogicznie otrzymać rozkład brzegowy tej zmiennej czyli $\mathcal{P}(\lambda_1)$.

W pracy skorzystano z parametryzacji $\lambda_i = \alpha_1 \beta_j \gamma$, $\lambda_j = \alpha_j \beta_i$, które interpretowano jako:

α_i, α_j - zdolność ataku drużyny gospodarzy/gości,

β_i, β_j - zdolność obrony drużyny gospodarzy/gości,
 γ - jednorodny multiplikatywny współczynnik "przewagi własnego boiska",
 ρ - jednorodna korelacja.

Parametry α_j, β_j estymowano dla każdej drużyny z ligi metodą maksymalizacji funkcji:

$$L_t = \prod_{k:t_k < t} \mathbb{P}[(X_{i_k}, Y_{j_k}) = (x_{i_k}, y_{j_k})]^{\phi(t-t_k)}, \quad (3.7)$$

przy następującym warunku gwarantującym możliwość identyfikacji parametrów modelu:

$$\frac{1}{\nu} \sum_{j=1}^{\nu} \alpha_j = 1,$$

gdzie ν to liczba drużyn w lidze, t_k to moment w czasie w którym rozegrany został k -ty mecz a i_k i j_k to odpowiednio indeksy drużyny gospodarzy i gości w k -tym meczu i ψ to "funkcja wagowa". Funkcje L_t maksymalizowane są dla każdego t i w artykule za funkcję ϕ przyjęto $t \mapsto \exp(\xi t)$ gdzie ξ jest hiper-parametrem (w oryginalnej pracy wybrana została wartość $\xi \approx 0.0065$). Podejście takie pozwoliło na otrzymanie dynamicznych (zależnych od czasu) parametrów $\alpha_j, \beta_j, \gamma, \rho$.

Wadą tego modelu jest to, że zastosowanie wy-estymowanych na podstawie zaobserwowanych danych współczynników $\alpha_j, \beta_j, \gamma, \rho$ jest ograniczone do jednej ligi (w kontekście artykułu Dixon and Coles 2002 ligi angielskiej) oraz do jednego sezonu.

Spróbujemy zastosować model (3.4) w którym parametry λ_1, λ_2 wyjaśniane będą przez średnie z statystyk pomeczowych z ostatnich 3 meczy ekstraklasy zespołów gospodarzy i gości (które reprezentować mają formę meczową). Porównamy go z dwuwymiarowym modelem Poissona z tymi samymi zmiennymi objaśniającymi (i za pomocą analizy dewiancji sprawdzimy, które zmienne objaśniające są istotne) oraz z oryginalnym podejściem z pracy Dixon and Coles 2002 oraz tym samym podejściem rozszerzonym o uwzględnienie zmiennych niezależnych.

Model Dixona-Colesa z oryginalną metodą estymacji także może zostać przedstawiony jako VGLM ale istnieje już implementacja w języku R w pakiecie `goalmodel` (Lindström 2023). Na repozytorium `GitHub` pakietu została opisane rozszerzenie metody estymacji (3.7) dla dodatkowych zmiennych niezależnych. Jeżeli testowany model nie będzie istotnie gorszy, to będzie można także zbadać, czy uzasadnione jest "przeniesienie" modelu na inne ligi.

3.2.1. Opis modelu i sposobu porównania modeli

Algorytm IRLS wymaga znajomości gradientu i hesjanu względem β oraz macierzy informacji Fishera (względem predyktorów liniowych). Ponieważ za-

zwyczaj hesjan jak i macierz informacji mają podobną postać oraz jeżeli przyjmujemy oznaczenie $g_k(\theta_k) = \eta_k$ $k = 1, 2, \dots$ dla funkcji łączących to stosując zasadę łańcuchową dla pochodnych cząstkowych dostajemy:

$$\begin{aligned}\frac{\partial f}{\partial \eta_k} &= \frac{\partial f}{\partial \theta_k} \frac{\partial g_k^{-1}}{\partial \eta_k} \\ \frac{\partial^2 f}{\partial \eta_k^2} &= \frac{\partial^2 f}{\partial \theta_k^2} \left(\frac{\partial g_k^{-1}}{\partial \eta_k} \right)^2 + \frac{\partial f}{\partial \theta_k} \frac{\partial^2 g_k^{-1}}{\partial \eta_k^2} \\ \frac{\partial^2 f}{\partial \eta_{k_1} \partial \eta_{k_2}} &= \frac{\partial^2 f}{\partial \theta_{k_1} \partial \theta_{k_2}} \frac{\partial g_{k_1}^{-1}}{\partial \eta_{k_1}} \frac{\partial g_{k_2}^{-1}}{\partial \eta_{k_2}},\end{aligned}\tag{3.8}$$

w pracy przy opisie konkretnych modeli ograniczymy się do podania pierwszych pochodnych cząstkowych względem parametrów rozkładu oraz macierzy informacji Fishera (względem parametrów rozkładu).

Przy wyprowadzaniu macierzy informacji korzystamy z własności estymatorów największej wiarygodności $\hat{\theta}$:

$$\mathbb{E} \left[\frac{\partial \ell}{\partial \theta} (\hat{\theta}) \right] = \mathbf{0}.$$

Logarytm funkcji wiarygodności dla rozkładu zdefiniowanego w (3.4) to:

$$\begin{aligned}\ell &= x \ln \lambda_1 - \lambda_1 - \ln(\Gamma(x+1)) \\ &\quad + y \ln \lambda_2 - \lambda_2 - \ln(\Gamma(y+1)) + \ln(\tau(x, y; \lambda_1, \lambda_2, \rho)),\end{aligned}$$

Pochodne cząstkowe i elementy macierzy informacji wyrażają się przez:

$$\begin{aligned}\frac{\partial \ell}{\partial \lambda_1} &= \frac{x}{\lambda_1} - \rho \lambda_2 \mathcal{I}_{\{(0,0)\}}(x, y) + \rho \mathcal{I}_{\{(0,1)\}}(x, y), \\ \frac{\partial \ell}{\partial \lambda_2} &= \frac{y}{\lambda_2} - \rho \lambda_1 \mathcal{I}_{\{(0,0)\}}(x, y) + \rho \mathcal{I}_{\{(1,0)\}}(x, y), \\ \frac{\partial \ell}{\partial \rho} &= \frac{1}{\tau(x, y; \lambda_1, \lambda_2, \rho)} \cdot \begin{cases} -\lambda_1 \lambda_2 & \text{Gdy: } x = 0 = y \\ \lambda_1 & \text{Gdy: } x = 0 \wedge y = 1 \\ \lambda_2 & \text{Gdy: } x = 1 \wedge y = 0 \\ -1 & \text{Gdy: } x = 1 \wedge y = 1 \\ 0 & \text{W.p.p} \end{cases}\end{aligned}$$

$$\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \lambda_1^2} \right] = \frac{1}{\lambda_1} + e^{-\lambda_1 - \lambda_2} \left(\frac{\lambda_2 \rho^2}{1 + \rho \lambda_1} + \frac{\lambda_2^2 \rho^2}{1 - \rho \lambda_1 \lambda_2} \right),$$

$$\begin{aligned}
\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \lambda_1 \partial \lambda_2} \right] &= e^{-\lambda_1 - \lambda_2} \left(\rho + \frac{\rho^2 \lambda_1 \lambda_2}{1 - \rho \lambda_1 \lambda_2} \right), \\
\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \lambda_1 \partial \rho} \right] &= e^{-\lambda_1 - \lambda_2} \left(\frac{\rho \lambda_1 \lambda_2}{1 + \lambda_1 \rho} + \frac{\rho \lambda_1 \lambda_2^2}{1 - \rho \lambda_1 \lambda_2} \right), \\
\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \lambda_2^2} \right] &= \frac{1}{\lambda_2} + e^{-\lambda_1 - \lambda_2} \left(\frac{\lambda_1 \rho^2}{1 + \rho \lambda_2} + \frac{(\lambda_1 \rho)^2}{1 - \rho \lambda_1 \lambda_2} \right), \\
\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \lambda_2 \partial \rho} \right] &= e^{-\lambda_1 - \lambda_2} \left(\frac{\lambda_1 \lambda_2 \rho}{1 + \lambda_2 \rho} + \frac{\lambda_1^2 \lambda_2 \rho}{1 - \rho \lambda_1 \lambda_2} \right), \\
\mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \rho^2} \right] &= e^{-\lambda_1 - \lambda_2} \left(\frac{\lambda_1 \lambda_2}{1 - \rho \lambda_1 \lambda_2} + \frac{\lambda_1 \lambda_2}{1 + \rho \lambda_1} + \frac{\lambda_1 \lambda_2}{1 + \rho \lambda_2} + \frac{\lambda_1 \lambda_2}{1 - \rho} \right).
\end{aligned}$$

W celu porównania modelu z pracy Dixon and Coles 2002 z modelem opisanym wyżej potraktujemy zagadnienie przewidywania wyniku meczu (wygrana gospodarzy, remis, wygrana gości) jako problem klasyfikacji i porównamy, który z modeli indukuje lepszy klasyfikator. Żaden z rozważanych modeli nie przewidział, że wynik remisowy jest najbardziej prawdopodobny w którymkolwiek z meczy w zbiorze danych. Prawdopodobieństwo wystąpienia remisu wydaje się być niskie z powodów "naturalnych" dlatego zamiast z błędu ponownego podstawienia dla oceny klasyfikatorów z log-loss jako funkcji straty:

$$\begin{aligned}
&\sum_{k=1}^n \left(\ln \left(\hat{\mathbb{P}}[X > Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k > y_k\}} \right. \\
&\quad + \ln \left(\hat{\mathbb{P}}[X = Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k = y_k\}} \\
&\quad \left. + \ln \left(\hat{\mathbb{P}}[X < Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k < y_k\}} \right)
\end{aligned} \tag{3.9}$$

oraz z entropii krzyżowej:

$$\begin{aligned}
&-\sum_{k=1}^n \left(\log_2 \left(\hat{\mathbb{P}}[X > Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k > y_k\}} \right. \\
&\quad + \log_2 \left(\hat{\mathbb{P}}[X = Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k = y_k\}} \\
&\quad \left. + \log_2 \left(\hat{\mathbb{P}}[X < Y | \{(x_t, y_t)_{t=1, \dots, n}\}] \right) \mathcal{I}_{\{x_k < y_k\}} \right)
\end{aligned} \tag{3.10}$$

gdzie $\hat{\mathbb{P}}[X > Y | \{(x_t, y_t)_t\}]$, $\hat{\mathbb{P}}[X = Y | \{(x_t, y_t)_t\}]$, $\hat{\mathbb{P}}[X < Y | \{(x_t, y_t)_t\}]$ są estymowanymi brzegowymi prawdopodobieństwami odpowiednio wygranej zespołów gospodarzy, remisu i wygranej gości.

3.2.2. Wyniki

W porównaniu modelu odpowiadającego mierze probabilistycznej (3.4) do dwuwymiarowego modelu Poissona z założeniem niezależności X i Y , po dobraniu zmiennych przez kryterium informacyjne Akaike (test A3 w tabeli poniżej)

Test	χ^2	Stopnie swobody	P-wartość
A1	24.91	34	0.87
A2	29.03	36	0.79
A3	11.77	3	$8.2 \cdot 10^{-3}$

okazuje się, że założenie niezależności nie jest odpowiednie, w przeciwieństwie do tego co sugeruje test χ^2 .

Test A1 w powyższej tabeli to test porównujący model regresji zakładający miarę (3.4) i uwzględniający wszystkie zmienne niezależne do modelu (3.4) po doborze zmiennych niezależnych, podobnie test A2 porównuje dwuwymiarowy niezależny model Poissona przed i po doborze zmiennych.

Prawdopodobieństwo remisu względem rozkładu (3.4) to:

$$\begin{aligned}
\mathbb{P}[X = Y] &= e^{-\lambda_1 - \lambda_2} \left(\sum_{k=2}^{\infty} \frac{\lambda_1^k \lambda_2^k}{k!^2} + (1 - \rho)\lambda_1\lambda_2 + (1 - \lambda_1\lambda_2\rho)\lambda_1\lambda_2 \right) \\
&= e^{-\lambda_1 - \lambda_2} \left(I_0(2\sqrt{\lambda_1\lambda_2}) - \lambda_1\lambda_2 - 1 \right) \\
&\quad + e^{-\lambda_1 - \lambda_2} ((1 - \rho)\lambda_1\lambda_2 + (1 - \lambda_1\lambda_2\rho)\lambda_1\lambda_2),
\end{aligned}$$

gdzie $I_\nu(t)$ to zmodyfikowana funkcja Bessela pierwszego rodzaju. Prawdopodobieństwo $\mathbb{P}[X \geq Y]$ dla danych $\lambda_1, \lambda_2, \rho$ przybliżone zostanie numerycznie za pomocą standardowego algorytmu. Poniższa tabela prezentuje wartości entropii krzyżowych zdefiniowanych w (3.10) oraz log-loss dla klasyfikatorów zbudowanych na podstawie rozważanych modeli:

Kryterium	Oryginalny model Dixona-Colesa z dobiorem zmiennych	Oryginalny model Dixona-Colesa	Model (3.4)
log-loss	-2867.6	-2864.2	-2922.5
cross-entropy	4137.0	4132.2	4216.3
accuracy	0.4707	0.4729	0.4537

Tabela 3.2. Porównanie klasyfikatorów wynikających z rozważanych modeli.

Oryginalny model Dixona-Colesa wypada najlepiej w tym porównaniu i co ciekawe lepiej niż ten sam model z dodatkowymi optymalnie dobranymi (względem kryterium AIC) zmiennych objaśniających.

Klasyfikatory zostały przetestowane na danych uczących, należy zatem zastanowić się nad tym czy porównanie nie jest niepoprawne w wyniku wystąpienia przeuczenia modelu. Niemniej jednak różnica w tabeli jest na tyle duża,

że rozsądnym wydaje się uznanie modelu Dixona-Colesa za najlepszy z tutaj rozważanych. Dodatkowo nie jest oczywistym, czy model (3.4) ze statystykami pomeczowymi jest faktycznie modelem "prostszy".

