Metody stochastyczne w ekonometrii przestrzennej
– nowoczesna analiza asymptotyczna

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Alicja Olejnik, Jakub Olejnik

Metody stochastyczne w ekonometrii przestrzennej – nowoczesna analiza asymptotyczna

WYDAWNICTWO UNIWERSYTETU ŁÓDZKIEGO

Łódź 2020

Alicja Olejnik – Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny
Instytut Gospodarki Przestrzennej, Katedra Ekonometrii Przestrzennej
90-255 Łódź, ul. POW 3/5

Jakub Olejnik – Uniwersytet Łódzki, Wydział Matematyki i Informatyki
Katedra Informatyki Stosowanej, 90-238 Łódź, ul. Banacha 22

RECENZENT
Jan Hauke

REDAKTOR INICJUJĄCY
Beata Koźniewska

OPRACOWANIE REDAKCYJNE
Joanna Balcerak

SKŁAD I ŁAMANIE
Jakub Olejnik, *Alicja Olejnik*

KOREKTA TECHNICZNA
Leonora Gralka

PROJEKT OKŁADKI
Polkadot Studio Graficzne Aleksandra Woźniak, *Hanna Niemierowicz*

Wydrukowano z gotowych materiałów dostarczonych do Wydawnictwa UŁ

© Copyright by Alicja Olejnik, Jakub Olejnik, Łódź 2020
© Copyright for this edition by Uniwersytet Łódzki, Łódź 2020

https://doi.org/10.18778/8220-438-4

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.10085.20.0.K

Ark. druk. 10,5

ISBN 978-83-8220-438-4
e-ISBN 978-83-8220-437-7

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. 42 665 58 63

SPIS TREŚCI

Wprowadzenie

Współcześnie, w literaturze światowej, wiele miejsca poświęca się problemom analizy i modelowania zjawisk o charakterze przestrzennym. Jedno z podstawowych narzędzi analiz regionalnych stanowi modelowanie przestrzenne. Termin „ekonometria przestrzenna" zaproponował Paelinck we wczesnych latach siedemdziesiątych XX wieku. Można przyjąć, że to właśnie wówczas powstał nowy dział ekonometrii, opisujący problemy specyfikacji i estymacji modeli ekonometrycznych, wynikające z występowania autokorelacji przestrzennej. Ekonometria przestrzenna umożliwia badanie zależności przestrzennych, a także uwzględnianie ich w modelach ekonometrycznych. Z punktu widzenia ekonomii, ekonometria przestrzenna otwiera drogę do lepszego zrozumienia związków i zależności między regionami, co umożliwia lepsze opisywanie systemów ekonomicznych. Dziś ekonometria przestrzenna bywa rozumiana szeroko jako zestaw metod i narzędzi statystycznych oraz ekonometrycznych do przestrzennej analizy danych, uwzględniających wielorakie efekty przestrzenne, obecne zarówno w danych dyskretnych, jak i ciągłych.

Przez wiele lat ekonometria przestrzenna pozostawała w cieniu głównych nurtów ekonometrii. W 1988 roku Anselin pisał o niej, że „jest ignorowana w większości klasycznych podręczników ekonometrii" (por. Anselin, 1988a: 1, tłumaczenie własne). Dekadę później Anselin i Bera podkreślali, że ekonometria przestrzenna nadal „nie jest obecna w głównym nurcie ekonometrii" (por. Anselin, Bera, 1998: 237–238, tłumaczenie własne). W swojej pracy zauważali, że badacze zajmujący się analizami empirycznymi dostrzegają potrzebę zmierzenia się z problemami związanymi z obecnością autokorelacji przestrzennej danych używanych w modelowaniu regionalnym.

Warto zauważyć znaczący udział w rozwoju ekonometrii przestrzennej nie wydawnictw *stricte* ekonometrycznych, lecz czasopism związanych z naukami regionalnymi, takich jak: „Journal of Regional Science", „Regional Science and Urban Economics", „Papers in Regional Science", „International Regional Science Review", „Geographical Analysis", „Journal of Geographical Systems". Dopiero w nowym milenium sytuacja ta zaczęła się szybko zmieniać i ekonometria prze-

strzenna trafiła do głównego nurtu ekonometrii, znajdując swoje miejsce w najlepszych wydawnictwach ekonometrycznych, między innymi: „Econometrica", „Econometric Reviews", „Econometric Theory", „Journal of Applied Econometrics", „Journal of Business and Economic Statistics", „Journal of Econometrics", „Review of Economics and Statistics".

Lata osiemdziesiąte i dziewięćdziesiąte XX wieku to na arenie międzynarodowej intensywny rozwój ekonometrii przestrzennej. Powstały wówczas prace: *Spatial Processes: Models and Applications* Cliffa i Orda (1981), *Spatial Econometrics: Methods and Models* Anselina (1988a), nazywana „biblią" ekonometrii przestrzennej oraz *New Directions in Spatial Econometrics* autorstwa Floraxa i Anselina (1995).

Na rozwój ekonometrii przestrzennej wpłynęło również opracowanie tzw. nowej ekonomii geograficznej, NEG (ang. *New Economic Geography*; por. Krugman, 1991a, b; Fujita et al., 1999a, b). Jej twórca — Paul Krugman — został uhonorowany Nagrodą Nobla w roku 2008. Prace Krugmana uzasadniały na gruncie teorii ekonomii wykorzystanie analizy przestrzennej w badaniach regionalnych. Dawały również teoretyczne podstawy modelowania regionalnej konwergencji oraz koncentracji przestrzennej aktywności ekonomicznej. Obecnie, zgodnie z obowiązującymi trendami, badanie wzrostu gospodarczego w kontekście teorii NEG jednoznacznie wymaga zastosowania narzędzi ekonometrii przestrzennej.

Najwięcej opracowań książkowych z dziedziny ekonometrii przestrzennej powstało jednak w pierwszej dekadzie XXI wieku. Można tu wymienić *Spatial Autocorrelation and Spatial Filtering. Gaining Understanding Through Theory and Scientific Visualization* Griffitha (2003), *Advances in Spatial Econometrics: Methodology, Tools and Applications* Floraxa i Anselina (2004), *Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence* Arbii (2006), *Spatial Econometrics. Methods and Applications* pod redakcją Arbii i Baltagiego (2009) oraz *Introduction to Spatial Econometrics* LeSage'a i Pace'a (2009). Na uwagę zasługuje monografia *Handbook of Regional Science* pod redakcją Nijkampa i Fischera (2014), która dotyczy regionalistyki, ale zawiera 150 stron poświęconych ekonometrii przestrzennej, a także książka o charakterze aplikacyjnym, *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL* Anselina i Reya (2014).

Wśród nowszych opracowań zagranicznych można wymienić między innymi *Spatial Econometrics: Qualitative and Limited Dependent Variables* pod redakcją Baltagiego, LeSage'a i Pace'a (2016) oraz *Spatial Econometrics* autorstwa Kelejiana i Pirasa (2017). Pierwsze jest zbiorem artykułów, poświęconych głównie metodom szacowania modeli dla dyskretnych zmiennych zależnych z zależnościami przestrzennymi (przy użyciu metody największej wiarogodności) oraz dla binarnych i licznikowych zmiennych zależnych (przy użyciu metod bayesowskich). Z kolei drugie, dzięki rzetelnemu wprowadzeniu i formalizacji założeń

i zasad, stanowi podstawowe opracowanie współczesnych osiągnięć ekonometrii przestrzennej. Obejmuje modele regresji przestrzennej, macierze wag, procedury szacowania (ze szczególnym uwzględnieniem uogólnionej metody momentów, zaawansowanych procedur przedtestowych i bayesowskiej analizy danych) oraz omawia komplikacje związane z ich wykorzystaniem.

W Polsce, oprócz przekładów z lat 1982–83 prekursorskich monografii Paelincka i Klassena (1979) oraz Klaassena i innych (1979), ukazały się nieliczne publikacje książkowe z tego zakresu. Jako pierwsza została wydana *Ekonometria przestrzenna* pod redakcją Zeliasia (1991), później *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN* Kopczewskiej (2007), a także *Ekonometryczna analiza wielowymiarowych procesów gospodarczych* Szulc (2007), poświęcona wybranym elementom ekonometrii przestrzennej i tematyce pól losowych. Dopiero wieloautorska monografia *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych* pod redakcją Sucheckiego (2010), w tak szerokim zakresie omawiała nowoczesne metody i modele ekonometrii przestrzennej. Wkrótce po niej, w 2012 roku, ukazało się opracowanie pod tą samą redakcją pt. *Ekonometria przestrzenna II. Modele zaawansowane*, zawierające opis współczesnych zaawansowanych metod i modeli ekonometrycznych (wraz z przykładami ich zastosowań). W roku 2016 ukazała się kolejna pozycja w tej serii, *Ekonometria przestrzenna III. Modele wielopoziomowe — teoria i zastosowania* autorstwa Łaszkiewicz, będąca omówieniem zasad modelowania wielopoziomowego z perspektywy analizy danych zlokalizowanych przestrzennie.

W literaturze widoczne jest ciągłe poszerzanie zakresu zastosowań modeli autoregresji przestrzennej. Opracowania opisujące badania empiryczne potwierdzające wagę efektów przestrzennych powstają między innymi w naukach społecznych, geografii, biologii, ochronie środowiska, a także w ekonomii. Problem efektów przestrzennych coraz częściej uwzględnia się również w badaniach rozwoju regionalnego. Rozważane są, między innymi, przestrzenne aspekty konwergencji regionalnej, infrastruktury regionów, a nawet demografii.

Ze względu na rosnącą popularność stosowania metod ekonometrii przestrzennej w badaniach empirycznych oraz dynamiczny rozwój związanej z tym metodologii, powstaje potrzeba tworzenia spójnych, uzasadnionych matematycznie podstaw wnioskowania ekonometrycznego. Wiele z dostępnych w Polsce i na świecie opracowań nie traktuje tego aspektu ekonometrii z dostateczną uwagą. Prowadzi to czasem do nieścisłości, a w efekcie może niestety nie pozostawać bez znaczenia dla wyciąganych wniosków. Za pośrednictwem naszego opracowania staramy się wypełnić lukę istniejącą na rodzimym rynku wydawniczym i przedstawić polskim czytelnikom publikację o formalnych podstawach metod ekonometrii przestrzennej. Przedstawione treści (nieznacznie uzupełnione) mogą stanowić bazę wykładu kursowego z ekonometrii przestrzennej dla studentów matematyki, statystyki czy ekonometrii. Pozycja ta może też być punktem

wyjścia dla osób rozwijających teorię ekonometrii. W szczególności czytelników zainteresowanych teorią asymptotyczną estymatorów w kontekście ekonometrii przestrzennej.

Tematyka podjęta w monografii

Niniejsza monografia prezentuje najnowsze rezultaty z zakresu teorii asymptotycznych dla modeli stochastycznych ekonometrii przestrzennej. Omówione w książce wyniki pracy naukowej autorów poprzedzone są przeglądem wybranych klasycznych zagadnień tej dziedziny ekonometrii, przedstawionych w nowoczesnym ujęciu. Przeprowadzone rozumowania osadzone są w ramach precyzyjnego wywodu matematycznego, a tym samym dają czytelnikom solidne podstawy do empirycznych badań ekonomicznych oraz do dalszych rozważań metodologicznych. Opracowanie to ma na celu przybliżenie podstawowych i wybranych zaawansowanych metod stochastycznych ekonometrii przestrzennej, ze szczególnym uwzględnieniem ich własności asymptotycznych.

Elementem centralnym prezentowanej teorii jest zagadnienie asymptotyki przestrzennej macierzy wag. Od lat dziewięćdziesiątych XX wieku wiadomo, że zachowanie macierzy wag — przy rosnącym do nieskończoności rozmiarze próby — ma decydujące znaczenie dla własności modeli ekonometrii przestrzennej (chociażby identyfikowalności parametrów czy własności związanych z nimi estymatorów). W niniejszej publikacji autorzy podjęli temat rozszerzalności znanych teorii ekonometrii przestrzennej na bardziej obszerne klasy schematów interakcji przestrzennych. Dzięki konstrukcji tzw. niesumowalnych macierzy wag, w specyfikacji modelu zjawiska można uwzględnić bardziej złożone zależności między badanymi jednostkami.

Problem macierzy niesumowalnych w kontekście ekonometrii przestrzennej jako pierwsi dostrzegli Gupta i Robinson (2018). Uzyskali oni wynik o zgodności pewnych estymatorów opartych na metodzie największej wiarogodności. Pytanie o rozkład asymptotyczny oszacowań pozostało jednak otwarte. Dopiero w pracy Olejnik i Olejnik (2020) sformułowano teorię matematyczną podającą kompletną odpowiedź w kontekście procesów z autoregresją zmiennej zależnej. W niniejszej monografii rozszerzamy tę teorię. W szczególności zajmujemy się specyfikacją modeli ekonometrycznych z autoskorelowanym składnikiem losowym, w których dopuszczona jest możliwość zależności przestrzennych wyższego rzędu z wektorowym współczynnikiem zależności. Przedstawione tutaj wyniki nie były wcześniej publikowane. Uzyskane przez nas centralne twierdzenie graniczne stosujemy również do udowodnienia zbieżności według dystrybuanty standaryzowanych statystyk: I Morana, testu mnożników Lagrange'a oraz informanty Rao, przy zastosowaniu niesumowalnych macierzy wag przestrzennych.

W rozdziale I wprowadzamy podstawowe pojęcia i oznaczenia, wykorzystywane w dalszej części pracy. Definiujemy formalnie pojęcie przestrzennego procesu stochastycznego i przedstawiamy ideę przestrzennej macierzy wag (wraz z przykładami). Przedstawiamy również podstawowe dla ekonometrii przestrzennej pojęcie autokorelacji przestrzennej oraz popularne statystyki stosowane do testowania obecności tego zjawiska.

W kolejnym rozdziale dokonujemy przeglądu popularnych specyfikacji modeli ekonometrii przestrzennej. Prezentację zaczynamy od tych najprostszych (model czystej autoregresji przestrzennej, autoregresji przestrzennej SAR, model z przestrzennie autoskorelowanym składnikiem losowym SEM), a kończymy na najbardziej złożonych, rozważanych często tylko teoretycznie, modelach wyższych rzędów klasy SARAR, czy modeli ze średnią ruchomą SARARMA. W dalszej części omawiamy teorię pozwalającą na poprawną interpretację oszacowań parametrów modeli autoregresji przestrzennej, opartą na przekształceniu równania specyfikacji modelu do postaci jawnej w celu wyliczenia tzw. efektów bezpośrednich i pośrednich. Znaczna część tego rozdziału została poświęcona podstawom matematycznym wybranych procedur estymacji. W szczególności pokazujemy, że mimo niezgodności estymatora najmniejszych kwadratów w przypadku ogólnym, dla pewnych klas macierzy wag (w przeciwieństwie do innych metod) może on oferować dobrej jakości oszacowania. Zaprezentowane metody estymacji obejmują również metody: zmiennych instrumentalnych, największej wiarogodności, uogólnioną metodą najmniejszych kwadratów oraz uogólnioną metodę momentów.

W rozdziale III badamy własności statystyk testowych, a w szczególności statystyki Morana. Przeprowadzamy formalne rozumowania prowadzące do uzyskania właściwych rozkładów asymptotycznych, przy zastosowaniu niesumowalnych macierzy wag przestrzennych. Rozważamy również problem tzw. niestacjonarności oraz — na bazie znanego testu niestacjonarności przestrzennej — prezentujemy procedurę testową Kosfelda–Lauridsena–Olejnika.

W rozdziale IV przedstawiamy kompletny wywód formalny, wykazujący zgodność estymatorów quasi-największej wiarogodności (QNW) dla modeli autoregresyjnych. Elementem nowatorskim jest ominięcie w sformułowanej teorii założenia o sumowalności przestrzennej macierzy wag, co poszerza stosowalność przestrzennych modeli ekonometrycznych. W prezentowanych rozumowaniach stosujemy złagodzone wymagania, dotyczące rozkładu składnika losowego modelu. W szczególności, rozkład ten nie musi być gaussowski, a elementy wektora zaburzeń nie muszą być niezależne. Opuszczamy również założenie równości rozkładów między elementami. Co więcej, dopuszczamy funkcje gęstości odpowiednich rozkładów o ogonach grubszych niż zakłada standardowa teoria asymptotyczna.

W rozdziale V monografii zamieszczamy autorskie centralne twierdzenie graniczne dla form liniowo-kwadratowych. Przy pomocy tego wyniku uzyskujemy twierdzenia o rozkładzie granicznym oszacowań estymatorów QNW. W szczególności wykazujemy asymptotyczną normalność oszacowań pochodzących z gaussowskiej estymacji QNW, przy niejednorodnych i niegaussowskich zaburzeniach. Istotnym elementem jest rozluźnienie warunków nakładanych na asymptotyczne zachowanie się macierzy wag użytych w specyfikacji modelu przestrzennego.

Rozdział I

Wprowadzenie do modelowania przestrzennego

Wstęp

Regresja liniowa jest jedną z najpopularniejszych technik eksploracji danych, zakłada jednak, że obserwacje w próbie są warunkowo niezależne. To założenie niekoniecznie jest spełnione w przypadku danych przestrzennych, np. danych geograficznych lub w szczególnym przypadku danych pochodzących z badań zjawisk zachodzących w przestrzeni. Wówczas na przebieg obserwowanego procesu często wpływa relacja sąsiedztwa, a ogólniej — układ odległości między jednostkami podlegającymi badaniu. W takim przypadku mamy do czynienia z tzw.

autokorelacją przestrzenną. Ze swojej natury, autokorelacja przestrzenna jest ściśle związana z rozmieszczeniem jednostek oraz interakcjami przestrzennymi, a jej idea wykorzystuje pierwsze prawo geografii, sformułowane w 1970 roku przez Waldo Toblera: *Everything is related to everything else, but near things are more related than distant things* (Tobler, 1970: 236).

Reguła Toblera mówi o tym, że siła oddziaływań między obiektami blisko położonymi w przestrzeni jest większa, niż pomiędzy obiektami znajdującymi się w dużym oddaleniu. W naszym opracowaniu słowo „przestrzeń" będziemy rozumieć w szerokim sensie. Zauważmy, że możemy mieć do czynienia nie tylko z przestrzenią fizyczną czy geograficzną. Zbiór jednostek, z których pochodzą dane, może być właściwie dowolnym zbiorem, o ile określona jest relacja sąsiedztwa bądź metryka reprezentująca odległość między jego elementami. W szczególności, przestrzenią może być oś punktów czasowych (tak jak w teorii szeregów czasowych), wycinek mapy reprezentujący obszar geograficzny, a także czasoprzestrzeń powstająca jako iloczyn kartezjański dowolnej przestrzeni abstrakcyjnej i osi czasowej (tak jak w badaniach panelowych).

Uwzględnienie prawa Toblera w analizach empirycznych wymaga precyzyjnego zdefiniowania pojęć bliskości (sąsiedztwa) lub odległości. W takim przypadku możliwym będzie uwzględnienie w modelu zjawiska zależności, na przykład korelacyjnej bądź regresyjnej, zmiennych losowych reprezentujących obserwacje pochodzące z jednostek bliskich sobie w przestrzeni. Zauważmy, że w przypadku szeregów czasowych kierunek uporządkowania jest naturalny. Przykład prostej autoregresji $y_t = \rho y_{t-1} + \varepsilon_t$, gdzie $t = 1, 2, \ldots, T$, pokazuje, że mamy tu do czynienia z jednokierunkową zależnością o zwrocie zgodnym z upływem czasu, a (niesymetryczna ze względu na przyczynowość) relacja bliskości ograniczona jest do sąsiednich okresów. Ogólnie w przestrzeni nie ma jednak takiego intuicyjnego uporządkowania, zależności mogą być wielokierunkowe, a ich zwrot nie musi być jednoznacznie określony. Niemniej jednak, siła zależności przestrzennych będzie zależna od odległości pomiędzy obiektami.

W niniejszym rozdziale wprowadzone zostaną podstawowe pojęcia i oznaczenia wykorzystywane w dalszej części pracy. W części pierwszej przedstawiona zostanie formalna definicja przestrzennego procesu stochastycznego. W kolejnej sekcji zaprezentujemy ideę przestrzennej macierzy wag. Zdefiniowanie tego pojęcia umożliwi wprowadzenie do modeli matematycznych informacji *a priori* o zależnościach przestrzennych. Ostatnia część rozdziału poświęcona będzie definicji autokorelacji przestrzennej. Zaprezentujemy tam również przegląd popularnych testów statystycznych, stosowanych do wykrywania obecności autokorelacji przestrzennej w analizowanych danych. W szczególności rozważamy popularną statystykę I Morana na poziomie zarówno globalnym jak i lokalnym. Ścisłe matematycznie rozumowanie dotyczące rozkładów asymptotycznych tej statystyki zaprezentowane zostanie jednak dopiero w rozdziale III.

1. Procesy stochastyczne w przestrzeni

1.1. Interakcje przestrzenne

W przypadku analiz procesów ekonomicznych obserwowanych w przestrzeni można założyć, że jednostki przestrzenne, np. regiony, nie stanowią niezależnych, odizolowanych gospodarek, ale wzajemnie na siebie oddziałują. Zgodnie z prawem Toblera, im bliżej położone są dwa obiekty w przestrzeni, tym większa jest siła interakcji między nimi. Występowanie zależności przestrzennych wynika najczęściej z charakteru procesów ekonomicznych, które są ograniczone w przestrzeni, a ich intensywność stanowi funkcję odległości. W efekcie, zależności przestrzenne występują wówczas, gdy wartość zmiennej z jednej lokalizacji jest uzależniona od jej wartości w innych lokalizacjach. Oddziaływania te mogą wynikać z heterogeniczności przestrzennej (ang. *spatial heterogeneity*), czyli przestrzennego zróżnicowania procesu ekonomicznego, bądź też z tzw. efektów przestrzennych (ang. *spatial effects*), wśród których wyróżniamy efekty zewnętrzne (ang. *externalities*) oraz efekty rozprzestrzeniania się (ang. *spillover effects*).

Heterogeniczność przestrzenna wiąże się z problemem niestabilności relacji ekonomicznych w przestrzeni geograficznej, co może skutkować niestabilnością parametrów strukturalnych w modelu ekonometrycznym (por. Anselin, 1988a; Suchecki [red.] 2010). W przypadku występowania heterogeniczności przestrzennej wskazuje się na występowanie tzw. reżimów przestrzennych (ang. *spatial regimes*), czyli obszarów, które ze względu na przebieg procesów ekonomicznych są wewnętrznie spójne. Zauważmy bowiem, że zjawiska ekonomiczne mogą inaczej przebiegać np. w regionach centralnych i peryferyjnych, w dużych aglomeracjach i na obszarach wiejskich czy w krajach tzw. starej i nowej Unii Europejskiej.

Przestrzenne efekty zewnętrzne są zjawiskiem polegającym na przeniesieniu części korzyści ekonomicznych wynikających z działalności jednego podmiotu na podmioty sąsiednie, bez odpowiedniej rekompensaty. Na występowanie efektów zewnętrznych mogą wpływać efekty występujące między producentami i konsumentami lub grupami producentów. Przykładem — właściciel przedsiębiorstwa, zlokalizowanego w pobliżu miasta uniwersyteckiego, który będzie odnosił korzyści w postaci możliwości zatrudnienia lepiej wykształconej kadry pracowniczej, nie ponosząc kosztów na rzecz ich edukacji czy samego uniwersytetu.

Rozprzestrzenianie się jest w pewnym sensie jednym z efektów zewnętrznych, choć dotyczy sytuacji, w której badane zjawisko wpływa samo na siebie, powodując zależności przestrzenne. Na przykład, przedsiębiorca otwierający sklep, klub lub restaurację może zdecydować się na lokalizację obok sklepu tej samej branży. Wówczas tworzy się pewnego rodzaju klaster podmiotów, oferujący szeroki wybór towarów i usług, przyciągający większą grupę konsumentów. Zauważmy, że zarówno w przypadku występowania efektów zewnętrznych jak i efektów

rozprzestrzeniania się obserwujemy kumulowanie się aktywności ekonomicznej w przestrzeni, jednak jego przyczyny są różne.

Niezależnie od procesów ekonomicznych leżących u podstaw badanego zjawiska, obserwowane efekty przestrzenne, nazywane również efektami zarażania (ang. *contagion effects*), mogą mieć dwojaką naturę. Powiemy, że ma miejsce właściwy efekt zarażania (ang. *true contagion effect*), gdy poziom określonego procesu ekonomicznego na pewnym obszarze ma dodatni wpływ na jego rozwój na obszarach sąsiednich (przestrzenne dodatnie sprzężenie zwrotne). O pozornym efekcie zarażania (ang. *spurious contagion effect*) mówimy, gdy zamiast rzeczywistego wpływu obserwujemy występowanie pewnych wspólnych, sprzyjających warunków zewnętrznych. Dla przykładu można rozważyć grupowanie się przedsiębiorstw, wynikające z faktu istnienia wspólnych korzyści ze współdzielenia lokalizacji, np. współpraca (właściwy efekt zarażania) lub korzystania z określonej właściwości samej lokalizacji, np. ulgi podatkowe, dogodna infrastruktura komunikacyjna (pozorny efekt zarażania).

Rozróżnienie analityczne powyższych efektów może mieć miejsce w doborze odpowiedniej postaci algebraicznej przestrzennego modelu ekonometrycznego (por. przegląd specyfikacji w rozdziale II). Niemniej jednak, podanie prostego przepisu na reprezentację zjawiska ekonomicznego poprzez specyfikację modelu nie jest możliwe. Należy zauważyć, że w przypadku rzeczywistych procesów mamy najczęściej do czynienia z kombinacją różnych efektów, działających jawnie lub w sposób niebezpośredni, prowadzących wspólnie do autokorelacji przestrzennej danych.

Poza opisanymi wcześniej mechanizmami, na występowanie zależności przestrzennych może mieć również wpływ pewien aspekt czysto techniczny. Jak wyjaśnia Olejnik (Suchecki [red.] 2010), na etapie przygotowywania analiz dane są grupowane pod względem przynależności do jednostek administracyjnych (takich jak gminy, powiaty, województwa), a nie struktury badanego zjawiska. W konsekwencji, jeśli wartości analizowanej zmiennej wykraczają poza ustalone granice, między sąsiadującymi ze sobą obiektami obserwujemy występowanie interakcji. Zagadnienie to nazywane jest problemem MAUP (ang. *Modifiable Areal Unit Problem*, patrz Arbia, 1989) i jest kategoryzowane jako błąd pomiaru, wynikający z rozbieżności między jednostką administracyjną poddawaną pomiarowi a obszarem, na którym zjawisko faktycznie ma miejsce. Problem ten jest częstą przyczyną obciążeń analiz statystycznych, w których odgórne nadanie regionom granic administracyjnych powoduje, że zależności przestrzenne i heterogeniczność danych przestrzennych mogą nawet być generowane sztucznie. Powszechnie uznaje się, iż jeśli nie ma możliwości pozyskania danych nieobarczonych błędem MAUP, wówczas pewnym sposobem na eliminację jego skutków jest uwzględnienie autokorelacji przestrzennej w postaci modelu ekonometrycznego.

1.2. Definicja przestrzennego procesu stochastycznego

Rozważmy zbiór obiektów w przestrzeni lokalizacji $P = 1, \ldots, N$. W dalszej części opracowania dla uproszczenia będziemy nazywać je lokalizacjami, jednostkami przestrzennymi bądź — dla lepszego zilustrowania omawianego pojęcia — regionami. Poszczególne jednostki będziemy identyfikować z ich indeksami: $i = 1, \ldots, N$.

W praktyce, w ekonometrii przestrzennej, przez przestrzenny proces stochastyczny rozumie się (najczęściej kolumnowy) wektor losowy

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix},$$

w którym każda ze współrzędnych x_i, $1 \leqslant i \leqslant N$, jest zmienną losową określoną na jednej wspólnej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ reprezentującą wartość procesu, obserwowaną w odpowiednich lokalizacjach $i \in P$. Struktura losowości, a tym samym struktura zależności pomiędzy poszczególnymi zmiennymi procesu, modelowana jest poprzez dystrybuantę rozkładu łącznego. Tak więc, jeśli $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N]^{\mathsf{T}} \in \mathbb{R}^N$ jest wektorem liczbowym potencjalnych realizacji procesu $\mathbf{x}$, to dystrybuanta stochastycznego procesu przestrzennego dana jest przez zależność

$$F_{\mathbf{x}}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N) = \sqrt{(2\pi)^{-N} \det \mathbf{V}} \int_{\{\boldsymbol{x} \leqslant \tilde{\mathbf{x}}\}} e^{-\frac{1}{2} Q_{\mathbf{V}}(\boldsymbol{x} - \boldsymbol{\mu}_0)} \mathrm{d}\boldsymbol{x},$$

gdzie $\boldsymbol{\mu}_0$ jest wektorem odpowiednich wartości oczekiwanych, a $\mathbb{R}^N \ni \boldsymbol{x} \mapsto Q_{\mathbf{V}}(\boldsymbol{x}) \in \mathbb{R}$ jest formą kwadratową o dodatnio określonej (symetrycznej) macierzy $\mathbf{V}$, której macierzowa odwrotność reprezentuje strukturę zależności procesu przestrzennego.

Zauważmy, że jeśli rozważamy ustalony i skończony zbiór lokalizacji $P = \{1, \ldots, N\}$, to — z punktu widzenia aparatu matematycznego — teoria przestrzennych procesów stochastycznych może być sprowadzona do teorii wektorów losowych ustalonego, skończonego wymiaru. Aby rozważać własności o charakterze asymptotycznym, musimy jednak rozszerzyć definicję procesu przestrzennego. W praktyce interesują nas często wartości pewnej liczbowej (losowej) charakterystyki $\theta(\mathbf{x})$ obserwowanego procesu $\mathbf{x}$. Taką charakterystyką może być pewna statystyka (funkcja obserwacji) procesu lub estymator nieznanego parametru. W teorii ekonometrii w naturalny sposób pojawiają się wtedy takie pojęcia jak zgodność czy asymptotyczna normalność, które wymagają (chociażby potencjalnej) możliwości nieograniczonego zwiększania rozmiaru próby. W praktyce

nie dysponujemy jednak wieloma realizacjami tego samego procesu losowego, a raczej jedną jego trajektorią. Zatem wnioskowanie opiera się na założeniu, że to dziedzina przestrzenna procesu potencjalnie rośnie nieograniczenie. Innymi słowy, asymptotyka zjawiska przestrzennego obserwowana jest nie przez wielokrotne próbkowanie trajektorii procesu, a raczej przez zwiększanie rozmiaru dziedziny przestrzennej N do nieskończoności.

Definicja

Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie ustaloną przestrzenią probabilistyczną. Przez **przestrzenny proces stochastyczny** będziemy rozumieć funkcję losową $\mathbf{x} = \mathbf{x}(N)$ postaci

$$\mathbb{N} \ni N \mapsto \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \bigtimes_{i=1}^{N} L_0(\Omega, \mathcal{F}, \mathbb{P}),$$

W przypadku standardowych modeli statystki i ekonometrii, w których elementy próby są niezależne, potencjalny wzrost rozmiaru próby można łatwo interpretować jako np. uzupełnienie/rozszerzenie próby o dodatkowe, niezależne obserwacje o analogicznej strukturze losowości. Mianowicie moglibyśmy dokonać dodatkowego losowego wyboru elementów populacji lub iteratywnie wykonywać eksperyment i rozszerzać próbę o zarejestrowane wyniki pomiarów.

Gdy próba ma jednak charakter przestrzenny, a obserwacje pochodzą z fizycznych lokalizacji, wówczas zbudowanie podobnej interpretacji może nastręczać pewne trudności. Podstawowym problemem jest brak fizycznej możliwości rozszerzenia próby. Jeśli, dla przykładu, badanie dotyczy krajów członkowskich Unii Europejskiej, wówczas wzrost próby mógłby nastąpić tylko w przypadku rozszerzania wspólnoty. Pozostaje zatem pytanie o sposób rozumienia takiego hipotetycznego rozszerzenia, a w efekcie również, o sposób interpretowania asymptotycznego zachowania się statystyki $\mathbb{N} \ni N \mapsto \theta(\mathbf{x})$, przy rosnącym $N \to \infty$. Poniżej przedstawimy krótko trzy możliwe interpretacje, oparte na różnych wzorcach hipotetycznego wzrostu rozmiaru próby.

Najprostsze podejście polega na przyjęciu asymptotyki trywialnej (ang. *simple asymptotics*), polegającej na założeniu, że jesteśmy w stanie potencjalnie przeprowadzić cały eksperyment ponownie i, tym samym, zebrać kolejne, niezależne próbki wartości całego obserwowanego wektora. Takie podejście ma oczywiste ograniczenie. Ponieważ, *de facto*, nigdy takiego ponownego losowania i pomiaru nie wykonujemy, w efekcie dokonujemy wnioskowania statystycznego na podstawie tylko jednej realizacji procesu przestrzennego. Co za tym idzie, taka hipotetyczna asymptotyka nie pozostawia miejsca na wzrost ilości informacji o zależnościach przestrzennych pomiędzy obserwacjami. W prawdzie można sformuło-

wać twierdzenia o własnościach asymptotycznych statystyk w tym modelu, lecz w praktyce trudno jest obronić zasadność użycia takiego wyniku do wnioskowania dla próby skończonej — efektywnie jednoelementowej! Dodatkowo, hipotetyczna wielkość próby zawsze stanowi wielokrotność wyjściowej wartości N.

Drugim modelem jest tzw. asymptotyka rosnącej dziedziny (ang. *increasing domain asymptotics*). Zakłada ona, że jesteśmy w stanie zwiększać nieograniczenie rozmiar próby poprzez rozszerzanie fizycznej dziedziny procesu. Inaczej, w efekcie zwiększania średnicy dziedziny przestrzennej zaliczane są do niej coraz to nowe jednostki. Istotne jest, aby nowo dołączone obszary charakteryzowały się podobną strukturą zależności przestrzennych procesu. Ostatecznie, istnieje pewna liczba $\eta > 0$ ograniczająca od dołu minimalną odległość między dowolnymi jednostkami przestrzennymi.

Podejściem w pewnym sensie odwrotnym w stosunku do powyższego jest koncepcja asymptotyki wypełniania (ang. *infill asymptotics*), w której zakłada się ograniczenie średnicy dziedziny przestrzennej. W takim przypadku nowe jednostki przestrzenne pojawiają się w sposób równomierny, np. pomiędzy istniejącymi lokacjami w próbie. W efekcie, minimalna odległość pomiędzy jednostkami przestrzennymi dąży do zera.

2. Przestrzenna macierz wag

2.1. Definicja i przykłady

Aby opisać i uwzględnić w modelu ekonometrycznym przestrzenną strukturę sąsiedztwa, stosuje się tzw. macierz wag przestrzennych (ang. *spatial weight matrix*), której elementy mają interpretację mnożników nadających wagi składnikom w zależności linowej. Zwykle przyjmowana jest ona *a priori* i uważa się, że ma charakter egzogeniczny. Zauważmy, że w polskiej terminologii zamiennie używany jest termin przestrzenna macierz wag, patrz Suchecki [red.] (2010).

Definicja

Przyjmijmy, że zbiór liczb naturalnych $\{1, \ldots, N\}$ indeksuje elementy w N-elementowej próbie jednostek przestrzennych podlegającej badaniu. Macierz kwadratową $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ o wymiarach $N \times N$, reprezentującą liczbowo moc zależności między obiektami, nazywamy **macierzą wag przestrzennych**. Dla dowolnych $1 \leqslant i \neq j \leqslant N$, element w_{ij} reprezentuje wpływ, jaki wywiera jednostka przestrzenna j na jednostkę i.

Należy tutaj zaznaczyć, że w większości przypadków (również w tym rozdziale) przyjmuje się, że macierz wag ma zerową przekątną, tj. $w_{ii} = 0$, dla wszystkich $1 \leqslant i \leqslant N$, a wszystkie elementy macierzy są nieujemne. W ogólności tak

jednak być nie musi. W przypadku macierzy wag przestrzennych, rozważanych na potrzeby przestrzennych modeli ekonometrycznych, których oryginalna specyfikacja podlega liniowemu przekształceniu (np. usuwaniu efektów stałych) to założenie nie musi być spełnione (patrz Olejnik, Olejnik, 2020).

Poniżej zamieszczamy przykłady macierzy wag przestrzennych, powszechnie stosowanych w badaniach ekonomicznych z geograficzną dziedziną procesu przestrzennego.

Przykład

Macierz $\mathbf{W} = [f(d_{ij})]_{1 \leqslant i,j \leqslant N}$ — gdzie d_{ij} oznacza odległość (najczęściej euklidesową) między lokalizacjami i oraz j (najczęściej centroidami obiektów przestrzennych), a $f\colon [0,\infty) \to [0,\infty)$ jest dowolną funkcją malejącą — nazywamy przestrzenną macierzą odległości. W praktyce najczęściej stosuje się funkcję wykładniczą lub potęgową. W przypadku funkcji wykładniczej przestrzenna macierz odległości przyjmuje postać $\mathbf{W} = [e^{-\alpha d_{ij}}]_{1 \leqslant i,j \leqslant N}$, gdzie parametr $\alpha > 0$ przyjmowany jest *a priori*. Dla funkcji potęgowej zaś $\mathbf{W} = [{d_{ij}}^{-\alpha}]_{1 \leqslant i,j \leqslant N}$, gdzie parametr α jest dodatni.

Należy tutaj wspomnieć, że do uzyskania pożądanych własności asymptotycznych modeli przestrzennych w przypadku macierzy odległości opartej na funkcji potęgowej, może okazać się konieczne zastosowanie tzw. punktu odcięcia. Zakłada się wtedy dodatkowo, że dla jednostek przestrzennych i, j takich, że d_{ij} przekracza określony poziom D, wzajemny wpływ jednostek jest znikomy, a więc $w_{ij} = 0$, dla $d_{ij} > D$. Alternatywnie rozważa się też możliwość przeskalowania macierzy $\mathbf{W}$ jej największą wartością własną (np. Elhorst, 2001; Vega, Elhorst, 2015; Olejnik, Olejnik, 2020; patrz również rozważenia dotyczące asymptotyki przestrzennej macierzy wag dalej w tym rozdziale).

Przykład

Przestrzenną macierzą ustalonego promienia nazywamy macierz $\mathbf{W}$, której elementy w_{ij} są równe 1, jeśli odległość d_{ij} pomiędzy dwoma lokalizacjami $1 \leqslant i, j \leqslant N$ nie przekracza pewnej, ustalonej *a priori* odległości d^*, a w przeciwnym wypadku są równe zero. Mamy zatem

$$w_{ij} = \begin{cases} 1, & 0 \leqslant d_{ij} \leqslant d^* \\ 0, & d_{ij} > d^* \end{cases}.$$

Przykład

Macierz, której elementy określają, czy dwa obiekty przestrzenne mają wspólną granicę, czy też nie, nazywamy przestrzenną macierzą wspólnych granic. Dokładniej, elementy macierzy $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ są zdefiniowane następująco

$$w_{ij} = \begin{cases} 1, & \text{jeśli obiekty } i \text{ i } j \text{ mają wspólną granicę} \\ 0, & \text{w pozostałych przypadkach} \end{cases}.$$

Przykład

Macierz $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$, której elementy określone są przynależnością obiektu j do zbioru k-najbliższych sąsiadów, NN_k, zgodnie ze wzorem

$$w_{ij} = \begin{cases} 1, & \text{dla } j \in NN_k \\ 0, & \text{w pozostałych przypadkach} \end{cases},$$

nazywamy przestrzenną macierz k-najbliższych sąsiadów.

W praktyce do najczęściej wykorzystywanych macierzy wag należą te oparte na odległości oraz sąsiedztwie (por. Anselin, 1988a), choć w przypadku niektórych analiz badacze widzą potrzebę uogólniania klasycznych definicji, czy wręcz poszukują nowych, bardziej elastycznych metod określania struktury przestrzennej. W szczególności rozważane są schematy anizotropowe, a nawet, co wiąże się z ryzykiem endogeniczności macierzy $\mathbf{W}$, oparte na wartościach zmiennych obserwowanych (por. Deng, 2008; Corrado, Fingleton, 2011; Kelejian, Piras, 2014; Olejnik i in., 2020).

Dyskusja na temat poprawnego określenia struktury przestrzennej badanego procesu poprzez odpowiednią macierz wag trwa od powstania omawianej dyscypliny. Poszukując dróg do jak najwierniejszego oddania zróżnicowań struktury przestrzennej, w literaturze można znaleźć liczne przykłady metod i procedur uwzględniających dodatkowe informacje wzbogacające specyfikację modelu. Przykładem może być tu praca Daceya (1968), który zbudował asymetryczną macierz wag, łączącą binarną macierz sąsiedztwa z wielkością regionu oraz długością wspólnych granic. Cliff i Ord (1981) zaproponowali macierz wag zawierającą kombinację miary odległości i długości wspólnych granic. Z kolei Bodson i Peeters (1975) przedstawili koncepcję macierzy dostępności, łączącą odległość pomiędzy poszczególnymi regionami z różnymi kanałami komunikacyjnymi. Na macierz odległości nałożyli oni wagi, związane z dostępnością środków transportu, takich jak drogi czy linie kolejowe.

Innym ciekawym przykładem jest przestrzenna macierz wag Besnera (2002), skonstruowana na podstawie miar podobieństw w zmiennych socjoekonomicznych. W pracy Getis i Aldstadt (2004) zaproponowano model, w którym przestrzenna macierz wag została skonstruowana w oparciu o lokalną statystykę Getisa-Orda G_i (por. Getis, Ord 1992; Ord, Getis 1995).

Z kolei w pracy Panak (2006) zaproponowano uwzględnienie w macierzy wag zarówno aspektów czysto geograficznych, jak i powiązań infrastruktury komunikacyjnej. Założono, iż czas podróży między obiektami przestrzennymi lepiej reprezentuje przestrzenną strukturę badanego procesu niż klasyczna macierz wag przestrzennych. Badając liczbę i jakość połączeń kolejowych i lotniczych oraz klasę dróg i autostrad wykazano, iż wybrane jednostki, odległe od siebie geograficznie, są tak dobrze skomunikowane, że są „bliżej" siebie niż wynikałoby

to z odległości euklidesowej i powinny być rozważane jako obiekty sąsiednie. Podobnie jak w przypadku wcześniej opisanych macierzy wag, konstrukcja wymagała ustalenia wielu nieznanych parametrów *a priori*.

W literaturze spotyka się również próby parametryzacji macierzy wag (np. Elhorst, Halleck, 2013), jednak należy zachować tu szczególną ostrożność. Jak zauważa Anselin (1988a), równoczesna estymacja parametrów macierzy wag z współczynnikami równania może prowadzić do problemów z efektywnością estymacji, a także z interpretacją tak otrzymanych wyników. Dodatkowo może wskazywać na istnienie zależności pozornych.

Problem właściwej specyfikacji macierzy wag przestrzennych pozostaje otwarty. Trudno jednak o jednoznaczne wytyczne ze względu na zróżnicowanie czynników wpływających na strukturę zależności przestrzennych i ich zależność od przedmiotu badań. Należy też zauważyć, że niektóre analizy empiryczne wymagają zastosowania niestandardowych koncepcji reprezentacji struktury przestrzennej. Wystarczy tu choćby wskazać problem obiektów brzegowych i odizolowanych przestrzennie, niepozostających w bezpośrednim sąsiedztwie z innym obiektem, a przecież nie stanowiących wyalienowanych gospodarek. W niektórych badaniach struktura przestrzenna powinna zatem uwzględniać również oddziaływania komunikacyjne, ekonomiczne i socjoekonomiczne (np. dojazdy do pracy, kontakty handlowe, a nawet powiązania etniczne).

Innym zagadnieniem jest problem dopuszczalnej ilości zależności wyrażonej w macierzy wag, tak aby modelowanie i wnioskowanie ekonometryczne pozostawało możliwe. Ten problem nakreślamy w następnym podrozdziale.

2.2. Asymptotyka macierzy wag

Zdefiniowanie przestrzennej macierzy wag umożliwia sformułowanie pojęcia opóźnienia przestrzennego, które jest analogiem operatora opóźnienia, znanego z teorii szeregów czasowych. Przypomnijmy, że dla procesu stochastycznego (x_t) operator opóźnienia L jest definiowany poprzez zależność $L[x_t] = x_{t-1}$. Określenie opóźnienia przestrzennego, ze względu na swoistą „wielokierunkowość" procesu, nie jest tak intuicyjne, jak w przypadku szeregów czasowych.

Definicja

Niech $\mathbf{x} = (x_i)_{i=1}^N$ będzie przestrzennym procesem stochastycznym, i niech $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ ustaloną macierzą wag. **Opóźnieniem przestrzennym** (ang. *spatial lag*) $L[x_i]$ procesu $\mathbf{x}$ w lokacji $1 \leqslant i \leqslant N$ nazywamy średnią ważoną wartości procesu $\mathbf{x}$ w lokacjach sąsiadujących

$$L[x_i] = \sum_{j=1}^{N} w_{ij} x_j.$$

Zatem sąsiedztwo, a także wagi, określone są przez odpowiednie elementy przestrzennej macierzy wag. Dokładniej, element w_{ij} można liczbowo interpretować jako siłę wpływu jednostki przestrzennej j na jednostkę i.

Interpretację wartości elementów w_{ij} jako wag przestrzennych ułatwia powszechność stosowania macierzy tzw. standaryzowanych (określenie zamienne: normalizowanych) wierszowo, o elementach nieujemnych, tj. $\sum_{j=1}^{N} w_{ij} = 1$, dla $1 \leqslant j \leqslant N$, oraz $w_{ij} \geqslant 0$, dla $1 \leqslant i, j \leqslant N$. W modelach ekonometrycznych rozważać będziemy przestrzenne opóźnienie zmiennej zależnej $\mathbf{Wy} = (Ly_i)_{i=1}^{N} = \left(\sum_{j=1}^{N} w_{ij} y_j\right)_{i=1}^{N}$, przestrzenne opóźnienie zmiennych egzogenicznych $\mathbf{WX} = (Lx_i)_{i=1}^{N} = \left(\sum_{j=1}^{N} w_{ij} x_j\right)_{i=1}^{N}$ oraz składnika losowego $\mathbf{W}\boldsymbol{\varepsilon} = (L\varepsilon_i)_{i=1}^{N} = \left(\sum_{j=1}^{N} w_{ij} \varepsilon_j\right)_{i=1}^{N}$. Umieszczenie w modelu opóźnienia przestrzennego pozwala na uwzględnienie w specyfikacji badanego zjawiska samozależności o charakterze przestrzennym.

Rozważając własności asymptotyczne, takie jak zgodność czy zbieżność rozkładów, pojęcie macierzy wag należy uzupełnić (podobnie jak w przypadku definicji przestrzennego procesu stochastycznego) o zależność od rozmiaru próby. Zatem przez macierz wag będziemy rozumieć nie pojedynczą macierz określonego rozmiaru, ale macierzową funkcję rozmiaru próby $\mathbb{N} \ni N \mapsto \mathbf{W} = \mathbf{W}_N$. Stosując się jednak do terminologii powszechnie przyjętej w literaturze ustalamy notację pomijającą indeks dolny N.

Pożądane własności statystyk badanego przestrzennego procesu stochastycznego można uzyskać tylko, wtedy gdy macierz $\mathbf{W}$, występująca w specyfikacji modelu statystycznego, spełnia pewne dodatkowe założenia. Aby uwzględnienie aspektu przestrzennego w modelu miało sens, siła interakcji między jednostkami przestrzennymi zawartych w macierzy wag nie może być zbyt mała, ani też zbyt duża. Nadmiar zależności przestrzennych między wartościami procesu może sprawić, że wzrost rozmiaru próby nie będzie skutkował dostatecznie dużym wzrostem informacji o estymowanych parametrach. Problem niedostatku zależności przestrzennych opisanych macierzą $\mathbf{W}$ wydaje się mniejszy. Dla modeli ekonometrycznych z opóźnieniem przestrzennym składnika losowego, jednostajna zbieżność elementów w_{ij}, $1 \leqslant i, j \leqslant N$ do zera może pozwolić na estymację nieznanych parametrów prostszą metodą estymacji, niż jest to możliwe w ogólnym przypadku (patrz Lee, 2002; Mynbaev, Ullah, 2008; Mynbaev, 2010 oraz podrozdział 2.1 w rozdziale II). W poniższych rozważaniach skupimy się zatem na problemie ograniczenia łącznej siły interakcji przestrzennych, zawartych w macierzy wag.

Typowym założeniem asymptotycznym używanym do ograniczenia zależności przestrzennych w macierzy wag jest wymaganie jednostajnej sumowalności wierszy i kolumn. Przypomnijmy, że dla dowolnej macierzy $\boldsymbol{A} = [a_{ij}]$ o roz-

miarach $N \times N$ zdefiniowane są następujące normy

$$\begin{aligned} \|\boldsymbol{A}\|_1 &= \max_{1 \leqslant j \leqslant N} \sum_{i=1}^{N} |w_{ij}|, \\ \|\boldsymbol{A}\|_\infty &= \max_{1 \leqslant i \leqslant N} \sum_{j=1}^{N} |w_{ij}|. \end{aligned} \tag{1.1}$$

Wówczas założenie o sumowalności wierszy i kolumn wymaga jednostajnej ze względu na rozmiar próby ograniczoności powyższych norm macierzy $\mathbf{W}$ (por. założenie II.C w rozdziale II), tj.

$$\sup_{N=1,2,\ldots} (\|\mathbf{W}_N\|_1 + \|\mathbf{W}_N\|_\infty) < \infty. \tag{1.2}$$

Jak argumentowano w pracy Olejnik i Olejnik (2020), takie założenie może okazać się zbyt restrykcyjne w przypadku wielu nawet dość naturalnych konstrukcji macierzy wag. Jest ono szczególnie problematyczne w przypadku asymptotyki wypełniania (patrz podrozdział 1.2), jednak złagodzenie warunku (1.2) może być konieczne również w przypadku asymptotyki rosnącej dziedziny. Na przykład, dla macierzy opartej na potędze odwróconej odległości (ang. *Inverse Distance Weighting*, IDW), czyli

$$w_{ij} = \frac{1}{\operatorname{dist}(i,j)^\alpha},$$

popularny przypadek interakcji newtonowskiej ($\alpha = 2$) (por. Anselin, 2002), przy równomiernie rozproszonej na płaszczyźnie dziedzinie, prowadzi do przestrzennej macierzy wag, która nie jest sumowalna w sensie (1.2). Ogólniej, dla ustalonego $1 \leqslant j \leqslant N$, przez $n(j,\delta)$ oznaczmy liczbę jednostek przestrzennych i pozostających pod wpływem jednostki j, dla których $\operatorname{dist}(i,j) \approx \delta$. Jeśli $n(j,\delta) \geqslant \text{const} \cdot \delta^{\alpha-1}$, jak jest najczęściej w α-wymiarowej przestrzeni euklidesowej, wówczas kolumny takiej macierz nie są sumowalne, gdyż

$$\lim_{\Delta \to \infty} \int_0^\Delta n(j,\delta) \cdot \delta^{-\alpha} \, \mathrm{d}\delta = \infty.$$

Jeśli natomiast $n(j,\delta) \leqslant \text{const} \cdot \delta^{2\alpha-1-\epsilon}$, dla pewnego $\epsilon > 0$, wówczas kolumny macierzy wag okażą się sumowalne z kwadratem, z uwagi na zbieżność

$$\lim_{\Delta \to \infty} \int_0^\Delta n(j,\delta) \cdot \left(\delta^{-\alpha}\right)^2 \, \mathrm{d}\delta = \infty.$$

Istnieje zatem potrzeba rozszerzania standardowej teorii asymptotycznej na przypadek macierzy nie koniecznie sumowalnych.