Dodatkowo najbardziej prawidłowym porównaniem powinno być wyuczenie modelu po każdej kolejce, następnie wyliczenia prognozy w.g. każdego z tych modeli i powtórzenie tej procedury tak długo jak dostępne są dane (zazwyczaj przewidywane są wyniki tylko na najbliższe spotkania).

Nawet pomimo tych zastrzeżeń wydaje się, że model Dixona-Colesa nie może zostać zastąpiony modelem, który dałoby się przenieść na inne ligi.

Jednym z celów tej analizy było zbadanie, które ze zmiennych objaśniających mają istotny wpływ na wyniki. Dobrane zmienne zaprezentowane są w poniższej tabeli (zmienne odrzucone ze wszystkich modeli nie zostały wymienione):

Zmienne tylko w modelu D-C	Zmienne wspólne	Zmienne w modelu (3.4)
Zespół gospodarzy Zespół gości Liczba czerwonych kartek gości	Posiadanie piłki gospodarzy Liczba rzutów różnych gospodarzy Liczba interwencji bramkarza gospodarzy Posiadanie piłki gości Liczba sytuacji krytycznych stworzonych przez gości Liczba bramek straconych przez gospodarzy Liczba bramek gospodarzy	Liczba strzałów gości Liczba bramek gości Liczba interwencji bramkarza gości Liczba bramek straconych przez gości

Pełna tabela analizy dewiancji dla modelu (3.4) znajduje się w materiałach dodatkowych B. Zmienne objaśniające są silnie skorelowane więc do ich istotności należy podchodzić sceptycznie.

3.3. Inne popularne godne uwagi przykłady

Z powodów objętościowych w pracy nie można omówić wszystkich interesujących przykładów modeli VGLM dlatego ograniczymy się do wymienienia kilku wybranych modeli i rozszerzeń VGLM:

- Zostając w temacie modelowania wyników w piłce nożnej proces INGARCH (ang. *Integer Generalized Autoregressive Conditional Heteroskedasticity*, *Integer GARCH*), który został opisany razem z innymi modelami dla szeregów czasowych jak np. GARCH w kontekście klasy VGLM w dysertacji doktorskiej Soberanis 2019 (rozdział drugi), został wykorzystany w pracy Cerqueti et al. 2022 do analizy skupień zespołów piłkarskich za pomocą

wyników spotkań. Proces INGARCH jest uogólnieniem procesu GARCH na przypadek szeregów czasowych dla danych dyskretnych.

- Modele o zredukowanym rzędzie (ang. *reduced rank*) opisane w piątym rozdziale książki Thomas W. Yee 2015 pozwalające na używanie większej liczby zmiennych objaśniających p niż danych na parametr rozkładu n/M . Dokonuje się tego przez przekształcenie określonej części zmiennych objaśniających za pomocą kombinacji liniowych w mniejszą ilość nowych zmiennych, które można traktować jak zmienne ukryte (ang. *latent variables*).
- Model o zmiennych współczynnikach regresji (ang. *Varying coefficient model*), w którym współczynniki regresji β mogą zależeć od dodatkowych zmiennych objaśniających \mathbf{u} , to znaczy pełna macierz predyktorów liniowych η spełnia:

$$\text{vect}(\eta) = \mathbf{X}_{\text{VLM}}\beta(\mathbf{u}) + \mathbf{o},$$

dokładny opis można znaleźć w artykule Hastie and Tibshirani 1993 oraz w rozdziale dziesiątym książki Thomas W. Yee 2015.

- Rozkłady ocenzurowane $\max(\min(Y, \tau_-), \tau^+)$ gdzie τ_-, τ^+ są traktowane jako hiper-parametry naturalnie wpisują się w teorię VGLM. Szczególnym przypadkiem jest na model *tobit* gdzie zaobserwowana jest zmienna:

$$\mathbf{Y} = \min(\mathbf{X}^T\beta + \varepsilon, \mathbf{0}), \quad \varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n),$$

gdzie minimum aplikowane jest do każdego elementu wektora. Tobit jest wykorzystywany np. przy badaniu wpływu różnych charakterystyk \mathbf{x} osób lub gospodarstw domowych na wydatki na "dobra luksusowe", które mogą być zakupione dopiero gdy dostęp do dóbr pierwszej potrzeby jest zapewniony.

- Różne podejścia do problemu regresji kwantylowej i ekspektylowej (ang. *expectile regression*) za pomocą VGLM są tematem 15 rozdziału książki Thomas W. Yee 2015.

Estymacja wielkości populacji

4.1. Wprowadzenie i motywacja

Wektorowe uogólnione modele liniowe można wykorzystać w zagadnieniu estymacji wielkości populacji na podstawie obserwacji tylko jej części, metodą *capture-recapture* zarówno w sytuacji gdy operujemy tylko jednym lub wieloma źródłami danych. W sytuacji wieloźródłowej zagadnienie to przedstawiane jest jako estymacja pełnej tabeli kontyngencji dla całej populacji zaprezentowanej w tabeli 4.1. W sytuacji gdy nie operujemy zmiennymi niezależnymi pełna (odpowiednio nie pełna) tabela kontyngencji jest statystyką dostateczną i zupełną dla całej skończonej populacji, rozumianej jako próba z super-populacji (odpowiednio dla próby zawierającej tylko jednostki zaobserwowane przynajmniej raz). Podobna sytuacja występuje gdy operujemy tylko zmiennymi kategorycznymi wtedy prezentuje się osobne tabele kontyngencji, dla każdego poziomu (wielowymiarowej) zmiennej niezależnej.

Źródło	Źródło pierwsze		I_1
drugie I_2	Z ($I_1 = 1$)	Nz ($I_1 = 0$)	Σ
Z ($I_2 = 1$)	n_{11} (p_{11})	n_{10} (p_{10})	$n_{1\cdot}$ ($p_{1\cdot}$)
Nz ($I_2 = 0$)	n_{01} (p_{01})	n_{00} (p_{00})	$n_{0\cdot}$ ($p_{0\cdot}$)
Σ	$n_{\cdot 1}$ ($p_{\cdot 1}$)	$n_{\cdot 0}$ ($p_{\cdot 0}$)	N

Tabela 4.1. Standardowa tabela kontyngencji, w badaniach capture-recapture n_{00} jest nieobserwowalne. Oznaczenia Z, Nz oznaczają zaobserwowaną/nie zaobserwowaną część populacji.

Jedną z najprostszych i najstarszych metod estymacji wielkości n_{00} i przez to też $N = n_{00} + n_{01} + n_{10} + n_{11}$ jest estymator Lincolna-Petersena³. Załóżmy, że pierwsze źródło zostało zebrane pierwsze w czasie i spośród N elementowej

³ Nazwa pochodzi od nazwisk dwóch statystyków, którzy wykorzystali ten estymator w swoich pracach ale sama idea wykorzystana była znacznie wcześniej na przykład przez Laplace'a w pracy Laplace 1786, (która została przetłumaczona na język angielski).

populacji zaobserwowane zostało n_1 . jednostek natomiast w źródle 2 zaobserwowaliśmy $n_{11} + n_{01} = n_{\cdot 1}$ jednostek n_{11} będących w źródle pierwszym wtedy przy założeniu równego prawdopodobieństwa obserwacji dla każdej jednostki przy próbkowaniu w każdym ze źródeł i założeniu niezależności pomiędzy inkluzją w pierwszym i drugim źródle otrzymujemy:

$$\frac{n_{11}}{n_{\cdot 1}} \approx \frac{n_{1\cdot}}{N} \implies N \approx \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n_{11}}. \quad (4.1)$$

Powyższy wynik można otrzymać w bardziej rygorystyczny sposób metodą największej wiarygodności. Estymator (4.1) ma znaczące problemy, nawet pomijając to, że nie istnieje z dodatnim prawdopodobieństwem, występuje tutaj nie realistyczne w większości zastosowań założenie niezależności inkluzji. Znaczna część literatury o metodzie capture-recapture poświęcona jest sposobom weryfikacji i metod estymacji nie wymagających tego założenia, bo jak się okazuje złamanie tego założenia ma znaczny wpływ na własności estymatora.

Skupimy się na metodzie wieloźródłowej ale warto zaznaczyć także, że w przypadku metody jednoźródłowej także wykorzystuje się VGLM. W tej metodzie brakuje bezpośrednich metod oszacowania prawdopodobieństwa inkluzji w próbie. jeżeli operujemy tylko informacją o inkluzji, dlatego w metodzie jednoźródłowej metodę tą wykorzystuje się gdy możliwe jest pojawianie się jednostek w próbie kilkakrotnie. To znaczy obserwujemy zmienną Y zliczającą ilość obserwacji w dostępnym źródle, a próbą jest ta część populacji dla której $Y \geq 1$, zobacz Cruyff and P. G. M. v. d. Heijden 2008; Peter GM van der Heijden et al. 2003; Böhning, Vidal-Diez, et al. 2013; Böhning and P. G. M. v. d. Heijden 2009; Chao 1989. Występuje jednak naturalnie problem związany z tym, że w przypadku gdy korzystamy na przykład z rejestrów policyjnych obserwacja jednostki może wiązać się z pozbawieniem jej wolności co skutkować będzie brakiem możliwości dalszej obserwacji tejże jednostki.

Najprostszym rozwiązaniem, rozważanym na przykład w Böhning and P. G. M. v. d. Heijden 2019; Tuoto, Di Cecco, and Tancredi 2022 i wielu innych, jest rozważenie nowej miary probabilistycznej \mathbb{P}^* z dodatkowym parametrem $\omega \in (0, 1)$ (nie koniecznie jednakowym w całej populacji) postaci:

$$\begin{aligned} \mathbb{P}^*[Y = y; \omega] &= \omega \mathcal{I}_{\{1\}}(y) + (1 - \omega) \mathbb{P}[Y = y] \text{ oraz gdy: } y > 0 : \\ \mathbb{P}^*[Y = y | Y > 0; \omega] &= \frac{\omega \mathcal{I}_{\{1\}}(y)}{1 - (1 - \omega) \mathbb{P}[Y = 0]} + \frac{(1 - \omega) \mathbb{P}[Y = y]}{1 - (1 - \omega) \mathbb{P}[Y = 0]}, \end{aligned}$$

gdzie \mathbb{P} oznacza miarę nie uwzględniającą nadmiaru jedynek.

4.2. Metoda wieloźródłowa

Zastosowania wektorowych uogólnionych modeli liniowych (oraz ich rozszerzeń addytywnych), w kontekście estymacji metodą capture-recapture zostały

omówione na przykład w pracy Thomas W. Yee, Stoklosa, and Huggins 2015, zagadnienie to przedstawione zostanie w podobny sposób. Opisana została także implementacja wybranych metod w pakiecie VGAM języka R (funkcje w pakiecie VGAM odpowiadające zaprezentowanym modelom to `posbernoulli.t`, `posbernoulli.b`, `posbernoulli.tb`).

4.2.1. Modele podstawowe

Wszystkie metody estymacji w modelach capture-recapture (w zamkniętych populacjach, bez urodzeń, śmierci, imigracji i emigracji) odwołują się w pewien sposób do czterech podstawowych modeli wyróżnionych w przeglądzie literatury Wolter 1986 (w kontekście estymacji wielkości populacji nie pokrytej spisem powszechnym) i uogólnione w artykule Thomas W. Yee, Stoklosa, and Huggins 2015 na przypadek większej ilości źródeł.

Niech τ oznacza liczbę źródeł a będzie \mathbf{Y} wektorem określającym "historię złapań", to znaczy obecność lub jej brak jednostki w odpowiednich źródłach danych. Podczas omawiania metody capture-recapture zakładamy, że inkluzje jednostek są warunkowo niezależne między sobą to znaczy rozkłady $\mathbf{Y}_{(i)}$, $\mathbf{Y}_{(j)}$ są niezależne warunkowo przy \mathbf{X} gdy $i \neq j$. Podstawowe modele w capture-recapture to:

\mathcal{M}_0 Model zerowy, w którym zakładamy, że inkluzja w źródle j jest niezależna od inkluzji w innych źródłach a prawdopodobieństwo inkluzji w każdym ze źródeł jest równe p dla każdej jednostki. Miara probabilistyczna odpowiadająca modelowi \mathcal{M}_0 (parametryzowana przez $p \in (0, 1)$):

$$\mathbb{P}[\mathbf{Y} = (y_1, \dots, y_\tau)] = \prod_{j=1}^{\tau} p^{y_j} (1-p)^{1-y_j} \quad (4.2)$$

\mathcal{M}_t Model z efektem czasowym, w którym nie zakładamy równości prawdopodobieństw inkluzji w każdym ze źródeł. Miara probabilistyczna odpowiadająca modelowi \mathcal{M}_t (parametryzowana przez $\mathbf{p} \in (0, 1)^\tau$):

$$\mathbb{P}[\mathbf{Y} = (y_1, \dots, y_\tau)] = \prod_{j=1}^{\tau} p_j^{y_j} (1-p_j)^{1-y_j} \quad (4.3)$$

\mathcal{M}_b Model z efektem behawioralnym, w którym nie zakładamy niezależności pomiędzy inkluzjami w poszczególnych źródłach. Miara probabilistyczna odpowiadająca modelowi \mathcal{M}_b (parametryzowana przez $\mathbf{p} \in (0, 1)^2$):

$$\mathbb{P}[\mathbf{Y} = (y_1, \dots, y_\tau)] = \prod_{j=1}^{\tau} p_{k_j}^{y_j} (1-p_{k_j})^{1-y_j} \quad (4.4)$$

gdzie:

$$k_j = \begin{cases} 1 & \text{gdy: } j = 1 \\ 1 & \text{gdy: } y_1 = \dots = y_{j-1} = 0 \\ 2 & \text{gdy: } \exists i < j : y_i > 0 \end{cases}$$

to znaczy prawdopodobieństwo ponownego zaobserwowania zależy od tego czy została już zaobserwowana przynajmniej raz.