Czytelnicy zaznajomieni z matematyczną teorią ekonometrii czy geostatystyki zauważają z pewnością powszechną obecność teorii przestrzeni Hilberta w tych dziedzinach. Można by więc oczekiwać, że w miejscu warunku (1.2) naturalnym ograniczeniem będzie nie bezwzględna sumowalność, ale sumowalność z kwadratem. Jak wynika z rozważań w Olejnik i Olejnik (2020), związek optymalnego warunku ograniczoności przestrzennej macierzy wag — mogący zastąpić warunek (1.2) — z teorią ciągów sumowalnych z kwadratem jest nieco bardziej subtelny. Okazuje się, że zamiast rozważać własności wierszy i kolumn macierzy $\mathbf{W}$ z osobna, należy ograniczyć normę operatora opóźnienia przestrzennego wyznaczonego przez $\mathbf{W}$. Innymi słowy, przestrzenna macierz wag traktowana jest nie jako zbiór wierszy i kolumn, a raczej operator na $\mathbb{R}^N$, którego norma spektralna podlega ograniczeniu

$$\sup_{N=1,2,\ldots} \|\mathbf{W}\| < \infty. \tag{1.3}$$

W rozdziale II, gdzie dokonujemy przeglądu klasycznych metod estymacji ekonometrycznych, odwołujemy się do standardowej teorii opartej na warunku (1.2). Z kolei w rozdziałach III, IV i V rozwijamy zapoczątkowaną w pracach Gupta i Robinson (2018) oraz Olejnik i Olejnik (2020) teorię własności asymptotycznych, opartych na warunku (1.3).

Nietrudno jest zauważyć, że warunek (1.2) implikuje warunek (1.3), co wynika ze znanej nierówności $\|\boldsymbol{A}\|^2 \leqslant \|\boldsymbol{A}\|_1 \|\boldsymbol{A}\|_\infty$, dla dowolnej macierzy $\boldsymbol{A}$ — patrz równanie (1.1). Łatwo również wskazać taki ciąg macierzy $\boldsymbol{A}_n$, $n = 1, 2, \ldots$, dla którego $\sup_{n\in\mathbb{N}}\|\boldsymbol{A}\|$ jest skończone i jednocześnie wartość $\|\boldsymbol{A}\|_1 + \|\boldsymbol{A}\|_\infty$ może być dowolnie duża. Nieoczywiste jest jednak wskazanie standaryzowanej wierszowo przestrzennej macierzy wag $\mathbf{W}$, w której liczba niesumowalnych kolumn rośnie do nieskończoności. Taką konstrukcję wykonujemy poniżej.

Przykład

Zdefiniujmy zbiory $B_1 = \{2\}$, $B_2 = \{3, 4\}$, $B_3 = \{5, 6, 7\}$, a dalej $B_k = \{l + \max B_{k-1}\}_{l=1}^{k}$ dla $k = 4, 5, \ldots$ Oczywiście, zbiory B_k, dla $k \geqslant 1$, są parami rozłączne oraz $\bigcup_{k=1}^{\infty} B_k = \mathbb{N} \setminus \{1\}$. Zatem każda liczba całkowita $i \geqslant 2$ jednoznacznie wyznacza parę liczb $(k(i), l(i))$ zdefiniowaną przez zależności

$$\begin{aligned} &i \in B_{k(i)}, \\ &i = \min B_{k(i)} - 1 + l(i). \end{aligned}$$

Innymi słowy, $k(i)$ jest numerem tego zbioru B_k, do którego należy i, a $l(i)$ jest numerem porządkowym liczby i w rosnącym ciągu elementów

zbioru $B_{k(i)}$. Zauważmy, że dla dowolnego $i \geqslant 2$ mamy $i > l(i)$. Zdefiniujmy również nieskończoną macierz $\widetilde{\mathbf{W}} = [\widetilde{w}_{ij}]_{1\leqslant i,j<\infty}$ w taki sposób, że wszystkie jej elementy są równe zero, poza elementami $\widetilde{w}_{1,2} = 1$ oraz $\widetilde{w}_{i,l(i)} = \frac{1}{k(i)}$, $\widetilde{w}_{i,i+1} = 1 - \frac{1}{k(i)}$, dla wszystkich $i \geqslant 2$.

Pokażemy, że żadna kolumna $\widetilde{\mathbf{W}}$ nie jest sumowalna. W tym celu zauważmy, że jeśli $j \geqslant 1$ jest numerem kolumny oraz $k \geqslant j$, wówczas istnieje $i \in B_k$ takie, że $j = l(i)$ i $\widetilde{w}_{ij} = \frac{1}{k}$. Zatem, dla dowolnego $i = 1, 2, \ldots$ mamy

$$\sum_{i=1}^{\infty} \widetilde{w}_{ij} \geqslant \sum_{k=j}^{\infty} \frac{1}{k} = \infty.$$

Niech $\|\widetilde{\mathbf{W}}\|$ będzie indukowaną normą spektralną macierzy $\widetilde{\mathbf{W}}$ — macierzy rozumianej jako operator na przestrzeni Hilberta l_2 ciągów nieskończonych, sumowalnych z kwadratem. Wówczas $\|\widetilde{\mathbf{W}}\| \leqslant 1 + \frac{\pi}{\sqrt{6}}$. Istotnie, przy oznaczeniach

$$\widetilde{\mathbf{W}}^{\mathsf{U}} = \left[\widetilde{w}_{ij}\mathbf{I}_{\{i<j\}}\right]_{1\leqslant i,j<\infty},$$
$$\widetilde{\mathbf{W}}_{\mathsf{L}} = \left[\widetilde{w}_{ij}\mathbf{I}_{\{i>j\}}\right]_{1\leqslant i,j<\infty},$$

gdzie funkcja indykatorowa $(i,j) \mapsto \mathbf{I}_{\{i<j\}}$ dana jest formułą

$$\mathbf{I}_{\{i<j\}} = \begin{cases} 1, & i < j \\ 0, & i \geqslant j \end{cases},$$

macierz $\widetilde{\mathbf{W}}$ możemy rozłożyć w następujący sposób: $\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^{\mathsf{U}} + \widetilde{\mathbf{W}}_{\mathsf{L}}$. Wystarczy więc pokazać, że $\|\widetilde{\mathbf{W}}^{\mathsf{U}}\| \leqslant 1$ oraz $\|\widetilde{\mathbf{W}}^{\mathsf{U}}\| \leqslant \frac{\pi}{\sqrt{6}}$. Niech F_N, dla $N \geqslant 1$, będzie podprzestrzenią l_2 zdefiniowaną przez $F_N = \left\{\boldsymbol{x} = (x_i)_{i=1}^{\infty} \in l_2 : x_i = 0 \text{ dla } i > N\right\}$. Oczywiście podprzestrzeń $F = \bigcup_{N=1}^{\infty} F_N$ nieskończonych ciągów o skończonej liczbie niezerowych elementów jest gęstym podzbiorem l_2. Zauważmy zatem, że

$$\begin{aligned}
\|\widetilde{\mathbf{W}}^{\mathsf{U}}\|^2 &= \sup_{\boldsymbol{x}\in l_2} \frac{\|\widetilde{\mathbf{W}}^{\mathsf{U}}\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2} = \sup_{\boldsymbol{x}\in F} \frac{\|\widetilde{\mathbf{W}}^{\mathsf{U}}\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2} \\
&= \sup_{N\in\mathbb{N}} \sup_{\boldsymbol{x}\in F_N} \frac{\|\widetilde{\mathbf{W}}^{\mathsf{U}}(N)\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2} = \sup_{N\in\mathbb{N}} \|\widetilde{\mathbf{W}}^{\mathsf{U}}(N)\|^2 \\
&\leqslant \sup_{N\in\mathbb{N}} \|\widetilde{\mathbf{W}}^{\mathsf{U}}(N)\|_1 \|\widetilde{\mathbf{W}}^{\mathsf{U}}(N)\|_\infty = 1.
\end{aligned}$$

Następnie oznaczmy przez c_j, $j \geqslant 1$, kolumny macierzy $\widetilde{\mathbf{W}}_{\mathsf{L}}$. Łatwo zauważyć, że wektory c_j, $j \geqslant 1$, są ortogonalne, gdyż odpowiadające im zbiory indeksów elementów niezerowych $l^{-1}(\{j\})$, $j \geqslant 1$, są parami rozłączne. Co więcej, mamy $\|c_j\|^2 = \sum_{k=j}^{\infty} \frac{1}{k^2} \leqslant \frac{\pi^2}{6}$. Z nierówności Bessela, dla dowolnego $\boldsymbol{x} \in l_2$ wnioskujemy, że

$$\|(\widetilde{\mathbf{W}}_{\mathsf{U}})^{\mathsf{T}}\|^2 = \sum_{j=1}^{\infty} |c_j^{\mathsf{T}} \boldsymbol{x}|^2 \leqslant \frac{\pi^2}{6} \|\boldsymbol{x}\|,$$

czyli w efekcie $\|\widetilde{\mathbf{W}}_{\mathsf{L}}\| \leqslant \frac{\pi}{\sqrt{6}}$.

Ostatecznie zdefiniujemy macierz $\mathbf{W} = \mathbf{W}(N) = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ w sposób następujący. Połóżmy $w_{ij} = \widetilde{w}_{ij}$, dla wszystkich $1 \leqslant i, j \leqslant N$, z wyjątkiem $w_{n,n-1} = 1 - \frac{1}{k(n)}$. Zauważmy, że prawdziwe jest ograniczenie

$$\|\mathbf{W}\| \leqslant \|\widetilde{\mathbf{W}}\| + 1.$$

Co więcej, macierz $\mathbf{W}$ jest standaryzowana wierszowo, gdyż $\sum_{j=1}^{N} w_{1j} = w_{1,2} = 1$ oraz $\sum_{j=1}^{N} w_{ij} = \frac{1}{k(i)} + 1 - \frac{1}{k(i)} = 1$, dla wszystkich $2 \leqslant i \leqslant N$.

Przykład

Z powyższego przykładu łatwo wywnioskować istnienie niesumowalnej symetrycznej przestrzennej macierzy wag o ograniczonej normie spektralnej. Istotnie, warunek normalizacji wierszowej łatwo zamienić na warunek symetrii stosując przekształcenie

$$\mathbf{W}_{\text{sym}} = \frac{\mathbf{W} + \mathbf{W}^{\mathsf{T}}}{2}.$$

3. Autokorelacja przestrzenna

W tym podrozdziale zdefiniujemy i omówimy pojęcie autokorelacji przestrzennej. Przyjrzymy się także klasycznym sposobom mierzenia i testowania obecności tego zjawiska na poziomie lokalnym jak i globalnym. Zatem przyjmijmy następującą definicję.

Definicja

Autokorelacją przestrzenną danych pochodzących z przestrzennego procesu losowego nazywamy tendencję do przyjmowania zbliżonych wartości w jednostkach sąsiadujących lub bliskich przestrzennie. Przestrzenny proces stochastyczny, dla którego obserwuje się takie zjawisko, nazywamy procesem przestrzennie autoskorelowanym.

Zauważmy, że przytoczona definicja nie jest ścisła i nie wskazuje stopnia autokorelacji procesu w sposób kwantytatywny. Formalnie można by opisać liczbowo autozależność procesu przestrzennego $\boldsymbol{x} = (x_i)_{i \in P}$ używając macierzy korelacji

$$\operatorname{Corr}[x_i, x_j], \qquad i, j \in P.$$

W praktyce jednak nie stosuje się takiego podejścia. Zamiast macierzy współczynników przyjmuje się różne jednoparametrowe liczbowe mierniki autokorelacji w zależności uszczegółowionej definicji autokorelacji. Wszystkie są jednak oparte na pewnego rodzaju zależności wartości zmiennej x_i, $1 \leqslant i \leqslant N$, od wartości opóźnienia przestrzennego $L[x_i] = \sum_{j=1}^{N} w_{ij} x_j$ (patrz definicja na s. 24).

Należy wyróżnić dwa rodzaje autokorelacji przestrzennej: dodatnią i ujemną. W przypadku autokorelacji dodatniej, wartości obserwowanej zmiennej z sąsiednich jednostek są do siebie podobne. Mamy wówczas do czynienia z przestrzennym grupowaniem się (w sensie lokalizacji) wysokich bądź niskich wartości obserwowanej zmiennej. Z kolei w przypadku ujemnej autokorelacji przestrzennej będziemy obserwować wysokie wartości zmiennej otoczone niskimi (i odwrotnie), układając się w ten sposób we wzór przypominający szachownicę. Badania empiryczne wskazują, że większość zjawisk ekonomicznych obserwowanych w przestrzeni charakteryzuje się dodatnimi oddziaływaniami, co jest zgodne z prawem Toblera. Autokorelacja ujemna jest obserwowana w praktyce dość rzadko.

Autokorelację przestrzenną możemy badać na poziomie lokalnym lub globalnym. Istnienie globalnej autokorelacji przestrzennej oznacza występowanie zależności przestrzennych w obrębie całego badanego obszaru, średnio dla wszystkich lokalizacji. Globalna autokorelacja przestrzenna uwidacznia się więc poprzez ogólną tendencję do grupowania się podobnych wartości w przestrzeni. Natomiast lokalną autokorelację przestrzenną definiuje się jako istnienie zależności przestrzennych danego obiektu z jego otoczeniem. Zatem jej identyfikacja umożliwia wskazanie położenia lokalnych klastrów wartości podobnych. Może też prowadzić do wychwycenia tzw. lokalizacji nietypowych (ang. *outliers*). W kolejnych podrozdziałach przedstawimy klasyczne metody testowania występowania zjawisk, zarówno globalnej, jak i lokalnej autokorelacji przestrzennej procesu losowego.

3.1. Testowanie globalnej autokorelacji przestrzennej

Istnieje kilka rodzajów wskaźników testujących grupowanie się danych. Wszystkie przedstawione poniżej statystyki mierzą stopień współzależności przestrzennych. Do najpowszechniejszych należą statystyki: I Morana, Geary'ego oraz statystyka $G(d)$. Jak dotąd, najpopularniejszą klasą testów służących wykrywaniu

autokorelacji przestrzennej są te oparte na pracy Morana (1950). Test I Morana może być także użyty do weryfikacji trafności doboru macierzy wag $\mathbf{W}$, reprezentującej przestrzenną strukturę zależności procesu losowego. Statystyka I Morana służy więc do oceny stopnia skorelowania przestrzennego pomiędzy sąsiadującymi lokalizacjami. Zmodyfikowana przez Cliffa i Orda (1973) pod kątem potrzeb ekonometrii przestrzennej procedura testowania jest przestrzennym analogiem testu Durbina-Watsona (Durbin, Watson, 1950; 1951). Statystyka I Morana służy do testowania obecności globalnej autokorelacji przestrzennej według schematu opisanego macierzą wag $\mathbf{W}$.

Definicja

Rozważmy proces przestrzenny $\mathbf{x} = (x_1, \ldots, x_N)^\mathsf{T}$. Wówczas wartość globalnej **statystyki** I **Morana** dla standaryzowanej wierszowo macierzy wag $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ wyraża się wzorem

$$I = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{\mathbf{x}})(x_j - \bar{\mathbf{x}})}{\sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})},$$

gdzie $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i$ oznacza średnią z realizacji badanego procesu. Ogólniej, jeśli przestrzenna macierz wag $\mathbf{W}$ nie jest wierszowo standaryzowana, a co za tym idzie $\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \neq N$, wówczas statystyka I Morana przyjmuje postać normalizowaną

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \cdot \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{\mathbf{x}})(x_j - \bar{\mathbf{x}})}{\sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})^2}. \tag{1.4}$$

Jeżeli macierz $\mathbf{W}$ opisuje stan rzeczywisty, tj. duże wagi odpowiadają rzeczywistym korelacjom, to wartość statystyki I Morana będzie miała tendencję do przyjmowania wartości dużych, co do wartości bezwzględnej. Można więc powiedzieć, że statystyka I Morana jest w pewnym sensie ważonym przestrzennie współczynnikiem (auto)korelacji, służącym do wykrywania odchyleń w losowym rozkładzie przestrzennym procesu $\mathbf{x}$.

Aby wykorzystać statystykę I Morana do ustalenia, czy sąsiadujące ze sobą wartości są bardziej do siebie podobne niż to wynika z losowości badanego zjawiska, rozpatrzmy następujące hipotezy testowe:

H_0: brak autokorelacji przestrzennej, przy hipotezie alternatywnej,
H_1: występowanie zależności przestrzennych.

Zwyczajowo, w przypadku, gdy statystyka I Morana przyjmuje wartości bliskie $I_0 = -1/(N-1)$, uważa się, że wartość I nie daje podstaw do odrzucenia

hipotezy zerowej. W przeciwnym przypadku zaś, odrzucając H_0 wnioskujemy o istnieniu pewnych istotnych statystycznie zależności przestrzennych. Zakładając brak heterogeniczności przestrzennej danych, przyjmuje się, że gdy $I > I_0$, obserwuje się autokorelację przestrzenną dodatnią, zaś dla $I < I_0$ autokorelację ujemną. Zauważmy, że dla dostatecznie dużych N, w przypadku braku autokorelacji przestrzennej, statystyka przyjmować będzie wartości bliskie zeru. Chociaż w praktyce wartość statystyki Morana często nie przekracza, co do modułu, wartości jeden, należy zaznaczyć, że w odróżnieniu od klasycznego współczynnika korelacji Pearsona nie jest to regułą. W rzeczywistości, jak sugeruje Kossowski (2010), za de Jong i inni (1984), mamy nierówności

$$\frac{N \cdot \lambda_{\min}}{2\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \leqslant I \leqslant \frac{N \cdot \lambda_{\max}}{2\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}},$$

gdzie $\lambda_{\min}$ i $\lambda_{\max}$ są odpowiednio najmniejszą i największą wartością własną iloczynu $\mathbf{M}(\mathbf{W}+\mathbf{W}^{\mathsf{T}})\mathbf{M}$, a macierz $\mathbf{M}$ jest operatorem rzutu na przestrzeń ortogonalną do podprzestrzeni wektorów stałych. Co więcej, wskazywana tradycyjnie wartość $I_0 = -1(N-1)$ nie zawsze jest poprawna, co wyjaśniamy w rozdziale III.

Powyżej przedstawiono popularną w literaturze postać procedury testowej Morana. Można jednak zwrócić uwagę na fakt, że w zasadzie poprawna hipoteza zerową powinna brzmieć następująco:

H_0: brak zależności przestrzennych w procesie $\mathbf{x}$.

Odrzucenie hipotezy zerowej nie informuje nas bowiem, czy przyczyną zależności przestrzennych jest autokorelacja przestrzenna procesu, czy heterogeniczność przestrzenna (por. Anselin, 1988a).

Poziom istotności testu Morana może być obliczony za pomocą standaryzowanej statystyki I Morana. Przy pewnych założeniach jej rozkład można przybliżyć standardowym rozkładem normalnym (patrz rozdział III), a dokładniej

$$\frac{I - \mathbb{E}(I)}{\sqrt{\mathrm{Var}(I)}} \simeq \mathcal{N}(0,1).$$

W praktyce często stosuje się również tzw. test randomizacyjny oparty na wartości statystyki I Moran obliczanych dla generowanych losowo permutacjach próby przestrzennej. Mianowicie, w ramach tej procedury próbkuje się przestrzeń wszystkich permutacji π zbioru $\{1,\ldots,N\}$. Dla każdego wylosowanego w ten sposób π oblicza się wartość statystyki Morana $I(\pi)$ dla procesu $\boldsymbol{x}^{\pi} = \mathbf{x}\circ\pi$, gdzie wartości procesu $\boldsymbol{x}^{\pi} = (x_1^{\pi},\ldots,x_N^{\pi})$ określone są przez równość $x_i^{\pi} = x_{\pi^{-1}(i)}$, dla wszystkich jednostek przestrzennych $1 \leqslant i \leqslant N$. Wówczas tzw. pseudowartość p (ang. *pseudo p-value*) określana jest jako iloraz liczby permutacji π, dla

których $I(\pi) \leqslant I(\mathrm{id}_\pi)$, gdzie id_π jest permutacją identycznościową, przez liczbę wszystkich permutacji w próbce.

Podobne zastosowanie ma przestawiona poniższej statystyka Geary'ego.

Definicja

Rozważmy proces przestrzenny $\mathbf{x} = (x_1, \ldots, x_N)^\mathsf{T}$ oraz macierz wag przestrzennych $\mathbf{W} = [w_{ij}]_{ij \leqslant N}$. Statystykę określoną wzorem

$$c = \frac{N-1}{2\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(x_i - x_j)^2}{\sum_{i=1}^N (x_i - \bar{\mathbf{x}})^2},$$

gdzie $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^N x_i$, nazywamy **globalną statystyką Geary'ego**.

W przypadku, gdy wartość tej statystyki spełnia $c < 1$, mamy do czynienia z autokorelacją przestrzenną dodatnią, natomiast dla $c > 1$ wnioskujemy o występowaniu autokorelacji ujemnej. Wartość statystyki $c \approx 1$ świadczy o braku autokorelacji. Podobnie jak w przypadku statystyki I Morana, zakres możliwych wartości statystyki Geary'ego jest pewną funkcją macierzy $\mathbf{W}$ (patrz de Jong i inni, 1984), chociaż w praktyce mieszczą się w przedziałach: $[0, 1)$ — dla autokorelacji dodatniej oraz $(1, 2]$ — dla autokorelacji ujemnej.

Kolejną popularną statystyką przestrzenną jest opracowana przez Getisa i Orda (1992) statystyka $G(d)$, mierząca siłę przestrzennego skorelowania pomiędzy poszczególnymi lokalizacjami, będącymi w obrębie ustalonego otoczenia d. Statystyka ta jest funkcją promienia d, a więc pozwala na ustalenie średniej siły zależności od przyjętego promienia oddziaływań.

Definicja

Niech $\mathbf{x} = (x_1, \ldots, x_N)^\mathsf{T}$ będzie procesem przestrzennym oraz niech $w_{ij}(d)$ oznacza elementy niestandaryzowanej przestrzennej macierzy wag dla ustalonego otoczenia d, tj. $w_{ij}(d) = 1$, gdy jednostki przestrzenne $1 \leqslant i \neq j \leqslant N$ są odległe od siebie o nie więcej niż d, oraz $w_{ij}(d) = 0$ w przeciwnym wypadku. Statystykę postaci

$$G(d) = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(d) x_i x_j}{\sum_{i=1}^N \sum_{j=1}^N x_i x_j},$$

nazywamy **globalną statystyką** $G(d)$.

Podobnie jak w przypadku statystki I Morana, wnioskowanie statystyczne opiera się na założeniu, że standaryzowana statystyka $G(d)$ ma w przybliżeniu standardowy rozkład normalny

$$\frac{G(d) - \mathbb{E}\left(G(d)\right)}{\sqrt{\mathrm{Var}\left(G(d)\right)}} \simeq \mathcal{N}(0, 1).$$

Wartości dwóch pierwszych momentów statystyki $G(d)$ mogą zostać wyznaczone przy dodatkowych założeniach dotyczących własności stochastycznych procesu $\mathbf{x}$. Na przykład, przy pewnych założeniach można przyjąć, że wartość pierwszego momentu statystyki $G(d)$ wyraża się wzorem

$$\mathbb{E}\left(G(d)\right) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(d)}{N(N-1)}.$$

Dodatnie wartości statystyki $G(d) - \mathbb{E}\,G(d)$ wskazują na przestrzenne grupowanie się wysokich (ang. *hot spots*), ujemne zaś niskich wartości badanej zmiennej (ang. *cold spots*); patrz Suchecki [red.] (2010).

3.2. Testowanie lokalnej autokorelacji przestrzennej

W praktyce badań ekonometrycznych nierzadko okazuje się, że zależności przestrzenne mogą nie mieć charakteru globalnego. Mogą być one obserwowane lokalnie — w pewnych rejonach dziedziny przestrzennej, natomiast w innych mogą występować w mniejszym natężeniu lub nie występować w ogóle. W takim wypadku w analizach empirycznych stosuje się testy i statystyki LISA (ang. *Local Indicator of Spatial Autocorrelation*). Za ich pomocą bada się lokalną autokorelację przestrzenną, czyli korelację wartości zmiennej w wybranej lokalizacji z jej sąsiadami (por. Anselin, 1988a: 284). W tym wypadku najczęściej wykorzystuje się lokalną statystykę I_i Morana, dla $i \in \{1, \ldots, N\}$, wyrażającą się wzorem

$$I_i = \frac{N}{\sum_{j=1}^{N}\sum_{k=1}^{N} w_{jk}} \cdot \frac{\sum_{j=1}^{N} w_{ij}(x_i - \bar{\mathbf{x}})(x_j - \bar{\mathbf{x}})}{\sum_{j=1}^{N}(x_j - \bar{\mathbf{x}})^2}.$$

Obliczenie statystyki I_i dla wszystkich obserwacji umożliwia wykrycie lokalnych zgrupowań porównywalnych wartości procesu przestrzennego.