\mathcal{M}_{tb} Model z efektem czasowym i efektem behawioralnym, który łączy modele \mathcal{M}_t i \mathcal{M}_b . Miara probabilistyczna odpowiadająca modelowi \mathcal{M}_b (parametryzowana przez dwa wektory $\mathbf{p}, \mathbf{p}' \in (0, 1)^\tau$, gdzie $p_1 = p'_1$):

$$\mathbb{P}[\mathbf{Y} = (y_1, \dots, y_\tau)] = \prod_{j=1}^{\tau} \begin{cases} (p'_j)^{y_j} (1 - p'_j)^{1-y_j} & \text{gdy: } \exists i < j : y_i > 0 \\ p_j^{y_j} (1 - p_j)^{1-y_j} & \text{w.p.p} \end{cases} \quad (4.5)$$

to znaczy oprócz różnego efektu czasowego efekt behawioralny różni się w czasie.

W modelach uwzględniających efekt behawioralny istotne (w kontekście interpretacji) może być poprawne uporządkowanie składowych wektora \mathbf{Y} względem czasu, szczególnie w zastosowaniach biologicznych. W zastosowaniach biologicznych metody źródła danych zazwyczaj są kolejnymi wyprawami badawczymi na dany obszar, w których zwierzęta danego gatunku są odnotowywane lub naznaczone (na przykład za pomocą obrączki u nogi) a celem jest estymacja ilości osobników danego gatunku (lub gatunków) w danym ekosystemie.

Nazwa modelu \mathcal{M}_b pochodzi właśnie z zastosowań biologicznych, jeżeli tak jak w przypadku danych administracyjnych źródeł danych nie da się uporządkować względem czasu modele \mathcal{M}_b i \mathcal{M}_{tb} nadal są bardzo użyteczne ale "zmiany" w prawdopodobieństwie z wartości p_1 na któreś p_j nie można interpretować jako efekt behawioralny. W takich sytuacjach modele $\mathcal{M}_b, \mathcal{M}_{tb}$ interpretuje się jako modele uwzględniające korelacje pomiędzy źródłami (to znaczy pomiędzy inkluzjami w źródłach danych).

Analogiczna sytuacja występuje w modelu \mathcal{M}_t z efektem czasowym, który przy braku uporządkowania czasowego można interpretować jako model z efektem konkretnego źródła. To znaczy gdzie prawdopodobieństwo inkluzji w j -tym źródle wynika ze sposobu próbkowania.

4.2.2. Podział metod estymacji częstościowej

Wyróżnimy trzy następujące ogólne metody estymacji w paradygmacie częstościowym bazujące na funkcji wiarygodności:

- Metody bazujące na "wiarygodności pełnej" (*ang. full likelihood*), w której modelowana jest funkcja wiarygodności pełnej skończonej populacji (rozumianej jako próba z super-populacji), przy N , lub równoważnie n_{00} , trakto-

wanym jako parametr rozkładu razem z prawdopodobieństwami a składowe statystyki dostatecznej z próby traktowane są jako dane.

Przykładowo dla modelu $\mathcal{M}_0, \tau = 2$ logarytm funkcji wiarygodności bez zmiennych niezależnych, to:

$$\ell(p, n_{00}) = (2n_{11} + n_{10} + n_{01}) \ln(p) + (n_{10} + n_{01} + 2n_{00}) \ln(1 - p) + \ln \binom{n_{11} + n_{10} + n_{01} + n_{00}}{n_{11}, n_{10}, n_{01}, n_{00}},$$

a pełna wielkość populacji szacowana jest za pomocą estymatora największej wiarygodności dla n_{00} .

- Metody bazujące na "wiarygodności obserwowanej" (*ang. conditional likelihood*) zwanej też "ocenzuowaną wiarygodnością" (*ang. censored likelihood*), w której jako próbę traktujemy zaobserwowaną populację:

$$\{(\mathbf{y}_i, \mathbf{x}_i) \mid \exists j = 1, \dots, \tau : y_{ij} > 0\} = \{(\mathbf{y}_i, \mathbf{x}_i) \mid \mathbf{y}_i \neq \mathbf{0}\},$$

na jej podstawie estymowane jest warunkowe prawdopodobieństwo inkluzji w próbie:

$$\mathbb{P}[\mathbf{Y}_{(i)} \neq \mathbf{0} \mid \mathbf{X}_{(i)}] =: G(\mathbf{X}_{(i)}),$$

za pomocą techniki największej wiarygodności, gdzie logarytm funkcji wiarygodności w modelu $\mathcal{M}_0, \tau = 2$ to bez zmiennych objaśniających to:

$$\ell(p) = n_{11} \ln \left(\frac{p^2}{1 - (1 - p)^2} \right) + (n_{10} + n_{01}) \ln \left(\frac{p(1 - p)}{1 - (1 - p)^2} \right).$$

Pełna populacja szacowana jest na podstawie estymatora Horwitza Thompsona Horvitz and Thompson 1952:

$$\hat{N} = \sum_{i=1}^N \frac{\mathcal{I}(\exists j = 1, \dots, \tau : y_{ij} > 0)}{1 - G(\mathbf{X}_{(i)})} = \sum_{i=1}^n \frac{1}{1 - G(\mathbf{X}_{(i)})}.$$

Dla modelu $\mathcal{M}_0, \tau = 2$ zmiennych objaśniających sprowadza się do:

$$\hat{N} = \frac{n_{11} + n_{10} + n_{01}}{(1 - (1 - \hat{p})^2)}, \quad \hat{p} = \frac{2n_{11}}{2n_{11} + n_{10} + n_{01}}.$$

- Modele logarytmiczno liniowe (*ang. log-linear models*) zaprezentowane po raz pierwszy w pracy Fienberg 1972 są alternatywą dla modeli wiarygodności pełnej, w których jako zmienne objaśniane traktujemy składowe statystyki dostatecznej, jako zmienne objaśniające traktujemy to czy dana składowa odnosi się do j -tego źródła a cały problem traktujemy jako regresję Poissona. To znaczy rozważamy w ogólności model:

$$n_{i_1 i_2, \dots, i_\tau} \sim \mathcal{P}(\lambda_{i_1, \dots, i_\tau})$$

$$\begin{aligned} \ln(\lambda_{i_1, \dots, i_\tau}) &= \beta_0 + \sum_{k=1}^{\tau} \beta_k \mathcal{I}_{\{i_k=1\}} + \sum_{j=2}^{\tau} \sum_{k=1}^{j-1} \beta_{jk} \mathcal{I}_{\{i_k=1\}} \mathcal{I}_{\{i_j=1\}} + \dots \\ &\quad + \beta_{123\dots\tau} \prod_{k=1}^{\tau} \mathcal{I}_{\{i_k=1\}}, \end{aligned} \quad (4.6)$$

a nieznana wartość $n_{00\dots 0}$ estymowana jest na podstawie prognozy:

$$\hat{n}_{00\dots 0} = \lambda_{00\dots 0} = \exp(\beta_0).$$

W ogólności model (4.6) zawiera więcej parametrów β_0, \dots , których jest:

$$1 + \tau + \sum_{j=2}^{\tau} (j-1) + \sum_{j=3}^{\tau} \sum_{k=2}^{j-1} (k-1) + \dots + 1,$$

niż danych $n_{i_1 i_2 \dots i_\tau}$, których mamy $2^\tau - 1$. Rola założeń parametrycznych w kontekście modeli log-liniowych to ograniczenie liczby estymowanych parametrów. Przykładowo w modelu $\mathcal{M}_0, \tau = 2$ używając przybliżenia rozkładu dwumianowego rozkładem Poissona dostajemy:

$$\begin{aligned} \lambda_{00} &= N(1-p)^2, \quad \lambda_{10} = Np(1-p) = \lambda_{01}, \quad \lambda_{11} = Np^2 \\ \beta_0 &= \ln(N) + 2\ln(1-p) \\ \beta_0 + \beta_1 &= \beta_0 + \beta_2 = \ln(N) + \ln(p) + \ln(1-p) \\ \beta_0 + \beta_{11} &= \ln(N) + 2\ln(p), \end{aligned}$$

czyli model sprowadza się do estymacji parametrów β_0, β_1 na podstawie danych n_{10}, n_{01}, n_{11} :

$$n_{10}, n_{01} \sim \mathcal{P}(\exp(\beta_0 + \beta_1)), \quad n_{11} \sim \mathcal{P}(\exp(\beta_0 + 2\beta_1^{-1})).$$

Pierwotnie najczęściej badanym modelem *capture-recapture* był model \mathcal{M}_t w przypadku $\tau = 2$ a sam przypadek $\tau = 2$ do dzisiaj pokrywa sporą część prac praktycznych z zakresu *capture-recapture*. Z powodu historycznej istotności modelu $\mathcal{M}_t, \tau = 2$ wyprowadzimy w jego kontekście estymator (4.1).

Wyprowadzenie. Dla pełnej populacji $\{\mathbf{y}_j | j = 1, \dots, N\}$ (której nie obserwujemy w całości) logarytm (pełnej) funkcji wiarygodności to (korzystając z notacji z tabeli 4.1):

$$\begin{aligned} \ell(p_1, p_2, n_{00}) &= \log \binom{N}{n_{11}, n_{10}, n_{01}, n_{00}} + n_{11} \log(p_1 p_2) + n_{10} \log(p_1 (1-p_2)) \\ &\quad + n_{01} \log((1-p_1) p_2) + n_{00} \log((1-p_1)(1-p_2)). \end{aligned}$$

Pamiętając, że N jest funkcją n_{00} oraz przy założeniu, że $n_{11}, n_{10}, n_{01} > 0$ otrzymujemy:

$$\begin{aligned}\frac{\partial \ell}{\partial p_1} &= \frac{n_{1.}}{p_1} - \frac{n_{0.}}{1-p_1} = 0 \Leftrightarrow \frac{p_1}{1-p_1} = \frac{n_{1.}}{n_{0.}} \Leftrightarrow p_1 = \frac{n_{1.}/n_{0.}}{1+n_{1.}/n_{0.}} = \frac{n_{1.}}{N} \\ \frac{\partial \ell}{\partial p_2} &= \frac{n_{.1}}{p_2} - \frac{n_{.0}}{1-p_2} = 0 \Leftrightarrow \frac{p_2}{1-p_2} = \frac{n_{.1}}{n_{.0}} \Leftrightarrow p_2 = \frac{n_{.1}/n_{.0}}{1+n_{.1}/n_{.0}} = \frac{n_{.1}}{N} \\ \frac{\partial \ell}{\partial n_{00}} &= \log(1-p_1) + \log(1-p_2) + \psi(N+1) - \psi(n_{00}+1) = 0 \\ &\Leftrightarrow \psi(N+1) - \psi(n_{00}+1) = 2\log(N) - \log(n_{0.}n_{.0}),\end{aligned}$$

ostatnie z tych równań jest bardzo trudne do rozwiązania metodami analitycznymi dlatego tradycyjnie używa się przybliżenia dla liczb harmoniczych (oznaczonych przez H_t):

$$\psi(N+1) - \psi(n_{00}+1) = H_N - H_{n_{00}} \approx \log(N) - \log(n_{00}),$$

rozwiązanie przybliżonego równania nie jest już problemem:

$$2\log(N) - \log(n_{0.}n_{.0}) = \log(N) - \log(n_{00}) \Leftrightarrow n_{00} = \frac{n_{10}n_{01}}{n_{11}},$$

co po uwzględnieniu, że $N = n_{00} + n_{11} + n_{01} + n_{10}$ daje nam estymator (4.1). Pozostało sprawdzić, że Hessian funkcji ℓ :

$$- \begin{pmatrix} \frac{n_{1.}}{p_1^2} + \frac{n_{0.}}{(1-p_1)^2} & 0 & -\frac{1}{1-p_1} \\ 0 & \frac{n_{.1}}{p_2^2} + \frac{n_{.0}}{(1-p_2)^2} & -\frac{1}{1-p_2} \\ -\frac{1}{1-p_1} & -\frac{1}{1-p_2} & \psi^{(1)}(n_{00}+1) - \psi^{(1)}(N+1) \end{pmatrix}$$

jest ujemnie określony. Ponieważ $\psi^{(1)}$ jest funkcją malejącą o wartościach dodatnich Hessian jest ujemnie określony dla dowolnych $p_1, p_2 \in (0, 1), n_{00} \in \mathbb{R}_+$.

Estymator (4.1) ma bardzo pożądane własności asymptotyczne o ile tylko założenie niezależności dwóch źródeł jest spełnione, co jest zaprezentowane w następującym twierdzeniu:

Twierdzenie. Estymator (4.1) jest estymatorem silnie zgodnym, w tym sensie, że:

$$\frac{\hat{N}}{N} \xrightarrow{a.s.} 1 \text{ przy: } N \rightarrow \infty. \quad (4.7)$$

Dowód. Skorzystamy z własności przestrzeni probabilistycznych (i ogólniej przestrzeni z miarami skończonymi):

$$X_n \xrightarrow{a.s.} X, Y_n \xrightarrow{a.s.} Y \implies X_n Y_n \xrightarrow{a.s.} XY.$$

Ponieważ wyrażenie w (4.7) dopuszcza przedstawienie:

$$\frac{\hat{N}}{N} = \frac{(n_{1\cdot}/N)(n_{\cdot 1}/N)}{n_{11}/N},$$

a z prawa wielkich liczb Bernoullego $n_{ij}/N \rightarrow p_{ij}$ prawie na pewno mamy więc:

$$\begin{aligned} \frac{\hat{N}}{N} &\xrightarrow{\text{a.s.}} \frac{p_{1\cdot}p_{\cdot 1}}{p_{11}} = \frac{p_{11}^2 + p_{11}p_{10} + p_{11}p_{01} + p_{10}p_{01}}{p_{11}} = p_{11} + p_{10} + p_{01} + \frac{p_{10}p_{01}}{p_{11}} \\ &= p_{11} + p_{10} + p_{01} + \frac{p_1(1-p_2)(1-p_1)p_2}{p_1p_2} = p_{11} + p_{10} + p_{01} + p_{00} = 1. \end{aligned}$$

□

Jest to także dobre miejsce by zaprezentować jedną z wad metody wiarygodności obserwowanej, czyli złożoność obliczeniową. W tym samym, dosyć prostym modelu probabilistycznym, z następującej postaci funkcji wiarygodności:

$$\begin{aligned} \ell &= n_{1\cdot} \ln(p_1) + n_{\cdot 1} \ln(p_2) + n_{01} \ln(1-p_1) + n_{10} \ln(1-p_2) \\ &\quad + (n_{11} + n_{10} + n_{01}) \ln(1 - (1-p_1)(1-p_2)) \\ \frac{\partial \ell}{\partial p_1} &= \frac{n_{1\cdot}}{p_1} - \frac{n_{01}}{1-p_1} + \frac{(n_{11} + n_{10} + n_{01})(1-p_2)}{1 - (1-p_1)(1-p_2)} \\ \frac{\partial \ell}{\partial p_2} &= \frac{n_{\cdot 1}}{p_2} - \frac{n_{10}}{1-p_2} + \frac{(n_{11} + n_{10} + n_{01})(1-p_1)}{1 - (1-p_1)(1-p_2)}, \end{aligned}$$

wynika, że estymatory dla parametrów p_1, p_2 to rozwiązanie następującego układu równań:

$$\begin{aligned} \hat{p}_1 &= \left(\left((n_{11}^2 + (2n_{10} + 4n_{01})n_{11} + n_{10}^2 + 4n_{01}n_{10} + 4n_{01}^2) \hat{p}_2^2 \right. \right. \\ &\quad + (-4n_{11}^2 + (-8n_{10} - 6n_{01})n_{11} - 4n_{10}^2 - 6n_{01}n_{10} - 4n_{01}^2) \hat{p}_2 \\ &\quad + 4n_{11}^2 + (8n_{10} + 4n_{01})n_{11} + 4n_{10}^2 + 4n_{01}n_{10} + n_{01}^2 \Big)^{1/2} \\ &\quad \left. + (3n_{11} + 3n_{10} + 2n_{01}) \hat{p}_2 - 2n_{11} - 2n_{10} - n_{01} \right) \\ &\quad \Big/ ((4n_{11} + 4n_{10} + 4n_{01}) \hat{p}_2 - 4n_{11} - 4n_{10} - 4n_{01}) \\ \hat{p}_2 &= \left((n_{11}^2 + (4n_{10} + 2n_{01})n_{11} + 4n_{10}^2 + 4n_{01}n_{10} + n_{01}^2) \hat{p}_1^2 \right. \\ &\quad \left. + (-4n_{11}^2 + (-6n_{10} - 8n_{01})n_{11} - 4n_{10}^2 - 6n_{01}n_{10} - 4n_{01}^2) \hat{p}_1 \right. \end{aligned} \tag{4.8}$$

$$\begin{aligned}
& + 4n_{11}^2 + (4n_{10} + 8n_{01})n_{11} + n_{10}^2 + 4n_{01}n_{10} + 4n_{01}^2 \\
& + (3n_{11} + 2n_{10} + 3n_{01})\hat{p}_1 - 2n_{11} - n_{10} - 2n_{01} \Big) \\
& \Big/ ((4n_{11} + 4n_{10} + 4n_{01})\hat{p}_1 - 4n_{11} - 4n_{10} - 4n_{01}),
\end{aligned}$$

z estymatorem:

$$\hat{N} = \frac{n}{1 - (1 - \hat{p}_1)(1 - \hat{p}_2)}.$$

4.2.3. Porównanie podejść w metodzie wieloźródłowej

- Metody log-liniowe oraz pełnej wiarygodności są równoważne co jest dobrze znanym faktem. Zgodność estymatorów bada się zazwyczaj w każdym modelu osobno, ponieważ nie spełnione są w niej warunki regularności Cramera-Rao.
- Własności asymptotyczne estymatorów największej wiarygodności pełnej zostały omówione dokładnie między innymi w pracy Sanathanan [1972](#) i można je znaleźć w większościach monografii i podręczników o capture-recapture np. w Bohning, Peter G.M. Heijden, and Bunge [2018](#).
- Estymacja wariancji estymatora wielkości populacji w metodzie "wiarygodności pełnej" jest dosyć problematyczne, uniwersalne estymatory w tej metodzie to tylko estymatory bootstrapowe (opisane np. w Norris and Pollock [1996](#); Zwane and Van der Heijden [2003](#)). Bootstrap (a w zasadzie kilka algorytmów bootstrapowych) jest zdecydowanie najczęściej wykorzystywaną metodą estymacji $\mathbb{V}[\hat{N}]$, inne metody wyprowadzane są najczęściej dla konkretnych postaci \hat{N} .
- Zgodność estymatorów zbudowanych na podstawie wiarygodności obserwowanej otrzymuje się za pomocą standardowego argumentu dla estymatorów największej wiarygodności:

Wyprowadzenie. *Estymator Horwitza-Thompsona przy znanych współczynnikach regresji jest nieobciążony:*

$$\begin{aligned}
\mathbb{E}[\hat{N}|\mathbf{X}] &= \mathbb{E}\left[\sum_{k=1}^N \frac{\mathcal{I}_k}{1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0}|\mathbf{X}_{(k)}]} \Big| \mathbf{X}\right] \\
&= \sum_{k=1}^N \frac{\mathbb{E}[\mathcal{I}_k|\mathbf{X}]}{1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0}|\mathbf{X}_{(k)}]} = N,
\end{aligned}$$

gdzie $\mathcal{I}_k = \mathcal{I}(\mathbf{Y}_{(k)} \neq \mathbf{0})$, więc aby zapewnić zgodność estymatora wystarczy wykazać, że warunkowa wariancja \hat{N}/N zbiega do zera:

$$\begin{aligned} \mathbb{V} \left[\frac{1}{N} \sum_{k=1}^N \frac{\mathcal{I}_k}{1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}]} \middle| \mathbf{X} \right] &= \frac{1}{N^2} \sum_{k=1}^N \frac{\mathbb{V}[\mathcal{I}_k | \mathbf{X}]}{(1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}])^2} \\ &= \frac{1}{N^2} \sum_{k=1}^N \frac{\mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}]}{(1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}])} = \frac{1}{N^2} \sum_{k=1}^N \frac{1}{1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}]} - \frac{1}{N}, \end{aligned}$$

a powyższy argument pokazuje, że jest to spełnione na przykład wtedy gdy:

$$\begin{aligned} \max_{k=1, \dots, N} (1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}])^{-1} &= o_{\mathbb{P}}(N) \quad (4.9) \\ \iff \max_{k=1, \dots, N} \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}] &= o_{\mathbb{P}}(1 - N^{-1}), \end{aligned}$$

gdzie $o_{\mathbb{P}}$ to małe o w probabilistycznej notacji asymptotycznej, lub ogólniej wtedy i tylko wtedy gdy:

$$\sum_{k=1}^N (1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}])^{-1} = o_{\mathbb{P}}(N^2), \quad (4.10)$$

natomiast warunek (4.9) jest łatwiejszy do sprawdzenia w praktyce.

Jeżeli $\boldsymbol{\theta}_k$ oznacza parametry rozkładu $\mathbf{Y}_{(k)} | \mathbf{Y}_{(k)} \neq \mathbf{0}, \mathbf{X}_{(k)}$ w modelu:

$$g_j(\theta_{ij}) = \mathbf{X}_{(i)} \boldsymbol{\beta}_j, \quad i = 1, \dots, n,$$

i dodatkowo estymatory dla wektorów współczynników regresji $\boldsymbol{\beta}_j$ są zgodne to przy założeniu, że funkcja:

$$G_k(\boldsymbol{\theta}) = \frac{1}{1 - \mathbb{P}[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)}]} = \frac{1}{\mathbb{P}[\mathbf{Y}_{(k)} \neq \mathbf{0} | \mathbf{X}_{(k)}]},$$

gdzie prawa strona wyrażenia zależy od $\boldsymbol{\theta}$ przez rozkład warunkowy $\mathbf{Y} | \mathbf{X}$, jest ciągła to twierdzenie o odwzorowaniu ciągłym gwarantuje, że estymator $G_k(\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}, \mathbf{X}_{(k)}))$ jest zgodny dla $G_k(\boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{X}_{(k)}))$ a, ponieważ:

$$\begin{aligned} \frac{1}{N} \left| \sum_{k=1}^N \mathcal{I}_k (G_k(\boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{X}_{(k)})) - G_k(\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}, \mathbf{X}_{(k)}))) \right| &\leq \\ \max_{k=1, \dots, N} |G_k(\boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{X}_{(k)})) - G_k(\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}, \mathbf{X}_{(k)}))|, \end{aligned}$$

gdzie ze zgodności ostatnie wyrażenie zbiega do 0 dla każdego indeksu.

Ostatecznie sprawdzenie zgodności estymatora metody wiarygodności obserwowanej sprowadza się do sprawdzenia tego jak szybko maleją prawdopodobieństwa jednostek w populacji oraz czy estymator współczynników regresji jest zgodny. Zauważmy, że własność (4.9) jest spełniona gdy jednostki wyróżniają się tylko poprzez zmienne kategoryczne.

- Wariancja estymatora zbudowanego na podstawie wiarygodności obserwowanej jest przybliżana za pomocą standardowego argumentu:

Wyprowadzenie. Niech \mathcal{F}_N oznacza σ -algebrę generowaną przez $\mathcal{I}_1, \dots, \mathcal{I}_N$, wtedy:

$$\mathbb{V}[\hat{N}] = \mathbb{E} \left[\mathbb{V} [\hat{N} | \mathcal{F}_N] \right] + \mathbb{V} \left[\mathbb{E} [\hat{N} | \mathcal{F}_N] \right],$$

wyrażenie $\mathbb{V} [\hat{N} | \mathcal{F}_N]$ przybliżane za pomocą metody δ jest przez formę kwadratową (wartość stałą):

$$\mathbb{V} [\hat{N} | \mathcal{F}_N] \approx \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}^T} \mathbb{V} [\hat{\boldsymbol{\beta}}] \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}},$$

natomiast $\mathbb{V} [\mathbb{E} [\hat{N} | \mathcal{F}_N]]$ estymowane jest przez naiwne podstawienie estymatorów dla G_k do wzoru dla estymatora Horwita-Thompsona:

$$\begin{aligned} \mathbb{V} \left[\mathbb{E} \left[\sum_{k=1}^N G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right) \mathcal{I}_k \mid \mathcal{F}_N \right] \right] &= \mathbb{V} \left[\sum_{k=1}^N G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right) \mathcal{I}_k \right] = \\ &= \sum_{k=1}^N G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^2 \mathbb{V} [\mathcal{I}_k] = \sum_{k=1}^N \frac{G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1}}{1 - G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1}} \\ &\approx \sum_{k=1}^N \mathcal{I}_k \frac{G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1}}{\left(1 - G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1} \right)^2}, \end{aligned}$$

gdzie ostatnia linijka jest estymatorem nieobciążonym o minimalnej wariancji dla poprzedniej. Ostatecznie uniwersalny estymator wariancji w metodzie wiarygodności obserwowanej to:

$$\begin{aligned} \widehat{\mathbb{V}[\hat{N}]} &= \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}^T} \mathbb{V} [\hat{\boldsymbol{\beta}}] \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \sum_{k=1}^N \mathcal{I}_k \frac{G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1}}{\left(1 - G_k \left(\boldsymbol{\theta} \left(\boldsymbol{\beta}, \mathbf{X}_{(k)} \right) \right)^{-1} \right)^2}, \\ &= \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}^T} \mathbb{V} [\hat{\boldsymbol{\beta}}] \frac{\partial \hat{N}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \sum_{k=1}^N \mathcal{I}_k \frac{1 - \hat{\mathbb{P}} \left[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)} \right]}{\hat{\mathbb{P}} \left[\mathbf{Y}_{(k)} = \mathbf{0} | \mathbf{X}_{(k)} \right]^2}. \end{aligned} \quad (4.11)$$

Estymator \hat{N} jest asymptotycznie normalny co pozwala na tworzenie przedziału ufności na podstawie (4.11) jednakże w próbach o "średniej wielkości" jego rozkład bardziej przypomina rozkład logarytmiczno-normalny a w niewielkich próbach nawet to przybliżenie jest wątpliwe dlatego bootstrap używany jest także w tej metodzie estymacji jeżeli (asymptotyczny) rozkład \hat{N} budzi wątpliwości.

4.3. Przykład wykorzystania na rzeczywistych danych

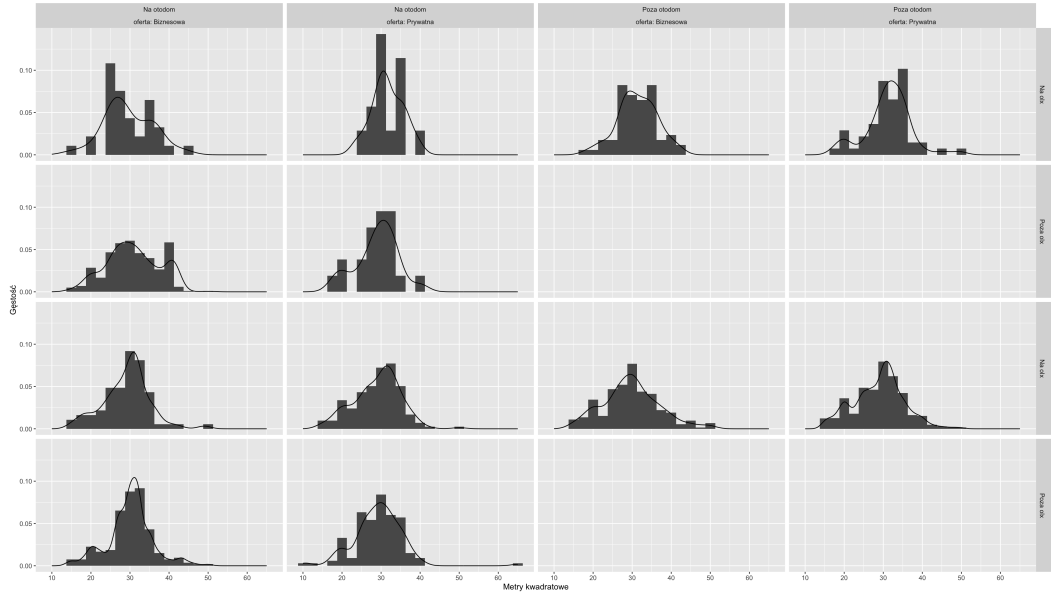
Jeżeli chodzi o zastosowania metody capture-recapture w kontekście polski wymienić należy przede wszystkim artykuł Beręsewicz, Gudaszewski, and Szymkowiak 2019, w którym modele log-liniowe zostały wykorzystane do estymacji ilości obcokrajowców przebywających na terenie Polski, z wykorzystaniem trzech źródeł danych. Artykuł ten zdecydowanie przyczynił się do spopularyzowania tej metody w Polsce. Istnieją dla niej ciekawe alternatywy w kontekście tego problemu jak na przykład metoda Zhanga i rozszerzenia Zhang 2008; Beręsewicz and Pawlukiewicz 2020 ale tym alternatywom na chwile pisania pracy brakuje solidnych podstaw teoretycznych.