Podobną statystyką jest lokalna statystyka Geary'ego o następującej postaci:

$$c_i = \frac{N-1}{2\sum_{j=1}^{N}\sum_{k=1}^{N} w_{jk}} \cdot \frac{\sum_{j=1}^{N} w_{ij}(x_i - x_j)^2}{\sum_{j=1}^{N}(x_j - \bar{\mathbf{x}})^2}.$$

Wartość statystyki przekraczająca 1 wskazuje na obecność ujemnej lokalnej autokorelacji przestrzennej, w przeciwnym wypadku mamy do czynienia z dodatnią lokalną autokorelacją.

W praktyce, zarówno dla celów analizy globalnej, jak i lokalnej, najpowszechniej jednak stosowane są statystyki Morana. Statystyka globalna oferuje jedną charakterystykę dla całej próby, bedącą średnią lokalnych statystyk I_i Morana. Należy zatem zauważyć, że w przypadku, gdy autokorelacja przestrzenna

nie występuje globalnie, statystyka I Morana nie jest z reguły w stanie wykryć obecności pojedynczych klastrów. Z kolei lokalne statystyki I_i Morana liczone są dla każdej lokalizacji osobno, co daje możliwość wykrycia lokalnych zgrupowań.

Należy również zwrócić uwagę na pewien sposób wizualnego określenia typu występującej autokorelacji przestrzennej, czyli tzw. moranowski wykres rozproszenia (ang. *Moran scatterplot*). Jak zaproponował Anselin (1996), aby w łatwy sposób wychwycić obecność zależności przestrzennych, w planarnym układzie współrzędnych można zaznaczyć punkty o współrzędnych

$$\left((x_i - \bar{\mathbf{x}}), \sum_{j=1}^{N} w_{ij}(x_j - \bar{\mathbf{x}})\right), \qquad i = 1, \ldots, N.$$

Na tak skonstruowanym wykresie, możliwy jest jeden z trzech przypadków. Jeśli obserwujemy przewagę punktów układających się w ćwiartkach I i III, wskazuje to obecność autokorelacji dodatniej. Gdy, z kolei, większość punktów wpada do ćwiartek II i IV, mamy so czynienia z autokorelacją ujemną. Brak widocznej przewagi w żadnej z par ćwiartek układu współrzędnych świadczy o braku autokorelacji przestrzennej.

Rozdział II

Modele ekonometryczne z zależnościami przestrzennymi

Wstęp

W wielu badaniach przestrzenny charakter struktury danych oraz ich przestrzenne zależności stanowią samodzielny przedmiot zainteresowań. Jednak w innych analizach autokorelacja przestrzenna stanowi problem poboczny, analogiczny do autokorelacji składnika losowego, występującej w szeregach czasowych. Zarówno w pierwszym, jak i w drugim przypadku nieuwzględnienie autokorelacji przestrzennej może prowadzić do oszacowań pozbawionych pożądanych własności, takich jak efektywność, nieobciążoność, a nawet zgodność. Efektem takiego podejścia może być wnioskowanie statystyczne oparte na niepełnej informacji. Dodatkowo, dwukierunkowa natura interakcji przestrzennych uniemożliwia przeniesienie rozwiązań problemu autokorelacji w teorii szeregów czasowych na nowy, przestrzenny grunt.

W modelach ekonometrii przestrzennej zjawisko autokorelacji przestrzennej uwzględnia się poprzez włączenie do postaci strukturalnej modelu składnika, najczęściej w formie liniowej, odpowiadającego przestrzennemu opóźnieniu wybranych zmiennych (patrz podrozdział 2.2 w rozdziale I). O autoregresji przestrzennej mówimy, gdy opóźnienie przestrzenne dotyczy zmiennej zależnej lub składnika losowego. Natomiast, jeśli rozważane jest opóźnienie przestrzenne zmiennych egzogenicznych, mamy do czynienia z modelem zawierającym przestrzenną regresją krzyżową.

W tym rozdziale dokonamy przeglądu najbardziej popularnych postaci strukturalnych modeli przestrzennych, a następnie przedstawimy wybrane metody estymacji ich parametrów. Zauważmy, że aby zagwarantować pożądane własności wymienionych estymatorów, koniecznym jest przyjęcie szeregu założeń dotyczących: zachowania asymptotycznego macierzy wag przestrzennych, własności zmiennych egzogenicznych oraz rozkładu prawdopodobieństwa zaburzeń losowych. W tej części pracy nie będziemy przedstawiać kompletnej matematycznie teorii dotyczącej prezentowanych metod. Podkreślimy jednak te z założeń, które odróżniają teorie standardowe od autorskiego podejścia, prezentowanego w rozdziałach kolejnych. W szczególności złagodzone zostaną wymagania wyrażone poniżej w założeniach II.B, II.C i II.D.

Założenie II.A

Macierz $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ jest standaryzowana wierszowo i ma zerową przekątną, tzn. dla każdego $1 \leqslant i \leqslant N$ mamy $\sum_{j=1}^{N} w_{ij} = 1$ oraz $w_{ii} = 0$.

Założenie II.B

Wiersze i kolumny macierzy $\mathbf{W}$ są jednostajnie bezwzględnie sumowalne, czyli istnieje (niezależna od N) stała $C > 0$, dla której

$$\max_{1\leqslant i,j\leqslant N}\left\{\sum_{k=1}^{N}|a_{ik}|+\sum_{k=1}^{N}|a_{kj}|\right\}<C,$$

jednostajnie dla wszystkich możliwych rozmiarów próby $N\in\mathbb{N}$.

Założenie II.C

Wartości zmiennych egzogenicznych $\mathbf{X}$ są jednostajnie ograniczone przy $N=1,2,\ldots$ Co więcej, istnieje asymptotyczna macierz kowariancji z próby dla zmiennych egzogenicznych $\frac{1}{N}\mathbf{X}^{\mathsf{T}}\mathbf{X}$, przy $N\to\infty$, i jest ona nieosobliwa.

Założenie II.D

Składnik zaburzeń losowych modelu przestrzennego ma N-wymiarowy rozkład normalny $\mathcal{N}(0,\sigma^2\mathbf{I})$.

Wprawdzie założenie II.A nie jest konieczne dla twierdzeń o asymptotycznych własnościach estymatorów, niemniej jednak ze względów aplikacyjno-interpretacyjnych (interpretacja elementów macierzy jako wag) oraz technicznych (patrz twierdzenie II.2) macierze wag stosowane w modelach ekonometrycznych często podlegają operacji standaryzacji wierszowej (ang. *row standardization*). Operacja standaryzacji wierszowej polega na przemnożeniu każdego elementu macierzy $\mathbf{W}=[w_{ij}]_{ij\leqslant N}$ w ustalonym wierszu przez odwrotność sumy elementów tego wiersza. Innymi słowy mamy

$$\mathbf{W}_{\text{std}}\coloneqq\left[\frac{w_{ij}}{\sum_{k=1}^{N}w_{ik}}\right]_{ij\leqslant N}.$$

Powszechnie uważa się, że standaryzacja macierzy wag umożliwia porównywanie oszacowań parametrów między modelami. Co prawda istnieją argumenty uzasadniające takie stwierdzenie, jednak nie jest ono pozbawione pewnych wad. Można zauważyć, że przemnożenie wierszy macierzy przez różne liczby automatycznie usuwa z niej informację o względny siłach interakcji w tych wierszach. W szczególności wielkości uzyskane w każdej kolumnie macierzy standaryzowanej wierszowo nie są już porównywalne. Idąc dalej, w przypadku macierzy odwróconej odległości, standaryzacja wierszowa może wręcz powodować problemy interpretacyjne. Jak twierdzi Anselin (1988a: 23–24, tłumaczenie własne): „normalizacja wierszowa macierzy wag, w której wagi są proporcjonalne do [potęgi — przyp. aut.] odwrotności odległości powoduje, że interpretacja ekonomiczna oparta na zaniku oddziaływań wraz z odległością nie jest już poprawna"; por. też Kelejian i Prucha (2010). Spotykaną w literaturze, choć mało spopularyzowaną alternatywą dla standaryzacji jest odpowiednie przeskalowywanie całej macierzy wag (patrz Elhorst, 2001; Vega, Elhorst, 2015; Olejnik, Olejnik, 2020). W takim

wypadku, co ważne, zachowywana jest względna wielkość wag w całej macierzy, przy jednoczesnym wprowadzeniu ograniczenia na całkowitą moc interakcji przestrzennych. W rozdziałach III, IV i V świadomie pomijamy założenie standaryzacji wierszowej, aby zachować ogólność rozważań.

1. Modele autoregresji przestrzennej

1.1. Przegląd specyfikacji

Niech $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ będzie ustaloną macierzą wag przestrzennych, o wymiarach $N \times N$. Przyjmijmy też, że kolumny macierzy $\mathbf{X}$ o wymiarach $N \times k$ są wektorami obserwacji zmiennych egzogenicznych (z uwzględnieniem wyrazu wolnego), wnoszących do modelu dodatkową zewnętrzną informację o badanym procesie. W najprostszym przypadku, przy braku zmiennych objaśniających, specyfikacja modelu może uwzględniać jedynie autozależności przestrzenne procesu generującego obserwacje (ang. *data generating process*). Taki model nazywamy modelem czystej autoregresji przestrzennej.

Definicja

Modelem czystej autoregresji przestrzennej (**pierwszego rzędu**) nazywamy model o specyfikacji opisanej równaniem

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}, \tag{2.1}$$

w którym $\mathbf{y} = [y_1, \ldots, y_N]^{\mathsf{T}}$ jest obserwowanym procesem przestrzennym, a $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_N]$ to składnik losowy o rozkładzie normalnym z zerową wartością oczekiwaną i stałą wariancją $\sigma^2\mathbf{I}$, gdzie σ^2 jest nieznanym parametrem. Parametr ρ nazywamy współczynnikiem autoregresji przestrzennej.

Zauważmy, że $\mathbf{W}\mathbf{y} = \left(\sum_{i=1}^{N} w_{ij} y_j\right)_{j=1}^{N}$ jest wektorem średnich wartości procesu w jednostkach sąsiednich, a więc równanie (2.1) możemy zapisać alternatywnie jako

$$y_i = \rho \sum_{i=1}^{N} w_{ij} y_j + \varepsilon_i.$$

Macierz $\mathbf{W}$ odzwierciedla zakładane zależności między lokalizacjami przestrzennymi (w_{ij} dla lokalizacji i-tej i j-tej), a tym samym pomiędzy poszczególnymi elementami procesu $\mathbf{y}$. Przestrzenny charakter procesu w modelu jest zatem uwzględniany poprzez składnik $\rho\mathbf{W}\mathbf{y}$.

W praktyce dostępne są pewne zmienne egzogeniczne, które objaśniają badany proces przestrzenny (w najprostszym przypadku składnik stały). Model,

którego specyfikacja uwzględnia dodatkowe zmienne $\mathbf{X}$ nazywany jest modelem autoregresji przestrzennej — SAR (ang. *Spatial Autoregressive Model*), jak również modelem opóźnień przestrzennych (ang. *Spatial Lag Model*).

Definicja

Przy przyjętych powyżej oznaczeniach, model oceniający zmianę poziomu procesu przestrzennego $\mathbf{y}$ w oparciu o wartość procesu w sąsiednich lokalizacjach oraz determinanty procesu $\mathbf{X}$, postaci

$$\begin{aligned} \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big), \end{aligned}$$

gdzie ρ jest współczynnikiem autoregresji przestrzennej, a $\boldsymbol{\beta}$ wektorem parametrów nachylenia (ang. *slope*), nazywamy **modelem autoregresji przestrzennej** (**zmiennej objaśnianej**) **— SAR**.

W przypadku, gdy zależność przestrzenna pojawia się wewnątrz procesu zakłóceń losowych, czyli błędy modelu dla poszczególnych lokalizacji są skorelowane z błędami w lokalizacjach sąsiednich, mówimy o modelu z przestrzennie autoskorelowanym składnikiem losowym — SEM (ang. *Spatial Error Model*).

Definicja

Przy przyjętych powyżej oznaczeniach, **modelem z przestrzennie autoskorelowanym składnikiem losowym** (**SEM**) nazywamy model postaci

$$\begin{aligned} y &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big), \end{aligned} \tag{2.2}$$

gdzie $\mathbf{u} = (u_1, \ldots, u_N)$ to N-wymiarowy wektor skorelowanych przestrzennie składników losowych, zaś $\boldsymbol{\varepsilon}$ to proces generujący błędy modelu o rozkładzie normalnym z zerową wartością oczekiwaną i wariancją równą $\sigma^2\mathbf{I}$. Parametr λ jest współczynnikiem przestrzennej korelacji składnika losowego.

Zakładając, że macierz $\mathbf{I}-\lambda\mathbf{W}$ jest nieosobliwa, model SEM dany w równaniu (2.2) można poddać przekształceniom

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon} \\ (\mathbf{I} - \lambda\mathbf{W})\,\mathbf{y} &= (\mathbf{I} - \lambda\mathbf{W})\,\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \mathbf{y} &= \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \end{aligned}$$

prowadzącym do specyfikacji równoważnej, zdefiniowanej poniżej.

Definicja

Przy przyjętych wcześniej oznaczeniach, model postaci

$$\begin{aligned}\mathbf{y} &= \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big),\end{aligned}$$

nazywamy **przestrzennym modelem Durbina** (SDM, ang. *Spatial Durbin Model*).

W literaturze można spotkać również ogólny model przestrzenny (SGM, ang. *Spatial General Model*), który łączy w sobie dwa powyższe modele. W tym wypadku mamy do czynienia zarówno z autoregresją przestrzenną, jak i z przestrzenną autokorelacją składnika losowego. Model ten w literaturze nazywany jest również modelem typu Cliffa-Orda (por. Cliff, Ord, 1973).

Definicja

Przy przyjętych powyżej oznaczeniach, **ogólnym modelem przestrzennym** nazywamy model postaci

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\\ \mathbf{u} &= \lambda\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big),\end{aligned}$$

gdzie $\mathbf{W}$ i $\mathbf{M}$ są potencjalnie różnymi macierzami wag przestrzennych.

W rozważaniach teoretycznych pojawia się również rozszerzenie przedstawionych specyfikacji na przypadek tzw. przestrzennych modeli wyższych rzędów (ang. *higher-order spatial models*). Co więcej, przez analogię do szeregów czasowych ARMA, wprowadza się również pojęcie przestrzennej średniej ruchomej składnika losowego, uzyskując specyfikację SARARMA(p, q, r), dla $p, q, r \geqslant 1$ (ang. *Spatial Autoregressive Autocorrelated Moving Average Model*).

Definicja

Niech $\mathbf{W}_1, \ldots, \mathbf{W}_p$ oraz $\mathbf{M}_1, \ldots, \mathbf{M}_q$ będą macierzami wag przestrzennych. Przy przyjętych wcześniej oznaczeniach, modelem **SARAR**(p, q) nazywamy model o specyfikacji

$$\begin{aligned}\mathbf{y} &= \rho_1\mathbf{W}_1\mathbf{y} + \rho_2\mathbf{W}_2\mathbf{y} + \ldots + \rho_p\mathbf{W}_p\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\\ \mathbf{u} &= \lambda_1\mathbf{M}_1\mathbf{u} + \lambda_2\mathbf{M}_2\mathbf{u} + \ldots + \lambda_q\mathbf{M}_q\mathbf{u} + \boldsymbol{\varepsilon},\end{aligned}$$

gdzie $\boldsymbol{\varepsilon} \sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big)$, a $\rho_1, \ldots, \rho_p$ oraz $\lambda_1, \ldots, \lambda_q$ są przestrzennymi parametrami autoregresyjnymi.

Definicja

Niech $\mathbf{W}_1, \ldots, \mathbf{W}_p$, $\mathbf{M}_1, \ldots, \mathbf{M}_q$ oraz $\mathbf{V}_1, \ldots, \mathbf{V}_r$ będą macierzami wag przestrzennych. Przy przyjętych wcześniej oznaczeniach, modelem typu **SARARAR** rzędu (p, q, r) nazywamy model o specyfikacji

$$\begin{aligned}
\mathbf{y} &= \rho_1 \mathbf{W}_1 \mathbf{y} + \rho_2 \mathbf{W}_2 \mathbf{y} + \ldots + \rho_p \mathbf{W}_p \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\
\mathbf{u} &= \lambda_1 \mathbf{M}_1 \mathbf{u} + \lambda_2 \mathbf{M}_2 \mathbf{u} + \ldots + \lambda_q \mathbf{M}_q \mathbf{u} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &= \psi + \phi_1 \mathbf{V}_1 \psi + \phi_2 \mathbf{V}_2 \psi + \ldots + \phi_r \mathbf{V}_r \psi \\
\psi &\sim \mathcal{N}\big(0, \sigma^2 \mathbf{I}\big),
\end{aligned}$$

gdzie $\phi_1, \ldots, \phi_r$ są parametrami średniej ruchomej.

Zaznaczmy, że w modelach wyższych rzędów SARAR i SARARMA wprowadzenie zbyt wielu macierzy wag przestrzennych może okazać się problematyczne. Trudność sprawia np. właściwe określenie tzw. przestrzeni parametrów (ang. *parameter space*) dla przestrzennych współczynników autoregresyjnych (pewne próby rozwiązania tego problemu były proponowane w publikacji Elhorst i inni (2012). Wówczas, aby zagwarantować identyfikowalność wszystkich parametrów modelu, konieczne może okazać się wprowadzanie dodatkowych założeń. Zatem, mimo że specyfikacje modeli wyższych rzędów dostarczają niewątpliwie ciekawych możliwości aplikacyjnych (por. Olejnik i inni, 2020), w praktyce wartości parametrów rzędu p, q oraz r najczęściej nie wykraczają poza 0, 1 bądź 2. W przypadku modeli o większej liczbie macierzy, problem identyfikowalności parametrów można złagodzić ostrożnie dobierając macierze wag. Na przykład, Gupta i Robinson rozważali model o nieskończonej (a dokładnie rosnącej wraz z rozmiarem próby) liczbie macierzy, stosując dla nich warunek pewnego rodzaju „ortogonalności" (por. Gupta, Robinson, 2015).

Analogicznie do modeli szeregów czasowych z rozkładami opóźnień (ang. *distributed lag models*), przestrzenne opóźnienie może dotyczyć również zmiennych egzogenicznych modelu. Mówimy wówczas o modelach regresji krzyżowej SCM (ang. *Spatial Cross-regressive Models*). Jak się okazuje, uwzględnienie przestrzennie opóźnionych wartości deterministycznych zmiennych objaśnianych w postaci strukturalnej modelu, nie powoduje dodatkowych trudności estymacyjnych. Rozszerzenie modelu SAR o regresję krzyżową prowadzi z kolei do specyfikacji SADL (ang. *Spatial Autoregressively Distributed Lag Model*).

Definicja

Przy przyjętych wcześniej oznaczeniach, **model autoregresji zmiennej objaśnianej z regresją krzyżową** (SADL) definiujemy poprzez specyfikację

$$\begin{aligned}
\mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2 \mathbf{I}\big),
\end{aligned} \tag{2.3}$$

gdzie $\boldsymbol{\gamma}$ jest współczynnikiem opóźnienia przestrzennego zmiennych objaśniających.

Zauważmy, że nieliniowy warunek ograniczający

$$\boldsymbol{\gamma} + \rho\boldsymbol{\beta} = 0, \tag{2.4}$$

powoduje, że model SADL sprowadza się do opisanej wcześniej specyfikacji SDM, równoważnej SEM. Należy tutaj zaznaczyć, iż w literaturze przedmiotu termin przestrzenny model Durbina bywa również używany do określenia specyfikacji (2.3) bez warunku (2.4). Jako przykład może tutaj posłużyć monografia LeSage i Pace (2009).

1.2. Interpretacja parametrów modeli autoregresji przestrzennej

W przypadku procesów, w których występuje autokorelacja przestrzenna, niewątpliwą zaletą uwzględnienia jej w modelu jest poprawienie własności statystycznych oszacowań parametrów. Dodatkowo, modele przestrzenne zawierające składnik odpowiadający za autoregresję zmiennej objaśnianej dostarczają badaczom dodatkowych możliwości interpretacyjnych. W tym podrozdziale przedstawiamy teorię pozwalającą na poprawną interpretację oszacowań parametrów nachylenia. Rozumowanie oparte jest na pomyśle zaprezentowanym w pracy LeSage i Pace (2009), polegającym na przekształceniu równania specyfikacji modelu SARAR$(1,0)$ do postaci jawnej.

W naszym uogólnieniu, przyjmijmy, że obserwujemy pewien proces przestrzenny $\mathbf{y} = (y_1, \ldots, y_N)$ zgodny ze specyfikacją SARAR(p,q), uwzględniającą dodatkowo efekt regresji krzyżowej. Bez straty ogólności możemy jednak założyć, że $q = 1$, a zatem

$$\begin{aligned} \mathbf{y} &= \sum_{n=1}^{p} \rho_n \mathbf{W}_n \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \mathbf{J}\boldsymbol{\delta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}, \end{aligned} \tag{2.5}$$

gdzie $\boldsymbol{\varepsilon} \sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big)$ jest wektorem zaburzeń losowych modelu, $\mathbf{X}$ i $\mathbf{J}$ są macierzami zmiennych objaśniających, $\rho_1, \ldots, \rho_p$ oraz λ parametrami autoregresyjnymi, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$ i $\boldsymbol{\delta}$ wektorami parametrów nachylenia. Przyjmijmy również, że kolumny macierzy $\mathbf{J}$ zawierają zmienne niepodlegające interpretacji w kategoriach wpływu, np. wyraz wolny czy zmienne indykatorowe dla efektów stałych. Oczywiście, jeżeli zmienna $\mathbf{X}_l$, $1 \leqslant l \leqslant k$, nie występuje w modelu *explicite* lub w formie opóźnionej przestrzennie, można przyjąć odpowiednio $\gamma_l = 0$ lub $\beta_l = 0$. Zakładając odwracalność macierzy $\mathbf{I} - \sum_{n=1}^{p} \rho_n \mathbf{W}_n$

oraz $\mathbf{I}-\lambda\mathbf{M}$, powyższe równanie możemy rozwiązać ze względu na $\mathbf{y}$, uzyskując równość

$$\left(\mathbf{I}-\sum_{n=1}^{p}\rho_n\mathbf{W}_n\right)\cdot\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{W}\mathbf{X}\boldsymbol{\gamma}+\mathbf{J}\boldsymbol{\delta}+\mathbf{u}.$$

Mamy zatem

$$\begin{aligned}\mathbb{E}\,\mathbf{y}=\sum_{l=1}^{k}\left(\mathbf{I}-\sum_{n=1}^{p}\rho_n\mathbf{W}_n\right)^{-1}\big(\beta_l\mathbf{I}+\gamma_l\mathbf{W}\big)\mathbf{X}_l\\ +\left(\mathbf{I}-\sum_{n=1}^{p}\rho_n\mathbf{W}_n\right)^{-1}\mathbf{J}\delta.\end{aligned}\tag{2.6}$$

Zauważmy, że w przypadku specyfikacji (2.5) prosta interpretacja wartości elementów parametru $\boldsymbol{\beta}$ jako krańcowych efektów poszczególnych zmiennych na wartości zmiennej $\mathbf{y}$ może być niewłaściwa. Istotnie, jakiekolwiek zaburzenie zmiennej objaśniającej w danej lokalizacji wywiera wpływ na wartość zmiennej zależnej w tej samej lokalizacji nie tylko bezpośredni, lecz także pośredni, poprzez lokalizacje sąsiednie. Wnioskowanie na podstawie uzyskanych oszacowań należy zatem oprzeć na równości (2.6).

Ustalmy dowolną liczbę $1\leqslant l\leqslant k$. Aby poprawnie ocenić marginalny wpływ zmiennej objaśniającej $\mathbf{X}_l$ na wartości procesu $\mathbf{y}$, należy obliczyć macierz efektów krańcowych jako pochodną wektorową wartości oczekiwanej $\mathbb{E}\,\mathbf{y}$. W ten sposób uzyskujemy macierz

$$\boldsymbol{E}_l(\mathbf{W})=\left[e_{ij}^{l}\right]_{1\leqslant i,j\leqslant N}\coloneqq\frac{\mathrm{d}(\mathbb{E}\,\mathbf{y})}{\mathrm{d}\mathbf{X}_l}=\left(\mathbf{I}-\sum_{n=1}^{p}\rho_n\mathbf{W}_n\right)^{-1}(\beta_l\mathbf{I}+\gamma_l\mathbf{W}),$$

gdzie e_{ij}^{l} wyraża efekt krańcowy zmiennej l w lokalizacji j na wartość y_i, który uwzględnia zależności przestrzenne. Rozszerzając terminologię wprowadzoną przez LeSage'a i Pace'a (2009), elementy macierzy $\boldsymbol{E}_l(\mathbf{W})$ leżące na głównej przekątnej nazwiemy efektami bezpośrednimi (ang. *direct impacts/effects*) zmiennej $\mathbf{X}_l$, a pozostałe elementy — efektami pośrednimi (ang. *indirect impacts/effects*). Przykładem zastosowania teorii efektów przestrzennych może być badanie opisane w pracy Olejnik i Olejnik (2019).

2. Estymacja modelu przestrzennego rzędu $(1,0)$

Dla modelu klasy SARAR$(1,0)$ wyrażonego specyfikacją

$$\begin{aligned}&\mathbf{y}=\rho\mathbf{W}\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\\ &\boldsymbol{\varepsilon}\sim\mathcal{N}\big(0,\sigma^2\mathbf{I}\big),\end{aligned}\tag{2.7}$$

estymator MNK jest w ogólności obciążony, a nawet niezgodny. Poniżej przedstawimy dwie alternatywne metody estymacji: metodę zmiennych instrumentalnych (MZI) oraz metodę największej wiarogodności (MNW). Wcześniej jednak sformułujemy pewną istotną uwagę dotyczącą dziedziny wartości parametru autoregresyjnego ρ. W tym celu przytoczymy tu twierdzenie Greszgorina. Wynikający z niego wniosek (twierdzenie II.2) jest kluczowy dla standaryzowanych wierszowo przestrzennych macierzy wag.

Twierdzenie II.1 (Greszgorin)
Niech $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$ będzie dowolną macierzą liczbową, a liczby R_i, dla $1 \leqslant i \leqslant N$, będą zdefiniowane jako sumy modułów elementów macierzy $\mathbf{W}$, poza przekątną, w odpowiednich wierszach, tj.

$$R_i = \sum_{j=1}^{i-1} |w_{ij}| + \sum_{j=i+1}^{N} |w_{ij}|.$$

Wówczas dla dowolnej, choćby zespolonej, wartości własnej λ istnieje $1 \leqslant i_\lambda \leqslant N$, takie, że

$$|\lambda - w_{i_\lambda i_\lambda}| \leqslant R_{i_\lambda}.$$

Dowód. Niech $\boldsymbol{x}^\lambda = \left(x_1^\lambda, \ldots, x_N^\lambda\right)$ będzie wektorem własnym macierzy $\mathbf{W}$, odpowiadającym dowolnej ustalonej wartości własnej λ. Wybierzmy jeden dowolny indeks $1 \leqslant i_\lambda \leqslant N$ spełniający warunek $|x_{i_\lambda}| \geqslant |x_i|$, dla wszystkich $1 \leqslant i \leqslant N$. Bezpośrednio z definicji wartości własnej mamy więc

$$\sum_{j=1}^{i_\lambda - 1} w_{ij} x_j^\lambda + \sum_{j=i_\lambda+1}^{N} w_{ij} x_j^\lambda = (\lambda - w_{i_\lambda i_\lambda}) \cdot x_{i_\lambda}^\lambda.$$

Zatem

$$|\lambda - w_{i_\lambda i_\lambda}| \leqslant \sum_{j=1}^{i_\lambda - 1} \left| w_{ij} \frac{x_j^\lambda}{x_{i_\lambda}^\lambda} \right| + \sum_{j=i_\lambda+1}^{N} \left| w_{ij} \frac{x_j^\lambda}{x_{i_\lambda}^\lambda} \right| \leqslant R_{i_\lambda}.$$

□

Jeśli za przestrzeń możliwych wartości parametru autoregresyjnego ρ przyjmiemy otwarty przedział $(-1, 1)$, wówczas macierz operatora różnicowania przestrzennego (ang. *spatial difference*) $\boldsymbol{\Delta}(\rho) = \mathbf{I} - \rho\mathbf{W}$ będzie nieosobliwa, dla dowolnej standaryzowanej macierzy $\mathbf{W}$ z zerową przekątną oraz dowolnej wartości parametru autoregresyjnego ρ.

Twierdzenie II.2

Niech $\mathbf{W}$ będzie macierzą standaryzowaną wierszowo o zerowej przekątnej. Wówczas dla dowolnego $\rho \in (-1,1)$ mamy

$$\det(\mathbf{I} - \rho\mathbf{W}) > 0$$

oraz

$$(\mathbf{I} - \rho\mathbf{W})^{-1} = I + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \dots$$

Dowód. Najpierw pokażemy nie wprost, że $\det(\mathbf{I} - \rho\mathbf{W}) \neq 0$, dla każdego $\rho \in (-1,1)$. W tym celu załóżmy przeciwnie, że istnieje pewne $-1 < \tilde{\rho} < 1$, dla którego $\det(\mathbf{I} - \tilde{\rho}\mathbf{W}) = 0$. Wówczas oczywiście $\tilde{\rho} \neq 0$ oraz dla $\lambda = 1/\tilde{\rho}$ mamy $\det(\lambda\mathbf{I} - \mathbf{W}) = 0$. Zatem λ jest wartością własną macierzy $\mathbf{W}$. Z twierdzenia Greszgorina (twierdzenie II.1) wynika więc istnienie indeksu i_λ, dla którego

$$|\lambda - w_{i_\lambda i_\lambda}| \leqslant R_{i_\lambda}.$$

Ponieważ na mocy założonych własności macierzy $\mathbf{W}$ mamy $w_{i_\lambda i_\lambda} = 0$ i $R_{i_\lambda} = 1$, więc $-1 < \lambda < 1$. Te nierówności z kolei pociągają za sobą alternatywę warunków

$$\tilde{\rho} < 1 \qquad \text{lub} \qquad 1 < \tilde{\rho},$$

z których każdy prowadzi do sprzeczności z założeniem, że $\tilde{\rho} \in (-1,1)$.

Zauważmy, że funkcja przypisująca dowolnej liczbie ρ wartość wyznacznika $\det(\mathbf{I} - \rho\mathbf{W})$ jest funkcją ciągłą. Dodatkowo, w punkcie $\rho_0 = 0$ przyjmuje ona wartość dodatnią, tj. $\det(\mathbf{I} - \rho_0\mathbf{W}) = \det\mathbf{I} = 1 > 0$. Zatem gdyby $\det(\mathbf{I} - \rho\mathbf{W}) < 0$ dla pewnego $\rho \in (-1,1)$, wówczas istniałaby pewna liczba ρ_1 leżąca pomiędzy ρ i ρ_0, dla której $\det(\mathbf{I} - \rho_1\mathbf{W}) = 0$, co ponownie prowadziłoby do sprzeczności. Wynika z tego, że prawdziwa jest nierówność $\det(\mathbf{I} - \rho\mathbf{W}) > 0$, dla wszystkich $\rho \in (-1,1)$.

Można wykazać, że funkcja, która przypisuje macierzy $\boldsymbol{A} = [a_{ij}]_{1 \leqslant i,j \leqslant N}$ wartość $\|\boldsymbol{A}\|_1 := \max_{1 \leqslant i \leqslant N} \sum_{j=1}^{N} |a_{ij}|$ jest multiplikatywną normą macierzową (patrz Horn, Johnson, 2013). Ponieważ $\|\rho\mathbf{W}\|_1 = \rho < 1$, macierz $\mathbf{I} - \rho\mathbf{W}$ jest odwracalna i zachodzi żądane rozwinięcie w szereg macierzowy. □

2.1. Estymacja metodą najmniejszych kwadratów

Najprostszym sposobem estymacji parametrów modelu ekonometrycznego jest metoda MNK. W przypadku modelu przestrzennego SAR nie gwarantuje ona jednak dobrej jakości oszacowań. Okazuje się, że dla pewnej klasy przestrzennych macierzy wag estymator MNK jest zgodny (por. Lee, 2002). Co więcej, inne metody estymacji, np. MNW, mogą prowadzić do oszacowań niezgodnych (Mynbaev, 2011). Fakt ten przybliżamy w obecnym podrozdziale.

Zauważmy najpierw, że zmienna $\mathbf{Wy}$ (występująca po prawej stronie równania opisującego specyfikację SAR) jest endogeniczna. Istotnie, przekształcając (2.7) otrzymamy

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon},$$

a zatem, uwzględniając fakt, iż $\boldsymbol{\varepsilon} \sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big)$, możemy wyliczyć

$$\begin{aligned}\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\mathbf{y} &= \mathbb{E}\left(\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\big((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\big)\right)\\ &= \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right] + \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\right] \qquad (2.8)\\ &= \sigma^2 \operatorname{tr}\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1},\end{aligned}$$

przy czym w ogólności nie jest prawdą, że $\operatorname{tr}\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1} = 0$. Przy oznaczeniach

$$\begin{aligned}\mathbf{Z} &= \begin{bmatrix}\mathbf{X} & \mathbf{Wy}\end{bmatrix},\\ \delta &= \begin{bmatrix}\boldsymbol{\beta}^{\mathsf{T}} & \rho\end{bmatrix}^{\mathsf{T}},\end{aligned}$$

estymator MNK $\hat{\delta}_{\text{MNK}}$ parametru δ ma postać

$$\hat{\delta}_{\text{MNK}} = \big(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\big)^{-1}\mathbf{Z}^{\mathsf{T}}\mathbf{y} = \delta_0 + \big(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\big)^{-1}\mathbf{Z}^{\mathsf{T}}\boldsymbol{\varepsilon}.$$

Jak wynika z ostatniej równości, właściwości estymatora MNK (takie jak nieobciążoność czy zgodność) zależą od zachowania asymptotycznego zmiennej losowej $(\mathbf{Z}^{\mathsf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathsf{T}}\boldsymbol{\varepsilon}$. Na przykład, jak pokazuje Lee (2002), przy pewnych naturalnych założeniach dotyczących zbieżności macierzy korelacji dla zmiennych objaśniających: $\frac{1}{N}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}$, estymator $\hat{\delta}_{\text{MNK}}$ jest asymptotycznie nieobciążony, gdy $\frac{1}{N}\,\mathbb{E}\,(\mathbf{Z}^{\mathsf{T}}\boldsymbol{\varepsilon})$ zbiega do zera. Ten warunek będzie z kolei spełniony, jeśli endogeniczny komponent $\mathbf{Z}$ (czyli zmienna $\mathbf{Wy}$) okaże się asymptotycznie egzogeniczny. Dokładniej, ponieważ $\frac{1}{N}\,\mathbb{E}\,(\mathbf{X}^{\mathsf{T}}\boldsymbol{\varepsilon}) = 0$, decydująca jest następująca zbieżność

$$\frac{1}{N}\,\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\mathbf{y} = \frac{\sigma^2}{N}\operatorname{tr}\left(\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1}\right) \xrightarrow{N\to\infty} 0.$$

Co więcej, powyższy warunek może również implikować zbieżność błędu estymatora $\hat{\delta}_{\text{MNK}}$ do zera według prawdopodobieństwa, a w efekcie zgodność $\hat{\delta}_{\text{MNK}}$.

Ostatecznie, pozostaje pytanie: dla jakich przestrzennych macierzy wag $\mathbf{W} = [w_{ij}]_{ij\leqslant N}$ można zastosować metodę najmniejszych kwadratów do estymacji modelu SAR. W oryginalnym rozumowaniu zawartym w pracy Lee (2002) rozważano macierze, będące wynikiem standaryzacji macierzy odległości (bez punktów

odcięcia) z założeniem asymptotyki rosnącej dziedziny (patrz s. 21). Okazuje się, że gdy wiersze takiej macierzy są niesumowalne, tj. $\lim_{N\to\infty}\sum_{j=1}^{N}|w_{ij}|=\infty$, dla każdego $i\geqslant 0$, to w wyniku operacji normalizacji wierszowej otrzymamy nową macierz $\mathbf{W}$, dla której można zastosować autorskie twierdzenie II.3.

Twierdzenie II.3

Niech $\mathbf{W}=[w_{ij}]_{ij\leqslant N}$ będzie macierzą standaryzowaną wierszowo o zerowej przekątnej (założenie II.A), dla której

$$\lim_{N\to\infty}\max_{1\leqslant i,j\leqslant N}|w_{ij}|=0.$$

Wówczas endogeniczność zmiennej $\mathbf{Wy}$ zanika asymptotycznie, szybciej niż odwrotność rozmiaru próby N^{-1}, a dokładniej zachodzi

$$\lim_{N\to\infty}\frac{1}{N}\,\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\mathbf{y}=0.$$

Dowód. Uwzględniając wyprowadzoną wcześniej równość (2.8), wystarczy pokazać, że

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{tr}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}=0.$$

W tym celu zauważmy, że jeśli pewne macierze $\mathbf{A}=[a_{ij}]_{1\leqslant i,j\leqslant N}$ i $\mathbf{B}=[b_{ij}]_{1\leqslant i,j\leqslant N}$ są standaryzowane wierszowo, wówczas macierzowy iloczyn $\mathbf{A}\cdot\mathbf{B}=[c_{ij}]_{1\leqslant i,j\leqslant N}$ jest również standaryzowana wierszowo. Istotnie,

$$\sum_{j=1}^{N}c_{ij}=\sum_{j=1}^{N}\sum_{k=1}^{N}a_{ik}b_{kj}=\sum_{k=1}^{N}a_{ik}\cdot\left(\sum_{j=1}^{N}b_{kj}\right)=1.$$

Wynika stąd, że macierz $\mathbf{W}$ oraz jej kolejne potęgi macierzowe $\mathbf{W}^2,\mathbf{W}^3,\ldots$ są standaryzowane wierszowo. Oznaczmy $\mathbf{V}=(\mathbf{I}-\rho\mathbf{W})^{-1}$, $\mathbf{V}=[v_{ij}]_{1\leqslant i,j\leqslant N}$ oraz $\mathbf{W}^k=\left[w_{ij}^{(k)}\right]_{1\leqslant i,j\leqslant N}$, dla $k\geqslant 0$. Na podstawie twierdzenia II.2 prawdziwe jest rozwinięcie $\mathbf{V}=\sum_{k=0}^{\infty}\rho^k\mathbf{W}^k$. Zatem, dla każdego $1\leqslant j\leqslant N$ wnioskujemy, że

$$\sum_{j=1}^{N}v_{ij}=\sum_{j=1}^{N}\sum_{k=0}^{\infty}\rho^k w_{ij}^{(k)}=\sum_{k=0}^{\infty}\rho^k\sum_{j=1}^{N}w_{ij}^{(k)}=\sum_{k=0}^{\infty}\rho^k=\frac{1}{1-\rho}.$$

Dla $\mathbf{G}=\mathbf{W}\cdot\mathbf{V}$, przy czym $\mathbf{G}=[g_{ij}]_{1\leqslant i,j\leqslant N}$, oraz $M_N=\max_{1\leqslant i,j\leqslant N}|w_{ij}|$ uzyskujemy ostatecznie

$$\frac{1}{N}\operatorname{tr}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}=\frac{1}{N}\sum_{i=1}^{N}g_{ii}=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}w_{ik}v_{ki}\leqslant\frac{M_N}{1-\rho},$$

Na mocy przyjętych założeń, prawa strona tej nierówności zbiega do zera wraz z $N\to\infty$. □

2.2. Estymacja metodą zmiennych instrumentalnych

Opisana poniżej procedura estymacji, oparta na metodzie MZI, wykorzystuje pomysł ze znanej pracy Kelejiana i Pruchy (1998). Polega ona na tym, by budowę instrumentów dla zmiennej $\mathbf{Wy}$ oprzeć na wartości oczekiwanej $\mathbb{E}\,\mathbf{Wy}$. Korzystając z twierdzenia II.2 mamy

$$\begin{aligned}\mathbb{E}\,\mathbf{Wy} &= \mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\,\mathbb{E}\,\boldsymbol{\varepsilon}\\ &= \mathbf{W}\big(\mathbf{I}+\rho\mathbf{W}+\rho^2\mathbf{W}^2+\dots\big)\mathbf{X}\boldsymbol{\beta}\\ &= \mathbf{WX}\boldsymbol{\beta} + \mathbf{W}^2\mathbf{X}\cdot\rho\boldsymbol{\beta} + \mathbf{W}^3\mathbf{X}\cdot\rho^2\boldsymbol{\beta}+\dots\end{aligned}$$

Zatem $\mathbb{E}\,\mathbf{Wy}$ jest pewną (nieskończoną) kombinacją liniową kolumn macierzy $\mathbf{WX}$, $\mathbf{W}^2\mathbf{X}$, $\mathbf{W}^3\mathbf{X}$, itd. Ponadto, optymalnym instrumentem dla macierzy zmiennych objaśniających $\mathbf{X}$ jest ona sama. Sugeruje to, że zbiór instrumentów można wybrać spośród liniowo niezależnych kolumn macierzy

$$\begin{bmatrix}\mathbf{X} & \mathbf{WX} & \mathbf{W}^2\mathbf{X} & \dots\end{bmatrix}.$$

Przyjmijmy zatem, że wybrana została macierz instrumentów $\mathbf{H}$. Załóżmy, że dla dostatecznie dużego N jest ona macierzą pełnego rzędu $P \geqslant K+1$, gdzie K to liczba kolumn w $\mathbf{X}$. Załóżmy, że $\mathbf{H}$ zawiera co najmniej liniowo niezależne kolumny macierzy $\begin{bmatrix}\mathbf{X} & \mathbf{WX}\end{bmatrix}$. Wówczas, przy oznaczeniach $\mathbf{Z} = \begin{bmatrix}\mathbf{X} & \mathbf{Wy}\end{bmatrix}$ oraz $\delta = \begin{bmatrix}\boldsymbol{\beta}^{\mathsf{T}} & \rho\end{bmatrix}^{\mathsf{T}}$, estymator MZI ma postać

$$\hat{\delta}_{\mathrm{MZI}} = \big(\hat{\mathbf{Z}}^{\mathsf{T}}\hat{\mathbf{Z}}\big)^{-1}\hat{\mathbf{Z}}^{\mathsf{T}}\mathbf{y},$$

gdzie $\hat{\mathbf{Z}} = \mathbf{P_H Z} = \begin{bmatrix}\mathbf{X} & \widehat{\mathbf{Wy}}\end{bmatrix}$, $\widehat{\mathbf{Wy}} = \mathbf{P_H Wy}$ oraz $\mathbf{P_H} = \mathbf{H}(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathsf{T}}$. Przy pewnych dodatkowych założeniach powyższy estymator jest zgodny, a jego wariancja jest wówczas asymptotycznie równa

$$\operatorname{Var}\hat{\delta}_{\mathrm{MZI}} = \hat{\sigma}^2\big(\mathbf{Z}^{\mathsf{T}}\mathbf{H}(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathsf{T}}\mathbf{Z}\big)^{-1},$$

gdzie $\hat{\sigma}^2 = \frac{1}{N}(\mathbf{y}-\mathbf{Z}\hat{\delta})^{\mathsf{T}}(\mathbf{y}-\mathbf{Z}\hat{\delta})$.

2.3. Estymacja metodą największej wiarogodności

Jedną z podstawowych metod estymacji modeli przestrzennych jest metoda największej wiarogodności. Tak jak w przypadku klasycznym (patrz Lehmann, Casella, 1998), jako wartość estymatora metoda ta wskazuje taką wartość parametru specyfikacji procesu generującego obserwacje (ang. *data generating process*), dla którego prawdopodobieństwo odnotowania faktycznie zgromadzonych wartości

próby jest największe. Wykorzystanie metody MNW w ekonometrii przestrzennej sugerował już w latach osiemdziesiątych Anselin (1988a) i do dziś pozostaje ona narzędziem często wybieranym przez praktyków. Należy też zwrócić uwagę na mnogość współczesnych opracowań teoretycznych, w których rozwijana jest metoda MNW. W efekcie, uzyskuje się narzędzie estymacyjne stosowalne do nawet bardzo rozbudowanych specyfikacji przestrzennych.

Metoda MNK bywała krytykowana w literaturze przedmiotu ze względu na związane z nią trudności obliczeniowe i konieczność zastosowania specjalistycznych numerycznych metod optymalizacji. Obecnie jednak, ze względu na gwałtowny rozwój technologii komputerowych, zwłaszcza obliczenia współbieżne, wzrost mocy obliczeniowej komputerów oraz szeroki wachlarz rozwiązań algorytmicznych (patrz, np. Kossowski, Hauke, 2011; Bivand i inni, 2013), obawy powtarzane jeszcze w książce Suchecki [red.] (2010) należy uznać za niewspółczesne.

Ponieważ z założenia składnik losowy $\boldsymbol{\varepsilon}$ ma wielowymiarowy rozkład normalny $\mathcal{N}(0, \sigma^2\mathbf{I})$, również $\mathbf{y}$, będący afiniczną transformacją $\boldsymbol{\varepsilon}$, ma rozkład gaussowski. Zatem funkcja gęstości $\mathbb{R}^N \ni \boldsymbol{y} \mapsto f_{\mathbf{y}}(\boldsymbol{y})$ rozkładu zmiennej $\mathbf{y}$ przyjmuje postać

$$f_{\mathbf{y}}(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Omega}_{\mathbf{y}}}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \mathbb{E}\,\mathbf{y})^{\mathsf{T}} \boldsymbol{\Omega}_{\mathbf{y}}^{-1} (\boldsymbol{y} - \mathbb{E}\,\mathbf{y})\right\},$$

gdzie $\mathbb{E}\,\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$ oraz $\boldsymbol{\Omega}_{\mathbf{y}} \coloneqq \operatorname{Var}\mathbf{y} = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}\big(\mathbf{I} - \rho\mathbf{W}^{\mathsf{T}}\big)^{-1}$. A zatem, rozwijając powyższe możemy zbudować funkcję wiarogodności z parametrami ρ, $\boldsymbol{\beta}$ i σ^2 przy obserwacji $\boldsymbol{y}$ zmiennej $\mathbf{y}$

$$L_{\boldsymbol{y}}(\rho, \boldsymbol{\beta}, \sigma^2) = \frac{\det \boldsymbol{\Delta}(\rho)}{\sqrt{(2\pi\sigma^2)^N}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\Delta}(\rho)\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\right)^{\mathsf{T}}\left(\boldsymbol{\Delta}(\rho)\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\},$$

gdzie $\boldsymbol{\Delta}(\rho) = \mathbf{I} - \rho\mathbf{W}$. Zauważmy, że przyłożenie do funkcji wiarogodności funkcji ściśle rosnącej nie zmienia argumentu maksymalizującego. W szczególności, w przypadku rozkładów z rodziny wykładniczej, do której należy rozkład normalny (patrz Lehmann, Casella, 1998), wygodnie jest użyć w tym celu funkcji logarytmicznej. A więc, logarytmując obustronnie i dokonując podstawienia $\boldsymbol{y} = \mathbf{y}$, otrzymujemy funkcję (zmienną losową) log-wiarogodności postaci

$$\ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln\det\boldsymbol{\Delta}(\rho) - \frac{1}{2\sigma^2}(\boldsymbol{\Delta}(\rho)\cdot\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{\Delta}(\rho)\cdot\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Oczywiście, odnalezienie estymatora MNW wymaga wskazania takich wartości parametrów ρ, $\boldsymbol{\beta}$ i σ^2, dla których $\ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)$ przyjmuje wartość możliwie najmniejszą, dla każdego ustalonego zestawu wartości obserwowanej zmiennej $\mathbf{y}$ z osobna. Należy jednak najpierw poczynić następujące obserwacje. Po pierwsze, funkcja log-wiarogodności $\ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)$ jest ciągła, ale jej dziedzina, czyli przestrzeń dopuszczalnych wartości parametrów, nie jest zbiorem zwartym. Co za tym idzie, należy sprawdzić, czy żądane maksimum jest w ogóle osiągane. Po drugie, gdyby nie składnik $\ln\det \boldsymbol{\Delta}(\rho)$, maksymalizacja $\ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)$ mogłaby odbyć się zwykłą metodą analityczną, podobnie jak ma to miejsce w przypadku modelu nieprzestrzennego. Zatem, argument optymalny, jeśli istnieje, jest odnajdywany metodami numerycznymi. Obliczanie wartości wyznacznika macierzy $\boldsymbol{\Delta}(\rho)$ dla dużych rozmiarów próby N okazuje się złożone obliczeniowo. Co więcej — w ogólności — składnik $\ln\det \boldsymbol{\Delta}(\rho)$, jako funkcja ρ, nie musi być nawet funkcją wklęsłą, a więc numeryczne metody optymalizacji wypukłej nie mają tu zastosowania.

Złożoność tego problemu można jednak zredukować poprzez zastosowanie pewnej metody prowadzącej do usunięcia zmiennych $\boldsymbol{\beta}$ i σ^2 z optymalizowanej funkcji celu. Metoda ta nosi nazwę metody wyrugowywania parametrów przez koncentrację (ang. *concentrating out*). Obrazowo można powiedzieć, że prowadzi ona do zmniejszenia liczby parametrów problemu optymalizacyjnego poprzez wycięcie z dziedziny funkcji celu pewnej krzywej (a ogólnie hiperpowierzchni), określającej zależność między optymalnymi parametrami, a więc przechodzącej tym samym przez punkty optymalne. Taką krzywą znajduje się najczęściej analitycznie, stosując warunki różniczkowe pierwszego rzędu. W naszym przypadku możemy wyliczyć następujące pochodne cząstkowe funkcji celu

$$\begin{aligned}
\frac{\partial \ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= \frac{1}{2\sigma^2} \cdot \frac{\partial}{\partial \boldsymbol{\beta}} \big(\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\big)^{\mathsf{T}} \big(\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\big) \\
&= \frac{1}{\sigma^2}\big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}^{\mathsf{T}}\mathbf{X}\boldsymbol{\beta}\big), \\
\frac{\partial \ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{1}{2\sigma^4}\big(\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\big)^{\mathsf{T}}\big(\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\big) - \frac{N}{2\sigma^2}.
\end{aligned}$$

Zauważmy, że uzyskane wyrażenia nie zawierają problematycznego wyznacznika $\ln\det \boldsymbol{\Delta}(\rho)$. Przyrównując jednocześnie do zera powyższe pochodne, otrzymujemy żądane zależności optymalnych wartości $\boldsymbol{\beta}$ i σ^2 od ρ, mianowicie

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(\rho) &= \big(\mathbf{X}^{\mathsf{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}(\mathbf{I} - \rho\mathbf{W})\mathbf{y} \\
\hat{\sigma}^2(\rho) &= \frac{1}{N}\big(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\rho)\big)^{\mathsf{T}}\big(\mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\rho)\big).
\end{aligned}$$

Uwzględniając powyższe zależności w oryginalnej funkcji celu $\ln L_{\mathbf{y}}(\rho, \boldsymbol{\beta}, \sigma^2)$ uzyskujemy funkcję zależną tylko od jednego parametru ρ, tj.

$$\ln L_{\mathbf{y}}\big(\rho, \hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)\big) = -\frac{N}{2}\left(\ln\left(2\pi \cdot \hat{\sigma}^2(\rho)\right) + 1\right) + \ln\det \Delta(\rho). \tag{2.9}$$

Rozważmy teraz problem istnienia rozwiązania wskazanego problemu optymalizacyjnego. Ponieważ funkcja celu $\rho \mapsto \ln L_{\mathbf{y}}\big(\rho, \hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)\big)$ jest ciągła na całej przestrzeni dopuszczalnych wartości parametru ρ, wiadomo, że osiąga ona swoje ekstremum na każdym zbiorze (zwartym) postaci przedziału domkniętego $[-1+\epsilon, 1-\epsilon] \subset (-1,1)$, dla dowolnie małej liczby $\epsilon > 0$. Możemy zatem zauważyć, że wystarczy zbadać asymptotyczne zachowanie funkcji celu w sąsiedztwie punktów skrajnych -1 i 1.