My zaprezentujemy modele $\mathcal{M}_t, \mathcal{M}_b, \mathcal{M}_{tb}$ na przykładzie danych z pracy licencjackiej Czichos 2021, dotyczących rynku kawalerek na wynajem na terenie Poznania badanego poprzez publicznie dostępne ogłoszenia na stronach [olx.pl](https://www.olx.pl) oraz [otodom.pl](https://www.otodom.pl), na których właściciele mieszkań wystawiają oferty wynajmu. Kwestie poprawnej definicji docelowej populacji i interpretacji statystycznej zostały rozwiązane w oryginalnej pracy skupimy się dlatego tylko na wynikach.

		Typ oferty				
		Biznesowa		Prywatna		Łącznie
		Na otodom	Poza otodom	Na otodom	Poza otodom	
Umeblowane	Na Olx	74	291	166	1357	1888
	Poza Olx	214	?	133	?	?
Nie	Na Olx	37	68	14	55	174
Umeblowane	Poza Olx	410	?	21	?	?
Łącznie		735	?	334	?	2840 + ?

Tabela 4.2. Zaobserwowana tabela kontyngencji.

Zebrane zostały dane zawierające informacje o ilości dostępnych kawalerek podzielone na którejkolwiek z rozważanych stron i dodatkowo podzielone ze względu na typ oferty (biznesowa lub prywatna), umeblowanie mieszkania (z tabelą kontyngencji 4.2) oraz metraż mieszkania (cecha ciągła) rozkład w zależności od zmiennych kategoriycznych przedstawiony na ilustracji 4.3. Łączna liczba zaobserwowanych mieszkań to $n = 2840$.



Rysunek 4.1. Wykres opisujący rozkład zmiennej metrów kwadratowych dla każdego zaobserwowanego poziomu zmiennych kategorycznych.

Do danych dopasowane zostały modele $\mathcal{M}_t, \mathcal{M}_b, \mathcal{M}_{tb}$ zawierające każdą kombinację zmiennych objaśniających, których pełny opis łącznie z wybranymi macierzami ograniczeń znajduje się w materiałach dodatkowych C. Dopasowanie modeli ocenimy na podstawie kryteriów informacji oraz statystyki G^2 zdefiniowanej jako:

$$G^2 = 2 \sum_{i,j,t,m} n_{ijtm} \ln(n_{ijtm}/\hat{n}_{ijtm}),$$

gdzie n_{ijtm} oznacza liczbę mieszkań w odpowiedniej komórce tabeli kontyngencji 4.2 a \hat{n}_{ijtm} jest wielkością dopasowaną. Im mniejsza wartość statystyki G^2 tym lepiej dopasowany jest model.

Biorąc pod uwagę te kryteria wyróżnić możemy dwa najlepsze modele \mathcal{M}_{t5} i \mathcal{M}_{tb5} , w którym prawdopodobieństwa zależą od typu oferty i umeblowania (ale nie od metrażu). Model \mathcal{M}_{t5} cechuje się niższą wartością bayesowskiego kryterium informacji (BIC = 3984.4) i wyższą statystyką G^2 (= 6.370) oraz kryterium Akaike (AIC = 3948.7), niż model \mathcal{M}_{tb5} (AIC = 3948.4, BIC = 4001.9, G^2 = 0.005).

		Typ oferty Biznesowa		Prywatna		Łącznie
		Na otodom	Poza otodom	Na otodom	Poza otodom	
Umeblowane	Na Olx	72.1	284.1	167.9	1363.9	1888.0
	Poza Olx	222.8	878.0	124.2	1008.6	2233.5
Nie Umeblowane	Na Olx	38.9	74.9	12.1	48.1	174.0
	Poza Olx	401.2	772.8	29.8	118.6	1322.3
	Łącznie	735.0	2009.8	334	2539.2	5617.9

Tabela 4.3. Dopasowana tabela kontyngencji dla modelu \mathcal{M}_{t5} .

		Typ oferty Biznesowa		Prywatna		Łącznie
		Na otodom	Poza otodom	Na otodom	Poza otodom	
Umeblowane	Na Olx	74.2	290.8	165.8	1357.2	1888.0
	Poza Olx	214.0	18.4	133.0	127.9	493.3
Nie Umeblowane	Na Olx	36.8	68.2	14.2	54.8	174.0
	Poza Olx	410.0	129.6	21.0	74.4	635.0
	Łącznie	735.0	507	334	1614.3	3190.3

Tabela 4.4. Dopasowana tabela kontyngencji dla modelu \mathcal{M}_{tb5} .

Model	\hat{N}	$\sqrt{\hat{V}[\hat{N}]}$	\mathcal{N}^-	\mathcal{N}^+	$\ln \mathcal{N}^-$	$\ln \mathcal{N}^+$
\mathcal{M}_{t5}	5617.9	246.6	5134.521	6101.306	5175.039	6144.785
\mathcal{M}_{tb5}	3190.3	671.5	2840.0	4506.4	2870.719	6834.595

Tabela 4.5. Wyniki estymacji dwóch najlepszych modeli, \mathcal{N} oznacza przedział ufności na bazie rozkładu normalnego a $\ln \mathcal{N}$ na podstawie rozkładu log-normalnego.

Dopasowane tabele kontyngencji tych modeli to odpowiednio tabele 4.3, 4.4 wskazują, że model \mathcal{M}_{tb5} zdecydowanie lepiej pasuje do zaobserwowanych danych ale jak sugeruje tabela C występują w nim obserwacje bardzo wpływowe w przeciwieństwie do \mathcal{M}_{t5} , lub ma poważne problemy numeryczne gdy usuniemy konkretne jednostki ze zbioru danych, a test χ^2 dla statystyki G^2 wskazuje, że obie dopasowane tabele kontyngencji nie różnią się istotnie od zaobserwowanych (p wartość testu dla \mathcal{M}_{t5} to 0.383). Dodatkowo wyniki z tabeli 4.3 wydają się być bardziej realistyczne w modelu \mathcal{M}_{t5} .

Podsumowanie

Omawiana klasa wektorowych uogólnionych modeli liniowych jest użytecznym narzędziem w nowoczesnej analizie danych ze względu na szerokie pole potencjalnych zastosowań i względną przystępność ogólnej postaci modelu, która nie odbiega znacząco od uogólnionych modeli liniowych, a mimo to pozwala na istotne rozszerzenia co w dużym stopniu spowodowane jest wprowadzeniem do teorii macierzy ograniczeń.

Dzięki elastyczności i wszechstronności, wektorowe uogólnione modele liniowe oferują użyteczne narzędzie do analizy złożonych zależności w danych, co zostało zademonstrowane w przykładach zastosowań w testowaniu hipotez i modelowaniu procesów wielowymiarowych, które istotnie wykorzystują możliwość wykroczenia poza teorię modeli uogólnionych modeli liniowych. Mimo że opisane przykłady zastosowań są różnorodne, nie wyczerpują one wszystkich najważniejszych praktycznych zastosowań tej klasy modeli.

W pracy przedstawiono również teorię estymacji wielkości zamkniętych populacji metodą capture-recapture jako specjalny przypadek wektorowego uogólnionego modelu liniowego, co ukazuje potencjał tych modeli w różnych dziedzinach badań.

Z powodów objętościowych w pracy nie omówiono szerszej klasy wektorowych uogólnionych modeli addytywnych (VGAM), w których funkcjonal identyfikowany jest nie poprzez "statyczny" wektor z przestrzeni euklidesowej a przez wektor z przestrzeni funkcji gładkich co pozwala na jeszcze większą ogólność modelu kosztem bardziej skomplikowanych obliczeń.

Podsumowując, wektorowy uogólniony model liniowy to wszechstronne narzędzie statystyczne, które znacząco rozszerza możliwości analizy danych w porównaniu do tradycyjnych modeli uogólnionych modeli liniowych.

Bibliografia

- Beręsewicz, Maciej, Grzegorz Gudaszewski, and Marcin Szymkowiak (Nov. 2019). “Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture-recapture”. In: *Wiadomości statystyczne (Warsaw, Poland: 1956)* 64, pp. 7–35.
- Beręsewicz, Maciej and Katarzyna Pawlukiewicz (2020). *Estimation of the number of irregular foreigners in Poland using non-linear count regression models*. arXiv: [2008.09407 \[stat.AP\]](https://arxiv.org/abs/2008.09407).
- Bohning, Dankmar, Peter G.M. Heijden, and John Bunge (Jan. 2018). *Capture-Recapture Methods for the Social and Medical Sciences*, pp. 1–462. DOI: <https://doi.org/10.4324/9781315151939>.
- Böhning, Dankmar and Peter G. M. van der Heijden (2009). “A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations”. In: *The Annals of Applied Statistics* 3.2, pp. 595–610. DOI: [10.1214/08-AOAS214](https://doi.org/10.1214/08-AOAS214).
- (2019). “The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain”. In: *The Annals of Applied Statistics* 13.2, pp. 1198–1211. DOI: [10.1214/18-AOAS1232](https://doi.org/10.1214/18-AOAS1232).
- Böhning, Dankmar, Alberto Vidal-Diez, et al. (2013). “A Generalization of Chao’s Estimator for Covariate Information”. In: *Biometrics* 69.4, pp. 1033–1042. DOI: <https://doi.org/10.1111/biom.12082>.
- Bradley, Ralph Allan and Milton E. Terry (1952). “Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons”. In: *Biometrika* 39.3/4, pp. 324–345. ISSN: 00063444. (Visited on 04/06/2024).
- Cerqueti, Roy et al. (2022). “INGARCH-based fuzzy clustering of count time series with a football application”. In: *Machine Learning with Applications* 10, p. 100417. ISSN: 2666-8270. DOI: <https://doi.org/10.1016/j.mlwa.2022.100417>.
- Chao, Anne (1989). “Estimating population size for sparse data in capture - recapture experiments”. In: *Biometrics*, pp. 427–438.
- Cruyff, Maarten J. L. F. and Peter G. M. van der Heijden (2008). “Point and Interval Estimation of the Population Size Using a Zero-Truncated Negative

- Binomial Regression Model”. In: *Biometrical Journal* 50.6, pp. 1035–1050. DOI: <https://doi.org/10.1002/bimj.200810455>.
- Czichos, Fabian (2021). “Estymacja liczby kawalerek na wynajem w Poznaniu z wykorzystaniem źródeł internetowych i metod capture-recapture”. In: URL: <https://github.com/DepartmentOfStatisticsPUE/prace-dyplomowe/blob/master/licencjat/2021-czichos.pdf>.
- Dixon, Mark J. and Stuart G. Coles (Jan. 2002). “Modelling Association Football Scores and Inefficiencies in the Football Betting Market”. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 46.2, pp. 265–280. ISSN: 0035-9254. DOI: [10.1111/1467-9876.00065](https://doi.org/10.1111/1467-9876.00065).
- Eicker, F. (1963). “Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions”. In: *The Annals of Mathematical Statistics* 34.2, pp. 447–456. DOI: [10.1214/aoms/1177704156](https://doi.org/10.1214/aoms/1177704156).
- Fienberg, Stephen E. (1972). “The Multiple Recapture Census for Closed Populations and Incomplete 2k Contingency Tables”. In: *Biometrika* 59.3, pp. 591–603. ISSN: 00063444. DOI: [10.2307/2334810](https://doi.org/10.2307/2334810). URL: <http://www.jstor.org/stable/2334810>.
- Hastie, Trevor J. and Robert Tibshirani (1993). “Varying-Coefficient Models”. In: *Journal of the royal statistical society series b-methodological* 55, pp. 757–779.
- Heijden, Peter GM van der et al. (2003). “Point and interval estimation of the population size using the truncated Poisson regression model”. In: *Statistical Modelling* 3.4, pp. 305–322. DOI: [10.1191/1471082X03st057oa](https://doi.org/10.1191/1471082X03st057oa).
- Horvitz, Daniel G and Donovan J Thompson (1952). “A generalization of sampling without replacement from a finite universe”. In: *Journal of the American statistical Association* 47.260, pp. 663–685.
- Laplace, Pierre-Simon (1786). “Sur les naissances, les mariages et les morts”. In: *Histoire de L’Acad’Emie Royale Des Sciences*. URL: <http://www.probabilityandfinance.com/pulskamp/Laplace/naissances.pdf>.
- Lindstrøm, Jonas Christoffer (2023). *goalmodel: Prediction Models For Goals Scored*. R package version 0.6.3. URL: <https://github.com/opisthokonta/goalmodel>.
- Maher, M. J. (1982). “Modelling association football scores”. In: *Statistica Neerlandica* 36.3, pp. 109–118. DOI: <https://doi.org/10.1111/j.1467-9574.1982.tb00782.x>.
- Norris, James L and Kenneth H Pollock (1996). “Including model uncertainty in estimating variances in multiple capture studies”. In: *Environmental and Ecological Statistics* 3.3, pp. 235–244.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