Przyjmijmy oznaczenie $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$. Jest oczywiste, że funkcja

$$\begin{aligned}
\rho \mapsto \hat{\sigma}^2(\rho) &= \frac{1}{N}\big((\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\rho)\big)^{\mathsf{T}}\big((\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\rho)\big) \\
&= \frac{1}{N}\big(\mathbf{M}_{\mathbf{X}}(\mathbf{I} - \rho\mathbf{W})\mathbf{y}\big)^{\mathsf{T}}\big(\mathbf{M}_{\mathbf{X}}(\mathbf{I} - \rho\mathbf{W})\mathbf{y}\big) \\
&= a \cdot \rho^2 + b \cdot \rho + c,
\end{aligned}$$

będąca trójmianem kwadratowym zmiennej ρ, przy czym

$$\begin{aligned}
a &= \frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\mathbf{M}_{\mathbf{X}}\mathbf{W}\mathbf{y} \\
b &= \frac{2}{N}\mathbf{y}^{\mathsf{T}}\mathbf{M}_{\mathbf{X}}\mathbf{W}\mathbf{y} \\
c &= \frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{M}_{\mathbf{X}}\mathbf{y},
\end{aligned}$$

przyjmuje wartości nieujemne. Co więcej, wartość zero w punktach $\rho = -1$ i $\rho = 1$ może ona przyjąć tylko wtedy, gdy zachodzi jeden z warunków

$$\mathbf{y} = \mathbf{W}\mathbf{y} + \mathbf{X}\hat{\boldsymbol{\beta}}(1),$$

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

z których każdy implikuje, w praktyce nierealistyczny, a w teorii mało ciekawy, warunek idealnego dopasowania danych. Możemy zatem przyjąć, że skończone są następujące granice

$$\limsup_{\rho \to -1}\left(-\frac{N}{2}\left(\ln\left(2\pi\hat{\sigma}^2(\rho)\right) + 1\right)\right) < \infty$$

oraz

$$\limsup_{\rho\to 1}\left(-\frac{N}{2}\left(\ln\left(2\pi\hat{\sigma}^2(\rho)\right)+1\right)\right)<\infty.$$

Przechodząc do drugiego składnika w formule (2.9), opisującej skoncentrowaną log-wiarogodność, zauważmy, że wyrażenie $\ln\det\mathbf{\Delta}(\rho)$ jest (ujemnie) koersywne na krańcach przestrzeni dopuszczalnych wartości dla parametru ρ w następującym sensie

$$\lim_{\rho\to 1}\left(\ln\det\mathbf{\Delta}(\rho)\right)=\lim_{\rho\to -1}\left(\ln\det\mathbf{\Delta}(\rho)\right)=-\infty.$$

Zatem, również dla funkcji $\rho\mapsto\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big)$ mamy

$$\lim_{\rho\to 1}\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big)=-\infty$$

oraz

$$\lim_{\rho\to 1}\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big)=-\infty,$$

z czego wynika, że przyjmuje ona swoje maksimum w otwartym przedziale $(-1,1)$.

Maksymalizacja skoncentrowanej log-wiarogodności $\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big)$ prowadzi do estymatora największej wiarogodności parametru autoregresyjnego, tj.

$$\hat{\rho}_{\mathrm{MNW}}=\operatorname*{arg\,max}_{-1<\rho<1}\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big).$$

Pozostaje jednak kwestia jednoznaczności wartości tak określonego estymatora stanowiąca osobne zagadnienie. Problem identyfikowalności parametru autoregresyjnego ρ formalnie rozwiążemy dopiero w rozdziale III. Tymczasem przyjmijmy, że operacja $\arg\max(\cdot)$ wybiera wartość argumentu maksymalizującego dowolnie, ale w sposób mierzalny, tzn. tak, że $\hat{\rho}_{\mathrm{MNW}}$ jest dobrze określoną zmienną losową. Należy tu zaznaczyć, że w praktyce oszacowanie $\hat{\rho}_{\mathrm{MNW}}$ jest uzyskiwane metodami numerycznymi, chociaż niektóre polskojęzyczne opracowania mogą wprowadzać w błąd, sugerując procedury oparte na przyrównywaniu pierwszej pochodnej funkcji $\rho\mapsto\ln L_{\mathbf{y}}\big(\rho,\hat{\boldsymbol{\beta}}(\rho),\hat{\sigma}^2(\rho)\big)$ do zera; por. Suchecki [red.], 2010. Oszacowania pozostałych parametrów uzyskujemy z wcześniejszych warunków różniczkowych pierwszego rzędu, czyli

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{MNW}}&=\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\cdot\left(\mathbf{I}-\hat{\rho}_{\mathrm{MNW}}\mathbf{W}\right)\mathbf{y},\\ \hat{\sigma}^2_{\mathrm{MNW}}&=\frac{1}{N}\left\|\mathbf{y}-\hat{\rho}_{\mathrm{MNW}}\mathbf{W}\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\rho}_{\mathrm{MNW}})\right\|^2\\ &=\frac{1}{N}\left\|\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}}\right)\cdot\left(\mathbf{I}-\hat{\rho}_{\mathrm{MNW}}\mathbf{W}\right)\mathbf{y}\right\|^2,\end{aligned}$$

gdzie $\mathbf{P_X}$ jest macierzą operatora rzutu ortogonalnego na przestrzeń liniową rozpiętą przez kolumny macierzy zmiennych objaśniających $\mathbf{X}$. Można zauważyć, że efektywnie metoda MNW maksymalizuje (skoncentrowaną) wiarogodność w celu znalezienia wartości parametru autoregresyjnego, a pozostałe parametry wyliczane są najmniejszych kwadratów.

3. Estymacja modelu przestrzennego rzędu $(0,1)$

Pierwszym zagadnieniem ekonometrii przestrzennej, rozważanym w literaturze dotyczącej problemów regionalnych, była analiza efektów zależności przestrzennych składnika losowego w liniowym modelu regresji. Początki opisu problemu oraz pierwsze próby jego rozwiązania przypadają na lata siedemdziesiąte XX wieku (por. Fisher, 1971; Cliff, Ord, 1973; Hordijk, 1974). Opracowania z tamtego okresu zaowocowały wieloma dalszymi ocenami własności różnorodnych estymatorów oraz testów statystycznych.

Przypomnijmy, że zależności przestrzenne składnika losowego w liniowym modelu regresji opisuje specyfikacja SARAR$(0,1)$

$$\begin{aligned}
&\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\\
&\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}\\
&\boldsymbol{\varepsilon} \sim \mathcal{N}\big(0, \sigma_{\varepsilon}^2\mathbf{I}\big)\\
&\lambda \in (-1,1).
\end{aligned}$$

Zakładając odpowiednią postać macierzy wag (np. standaryzowanie wierszowe, czyli założenie II.A) i uwzględniając twierdzenie II.2, otrzymujemy $\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$, zatem możemy wnioskować o normalności rozkładu wektora reszt $\mathbf{u}$, a w szczególności o jego momentach. Mamy zatem

$$\mathbb{E}\,\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\,\mathbb{E}\,\boldsymbol{\varepsilon} = 0$$

oraz

$$\begin{aligned}
\operatorname{Var}(\mathbf{u}) &= \mathbb{E}\,\mathbf{u}\mathbf{u}^{\mathsf{T}} - \mathbb{E}\,\mathbf{u}\,\mathbb{E}\,\mathbf{u}^{\mathsf{T}}\\
&= \mathbb{E}\,(\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}}\big(\mathbf{I} - \lambda\mathbf{W}^{\mathsf{T}}\big)^{-1}\\
&= (\mathbf{I} - \lambda\mathbf{W})^{-1}\cdot\sigma_{\varepsilon}^2\mathbf{I}\cdot\big(\mathbf{I} - \lambda\mathbf{W}^{\mathsf{T}}\big)^{-1}\\
&= \sigma_{\varepsilon}^2(\mathbf{I} - \lambda\mathbf{W})^{-1}\big(\mathbf{I} - \lambda\mathbf{W}^{\mathsf{T}}\big)^{-1}.
\end{aligned}$$

Przy oznaczeniu

$$\boldsymbol{\Omega}_{\mathbf{u}}(\lambda) = \sigma_{\varepsilon}^2\,(\mathbf{I} - \lambda\mathbf{W})^{-1}\,\big(\mathbf{I} - \lambda\mathbf{W}^{\mathsf{T}}\big)^{-1}, \tag{2.10}$$

otrzymujemy $\mathbf{u} \sim \mathcal{N}(0, \mathbf{\Omega_u}(\lambda))$, a w efekcie możemy wnioskować o normalności rozkładu zmiennej zależnej $\mathbf{y}$, tj. $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{\Omega}_u(\lambda))$.

3.1. Nieadekwatność uogólnionej metody najmniejszych kwadratów

Ze względu na zaobserwowaną w wariancji zmiennej zależnej nieznaną macierz $\mathbf{\Omega_u}(\lambda)$, będącą funkcją estymowanego parametru, klasyczny estymator UMNK nie ma zastosowania w przypadku modelu z autoskorelowanym składnikiem losowym. W przypadku tego modelu metoda zmiennych instrumentalnych również nie gwarantuje pożądanych własności oszacowań parametrów. Podobny problem dotyczy wszystkich ogólniejszych modeli z autoskorelowanym składnikiem losowym, czyli również modeli SARAR$(1,1)$ i modeli wyższych rzędów oraz modeli klasy Durbina.

3.2. Estymacja metodą największej wiarogodności

Jak zauważyliśmy wcześniej, zmienna zależna ma rozkład gaussowski, a dokładniej $\mathbf{y} \sim \mathcal{N}\big(\mathbf{X}\boldsymbol{\beta}, \sigma_{\varepsilon}^2 \mathbf{\Omega}_u(\lambda)\big)$. Związana z nią funkcja gęstości $\mathbb{R}^N \ni \boldsymbol{y} \mapsto f_{\mathbf{y}}(\boldsymbol{y})$ przyjmuje postać

$$f_{\mathbf{y}}(\boldsymbol{y}) = \frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{y} - \mathbb{E}\,\mathbf{y}\right)^{\mathsf{T}} \mathbf{\Omega_u}(\lambda)^{-1}\left(\boldsymbol{y} - \mathbb{E}\,\mathbf{y}\right)\right\}}{\sqrt{(2\pi)^N \det \mathbf{\Omega_u}(\lambda)}}, \tag{2.11}$$

gdzie $\mathbb{E}\,\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, a $\operatorname{Var}\mathbf{y} = \mathbf{\Omega_u}(\lambda)$ dana jest w (2.10). Rozwijając (2.11) możemy zbudować funkcję wiarogodności z parametrami $\boldsymbol{\beta}$, λ i σ_{ε}^2, przy obserwacji $\boldsymbol{y}$ zmiennej $\mathbf{y}$

$$L_{\boldsymbol{y}}(\boldsymbol{\beta}, \lambda, \sigma_{\varepsilon}^2) = \frac{\det \mathbf{\Gamma}(\lambda)}{\sqrt{\left(2\pi\sigma_{\varepsilon}^2\right)^N}} \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^2} \left\|\mathbf{\Gamma}(\lambda)\boldsymbol{y} - \mathbf{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\right\|^2\right\},$$

gdzie $\mathbf{\Gamma}(\lambda) \coloneqq \mathbf{I} - \lambda\mathbf{W}$. Podobnie jak w przypadku procedury estymacji największej wiarogodności przestrzennego modelu autoregresyjnego SAR, logarytmując obustronnie i podstawiając $\mathbf{y} = \boldsymbol{y}$ otrzymujemy funkcję log-wiarogodności postaci

$$\begin{aligned}\ln L_{\mathbf{y}}(\boldsymbol{\beta}, \lambda, \sigma_{\varepsilon}^2) = -\frac{N}{2}\ln\left(2\pi\sigma_{\varepsilon}^2\right) + \ln\det\mathbf{\Gamma}(\lambda) \\ - \frac{1}{2\sigma_{\varepsilon}^2}\big(\mathbf{\Gamma}(\lambda)\mathbf{y} - \mathbf{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big)^{\mathsf{T}}\big(\mathbf{\Gamma}(\lambda)\mathbf{y} - \mathbf{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big).\end{aligned}$$

Ponownie, odnalezienie estymatora MNW wymaga znalezienia takich wartości parametrów $\boldsymbol{\beta}$, λ i σ_{ε}^2, dla których $\ln L_{\mathbf{y}}(\boldsymbol{\beta}, \lambda, \sigma_{\varepsilon}^2)$ przyjmuje wartość możliwie

najmniejszą, dla każdego ustalonego zestawu wartości obserwowanej zmiennej $\mathbf{y}$ z osobna. W celu wyrugowania zmiennych $\boldsymbol{\beta}$ i $\sigma_{\boldsymbol{\varepsilon}}^2$ z optymalizowanej funkcji celu wyprowadzamy następujące pochodne cząstkowe

$$\begin{aligned}\frac{\partial \ln L_{\mathbf{y}}}{\partial \boldsymbol{\beta}} &= \frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2} \cdot \frac{\partial}{\partial \boldsymbol{\beta}} \big(\boldsymbol{\Gamma}(\lambda)\mathbf{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big)^{\mathsf{T}} \big(\boldsymbol{\Gamma}(\lambda)\mathbf{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big) \\ &= \frac{1}{\sigma_{\boldsymbol{\varepsilon}}^2}\big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)\mathbf{y} - \mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)\mathbf{X}\boldsymbol{\beta}\big), \\ \frac{\partial \ln L_{\mathbf{y}}}{\partial \sigma_{\boldsymbol{\varepsilon}}^2} &= -\frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^4}\big(\boldsymbol{\Gamma}(\lambda)\mathbf{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big)^{\mathsf{T}}\big(\boldsymbol{\Gamma}(\lambda)\mathbf{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\big) - \frac{N}{2\sigma_{\boldsymbol{\varepsilon}}^2}.\end{aligned}$$

Przyrównując powyższe pochodne jednocześnie do zera, otrzymujemy żądane zależności optymalnych wartości $\boldsymbol{\beta}$ i $\sigma_{\boldsymbol{\varepsilon}}^2$ od λ, mianowicie

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\lambda) &= \big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)\mathbf{y}, \\ \hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\lambda) &= \frac{1}{N}\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\lambda)\big)^{\mathsf{T}}\boldsymbol{\Omega}_u(\lambda)\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\lambda)\big).\end{aligned}$$

Powyższe równania można zapisać za pomocą przestrzennego odpowiednika klasycznej transformacji Cochrana-Orcutta, znanej z analizy szeregów czasowych. Mianowicie, ustalając notację $\widetilde{\mathbf{y}} = \mathbf{y} - \lambda\mathbf{W}\mathbf{y}$ oraz $\widetilde{\mathbf{X}} = \mathbf{X} - \lambda\mathbf{W}\mathbf{X}$, możemy zapisać

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\lambda) &= \big(\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\mathbf{y}}, \\ \hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\lambda) &= \frac{1}{N}\big(\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}(\lambda)\big)^{\mathsf{T}}\big(\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}(\lambda)\big).\end{aligned}$$

Uwzględniając wskazane zależności w oryginalnej funkcji celu $\ln L_{\mathbf{y}}(\boldsymbol{\beta}, \lambda, \sigma_{\boldsymbol{\varepsilon}}^2)$, uzyskujemy funkcję zależną od jednego tylko parametru, tj.

$$\lambda \mapsto \ln L_{\mathbf{y}}\big(\hat{\boldsymbol{\beta}}(\lambda), \lambda, \hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\lambda)\big) = -\frac{N}{2}\left(\ln\left(2\pi\hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\lambda)\right) + 1\right) + \ln\det\boldsymbol{\Gamma}(\lambda).$$

Dla modelu SEM, podobnie jak w przypadku modelu SAR, można wykazać, że powyższa (zredukowana przez rugowanie) funkcja log-wiarogodności osiąga swoje maksimum po zbiorze argumentów $\lambda \in (-1, 1)$. Argument maksymalizujący, to estymator największej wiarogodności parametru autoregresyjnego, czyli

$$\hat{\lambda}_{\text{MNW}} \coloneqq \underset{-1<\lambda<1}{\arg\max} \ln L_{\mathbf{y}}\big(\hat{\boldsymbol{\beta}}(\lambda), \lambda, \hat{\sigma}^2\lambda\big).$$

Oszacowania pozostałych parametrów uzyskujemy z wcześniejszych warunków różniczkowych pierwszego rzędu, czyli

$$\hat{\boldsymbol{\beta}}_{\text{MNW}} \coloneqq \hat{\boldsymbol{\beta}}\big(\hat{\lambda}_{\text{MNW}}\big) = \big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\hat{\lambda}_{\text{MNW}})\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\hat{\lambda}_{\text{MNW}})\mathbf{y},$$

$$\begin{aligned}\hat{\sigma}^2_{\text{MNW}} &\coloneqq \hat{\sigma}^2_{\varepsilon}\big(\hat{\lambda}_{\text{MNW}}\big)\\ &= \frac{1}{N}\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\lambda}_{\text{MNW}})\big)^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\hat{\lambda}_{\text{MNW}})\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\lambda}_{\text{MNW}})\big).\end{aligned}$$

Można zauważyć, że efektywnie metoda MNW maksymalizuje wiarogodność w celu znalezienia wartości parametru autoregresyjnego składnika losowego, a pozostałe parametry wyliczane są zgodnie z uogólnioną metodą najmniejszych kwadratów.

4. Estymacja modelu przestrzennego rzędu $(1, 1)$

Przyjmijmy następującą specyfikację modelu SARAR$(1, 1)$

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\\ \mathbf{u} &= \lambda\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2_{\varepsilon}\mathbf{I}),\end{aligned}$$

gdzie macierze wag przestrzennych $\mathbf{W}$ oraz $\mathbf{M}$ mogą, choć nie muszą, być sobie równe, a przestrzenie wartości parametrów autoregresyjnych wyznaczone są przez nierówności

$$-1 < \rho < 1 \quad \text{oraz} \quad -1 < \lambda < 1.$$

Na wstępie zauważmy, że $\mathbb{E}\,\mathbf{u} = (\mathbf{I} - \lambda\mathbf{M})^{-1}\,\mathbb{E}\,\boldsymbol{\varepsilon} = 0$, i dalej

$$\begin{aligned}\operatorname{Var}(\mathbf{u}) &= \mathbb{E}\,\mathbf{u}\mathbf{u}^{\mathsf{T}} - \mathbb{E}\,\mathbf{u}\,\mathbb{E}\,\mathbf{u}^{\mathsf{T}}\\ &= \mathbb{E}\,(\mathbf{I} - \lambda\mathbf{M})^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}}\big(\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}}\big)^{-1}\\ &= (\mathbf{I} - \lambda\mathbf{M})^{-1}\cdot\sigma^2_{\varepsilon}\mathbf{I}\cdot\big(\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}}\big)^{-1}\\ &= \sigma^2_{\varepsilon}(\mathbf{I} - \lambda\mathbf{M})^{-1}\big(\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}}\big)^{-1}.\end{aligned}$$

Zatem, przy oznaczeniu

$$\boldsymbol{\Omega}_{\mathbf{u}}(\lambda) = \big((\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}})(\mathbf{I} - \lambda\mathbf{M})\big)^{-1},$$

mamy $\mathbf{u} \sim \mathcal{N}\big(0, \boldsymbol{\Omega}_{\mathbf{u}}(\lambda)\big)$. Ponieważ

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \lambda\mathbf{M})^{-1}\boldsymbol{\varepsilon},$$

możemy wnioskować również o normalności rozkładu zmiennej zależnej oraz o jej momentach, które są równe

$$\begin{aligned}\mathbb{E}\,\mathbf{y} &= (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta},\\ \operatorname{Var}(\mathbf{y}) &= \sigma^2_{\varepsilon}(\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \lambda\mathbf{M})^{-1}\big(\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}}\big)^{-1}\big(\mathbf{I} - \rho\mathbf{W}^{\mathsf{T}}\big)^{-1}\\ &= \sigma^2_{\varepsilon}(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\Omega}_u(\lambda)\big(\mathbf{I} - \rho\mathbf{W}^{\mathsf{T}}\big)^{-1}.\end{aligned}$$

Ostatecznie wprowadzamy oznaczenie

$$\boldsymbol{\Omega}_{\mathbf{y}}(\rho,\lambda) \coloneqq \operatorname{Var}(\mathbf{y}) = (\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\Omega}_u(\lambda)\big(\mathbf{I}-\rho\mathbf{W}^{\mathsf{T}}\big)^{-1}. \tag{2.12}$$

Wykażemy, że w modelu SARAR$(1,1)$, podobnie jak w przypadku modelu SEM, zmienna $\mathbf{Wy}$ jest endogeniczna. Istotnie, mamy

$$\begin{aligned}\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\mathbf{y} &= \mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\lambda\mathbf{M})^{-1}\boldsymbol{\varepsilon}\\ &= \sigma_{\varepsilon}^2\operatorname{tr}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\lambda\mathbf{M})^{-1},\end{aligned}$$

przy czym, w ogólności, nie jest prawdą, że

$$\operatorname{tr}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\lambda\mathbf{M})^{-1} = 0,$$

zatem estymator MNK nie musi być zgodny.

4.1. Estymacja z wykorzystaniem uogólnionej metody momentów

Zauważmy, że chociaż w przypadku modelu SARAR$(1,1)$ można przeprowadzić rozumowanie dotyczące estymacji metodą zmiennych instrumentalnych, podobne do tego zaprezentowanego dla modelu SARAR$(1,0)$, to jednak estymator MZI nie uwzględnia skorelowania składnika losowego. Z tego właśnie powodu Kelejian i Prucha (1998) zaproponowali dwustopniowe rozszerzenie metody MNK (ang. *Generalized Spatial Two-Stage Least Squares*, GS2SLS), poprzez zastosowanie uogólnionej metody momentów (ang. *Generalized Method of Moments, GMM*). Poniżej zaprezentujemy w pewnym skrócie kolejne etapy tej procedury.

Etap 1

Estymujemy model regresyjno-autoregresyjny

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{2.13}$$

ignorując chwilowo autoskorelowanie składnika losowego. Stosujemy w tym celu opisaną wcześniej metodę zmiennych instrumentalnych. Przyjmijmy, że użytą macierzą instrumentów jest macierz

$$\mathbf{H} = \begin{bmatrix}\mathbf{X} & \mathbf{W}\mathbf{X} & \mathbf{W}^2\mathbf{X}\end{bmatrix}.$$

Wówczas, rzut macierzy $\mathbf{Z} = \begin{bmatrix}\mathbf{X} & \mathbf{W}\mathbf{y}\end{bmatrix}$ na macierz instrumentów $\mathbf{H}$ przyjmuje postać

$$\hat{\mathbf{Z}} = \mathbf{P}_{\mathbf{H}}\mathbf{Z} = \mathbf{P}_{\mathbf{H}}\cdot\begin{bmatrix}\mathbf{X} & \mathbf{W}\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{X} & \mathbf{P}_{\mathbf{H}}\mathbf{W}\mathbf{y}\end{bmatrix},$$

gdzie $\mathbf{P_H} = \mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}$. W rezultacie otrzymujemy estymator

$$\hat{\boldsymbol{\delta}}_{\text{MZI}} = \left(\hat{\mathbf{Z}}^\mathsf{T}\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^\mathsf{T}\mathbf{y},$$
$$\hat{\boldsymbol{\delta}}_{\text{MZI}} = \left[\hat{\boldsymbol{\beta}}^\mathsf{T}_{\text{MZI}} \quad \hat{\rho}_{\text{MZI}}\right]^\mathsf{T},$$

parametru cząstkowego $\boldsymbol{\delta} = \left[\boldsymbol{\beta}^\mathsf{T} \quad \rho\right]^\mathsf{T}$.