-
- Sanathanan, Lalitha (1972). “Estimating the Size of a Multinomial Population”. In: *The Annals of Mathematical Statistics* 43.1, pp. 142–152. DOI: [10.1214/aoms/1177692709](https://doi.org/10.1214/aoms/1177692709).
- Soberanis, Víctor Miranda (2019). “Vector Generalized Linear Time Series Models with an Implementation in R”. In: URL: <https://researchspace.auckland.ac.nz/bitstream/handle/2292/45778/whole.pdf?sequence=4>.
- Tuoto, Tiziana, Davide Di Cecco, and Andrea Tancredi (2022). “Bayesian analysis of one-inflated models for elusive population size estimation”. In: *Biometrical Journal* 64.5, pp. 912–933. DOI: <https://doi.org/10.1002/bimj.202100187>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202100187>.
- White, Halbert (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press.
- Wolter, Kirk M. (1986). “Some Coverage Error Models for Census Data”. In: *Journal of the American Statistical Association* 81.394, pp. 338–346. ISSN: 01621459. URL: <http://www.jstor.org/stable/2289222> (visited on 03/12/2024).
- Yee, T. W. (2023). *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.1-9. URL: <https://CRAN.R-project.org/package=VGAM>.
- Yee, Thomas W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. 1st. Springer Publishing Company, Incorporated.
- Yee, Thomas W., Jakub Stoklosa, and Richard M. Huggins (2015). “The VGAM Package for Capture-Recapture Data Using the Conditional Likelihood”. In: *Journal of Statistical Software* 65.5, pp. 1–33. DOI: [10.18637/jss.v065.i05](https://doi.org/10.18637/jss.v065.i05).
- Zeileis, Achim (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators”. In: *Journal of Statistical Software* 11.10, pp. 1–17. DOI: [10.18637/jss.v011.i10](https://doi.org/10.18637/jss.v011.i10).
- (2006). “Object-oriented Computation of Sandwich Estimators”. In: *Journal of Statistical Software* 16.9, pp. 1–16. DOI: [10.18637/jss.v016.i09](https://doi.org/10.18637/jss.v016.i09).
- Zhang, Li-Chun (2008). “Developing methods for determining the number of unauthorized foreigners in Norway”. In: *Statistics Norway (SSB), Division for Statistical Methods and Standards (accessed July 28, 2008)*.
- Zimmer, Felix, Clemens Draxler, and Rudolf Debelak (Aug. 2022). “Power Analysis for the Wald, LR, Score, and Gradient Tests in a Marginal Maximum Likelihood Framework: Applications in IRT”. In: *Psychometrika* 88. DOI: [10.1007/s11336-022-09883-5](https://doi.org/10.1007/s11336-022-09883-5).

Zwane, EN and PGM Van der Heijden (2003). “Implementing the parametric bootstrap in capture–recapture models with continuous covariates”. In: *Statistics & probability letters* 65.2, pp. 121–125.

Wyniki symulacji z rozdziału 3

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0630	0.0789
lm.HC0	D1/D3	0.0952	0.1105
lm.HC1	D1/D3	0.0793	0.0920
lm.HC2	D1/D3	0.0738	0.0873
lm.HC3	D1/D3	0.0561	0.0658
lm.HC4	D1/D3	0.0614	0.0726
lm.HC4m	D1/D3	0.0522	0.0620
lm.HC5	D1/D3	0.0759	0.0908
vlm	D1/D3	0.0871	0.1068
vlm.HC0	D1/D3	0.1027	0.1181
vlm.HC1	D1/D3	0.0875	0.1001
vlm.HC2	D1/D3	0.0806	0.0934
vlm.HC3	D1/D3	0.0635	0.0740
vlm.HC4	D1/D3	0.0636	0.0738
vlm.HC4m	D1/D3	0.0572	0.0681
vlm.HC5	D1/D3	0.0823	0.0946
lm	D2/D4	0.0531	0.0519
lm.HC0	D2/D4	0.0585	0.0581
lm.HC1	D2/D4	0.0441	0.0452
lm.HC2	D2/D4	0.0401	0.0416
lm.HC3	D2/D4	0.0241	0.0266
lm.HC4	D2/D4	0.0294	0.0325
lm.HC4m	D2/D4	0.0219	0.0249
lm.HC5	D2/D4	0.0426	0.0433
vlm	D2/D4	0.0772	0.0763
vlm.HC0	D2/D4	0.0650	0.0667
vlm.HC1	D2/D4	0.0522	0.0540
vlm.HC2	D2/D4	0.0463	0.0480
vlm.HC3	D2/D4	0.0289	0.0341
vlm.HC4	D2/D4	0.0310	0.0352
vlm.HC4m	D2/D4	0.0254	0.0293
vlm.HC5	D2/D4	0.0468	0.0479

Tabela A.1. Wyniki symulacji dla rozmiaru próby $n = 25$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0575	0.0889
lm.HC0	D1/D3	0.0653	0.1064
lm.HC1	D1/D3	0.0589	0.0983
lm.HC2	D1/D3	0.0564	0.0948
lm.HC3	D1/D3	0.0494	0.0838
lm.HC4	D1/D3	0.0522	0.0892
lm.HC4m	D1/D3	0.0475	0.0818
lm.HC5	D1/D3	0.0592	0.0981
vlm	D1/D3	0.0634	0.1048
vlm.HC0	D1/D3	0.0693	0.1099
vlm.HC1	D1/D3	0.0622	0.1020
vlm.HC2	D1/D3	0.0603	0.0983
vlm.HC3	D1/D3	0.0518	0.0882
vlm.HC4	D1/D3	0.0536	0.0903
vlm.HC4m	D1/D3	0.0488	0.0842
vlm.HC5	D1/D3	0.0612	0.0998
lm	D2/D4	0.0465	0.0495
lm.HC0	D2/D4	0.0473	0.0503
lm.HC1	D2/D4	0.0397	0.0435
lm.HC2	D2/D4	0.0383	0.0418
lm.HC3	D2/D4	0.0313	0.0350
lm.HC4	D2/D4	0.0347	0.0387
lm.HC4m	D2/D4	0.0305	0.0338
lm.HC5	D2/D4	0.0408	0.0436
vlm	D2/D4	0.0548	0.0593
vlm.HC0	D2/D4	0.0495	0.0528
vlm.HC1	D2/D4	0.0440	0.0477
vlm.HC2	D2/D4	0.0418	0.0452
vlm.HC3	D2/D4	0.0341	0.0382
vlm.HC4	D2/D4	0.0367	0.0393
vlm.HC4m	D2/D4	0.0323	0.0352
vlm.HC5	D2/D4	0.0428	0.0463

Tabela A.2. Wyniki symulacji dla rozmiaru próby $n = 50$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0560	0.1010
lm.HC0	D1/D3	0.0648	0.1132
lm.HC1	D1/D3	0.0603	0.1053
lm.HC2	D1/D3	0.0580	0.1037
lm.HC3	D1/D3	0.0521	0.0948
lm.HC4	D1/D3	0.0548	0.0994
lm.HC4m	D1/D3	0.0511	0.0938
lm.HC5	D1/D3	0.0603	0.1058
vlm	D1/D3	0.0612	0.1109
vlm.HC0	D1/D3	0.0661	0.1158
vlm.HC1	D1/D3	0.0620	0.1096
vlm.HC2	D1/D3	0.0599	0.1057
vlm.HC3	D1/D3	0.0535	0.0980
vlm.HC4	D1/D3	0.0551	0.0997
vlm.HC4m	D1/D3	0.0524	0.0962
vlm.HC5	D1/D3	0.0613	0.1080
lm	D2/D4	0.0451	0.0490
lm.HC0	D2/D4	0.0450	0.0474
lm.HC1	D2/D4	0.0403	0.0424
lm.HC2	D2/D4	0.0391	0.0414
lm.HC3	D2/D4	0.0333	0.0363
lm.HC4	D2/D4	0.0368	0.0385
lm.HC4m	D2/D4	0.0324	0.0351
lm.HC5	D2/D4	0.0412	0.0428
vlm	D2/D4	0.0516	0.0562
vlm.HC0	D2/D4	0.0481	0.0497
vlm.HC1	D2/D4	0.0426	0.0451
vlm.HC2	D2/D4	0.0412	0.0431
vlm.HC3	D2/D4	0.0353	0.0379
vlm.HC4	D2/D4	0.0370	0.0393
vlm.HC4m	D2/D4	0.0339	0.0366
vlm.HC5	D2/D4	0.0424	0.0437

Tabela A.3. Wyniki symulacji dla rozmiaru próby $n = 75$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0545	0.1309
lm.HC0	D1/D3	0.0596	0.1439
lm.HC1	D1/D3	0.0553	0.1386
lm.HC2	D1/D3	0.0545	0.1366
lm.HC3	D1/D3	0.0495	0.1282
lm.HC4	D1/D3	0.0523	0.1331
lm.HC4m	D1/D3	0.0480	0.1267
lm.HC5	D1/D3	0.0557	0.1387
vlm	D1/D3	0.0582	0.1437
vlm.HC0	D1/D3	0.0600	0.1476
vlm.HC1	D1/D3	0.0573	0.1426
vlm.HC2	D1/D3	0.0552	0.1400
vlm.HC3	D1/D3	0.0514	0.1310
vlm.HC4	D1/D3	0.0526	0.1341
vlm.HC4m	D1/D3	0.0504	0.1283
vlm.HC5	D1/D3	0.0565	0.1412
lm	D2/D4	0.0454	0.0499
lm.HC0	D2/D4	0.0458	0.0500
lm.HC1	D2/D4	0.0415	0.0461
lm.HC2	D2/D4	0.0405	0.0451
lm.HC3	D2/D4	0.0369	0.0403
lm.HC4	D2/D4	0.0385	0.0430
lm.HC4m	D2/D4	0.0361	0.0395
lm.HC5	D2/D4	0.0419	0.0467
vlm	D2/D4	0.0506	0.0542
vlm.HC0	D2/D4	0.0460	0.0511
vlm.HC1	D2/D4	0.0432	0.0481
vlm.HC2	D2/D4	0.0419	0.0467
vlm.HC3	D2/D4	0.0379	0.0420
vlm.HC4	D2/D4	0.0391	0.0435
vlm.HC4m	D2/D4	0.0372	0.0412
vlm.HC5	D2/D4	0.0423	0.0470

Tabela A.4. Wyniki symulacji dla rozmiaru próby $n = 100$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0536	0.1369
lm.HC0	D1/D3	0.0572	0.1446
lm.HC1	D1/D3	0.0549	0.1387
lm.HC2	D1/D3	0.0541	0.1375
lm.HC3	D1/D3	0.0497	0.1316
lm.HC4	D1/D3	0.0528	0.1348
lm.HC4m	D1/D3	0.0486	0.1299
lm.HC5	D1/D3	0.0550	0.1393
vlm	D1/D3	0.0556	0.1460
vlm.HC0	D1/D3	0.0578	0.1468
vlm.HC1	D1/D3	0.0552	0.1420
vlm.HC2	D1/D3	0.0545	0.1398
vlm.HC3	D1/D3	0.0514	0.1338
vlm.HC4	D1/D3	0.0527	0.1356
vlm.HC4m	D1/D3	0.0502	0.1317
vlm.HC5	D1/D3	0.0549	0.1411
lm	D2/D4	0.0491	0.0513
lm.HC0	D2/D4	0.0446	0.0483
lm.HC1	D2/D4	0.0412	0.0457
lm.HC2	D2/D4	0.0409	0.0453
lm.HC3	D2/D4	0.0380	0.0417
lm.HC4	D2/D4	0.0397	0.0435
lm.HC4m	D2/D4	0.0377	0.0412
lm.HC5	D2/D4	0.0418	0.0462
vlm	D2/D4	0.0531	0.0542
vlm.HC0	D2/D4	0.0454	0.0500
vlm.HC1	D2/D4	0.0424	0.0478
vlm.HC2	D2/D4	0.0418	0.0468
vlm.HC3	D2/D4	0.0394	0.0441
vlm.HC4	D2/D4	0.0406	0.0447
vlm.HC4m	D2/D4	0.0384	0.0427
vlm.HC5	D2/D4	0.0424	0.0476

Tabela A.5. Wyniki symulacji dla rozmiaru próby $n = 125$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0580	0.1574
lm.HC0	D1/D3	0.0613	0.1701
lm.HC1	D1/D3	0.0598	0.1651
lm.HC2	D1/D3	0.0593	0.1639
lm.HC3	D1/D3	0.0556	0.1571
lm.HC4	D1/D3	0.0580	0.1617
lm.HC4m	D1/D3	0.0552	0.1553
lm.HC5	D1/D3	0.0598	0.1658
vlm	D1/D3	0.0595	0.1660
vlm.HC0	D1/D3	0.0619	0.1711
vlm.HC1	D1/D3	0.0598	0.1677
vlm.HC2	D1/D3	0.0590	0.1661
vlm.HC3	D1/D3	0.0562	0.1593
vlm.HC4	D1/D3	0.0571	0.1614
vlm.HC4m	D1/D3	0.0556	0.1572
vlm.HC5	D1/D3	0.0594	0.1675
lm	D2/D4	0.0465	0.0523
lm.HC0	D2/D4	0.0474	0.0495
lm.HC1	D2/D4	0.0447	0.0462
lm.HC2	D2/D4	0.0444	0.0454
lm.HC3	D2/D4	0.0422	0.0428
lm.HC4	D2/D4	0.0434	0.0446
lm.HC4m	D2/D4	0.0418	0.0421
lm.HC5	D2/D4	0.0448	0.0464
vlm	D2/D4	0.0498	0.0538
vlm.HC0	D2/D4	0.0491	0.0502
vlm.HC1	D2/D4	0.0463	0.0480
vlm.HC2	D2/D4	0.0456	0.0469
vlm.HC3	D2/D4	0.0439	0.0429
vlm.HC4	D2/D4	0.0446	0.0442
vlm.HC4m	D2/D4	0.0434	0.0423
vlm.HC5	D2/D4	0.0461	0.0476