Etap 2

Za pomocą $\hat{\boldsymbol{\delta}}_{\text{MZI}}$ z etapu pierwszego możemy wyznaczyć oszacowanie reszt $\mathbf{u}$ modelu regresyjno-autoregresyjnego (2.13), tj.

$$\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\cdot\hat{\boldsymbol{\delta}}_{\text{MZI}} = \boldsymbol{y} - \hat{\rho}_{\text{MZI}}\mathbf{W}\mathbf{y} - \mathbf{X}\cdot\hat{\boldsymbol{\beta}}_{\text{MZI}}.$$

Celem tego etapu jest oszacowanie dodatkowego parametru autoregresji $\mathbf{u}$, czyli λ, z zastosowaniem pewnego estymatora metody momentów. Dokładniej, rozważmy zależność

$$\mathbf{u} = \lambda\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon},$$
$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(0, \sigma^2_{\boldsymbol{\varepsilon}}\mathbf{I}\right).$$

Oczywiście, bezpośrednie zastosowanie wartości oczekiwanej, prowadzi do pustego informacyjnie równania $0 = \mathbb{E}\,\mathbf{u} = \lambda\mathbf{M}\cdot\mathbb{E}\,\mathbf{u} + \mathbb{E}\,\boldsymbol{\varepsilon} = \lambda\cdot 0 + 0$, gdyż $\hat{\mathbf{Z}}$, tak jak $\mathbf{Z}$ i $\mathbf{X}$, zawiera składnik stały. Pomysł Kelejiana i Pruchy polega na wykorzystaniu drugich momentów układu

$$\begin{aligned}\boldsymbol{\varepsilon} &= \mathbf{u} - \lambda\mathbf{M}\mathbf{u}\\ \mathbf{M}\boldsymbol{\varepsilon} &= \mathbf{M}\mathbf{u} - \lambda\mathbf{M}^2\mathbf{u}.\end{aligned} \tag{2.14}$$

Stosując wektorowy wzór skróconego mnożenia, z pierwszego równania uzyskujemy

$$\boldsymbol{\varepsilon}^\mathsf{T}\boldsymbol{\varepsilon} = \mathbf{u}^\mathsf{T}\mathbf{u} - 2\cdot\lambda\mathbf{u}^\mathsf{T}\mathbf{M}\mathbf{u} + \lambda^2\mathbf{u}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}\mathbf{u},$$

z drugiego

$$\boldsymbol{\varepsilon}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}\boldsymbol{\varepsilon} = (\mathbf{M}\mathbf{u})^\mathsf{T}\mathbf{M}\mathbf{u} - 2\cdot\lambda\mathbf{u}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}^2\mathbf{u} + \lambda^2\left(\mathbf{M}^2\mathbf{u}\right)^\mathsf{T}\mathbf{M}^2\mathbf{u},$$

a krzyżowy iloczyn skalarny odpowiadających sobie stron równań układu (2.14) daje

$$\boldsymbol{\varepsilon}^\mathsf{T}\cdot\mathbf{M}\boldsymbol{\varepsilon} = \mathbf{u}^\mathsf{T}\mathbf{M}\mathbf{u} - \lambda\mathbf{u}^\mathsf{T}\mathbf{M}^2\mathbf{u} - \lambda\mathbf{u}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}\mathbf{u} + \lambda^2\mathbf{u}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}^2\mathbf{u}.$$

Z założenia o rozkładzie $\boldsymbol{\varepsilon}$ wynika, że $\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon} = \sigma_{\varepsilon}^2 \operatorname{tr}\boldsymbol{A}$, dla dowolnej macierzy $\boldsymbol{A}$, por. lemat III.1. Przykładając zatem do powyższych równań operator średniej wartości oczekiwanej $\frac{1}{N}\mathbb{E}$ oraz grupując wyrażenia, otrzymujemy układ z niewiadomymi λ, λ^2, σ_{ε}^2. Układ ten może być dalej użyty na potrzeby metody momentów, przy czym wartości zaburzenia $\mathbf{u}$ będą aproksymowane resztami $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\cdot\hat{\boldsymbol{\delta}}_{\mathrm{MZI}}$ z etapu pierwszego. Zatem, uwzględniając dodatkowo założenie o przekątnej macierzy $\mathbf{M}$, tj. $\operatorname{tr}\mathbf{M} = 0$, i oznaczając za Kelejianem i Pruchą (1998)

$$\boldsymbol{G} = \begin{bmatrix} \frac{2}{N}\widetilde{\mathbf{u}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} & -\frac{1}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} & 1 \\ \frac{2}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} & -\frac{1}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} & \operatorname{tr}\left(\mathbf{M}^{\mathsf{T}}\mathbf{M}\right) \\ \frac{1}{N}\widetilde{\mathbf{u}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} + \frac{1}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\mathbf{u}} & -\frac{1}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} & 0 \end{bmatrix}$$

oraz

$$\boldsymbol{g} = \begin{bmatrix} \frac{1}{N}\widetilde{\mathbf{u}}^{\mathsf{T}}\widetilde{\mathbf{u}} \\ \frac{1}{N}\widetilde{\widetilde{\mathbf{u}}}^{\mathsf{T}}\widetilde{\mathbf{u}} \\ \frac{1}{N}\widetilde{\mathbf{u}}^{\mathsf{T}}\widetilde{\widetilde{\mathbf{u}}} \end{bmatrix},$$

gdzie $\widetilde{\widetilde{\mathbf{u}}} \coloneqq \mathbf{M}\cdot\widetilde{\mathbf{u}}$ i $\widetilde{\widetilde{\widetilde{\mathbf{u}}}} \coloneqq \mathbf{M}^2\cdot\widetilde{\mathbf{u}}$ są obrazami przestrzennymi reszt modelu, pozostaje nam rozwiązanie równania

$$\boldsymbol{G}\cdot\begin{bmatrix}\lambda \\ \lambda^2 \\ \sigma_{\varepsilon}^2\end{bmatrix} = \boldsymbol{g}.$$

Ponieważ układ ten jest nadmiernie identyfikowany, estymatory uogólnionej metody momentów uzyskujemy przybliżając strony równania metodą najmniejszych kwadratów. Oznacza to, że

$$\begin{bmatrix}\hat{\lambda}_{\mathrm{UMM}} & \hat{\sigma}^2_{\mathrm{UMM}}\end{bmatrix} = \underset{\lambda,\,\sigma_{\varepsilon}^2}{\arg\min} \left[\boldsymbol{g} - \boldsymbol{G}\begin{bmatrix}\lambda \\ \lambda^2 \\ \sigma_{\varepsilon}^2\end{bmatrix}\right]^{\mathsf{T}} \left[\boldsymbol{g} - \boldsymbol{G}\begin{bmatrix}\lambda \\ \lambda^2 \\ \sigma_{\varepsilon}^2\end{bmatrix}\right].$$

Zauważmy, że wskazany problem optymalizacyjny jest wypukły, a zatem jego rozwiązanie nie nastręcza żadnych trudności.

Jak wynika z twierdzeń Kelejiana i Pruchy (1998), przy pewnych naturalnych założeniach, które przytoczymy na końcu podrozdziału, zarówno estymator $\hat{\lambda}_{\mathrm{UMM}}$, jak i jego prostsza wersja, tj. pierwsza współrzędna macierzy $\boldsymbol{G}^{-1}\boldsymbol{g}$, są zgodne, czyli dla dowolnie małej liczby $\xi > 0$ mamy

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\hat{\lambda}_{\mathrm{UMM}} - \lambda_0\right| < \xi\right) = 0,$$

gdzie λ_0 jest prawdziwą wartością parametru autoregresji składnika losowego $\mathbf{u}$. Pierwszy z tych estymatorów okazuje się jednak bardziej efektywny.

Etap 3

W ostatnim kroku należy dokonać przestrzennej transformacji Cochrana-Orcutta modelu (2.13), tj. przyjąć

$$\begin{aligned}\mathbf{y}^* &= \mathbf{y} - \hat{\lambda}_{\text{UMM}} \cdot \mathbf{W}\mathbf{y},\\ \mathbf{X}^* &= \mathbf{X} - \hat{\lambda}_{\text{UMM}} \cdot \mathbf{W}\mathbf{X},\\ \mathbf{Z}^* &= \mathbf{Z} - \hat{\lambda}_{\text{UMM}} \cdot \mathbf{W}\mathbf{Z},\end{aligned}$$

przy czym $\mathbf{Z}^* = \begin{bmatrix}\mathbf{X}^* & \mathbf{W}\mathbf{y}^*\end{bmatrix}$. Przekształcony model przybiera zatem postać

$$\mathbf{y}^* = \mathbf{Z}^*\delta + \boldsymbol{\varepsilon}.$$

Powyższe równanie estymujemy ponownie, stosując procedurę MZI, a uzyskany estymator nazywamy, za Kelejianem i Pruchą, uogólnionym przestrzennym dwustopniowym estymatorem najmniejszych kwadratów (GS2SLS). Wynik etapu trzeciego w postaci zamkniętej można zapisać jako

$$\hat{\delta}_{\text{GS2SLS}} = \left((\widehat{\mathbf{Z}^*})^{\mathsf{T}}(\widehat{\mathbf{Z}^*})\right)^{-1}(\widehat{\mathbf{Z}^*})^{\mathsf{T}}\mathbf{y}^*,$$

gdzie $\widehat{\mathbf{Z}^*}$ jest rzutem $\mathbf{Z}^*$ na macierz $\mathbf{H}$ (patrz etap pierwszy) o następującej postaci

$$\widehat{\mathbf{Z}^*} = \mathbf{P}_{\mathbf{H}}\mathbf{Z}^* = \begin{bmatrix}\mathbf{X} - \hat{\lambda}_{\text{UMM}} \cdot \mathbf{W}\mathbf{X} & \mathbf{P}_{\mathbf{H}}\mathbf{W}\mathbf{y} - \hat{\lambda}_{\text{UMM}} \cdot \mathbf{P}_{\mathbf{H}}\mathbf{W}^2\mathbf{y}\end{bmatrix}. \quad (2.15)$$

4.2. Własności asymptotyczne estymatora GS2SLS

Za Kelejianem i Pruchą (1998) przyjmijmy następujące założenia.

- Macierze $\mathbf{W}$ oraz $\mathbf{M}$ mają zerowe przekątne i są standaryzowane wierszowo.
- Macierze $\mathbf{W}$, $\mathbf{M}$, jak i $(\mathbf{I} - \rho\mathbf{W})^{-1}$ oraz $(\mathbf{I} - \lambda\mathbf{M})^{-1}$ są jednostajnie bezwzględnie sumowalne, tzn. spełniony jest dla każdej z nich z osobna warunek opisany w założeniu (II.B).
- Macierz zmiennych objaśniających $\mathbf{X}$ jest macierzą pełnego rzędu, a jej elementy są co do modułu jednostajnie ograniczone dla wszystkich $N \in \mathbb{N}$.
- Kolumny macierzy instrumentów $\mathbf{H}$ to pewien podzbiór łącznie liniowo niezależnych kolumn wybrany spośród kolumn macierzy

$$\begin{bmatrix}\mathbf{X} & \mathbf{W}\mathbf{X} & \mathbf{W}^2\mathbf{X} & \ldots & \mathbf{M}\mathbf{X} & \mathbf{M}\mathbf{W}\mathbf{X} & \mathbf{M}\mathbf{W}^2\mathbf{X} & \ldots\end{bmatrix}.$$

- Następujące granice istnieją i są macierzami pełnego rzędu:

$$\begin{aligned}
\mathbf{Q_{HH}} &= \lim_{N\to\infty} \frac{1}{N}\mathbf{H}^\mathsf{T}\mathbf{H},\\
\mathbf{Q_{HZ}} &= \lim_{N\to\infty} \frac{1}{N}\mathbf{H}^\mathsf{T}\mathbf{Z},\\
\mathbf{Q_{HMZ}} &= \lim_{N\to\infty} \frac{1}{N}\mathbf{H}^\mathsf{T}\mathbf{M}\mathbf{Z},\\
\mathbf{\Phi}(\lambda) &= \lim_{N\to\infty} \frac{1}{N}\mathbf{H}^\mathsf{T}\left(\mathbf{I}-\lambda\mathbf{M}\right)^{-1}\left(\mathbf{I}-\lambda\mathbf{M}^\mathsf{T}\right)^{-1}\mathbf{H},
\end{aligned}$$

  oraz macierz $\mathbf{Q_{HZ}} - \lambda\mathbf{Q_{HMZ}}$ jest pełnego rzędu, dla wszystkich wartości $\lambda \in (-1,1)$.

- Przy oznaczeniu

$$\mathbf{G}_* = \begin{bmatrix} \frac{2}{N}\mathbf{u}^\mathsf{T}\bar{\mathbf{u}} & -\frac{1}{N}\bar{\mathbf{u}}^\mathsf{T}\bar{\mathbf{u}} & 1\\ \frac{2}{N}\bar{\mathbf{u}}^\mathsf{T}\bar{\bar{\mathbf{u}}} & -\frac{1}{N}\bar{\bar{\mathbf{u}}}^\mathsf{T}\bar{\bar{\mathbf{u}}} & \operatorname{tr}\left(\mathbf{M}^\mathsf{T}\mathbf{M}\right)\\ \frac{1}{N}\mathbf{u}^\mathsf{T}\bar{\bar{\mathbf{u}}}+\frac{1}{N}\bar{\mathbf{u}}^\mathsf{T}\bar{\mathbf{u}} & -\frac{1}{N}\bar{\mathbf{u}}^\mathsf{T}\bar{\bar{\mathbf{u}}} & 0 \end{bmatrix},$$

  gdzie $\bar{\mathbf{u}} \coloneqq \mathbf{M}\mathbf{u}$ oraz $\bar{\bar{\mathbf{u}}} \coloneqq \mathbf{M}^2\mathbf{u}$, macierz $\mathbf{G}_*^\mathsf{T}\mathbf{G}_*$ jest odwracalna i norma macierzy $\left(\mathbf{G}_*^\mathsf{T}\mathbf{G}_*\right)^{-1}$ jest jednostajnie ograniczona przy $N \in \mathbb{N}$.

Wówczas, jak dowodzą autorzy owej pracy, prawdziwe jest następujące twierdzenie o zachowaniu asymptotycznym.

Twierdzenie II.4

Przy powyższych założeniach, oba estymatory GMM parametru λ: $\hat{\lambda}_{\text{UMM}}$ oraz $\boldsymbol{G}^{-1}\boldsymbol{g}$ są zgodne. Co więcej, zgodne są również związane z nimi estymatory parametru wariancji σ_ε^2. Zgodny i $\sqrt{n}$-asymptotycznie normalny jest także estymator GS2SLS, a dokładniej

$$\sqrt{n}\left(\hat{\delta}_{\text{GS2SLS}} - \delta_0\right) \to \mathcal{N}(0, \mathbf{\Phi}),$$

gdzie δ_0 jest prawdziwą wartością parametru δ, a macierz $\mathbf{\Phi}$ jest granicą według prawdopodobieństwa macierzy $\sigma_\varepsilon^2 \cdot \left(\frac{1}{N}(\widehat{\mathbf{Z}^*})^\mathsf{T}\widehat{\mathbf{Z}^*}\right)^{-1}$, zarówno dla $(\widehat{\mathbf{Z}^*})^\mathsf{T}$ danego przez (2.15) z użyciem $\hat{\lambda}_{\text{UMM}}$, jak i pierwszej współrzędnej $\boldsymbol{G}^{-1}\boldsymbol{g}$. Dla obu estymatorów wariancja składnika losowego z próby uzyskana w etapie trzecim, tj.

$$\frac{1}{N}\left(\mathbf{y}^* - \mathbf{Z}^*\hat{\delta}_{\text{GS2SLS}}\right)^\mathsf{T}\left(\mathbf{y}^* - \mathbf{Z}^*\hat{\delta}_{\text{GS2SLS}}\right)$$

jest zgodnym estymatorem parametru σ_ε^2.

4.3. Estymacja metodą największej wiarogodności

Jak zauważyliśmy wcześniej, zmienna zależna $\mathbf{y}$ ma rozkład gaussowski

$$\mathbf{y} \sim \mathcal{N}\big((\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta},\ \sigma_{\varepsilon}^2\boldsymbol{\Omega}_{\mathbf{y}}(\rho,\lambda)\big),$$

zatem związana z nią funkcja gęstości $\mathbb{R}^N \ni \boldsymbol{y} \mapsto f_{\mathbf{y}}(\boldsymbol{y})$ przyjmuje postać

$$f_{\mathbf{y}}(\boldsymbol{y}) = \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{y} - \mathbb{E}\,\mathbf{y})^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{y}}(\rho,\lambda)^{-1}(\boldsymbol{y} - \mathbb{E}\,\mathbf{y})\right\}}{\sqrt{(2\pi)^N \det\boldsymbol{\Omega}_{\mathbf{y}}(\rho,\lambda)}}, \tag{2.16}$$

gdzie $\mathbb{E}\,\mathbf{y} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$ oraz

$$\boldsymbol{\Omega}_{\mathbf{y}}(\rho,\lambda) = \mathrm{Var}(\mathbf{y}) = \sigma_{\varepsilon}^2(\mathbf{I} - \rho\mathbf{W})^{-1}\big(\mathbf{I} - \lambda\mathbf{M}\big)^{-1}\big(\mathbf{I} - \lambda\mathbf{M}^{\mathsf{T}}\big)^{-1}\big(\mathbf{I} - \rho\mathbf{W}^{\mathsf{T}}\big)^{-1}.$$

A zatem, uwzględniając powyższe w równości (2.16), możemy zbudować funkcję wiarogodności dla parametrów ρ, $\boldsymbol{\beta}$, λ i σ_{ε}^2, przy obserwacji $\boldsymbol{y}$ zmiennej $\mathbf{y}$

$$L_{\boldsymbol{y}}\big(\rho,\boldsymbol{\beta},\lambda,\sigma_{\varepsilon}^2\big) = \frac{\det\boldsymbol{\Theta}(\rho,\lambda)}{\sqrt{(2\pi\sigma_{\varepsilon}^2)^N}}\exp\left\{-\frac{\|\boldsymbol{\Theta}(\rho,\lambda)\boldsymbol{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma_{\varepsilon}^2}\right\},$$

gdzie dla uproszczenia zapisu

$$\boldsymbol{\Gamma}(\lambda) = \mathbf{I} - \lambda\mathbf{M}, \qquad \boldsymbol{\Delta}(\rho) = \mathbf{I} - \rho\mathbf{W},$$

$$\boldsymbol{\Theta}(\rho,\lambda) = \boldsymbol{\Gamma}(\lambda)\boldsymbol{\Delta}(\rho).$$

Podobnie jak w przypadku estymacji MNW modeli przestrzennych z pojedynczą autoregresją, logarytmując obustronnie i podstawiając $\boldsymbol{y} = \mathbf{y}$ otrzymujemy funkcję log-wiarogodności postaci

$$\begin{aligned}\ln L_{\mathbf{y}}\big(\rho,\boldsymbol{\beta},\lambda,\sigma_{\varepsilon}^2\big) = &-\frac{N}{2}\ln\left(2\pi\sigma_{\varepsilon}^2\right) + \ln\det\boldsymbol{\Theta}(\rho,\lambda)\\ &-\frac{1}{2\sigma_{\varepsilon}^2}\left\|\boldsymbol{\Theta}(\rho,\lambda)\boldsymbol{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta}\right\|^2.\end{aligned} \tag{2.17}$$

Konstrukcja estymatora MNW wymaga identyfikacji, dla każdego ustalonego zestawu wartości obserwowanej zmiennej $\mathbf{y}$ z osobna, takich wartości parametrów ρ, $\boldsymbol{\beta}$, λ i σ_{ε}^2, dla których $\ln L_{\mathbf{y}}(\rho,\boldsymbol{\beta},\lambda,\sigma_{\varepsilon}^2)$ przyjmuje wartość możliwie najmniejszą. Aby wyrugować zmienne $\boldsymbol{\beta}$ i σ_{ε}^2 z optymalizowanej funkcji celu,

wyprowadzamy następujące pochodne cząstkowe

$$\begin{aligned}\frac{\partial \ln L_{\mathbf{y}}}{\partial \boldsymbol{\beta}} &= \frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2} \cdot \frac{\partial}{\partial \boldsymbol{\beta}} \left\| \boldsymbol{\Theta}(\rho,\lambda)\widetilde{\boldsymbol{y}} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta} \right\|^2 \\ &= \frac{-1}{\sigma_{\boldsymbol{\varepsilon}}^2}\big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)^{-1}\boldsymbol{\Delta}(\rho)\mathbf{y} - \mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_{\mathbf{u}}(\lambda)^{-1}\mathbf{X}\boldsymbol{\beta}\big), \\ \frac{\partial \ln L_{\mathbf{y}}}{\partial \sigma_{\boldsymbol{\varepsilon}}^2} &= -\frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^4} \left\| \boldsymbol{\Theta}(\rho,\lambda)\boldsymbol{y} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta} \right\|^2 - \frac{N}{2\sigma_{\boldsymbol{\varepsilon}}^2}.\end{aligned}$$

Przyrównując powyższe formuły jednocześnie do zera, otrzymujemy żądane zależności optymalnych wartości $\boldsymbol{\beta}$ i σ^2 od ρ i λ

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\rho,\lambda) &= \big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_u(\lambda)^{-1}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Omega}_u(\lambda)^{-1}\boldsymbol{\Delta}(\rho)\mathbf{y}, \\ \hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\rho,\lambda) &= \frac{1}{N}\left\| \boldsymbol{\Theta}(\rho,\lambda)\widetilde{\boldsymbol{y}} - \boldsymbol{\Gamma}(\lambda)\mathbf{X}\boldsymbol{\beta} \right\|^2.\end{aligned}$$

Uwzględniając te zależności w oryginalnej funkcji celu (2.17), uzyskujemy funkcję zależną od dwóch parametrów: ρ i λ.

Dla modelu SARAR$(1,1)$, podobnie jak w przypadku modelu SAR, można wykazać, że zredukowana funkcja log-wiarogodności

$$(\rho,\lambda) \mapsto \ln L_{\mathbf{y}}\big(\rho, \hat{\beta}(\rho,\lambda), \lambda, \hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\rho,\lambda)\big)$$

osiąga swoje maksimum po zbiorze argumentów $(\rho,\lambda) \in (-1,1)\times(-1,1)$ prawie pewnie, a współrzędne argumentu maksymalizującego prowadzą do estymatorów największej wiarogodności parametrów autoregresyjnych, mianowicie

$$\big(\hat{\rho}_{\mathrm{MNW}}, \hat{\lambda}_{\mathrm{MNW}}\big) \coloneqq \underset{-1<\rho,\lambda<1}{\arg\max} \ln L_{\mathbf{y}}\big(\rho, \hat{\boldsymbol{\beta}}(\rho,\lambda), \lambda, \hat{\sigma}^2(\rho,\lambda)\big).$$

Oszacowania pozostałych parametrów uzyskujemy z wcześniejszych warunków różniczkowych pierwszego rzędu, czyli

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{MNW}} &= \hat{\boldsymbol{\beta}}\big(\hat{\rho}_{\mathrm{MNW}}, \hat{\lambda}_{\mathrm{MNW}}\big), \\ \hat{\sigma}^2_{\mathrm{MNW}} &= \hat{\sigma}^2\big(\hat{\rho}_{\mathrm{MNW}}, \hat{\lambda}_{\mathrm{MNW}}\big).\end{aligned}$$

Powyższe równania można przekształcić za pomocą przestrzennego odpowiednika klasycznej transformacji Cochrana-Orcutta, znanej z analizy szeregów czasowych. Ustalając notację

$$\begin{aligned}\widetilde{\mathbf{y}} &= \mathbf{y} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{M}\mathbf{y}, \\ \widetilde{\mathbf{X}} &= \mathbf{X} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{M}\mathbf{X}, \\ \widetilde{\Delta\mathbf{y}} &= \boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\mathbf{y} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{M}\boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\mathbf{y},\end{aligned}$$

możemy zapisać

$$\hat{\boldsymbol{\beta}}_{\mathrm{MNW}} = \left(\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\Delta\mathbf{y}},$$
$$\hat{\sigma}^2_{\mathrm{MNW}} = \frac{1}{N}\left(\widetilde{\Delta\mathbf{y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_{\mathrm{MNW}}\right)^{\mathsf{T}}\left(\widetilde{\Delta\mathbf{y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_{\mathrm{MNW}}\right).$$

W szczególnym przypadku, gdy $\mathbf{M} = \mathbf{W}$, mamy również

$$\begin{aligned}\widetilde{\Delta\mathbf{y}} &= \boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\mathbf{y} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{W}\boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\mathbf{y}\\ &= \mathbf{y} - \hat{\rho}_{\mathrm{MNW}}\mathbf{W}\mathbf{y} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{W}\mathbf{y} + \hat{\lambda}_{\mathrm{MNW}}\hat{\rho}_{\mathrm{MNW}}\mathbf{W}^2\mathbf{y}\\ &= \mathbf{y} - \hat{\lambda}_{\mathrm{MNW}}\mathbf{W}\mathbf{y} - \left(\hat{\rho}_{\mathrm{MNW}}\mathbf{W}\mathbf{y} - \hat{\rho}_{\mathrm{MNW}}\hat{\lambda}_{\mathrm{MNW}}\mathbf{W}^2\mathbf{y}\right)\\ &= \widetilde{\mathbf{y}} - \hat{\rho}_{\mathrm{MNW}}\mathbf{W}\widetilde{\mathbf{y}} = \boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\cdot\widetilde{\mathbf{y}},\end{aligned}$$

a zatem

$$\hat{\boldsymbol{\beta}}_{\mathrm{MNW}} = \left(\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^{\mathsf{T}}\boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\widetilde{\mathbf{y}},$$
$$\hat{\sigma}^2_{\mathrm{MNW}} = \frac{1}{N}\left\|\boldsymbol{\Delta}(\hat{\rho}_{\mathrm{MNW}})\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_{\mathrm{MNW}}\right\|^2.$$

Rozdział III

Testy statystyczne regresji przestrzennej

Wstęp

W tym rozdziale zostaną przedstawione wybrane testy statystyczne ekonometrii przestrzennej. Omówimy własności statystyki I Morana oraz tzw. testu mnożników Lagrange'a. W szczególności zaprezentujemy dokładne formuły określające

momenty rozkładów tych statystyk dla skończonej próby. Co więcej, używając naszego centralnego twierdzenia granicznego (twierdzenie V.1), wyprowadzimy ich rozkład asymptotyczny. Co istotne, przeprowadzone rozumowanie będzie stanowić ścisły matematycznie dowód. Wzorując się na pracy Kelejiana i Pruchy (2001), uwzględnimy przypadek, w którym elementy przestrzennej macierzy wag zależą od wielkości próby. Jednak — w przeciwieństwie do wspomnianej pracy — unikniemy ograniczającego założenia o sumowalności jej wierszy i kolumn. Takie podejście umożliwia stosowanie szerszej klasy przestrzennych macierzy wag.