Tabela A.6. Wyniki symulacji dla rozmiaru próby $n = 150$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0521	0.1726
lm.HC0	D1/D3	0.0538	0.1803
lm.HC1	D1/D3	0.0514	0.1756
lm.HC2	D1/D3	0.0512	0.1744
lm.HC3	D1/D3	0.0484	0.1689
lm.HC4	D1/D3	0.0501	0.1724
lm.HC4m	D1/D3	0.0483	0.1679
lm.HC5	D1/D3	0.0515	0.1763
vlm	D1/D3	0.0517	0.1803
vlm.HC0	D1/D3	0.0545	0.1814
vlm.HC1	D1/D3	0.0530	0.1780
vlm.HC2	D1/D3	0.0520	0.1766
vlm.HC3	D1/D3	0.0499	0.1720
vlm.HC4	D1/D3	0.0505	0.1739
vlm.HC4m	D1/D3	0.0492	0.1697
vlm.HC5	D1/D3	0.0526	0.1775
lm	D2/D4	0.0472	0.0528
lm.HC0	D2/D4	0.0460	0.0519
lm.HC1	D2/D4	0.0436	0.0492
lm.HC2	D2/D4	0.0431	0.0488
lm.HC3	D2/D4	0.0407	0.0470
lm.HC4	D2/D4	0.0425	0.0480
lm.HC4m	D2/D4	0.0399	0.0461
lm.HC5	D2/D4	0.0442	0.0494
vlm	D2/D4	0.0515	0.0546
vlm.HC0	D2/D4	0.0463	0.0521
vlm.HC1	D2/D4	0.0438	0.0506
vlm.HC2	D2/D4	0.0434	0.0498
vlm.HC3	D2/D4	0.0411	0.0472
vlm.HC4	D2/D4	0.0417	0.0478
vlm.HC4m	D2/D4	0.0406	0.0466
vlm.HC5	D2/D4	0.0435	0.0505

Tabela A.7. Wyniki symulacji dla rozmiaru próby $n = 175$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0548	0.1870
lm.HC0	D1/D3	0.0568	0.1944
lm.HC1	D1/D3	0.0552	0.1904
lm.HC2	D1/D3	0.0548	0.1895
lm.HC3	D1/D3	0.0531	0.1843
lm.HC4	D1/D3	0.0541	0.1871
lm.HC4m	D1/D3	0.0523	0.1840
lm.HC5	D1/D3	0.0553	0.1908
vlm	D1/D3	0.0547	0.1946
vlm.HC0	D1/D3	0.0570	0.1964
vlm.HC1	D1/D3	0.0559	0.1925
vlm.HC2	D1/D3	0.0554	0.1917
vlm.HC3	D1/D3	0.0534	0.1876
vlm.HC4	D1/D3	0.0540	0.1892
vlm.HC4m	D1/D3	0.0527	0.1864
vlm.HC5	D1/D3	0.0559	0.1922
lm	D2/D4	0.0495	0.0512
lm.HC0	D2/D4	0.0452	0.0508
lm.HC1	D2/D4	0.0434	0.0486
lm.HC2	D2/D4	0.0428	0.0483
lm.HC3	D2/D4	0.0414	0.0455
lm.HC4	D2/D4	0.0422	0.0474
lm.HC4m	D2/D4	0.0411	0.0453
lm.HC5	D2/D4	0.0435	0.0490
vlm	D2/D4	0.0502	0.0532
vlm.HC0	D2/D4	0.0462	0.0508
vlm.HC1	D2/D4	0.0448	0.0494
vlm.HC2	D2/D4	0.0440	0.0488
vlm.HC3	D2/D4	0.0422	0.0469
vlm.HC4	D2/D4	0.0430	0.0474
vlm.HC4m	D2/D4	0.0417	0.0461
vlm.HC5	D2/D4	0.0443	0.0491

Tabela A.8. Wyniki symulacji dla rozmiaru próby $n = 200$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0511	0.2057
lm.HC0	D1/D3	0.0534	0.2134
lm.HC1	D1/D3	0.0513	0.2092
lm.HC2	D1/D3	0.0510	0.2080
lm.HC3	D1/D3	0.0498	0.2042
lm.HC4	D1/D3	0.0506	0.2066
lm.HC4m	D1/D3	0.0491	0.2030
lm.HC5	D1/D3	0.0515	0.2101
vlm	D1/D3	0.0520	0.2119
vlm.HC0	D1/D3	0.0540	0.2151
vlm.HC1	D1/D3	0.0522	0.2116
vlm.HC2	D1/D3	0.0516	0.2103
vlm.HC3	D1/D3	0.0497	0.2061
vlm.HC4	D1/D3	0.0504	0.2072
vlm.HC4m	D1/D3	0.0494	0.2044
vlm.HC5	D1/D3	0.0519	0.2112
lm	D2/D4	0.0453	0.0511
lm.HC0	D2/D4	0.0443	0.0524
lm.HC1	D2/D4	0.0427	0.0508
lm.HC2	D2/D4	0.0421	0.0506
lm.HC3	D2/D4	0.0404	0.0486
lm.HC4	D2/D4	0.0415	0.0496
lm.HC4m	D2/D4	0.0400	0.0483
lm.HC5	D2/D4	0.0428	0.0510
vlm	D2/D4	0.0461	0.0533
vlm.HC0	D2/D4	0.0451	0.0528
vlm.HC1	D2/D4	0.0443	0.0514
vlm.HC2	D2/D4	0.0435	0.0509
vlm.HC3	D2/D4	0.0421	0.0496
vlm.HC4	D2/D4	0.0429	0.0500
vlm.HC4m	D2/D4	0.0417	0.0487
vlm.HC5	D2/D4	0.0439	0.0513

Tabela A.9. Wyniki symulacji dla rozmiaru próby $n = 225$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0477	0.2291
lm.HC0	D1/D3	0.0510	0.2355
lm.HC1	D1/D3	0.0499	0.2309
lm.HC2	D1/D3	0.0496	0.2307
lm.HC3	D1/D3	0.0480	0.2263
lm.HC4	D1/D3	0.0488	0.2290
lm.HC4m	D1/D3	0.0478	0.2255
lm.HC5	D1/D3	0.0500	0.2316
vlm	D1/D3	0.0492	0.2323
vlm.HC0	D1/D3	0.0510	0.2361
vlm.HC1	D1/D3	0.0502	0.2332
vlm.HC2	D1/D3	0.0498	0.2314
vlm.HC3	D1/D3	0.0487	0.2266
vlm.HC4	D1/D3	0.0492	0.2281
vlm.HC4m	D1/D3	0.0477	0.2257
vlm.HC5	D1/D3	0.0501	0.2326
lm	D2/D4	0.0514	0.0528
lm.HC0	D2/D4	0.0491	0.0524
lm.HC1	D2/D4	0.0475	0.0512
lm.HC2	D2/D4	0.0473	0.0510
lm.HC3	D2/D4	0.0460	0.0495
lm.HC4	D2/D4	0.0468	0.0504
lm.HC4m	D2/D4	0.0452	0.0489
lm.HC5	D2/D4	0.0476	0.0513
vlm	D2/D4	0.0523	0.0544
vlm.HC0	D2/D4	0.0499	0.0525
vlm.HC1	D2/D4	0.0491	0.0516
vlm.HC2	D2/D4	0.0487	0.0512
vlm.HC3	D2/D4	0.0465	0.0501
vlm.HC4	D2/D4	0.0475	0.0505
vlm.HC4m	D2/D4	0.0459	0.0498
vlm.HC5	D2/D4	0.0489	0.0513

Tabela A.10. Wyniki symulacji dla rozmiaru próby $n = 250$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0497	0.2456
lm.HC0	D1/D3	0.0525	0.2536
lm.HC1	D1/D3	0.0509	0.2502
lm.HC2	D1/D3	0.0506	0.2493
lm.HC3	D1/D3	0.0493	0.2449
lm.HC4	D1/D3	0.0499	0.2474
lm.HC4m	D1/D3	0.0489	0.2442
lm.HC5	D1/D3	0.0513	0.2507
vlm	D1/D3	0.0515	0.2523
vlm.HC0	D1/D3	0.0531	0.2545
vlm.HC1	D1/D3	0.0521	0.2520
vlm.HC2	D1/D3	0.0518	0.2509
vlm.HC3	D1/D3	0.0504	0.2471
vlm.HC4	D1/D3	0.0508	0.2482
vlm.HC4m	D1/D3	0.0503	0.2458
vlm.HC5	D1/D3	0.0519	0.2516
lm	D2/D4	0.0514	0.0539
lm.HC0	D2/D4	0.0496	0.0535
lm.HC1	D2/D4	0.0483	0.0529
lm.HC2	D2/D4	0.0478	0.0527
lm.HC3	D2/D4	0.0463	0.0502
lm.HC4	D2/D4	0.0473	0.0513
lm.HC4m	D2/D4	0.0459	0.0499
lm.HC5	D2/D4	0.0483	0.0530
vlm	D2/D4	0.0520	0.0558
vlm.HC0	D2/D4	0.0501	0.0538
vlm.HC1	D2/D4	0.0491	0.0523
vlm.HC2	D2/D4	0.0485	0.0519
vlm.HC3	D2/D4	0.0474	0.0499
vlm.HC4	D2/D4	0.0480	0.0508
vlm.HC4m	D2/D4	0.0467	0.0493
vlm.HC5	D2/D4	0.0489	0.0525

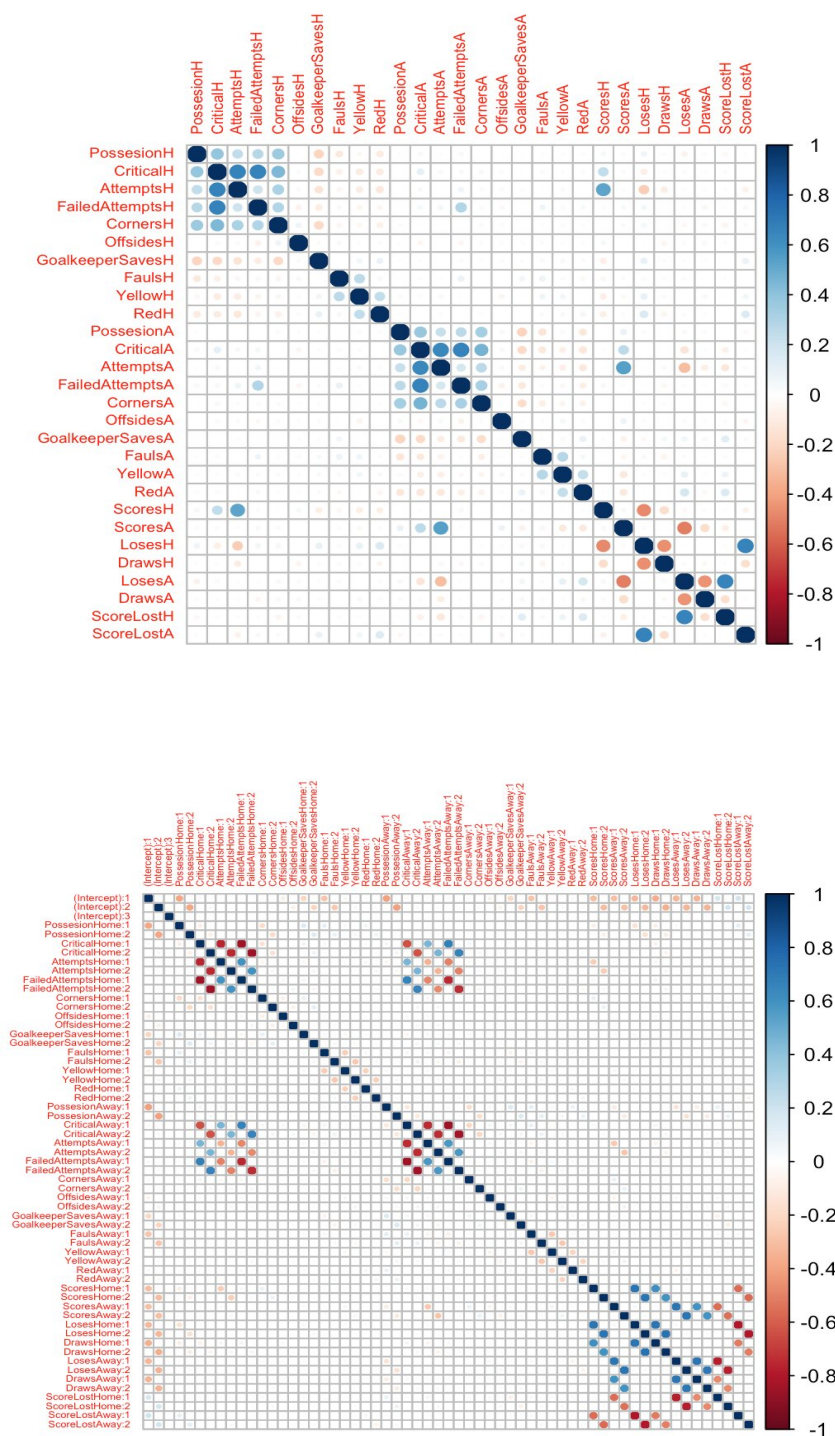
Tabela A.11. Wyniki symulacji dla rozmiaru próby $n = 275$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Nazwa testu	Warunki eksperymentalne	Błąd 1-szego rodzaju	Moc
lm	D1/D3	0.0500	0.2643
lm.HC0	D1/D3	0.0523	0.2696
lm.HC1	D1/D3	0.0515	0.2669
lm.HC2	D1/D3	0.0510	0.2661
lm.HC3	D1/D3	0.0498	0.2621
lm.HC4	D1/D3	0.0504	0.2641
lm.HC4m	D1/D3	0.0496	0.2611
lm.HC5	D1/D3	0.0515	0.2669
vlm	D1/D3	0.0511	0.2688
vlm.HC0	D1/D3	0.0527	0.2703
vlm.HC1	D1/D3	0.0515	0.2679
vlm.HC2	D1/D3	0.0513	0.2668
vlm.HC3	D1/D3	0.0500	0.2637
vlm.HC4	D1/D3	0.0506	0.2649
vlm.HC4m	D1/D3	0.0495	0.2628
vlm.HC5	D1/D3	0.0514	0.2676
lm	D2/D4	0.0451	0.0519
lm.HC0	D2/D4	0.0424	0.0532
lm.HC1	D2/D4	0.0416	0.0515
lm.HC2	D2/D4	0.0414	0.0514
lm.HC3	D2/D4	0.0402	0.0500
lm.HC4	D2/D4	0.0407	0.0509
lm.HC4m	D2/D4	0.0398	0.0495
lm.HC5	D2/D4	0.0416	0.0518
vlm	D2/D4	0.0459	0.0545
vlm.HC0	D2/D4	0.0427	0.0541
vlm.HC1	D2/D4	0.0419	0.0525
vlm.HC2	D2/D4	0.0412	0.0519
vlm.HC3	D2/D4	0.0405	0.0503
vlm.HC4	D2/D4	0.0407	0.0508
vlm.HC4m	D2/D4	0.0405	0.0497
vlm.HC5	D2/D4	0.0418	0.0523

Tabela A.12. Wyniki symulacji dla rozmiaru próby $n = 300$ błędy pierwszego rodzaju istotnie wyższe niż poziom nominalny 0.05 zostały wytłuszczone.

Analiza dewiancji z rozdziału 3

Nazwa zmiennej	Wartość statystyki testowej	Wartość logarytmu funkcji wiarygodności po usunięciu zmiennej	P-wartość df = 2
PossesionHome	17.72	-8043.13	0.0001
CriticalHome	0.63	-8034.58	0.7311
AttemptsHome	0.25	-8034.39	0.8815
FailedAttemptsHome	0.87	-8034.70	0.6473
CornersHome	6.01	-8037.27	0.0496
OffsidesHome	0.51	-8034.52	0.7768
GoalkeeperSavesHome	6.13	-8037.33	0.0466
FaultsHome	2.86	-8035.70	0.2390
YellowHome	2.48	-8035.51	0.2897
RedHome	0.43	-8034.48	0.8061
PossesionAway	10.63	-8039.58	0.0049
CriticalAway	2.77	-8035.65	0.2508
AttemptsAway	7.46	-8038.00	0.0240
FailedAttemptsAway	1.62	-8035.08	0.4452
CornersAway	0.01	-8034.27	0.9944
OffsidesAway	3.39	-8035.96	0.1834
GoalkeeperSavesAway	3.79	-8036.16	0.1503
FaultsAway	2.17	-8035.36	0.3371
YellowAway	2.07	-8035.30	0.3556
RedAway	0.66	-8034.60	0.7190
ScoresHome	3.04	-8035.79	0.2191
ScoresAway	5.59	-8037.06	0.0613
LosesHome	0.30	-8034.42	0.8617
DrawsHome	0.03	-8034.28	0.9860
LosesAway	2.99	-8035.76	0.2242
DrawsAway	2.73	-8035.63	0.2557
ScoreLostHome	6.83	-8037.68	0.0329
ScoreLostAway	2.88	-8035.71	0.2375



Rysunek B.1. Korelacje pomiędzy zmiennymi objaśniającymi (na górze) i estymatorami współczynników regresji dla zmiennych objaśniających (na dole).

Statystyki opisowe modeli z rozdziału 4

Zmienna	Minimum	Kwantyl 25%	Mediana	Średnia	Kwantyl 75%	Maximum
Model \mathcal{M}_{t5}						
(Intercept)1	-3.624e-03	-4.848e-04	5.216e-05	3.691e-06	5.216e-05	2.131e-02
(Intercept)2	-9.687e-03	-4.492e-04	2.031e-04	2.515e-06	2.031e-04	2.021e-02
offerprivate1	-1.158e-02	-3.638e-05	-3.638e-05	-1.485e-06	5.468e-04	1.408e-02
offerprivate2	-1.188e-02	-6.368e-04	-6.368e-04	-3.840e-07	4.147e-05	1.233e-02
furnished1	-2.381e-02	-5.773e-05	-5.769e-05	-2.392e-06	1.647e-03	9.962e-03
furnished2	2.732e-02	-2.761e-04	-2.761e-04	-1.132e-06	1.231e-04	9.371e-03
Model \mathcal{M}_{tb5}						
(Intercept)1	-2.17574	-0.92328	-0.55656	-0.39525	-0.09863	28.86639
(Intercept)2	-0.27719	0.02605	0.16482	0.16989	0.31128	1.05637
(Intercept)3	-28.86619	0.09694	0.55620	0.39526	0.92389	2.17574
offerprivate1	-28.70751	0.06298	0.37725	0.24860	0.56276	5.11750
offerprivate2	-1.767180	-0.053720	0.001746	-0.015273	0.054341	0.297646
offerprivate3	-5.11812	-0.56224	-0.37757	-0.24860	-0.06342	28.70688
furnished1	-26.181006	-0.000094	0.141195	0.087281	0.260169	0.935718
furnished2	-0.83766	-0.23752	-0.12430	-0.11979	-0.01868	1.42255
furnished3	-0.935718	-0.260146	-0.141450	-0.087283	0.000094	26.181006

Tabela C.1. Diagnostyka *leave one out* dla dwóch wybranych modeli.

- Macierze ograniczeń dla modeli \mathcal{M}_t to macierze trywialne tzn. I_2 dla każdej zmiennej.
- Macierze ograniczeń dla modeli \mathcal{M}_b to $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, dla wszystkich zmiennych z wyjątkiem modeli $\mathcal{M}_{b6}, \mathcal{M}_{b7}$, gdzie macierze dla odpowiednio umeblowania i typu ogłoszenia to $\begin{pmatrix} 1 & 1 \end{pmatrix}^T$ co zostało wymuszone przez błędny numeryczne w dopasowywaniu modeli.
- Macierze ograniczeń dla modeli \mathcal{M}_{tb} to $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$, dla wszystkich zmiennych z wyjątkiem modelu \mathcal{M}_{tb8} gdzie macierze dla wyrazu wolnego i umeblowania to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$

	\mathcal{M}_{t1}	\mathcal{M}_{t2}	\mathcal{M}_{t3}	\mathcal{M}_{t4}	\mathcal{M}_{t5}	\mathcal{M}_{t6}	\mathcal{M}_{t7}	\mathcal{M}_{t8}
(Intercept):1	-0.983 (0.069)	-1.7 (0.1)	-2.13 (0.15)	-0.24 (0.32)	-2.33 (0.15)	-1.26 (0.35)	-1.75 (0.37)	-2.05 (0.38)
(Intercept):2	-1.806 (0.063)	-1.17 (0.11)	-0.88 (0.17)	-1.8 (0.3)	-0.66 (0.17)	-1.03 (0.32)	-0.74 (0.34)	-0.44 (0.35)
offerprivate:1		1.88 (0.15)			1.43 (0.16)	1.87 (0.15)		1.42 (0.16)
offerprivate:2		-0.89 (0.13)			-0.72 (0.14)	-0.89 (0.13)		-0.73 (0.14)
furnished:1			1.77 (0.17)		1.20 (0.18)		1.75 (0.17)	1.19 (0.18)
furnished:2			-1.05 (0.18)		-0.72 (0.19)		-1.05 (0.18)	-0.72 (0.19)
$m^2:1$				-0.025 (0.011)		-0.016 (0.011)	-0.013 (0.011)	-0.0091 (0.0114)
$m^2:2$				-0.00095 (0.00987)		-0.005 (0.010)	-0.0045 (0.0099)	-0.007 (0.010)
n	2840	2840	2840	2840	2840	2840	2840	2840
AIC	5017.4	4182.1	4312.5	5007.7	3948.7	4183.6	4314.9	3952.0
BIC	5029.3	4205.9	4336.3	5031.5	3984.4	4219.3	4350.6	3999.7
RMSE	3.52	3.81	3.69	3.52	3.96	3.82	3.69	3.96
\hat{N}	7575	6066	6262	7580	5617.9	6077	6270	5624.9
$\sqrt{\widehat{\mathbb{V}}[\hat{N}]}$	351	276	282	355	246.6	279	283	247.9
G^2	1083	244	374	1059	6.4	241	373	6.4

Tabela C.2. Tabela ze statystykami opisowymi modeli tylko z efektem czasowym. Współczynniki, których nazwy kończą się na :1 (odpowiednio :2) odnoszą się do prawdopodobieństwa wystawienia mieszkania na olx (odpowiednio otodom). Liczby w nawiasie pod wartościami współczynników to błędy standardowe.

	\mathcal{M}_{b1}	\mathcal{M}_{b2}	\mathcal{M}_{b3}	\mathcal{M}_{b4}	\mathcal{M}_{b5}	\mathcal{M}_{b6}	\mathcal{M}_{b7}	\mathcal{M}_{b8}
(Intercept):1	-2.307 (0.093)	-3.058 (0.098)	-2.961 (0.094)	-3.44 (0.44)	2.78 (0.59)	-3.76 (0.46)	-3.68 (0.44)	2.8 (0.8)
(Intercept):2	0.501 (0.068)	-0.30 (0.13)	-0.99 (0.24)	1.67 (0.32)	-3.39 (0.58)	0.33 (0.35)	-0.33 (0.40)	-3.18 (0.74)
offerprivate		1.80 (0.13)				1.78 (0.13)		
furnished			2.23 (0.25)				2.22 (0.24)	
$m^2:1$				0.039 (0.015)		0.024 (0.015)	0.025 (0.015)	-0.0001 (0.0172)
$m^2:2$				-0.040 (0.011)		-0.021 (0.011)	-0.022 (0.011)	-0.0072 (0.0144)
offerprivate:1					-3.8 (0.3)			-3.8 (0.3)
offerprivate:2					3.03 (0.26)			3.01 (0.26)
furnished:1					-3.51 (0.56)			-3.54 (0.57)
furnished:2					2.76 (0.53)			2.78 (0.54)
n	2840	2840	2840	2840	2840	2840	2840	2840
AIC	5017.4	4747.0	4740.1	5007.6	4155.3	4746.8	4739.4	4158.5
BIC	5029.3	4764.9	4758.0	5031.4	4191.0	4776.5	4769.2	4206.1
RMSE	3.49	7.31	9.45	3.52	19.18	7.19	9.31	19.44
\hat{N}	3311	3442	3643	3343	10857	3453	3665	10963
$\sqrt{\hat{V}[\hat{N}]}$	52	93	198	59	4549	95	204	4662
G^2	1083	811	804	1059	213	804	798	213

Tabela C.3. Tabela ze statystykami opisowymi modeli tylko z efektem behawioralnym. Współczynniki, których nazwy kończą się na :1 (odpowiednio :2) odnoszą się do prawdopodobieństwa wystawienia mieszkania na olx i prawdopodobieństwa wystawienia na otodom jeżeli nie zostało wystawione na olx (odpowiednio prawdopodobieństwa wystawienia na otodom jeżeli jest wiemy, że zostało wystawione na olx). Liczby w nawiasie pod wartościami współczynników to błędy standardowe.

	\mathcal{M}_{tb1}	\mathcal{M}_{tb2}	\mathcal{M}_{tb3}	\mathcal{M}_{tb4}	\mathcal{M}_{tb5}	\mathcal{M}_{tb6}	\mathcal{M}_{tb7}	\mathcal{M}_{tb8}
(Intercept):1	-3.1 (1.7)	-2.4 (2.2)	-1.1 (1.9)	-1.8 (3.7)	-1.6 (3.7)	-5.6 (3.7)	-22 (30144)	-2.08 (0.35)
(Intercept):2	0.73 (0.35)	-0.53 (0.47)	-1.49 (0.85)	1.3 (1.8)	-1.7 (1.0)	0.72 (0.58)	-1.31 (0.42)	-0.39 (0.35)
(Intercept):3	1.3 (1.7)	1.3 (2.2)	0.23 (1.92)	0.063 (3.705)	0.99 (3.66)	4.5 (3.7)	21 (30144)	
offerprivate:1		0.33 (2.54)			1.8 (3.1)	0.011 (0.861)		-31 (17744)
offerprivate:2		2.20 (0.81)			1.16 (0.89)	2.34 (0.38)		3.2 (0.5)
offerprivate:3		-1.2 (2.5)			-2.5 (3.1)	-0.90 (0.85)		30 (17744)
furnished:1			-1.7 (2.7)		-2.0 (1.2)		23 (30144)	1.36 (0.17)
furnished:2			2.8 (1.0)		2.11 (0.74)		2.35 (0.28)	-0.80 (0.18)
furnished:3			0.63 (2.67)		1.2 (1.1)		-24 (30144)	
$m^2:1$				-0.065 (0.210)		0.103 (0.058)	-0.155 (0.094)	-0.016 (0.024)
$m^2:2$				-0.016 (0.032)		-0.043 (0.033)	0.013 (0.013)	-0.00034 (0.01198)
$m^2:3$				0.064 (0.210)		-0.107 (0.057)	0.149 (0.093)	0.009 (0.023)
n	2840	2840	2840	2840	2840	2840	2840	2840
AIC	5019.4	4186.1	4316.5	5011.6	3948.4	4187.7	4315.6	3949.6
BIC	5037.3	4221.8	4352.2	5047.3	4001.9	4241.3	4369.2	4009.1
RMSE	3.65	3.16	3.25	6.33	3.97	4.38	36.94	2.75
\hat{N}	3051	3164	3332	2954	3288	3185	2952	3838
$\sqrt{\widehat{\mathbb{V}}[\hat{N}]}$	349	428	716	823	764	689	128	738
G^2	1083	244	374	1059	0.005	239	371	0.28

Tabela C.4. Tabela ze statystykami opisowymi modeli z efektem czasowym i behawioralnym. Współczynniki, których nazwy kończą się na :1 (odpowiednio :2 i :3) odnoszą się do pierwotnego prawdopodobieństwa wystawienia mieszkania na pierwszej ze olx (odpowiednio pierwotnego prawdopodobieństwa wystawienia na otodom i prawdopodobieństwa wystawienia na otodom jeżeli jest wiemy, że zostało wystawione na olx). Liczby w nawiasie pod wartościami współczynników to błędy standardowe.