W podrozdziale trzecim omówimy test specyfikacji regresji krzyżowej, oparty na rozkładzie F, czyli ilorazie niezależnych rozkładów χ^2. W podrozdziale czwartym, przedstawiamy procedurę testową tzw. niestacjonarności przestrzennej Kosfelda–Lauridsena, w skorygowanej przez nas formie.

1. Testy oparte na asymptotycznym rozkładzie statystyki Morana

W rozdziale I przytoczyliśmy popularnie stosowaną w badaniach empirycznych postać statystyki I Morana, patrz równanie (1.4) w definicji na s. 31. Naturalnie można jednak oczekiwać, że właściwa postać statystyki testowej oceniającej obecność autokorelacji przestrzennej będzie zależna od rozkładu badanego procesu przestrzennego $\mathbf{y}$. Podobnie, zależny od niego będzie również rozkład samej statystki I, a w szczególności dwa pierwsze momenty (wartość oczekiwana i wariancja) potrzebne do jej normalizacji. W efekcie, postulowana wcześniej zbieżność rozkładów

$$\frac{I - \mathbb{E}\, I}{\sqrt{\operatorname{Var} I}} \simeq \mathcal{N}(0,1),$$

może mieć miejsce tylko wtedy, gdy poczynimy pewne założenia dotyczące specyfikacji modelu dla procesu $\mathbf{y}$, a więc i nałożymy odpowiednie ograniczenia na samą macierz wag przestrzennych $\mathbf{W}$.

Powyższe zagadnienie było rozważane przez wielu teoretyków ekonometrii. Jako pierwsi zajęli się nim Cliff i Ord (1972), którzy na grunt ekonometrii przestrzennej przenieśli rozumowanie Durbina i Watsona (1950, 1951). W swoich rozważaniach jako punkt wyjścia przyjęli oni iloraz norm

$$\frac{\|\mathbf{W}\boldsymbol{\varepsilon}\|^2}{\|\boldsymbol{\varepsilon}\|^2} = \frac{\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon}}$$

i badali rozkład asymptotyczny bardziej ogólnej statystyki postaci ilorazu form kwadratowych

$$Q(\boldsymbol{\varepsilon}, \mathbf{W}) = \frac{\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon}},$$

przy czym występujący w powyższych wzorach element ε oznaczał wektor reszt liniowego modelu ekonometrycznego.

Rozkłady asymptotyczne statystki I Morana i statystyk pochodnych były badane w kolejnych latach również przez Sena (1976), Pinkse'a (1999), Kelejiana i Pruchę (2001), a ostatnio przez Borna i Breitunga (2011). Niemniej jednak w niniejszej monografii przedstawiamy teorię własności asymptotycznych statystyki I Morana we własnym, oryginalnym ujęciu, którego unikatową cechą jest zastosowanie nowego centralnego twierdzenia granicznego. W rezultacie prezentowane przez nas własności są prawdziwe dla większej klasy macierzy wag przestrzennych, niż sugeruje literatura przedmiotu. O zasadności rozszerzania teorii asymptotycznych na przypadek macierzy niesumowalnych można przeczytać również w rozdziale IV.

Dla zaprezentowanych w tym rozdziale rozważań kluczowe będą następujące dwa lematy.

LEMAT III.1

Załóżmy, że $\boldsymbol{A}$ jest macierzą kwadratową o rozmiarze $N \times N$ i dowolnych elementach liczbowych a_{ij}, $1 \leqslant i, j \leqslant N$. Niech $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$ będzie przestrzennym procesem losowym o wielowymiarowym rozkładzie normalnym, z parametrami

$$\begin{aligned} \mathbb{E}\,\boldsymbol{\xi} &= \boldsymbol{\mu} = (\mu_1, \ldots, \mu_N), \\ \operatorname{Var}\boldsymbol{\xi} &= \sigma^2 \mathbf{I}, \end{aligned}$$

dla pewnego $\sigma^2 \geqslant 0$. Wówczas dla formy kwadratowej

$$\mathbb{R}^2 \ni \boldsymbol{x} \mapsto K_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{x} \tag{3.1}$$

mamy

$$\mathbb{E}\, K_{\boldsymbol{A}}(\boldsymbol{\xi}) = K_{\boldsymbol{A}}(\boldsymbol{\mu}) + \sigma^2 \operatorname{tr} \boldsymbol{A}.$$

Dodatkowo, jeśli $\mathbb{E}\,\boldsymbol{\xi} = \boldsymbol{\mu} = 0$, wtedy mamy

$$\operatorname{Var} K_{\boldsymbol{A}}(\boldsymbol{\xi}) = \sigma^4 \operatorname{tr}\left(\boldsymbol{A}^{\mathsf{T}} \boldsymbol{A} + \boldsymbol{A}^2\right).$$

DOWÓD. Przy powyższych oznaczeniach możemy wyliczyć

$$\begin{aligned} \mathbb{E}\, K_{\boldsymbol{A}}(\boldsymbol{\xi}) = \mathbb{E}(\boldsymbol{\xi}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{\xi}) &= \mathbb{E}\left(\sum_{i=1}^{N} \sum_{j=1}^{N} \xi_i a_{ij} \xi_j \right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}(\xi_i a_{ij} \xi_j) \\ &= \sum_{i \neq j} a_{ij} \mathbb{E}\xi_i \xi_j + \sum_{i=1}^{N} a_{ii} \mathbb{E}\xi_i^2 = \sum_{i \neq j} a_{ij} \mu_i \mu_j + \sum_{i=1}^{N} a_{ii}(1 + \mu_i^2) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \mu_i \mu_j + \sum_{i=1}^{N} a_{ii} = K_A(\boldsymbol{\mu}) + \operatorname{tr} \boldsymbol{A}. \end{aligned}$$

Załóżmy, że $\mathbb{E}\,\boldsymbol{\xi} = \boldsymbol{\mu} = 0$. Aby wyprowadzić żądane wyrażenie jako wartość wariancji zmiennej $K_{\boldsymbol{A}}(\boldsymbol{\xi})$, wyliczymy najpierw wartość drugiego momentu. I tak mamy

$$\begin{aligned}
\mathbb{E}\,K_{\boldsymbol{A}}^2(\boldsymbol{\xi}) &= \mathbb{E}\left(\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\xi}\right)^2 = \mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_i a_{ij}\xi_j\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_i a_{ij}\xi_j\right)\left(\sum_{k=1}^{N}\sum_{l=1}^{N}\xi_k a_{kl}\xi_l\right) \\
&= \mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\xi_i a_{ij}\xi_j\xi_k a_{kl}\xi_l \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}a_{ij}a_{kl}\mathbb{E}\xi_i\xi_j\xi_k\xi_l.
\end{aligned}$$

Wynik ostatniej równości można uprościć. Zauważmy, że wartość $\mathbb{E}\,\xi_i\xi_j\xi_k\xi_l$ jest różna od zera tylko wtedy, gdy wartość żadnego z indeksów i, j, k, l nie jest różna od wartości wszystkich pozostałych. Istotnie, gdyby — dla ustalenia uwagi — wartość indeksu i spełniała $i \notin \{j,k,l\}$, wówczas na mocy założeń mielibyśmy $\mathbb{E}\,\xi_i\xi_j\xi_k\xi_l = \mathbb{E}\,\xi_i \cdot \mathbb{E}\,\xi_j\xi_k\xi_l = 0$. Analogicznie mamy $j \in \{i,k,l\}$, $k \in \{i,j,l\}$ oraz $l \in \{i,j,k\}$. W ten sposób w zbiorze wszystkich indeksów $1 \leqslant i,j,k,l \leqslant N$ możemy wyróżnić cztery następujące przypadki:

a) $i \neq j$, $i = k$, $l = j$ — wówczas składniki w powyższej sumie są postaci

$$a_{ij}a_{ij}\,\mathbb{E}\,\xi_i^2\xi_j^2 = a_{ij}^2\,\mathbb{E}\,\xi_i^2\,\mathbb{E}\,\xi_j^2 = a_{ij}^2\sigma^4,$$

b) $i \neq j$, $i = l$, $k = j$ — wówczas składniki w powyższej sumie są postaci

$$a_{ij}a_{ji}\,\mathbb{E}\,\xi_i^2\xi_j^2 = \mathbb{E}\,\xi_i^2\,\mathbb{E}\,\xi_j^2 = a_{ij}a_{ji}\sigma^4,$$

c) $i = j$, $k = l \neq j$ — wówczas składniki w powyższej sumie są postaci

$$a_{ii}a_{jj}\,\mathbb{E}\,\xi_i^2\xi_j^2 = a_{ii}a_{jj}\,\mathbb{E}\,\xi_i^2\,\mathbb{E}\,\xi_j^2 = a_{ii}a_{jj}\sigma^4,$$

d) $i = j = k = l$ — wówczas składniki w powyższej sumie są postaci

$$a_{ii}a_{ii}\,\mathbb{E}\,\xi_i^4 = a_{ii}^2 \cdot 3\sigma^4.$$

Zatem uzyskujemy

$$\begin{aligned}\mathbb{E}\, K_{\boldsymbol{A}}^2(\boldsymbol{\xi}) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} a_{ij}a_{kl}\,\mathbb{E}\,\xi_i\xi_j\xi_k\xi_l \\ &= \sigma^4\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}a_{ij} + \sigma^4\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}a_{ji} + \sigma^4\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ii}a_{jj} + 3\sigma^4\sum_{i=1}^{N} a_{ii} \\ &= \sigma^4 \operatorname{tr}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} + \sigma^4 \operatorname{tr}\boldsymbol{A}^2 + \sigma^4(\operatorname{tr}\boldsymbol{A})^2.\end{aligned}$$

□

Lemat III.2

Niech $\boldsymbol{A}$ będzie dowolną macierzą kwadratową o rozmiarach $N \times N$ i niech $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$ będzie przestrzennym procesem losowym o wielowymiarowym rozkładzie normalnym $\mathcal{N}(0, \sigma^2\mathbf{P})$, gdzie $\mathbf{P}$ jest niezerową macierzą rzutu ortogonalnego. Wówczas dla dowolnej potęgi $p \in \mathbb{R}$ oraz formy kwadratowej $K_{\boldsymbol{A}}$, patrz równanie (3.1), mamy

$$\mathbb{E}\left(\frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})}\right)^p = \frac{\mathbb{E}K_{\boldsymbol{A}}^p(\boldsymbol{\xi})}{\mathbb{E}K_{\mathbf{I}}^p(\boldsymbol{\xi})},$$

gdzie $K_{\mathbf{I}}(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathsf{T}}\mathbf{I}\boldsymbol{\xi} = \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\xi}$.

Dowód. Oczywiście iloraz $\frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})} = \frac{\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\xi}}{\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\xi}}$ jest niezmienniczy ze względu na przeskalowanie argumentu $\boldsymbol{\xi}$, tj. dla dowolnego $c \neq 0$ mamy $\frac{K_{\boldsymbol{A}}(c\cdot\boldsymbol{\xi})}{K_{\mathbf{I}}(c\cdot\boldsymbol{\xi})} = \frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})}$. Można więc zauważyć, że zmienna losowa $\frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})}$ jest niezależna według prawdopodobieństwa od długości wektora $\boldsymbol{\xi}$, a zatem również od zmiennej $K_{\mathbf{I}}(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\xi}$. W rezultacie mamy

$$\mathbb{E}\, K_{\boldsymbol{A}}^p(\boldsymbol{\xi}) = \mathbb{E}\left(\left(\frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})}\right)^p \cdot K_{\mathbf{I}}^p(\boldsymbol{\xi})\right) = \mathbb{E}\left(\frac{K_{\boldsymbol{A}}(\boldsymbol{\xi})}{K_{\mathbf{I}}(\boldsymbol{\xi})}\right)^p \cdot \mathbb{E}\, K_{\mathbf{I}}^p(\boldsymbol{\xi}).$$

□

Lemat III.3

Niech $\xi = \xi(N)$ będzie zmienną losową zbieżną według rozkładu do zmiennej o rozkładzie normalnym $\mathcal{N}(0,1)$ oraz niech $\zeta = \zeta(N)$ będzie zmienną losową zbieżną według prawdopodobieństwa do liczby s (stałej), różnej od zera. Wówczas iloraz $\frac{\xi}{\zeta}$ jest zbieżny według rozkładu do zmiennej losowej o rozkładzie normalnym $\mathcal{N}(0, s^{-1})$.

Dowód powyższego lematu można łatwo przeprowadzić, opierając się na zadaniach 8.1.3 oraz 8.2.6 z książki Jakubowskiego i Sztencla (2001).

W dalszej części rozdziału wyprowadzimy dokładne formuły, określające momenty rozkładu dla małej próby oraz rozkłady asymptotyczne statystyki I Morana. W tym celu będziemy rozważać dwa modele procesu generującego dane SAR oraz SEM.

1.1. Rozkład statystyki Morana dla procesu czystej autoregresji

Rozważmy następujący proces generujący obserwacje $\mathbf{y} = (y_1, \ldots, y_N)$, którego losowość jest opisana przez model oparty na specyfikacji czystej autoregresji przestrzennej SAR

$$\begin{aligned} \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big). \end{aligned}$$

Zauważmy, że przy odpowiednich założeniach dotyczących przestrzeni dopuszczalnych wartości parametru ρ oraz macierzy wag $\mathbf{W}$, powyższe równanie, równoważne jest specyfikacji

$$\begin{aligned} \mathbf{y} &= (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big). \end{aligned}$$

Gdy przemianujemy parametr ρ na λ, ze względu na brak zmiennych objaśniających, uzyskamy specyfikację SEM, czyli

$$\begin{aligned} \mathbf{y} &= \mathbf{u} \\ \mathbf{u} &= (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0, \sigma^2\mathbf{I}\big). \end{aligned}$$

Hipoteza zerowa o braku autokorelacji przestrzennej procesu $\mathbf{y}$ sprowadza się do równości

$$H_0: \quad \rho = \lambda = 0.$$

W takim przypadku odpowiednia postać statystyki I Morana wyraża się wzorem

$$I = \frac{N}{W} \cdot \frac{\mathbf{y}^\mathsf{T}\mathbf{W}\mathbf{y}}{\mathbf{y}^\mathsf{T}\mathbf{y}} = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} y_i y_j}{\sum_{i=1}^n y_i^2},$$

gdzie, dla zwartości zapisu, przyjmujemy oznaczenie

$$W = \sum_{i=1}^N \sum_{j=1}^N w_{ij}. \tag{3.2}$$

Zauważmy, że H_0 jest hipotezą prostą, która jednoznacznie wyznacza rozkład prawdopodobieństwa wartości procesu przestrzennego $\mathbf{y}$. Korzystając z lematu III.2, a następnie lematu III.1, przy założeniu prawdziwości hipotezy zerowej, możemy wyliczyć wartość oczekiwaną i wariancję statystyki I Morana. Zatem

$$\mathbb{E}_0\, I = \mathbb{E}\left(\frac{N}{W}\cdot\frac{\mathbf{y}^\mathsf{T}\mathbf{W}\mathbf{y}}{\mathbf{y}^\mathsf{T}\mathbf{y}}\right) = \frac{N}{W}\cdot\frac{\mathbb{E}\,\mathbf{y}^\mathsf{T}\mathbf{W}\mathbf{y}}{\mathbb{E}\,\mathbf{y}^\mathsf{T}\mathbf{y}} = \frac{N}{W}\cdot\frac{\sigma^2\operatorname{tr}\mathbf{W}}{N\cdot\sigma^2},$$

a więc, przy założeniu o zerowej przekątnej macierzy $\mathbf{W}$, mamy $\mathbb{E}\, I = 0$. Podobnie możemy wyliczyć

$$\begin{aligned}\operatorname{Var}_0(I) = \mathbb{E}_0\, I^2 &= \mathbb{E}\left(\frac{N}{W}\cdot\frac{\mathbf{y}^\mathsf{T}\mathbf{W}\mathbf{y}}{\mathbf{y}^\mathsf{T}\mathbf{y}}\right)^2 = \frac{N^2}{W^2}\cdot\frac{\mathbb{E}\,(\mathbf{y}^\mathsf{T}\mathbf{W}\mathbf{y})^2}{\mathbb{E}\,(\mathbf{y}^\mathsf{T}\mathbf{y})^2}\\ &= \frac{N^2}{W^2}\cdot\frac{\sigma^2\operatorname{tr}\left(\mathbf{W}^\mathsf{T}\mathbf{W}+\mathbf{W}^2\right)}{(2N+N^2)\cdot\sigma^2} = \frac{N}{N+2}\cdot\frac{1}{W^2}\cdot\operatorname{tr}\left(\mathbf{W}^\mathsf{T}\mathbf{W}+\mathbf{W}^2\right).\end{aligned}$$

Aby ustalić rozkład graniczny statystyki I Morana, musimy poczynić pewne dodatkowe założenia dotyczące zachowania asymptotycznego macierzy wag przestrzennych $\mathbf{W}$. Mianowicie załóżmy, że zachodzi zbieżność

$$\lim_{N\to\infty}\frac{\|\mathbf{W}_{\text{sym}}\|}{\|\mathbf{W}_{\text{sym}}\|_\mathsf{F}} = 0,$$

gdzie $\|\mathbf{W}_{\text{sym}}\|$ jest wartością normy operatorowej (spektralnej) macierzy $\mathbf{W}_{\text{sym}} := \frac{\mathbf{W}+\mathbf{W}^\mathsf{T}}{2}$, czyli największą wartością osobliwą. Liczba $\|\mathbf{W}_{\text{sym}}\|_\mathsf{F}$ to tzw. norma Frobeniusa, czyli średnia kwadratowa wszystkich wartości własnych $\mathbf{W}_{\text{sym}}$, znana również jako norma Hilberta-Schimdta. Zauważmy, że powyższy warunek jest dość naturalny. Istotnie, w klasycznych centralnych twierdzeniach granicznych zakłada się, że żaden ze składników rozważanej sumy zmiennych losowych nie jest elementem dominującym. W naszym warunku wymagamy, aby żadna z wartości osobliwych macierzy $\mathbf{W}_{\text{sym}}$ nie dominowała w sumie kwadratów wszystkich tych wartości.

Wyliczone przez nas wartości momentów statystyki I Morana przy prawdziwości hipotezy zerowej są dokładne, jednakże nie dają nam pełnej informacji o kształcie jej rozkładu dla małej próby. Znajomość takiego rozkładu jest oczywiście potrzebna do wyznaczenia poziomów krytycznych dla wartości statystki testowej, które odpowiadają różnym poziomom istotności testu statystycznego. Zatem w praktyce do ich wyznaczenia wykorzystuje się aproksymację rozkładu znormalizowanej statystyki I Morana, tj. statystyki $I_{\text{norm}} = \frac{I-\mathbb{E}_0 I}{\sqrt{\operatorname{Var}_0 I}}$, rozkładem normalnym $\mathcal{N}(0,1)$. Fakt takiej zbieżności dystrybuant musi być jednak formalnie wykazany.

Zauważmy, że twierdzenie V.1 implikuje następującą zbieżność (według rozkładu, oznaczaną dalej symbolem $\xrightarrow{D}$) przy rozmiarze próby N rosnącym do nieskończoności

$$\xi := \frac{\frac{N}{W}\mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{y}}{\sqrt{\frac{1}{W^2}\operatorname{tr}(\mathbf{W}^{\mathsf{T}}\mathbf{W}+\mathbf{W}^2)}} \xrightarrow{D} \mathcal{N}(0,\sigma^2).$$

Jednocześnie klasyczne twierdzenia wielkich liczb implikują, że $\zeta = \frac{1}{N+2}\mathbf{y}^{\mathsf{T}}\mathbf{y}$ zbiega według prawdopodobieństwa do σ^2. Zatem — korzystając z lematu III.3 — wnioskujemy, że

$$I_{\text{norm}} = \frac{I}{\sqrt{\operatorname{Var}_0 I}} = \frac{N+2}{\sqrt{\operatorname{tr}(\mathbf{W}^{\mathsf{T}}\mathbf{W}+\mathbf{W}^2)}}\frac{\mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{y}}{\mathbf{y}^{\mathsf{T}}\mathbf{y}} = \frac{\xi}{\zeta} \xrightarrow{D} \mathcal{N}(0,1).$$

1.2. Rozkład statystyki Morana dla procesu autoregresji ze składową stałą

Rozważmy teraz model mechanizmu generującego obserwacje procesu przestrzennego $\mathbf{y}$, oparty na specyfikacji SAR ze składnikiem stałym

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + c\cdot\mathbf{1} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}\big(0,\sigma^2\mathbf{I}\big),\end{aligned}$$

gdzie $\mathbf{1}$ jest N-elementowym wektorem jedynek, a parametr $c \in \mathbb{R}$ parametrem określającym wyraz wolny modelu. Hipoteza zerowa o braku autokorelacji przestrzennej procesu $\mathbf{y}$ wyraża się równością

$$H_0: \quad \rho = 0.$$

Ponownie H_0 jest hipotezą prostą.

W przypadku obecnej specyfikacji modelu dla procesu $\mathbf{y}$ odpowiednia postać statystyki I Morana powinna również uwzględnić obecność składnika stałego. Ponieważ przy założeniu prawdziwości hipotezy zerowej mamy $\mathbb{E}\,\mathbf{y} = c$, możemy szacować wartość c przez $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. Zatem właściwą w tym wypadku postać statystyki I Morana możemy zapisać wzorem

$$I = \frac{N}{W}\cdot\frac{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}(\mathbf{y}-\bar{\mathbf{y}})},$$

gdzie $\bar{\mathbf{y}} = \bar{y}\cdot\mathbf{1}$, a W dane jest przez (3.2). Zauważmy, że powyższa formuła zgodna jest z definicją (1.4).

Aby wyliczyć parametry rozkładu, wprowadzimy dodatkowe oznaczenia. Zauważmy, że macierz $\mathbf{D} = \frac{1}{N}\mathbf{1}\cdot\mathbf{1}^{\mathsf{T}}$ wyznacza operator średniej, czyli w szczególności $\bar{\mathbf{y}} = \mathbf{D}\mathbf{y}$. Zatem $\mathbf{y} - \bar{\mathbf{y}} = (\mathbf{I} - \mathbf{D})\mathbf{y}$. Analogicznie jak w przypadku poprzedniego modelu, korzystając z lematów III.2 i III.1, możemy wyliczyć wartość oczekiwaną i wariancję statystyki I Morana przy założeniu prawdziwości hipotezy zerowej. I tak, ponieważ $(\mathbf{I} - \mathbf{D})\mathbf{y} = (\mathbf{I} - \mathbf{D})\boldsymbol{\varepsilon}$, uwzględniając (3.2) mamy

$$\begin{aligned}\mathbb{E}_0\, I &= \mathbb{E}\left(\frac{N}{W}\cdot\frac{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}(\mathbf{y}-\bar{\mathbf{y}})}\right) = \frac{N}{W}\cdot\frac{\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\mathbf{D})\boldsymbol{\varepsilon}}{\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}(\mathbf{I}-\mathbf{D})\boldsymbol{\varepsilon}}\\ &= \frac{N}{W}\cdot\frac{\sigma^2\operatorname{tr}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\mathbf{D})}{\sigma^2\operatorname{tr}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}(\mathbf{I}-\mathbf{D})}.\end{aligned}$$

Nietrudno zaobserwować, że macierz $\mathbf{D}$ jest symetryczna ($\mathbf{D}^{\mathsf{T}} = \mathbf{D}$) i idempotentna, gdyż $\mathbf{D}^2\boldsymbol{x} = \bar{\bar{\boldsymbol{x}}} = \bar{\boldsymbol{x}} = \mathbf{D}\boldsymbol{x}$, dla dowolnego $\boldsymbol{x}\in\mathbb{R}^N$, a więc $\mathbf{D}^2 = \mathbf{D}$.

Przy założeniu o zerowej przekątnej macierzy $\mathbf{W}$, mamy

$$\operatorname{tr}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\mathbf{D}) = \operatorname{tr}(\mathbf{W}-\mathbf{W}\mathbf{D}) = -\operatorname{tr}\mathbf{D} = -\sum_{i=1}^{N}\frac{1}{N} = -1.$$

Podobnie uzyskujemy wartość z mianownika

$$\operatorname{tr}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}(\mathbf{I}-\mathbf{D}) = \operatorname{tr}(\mathbf{I}-\mathbf{D}) = \sum_{i=1}^{N}\left(1-\frac{1}{N}\right) = N-1,$$

co daje

$$\mathbb{E}_0\, I = -\frac{N}{N-1}\cdot\frac{1}{W},$$

a w przypadku, gdy macierz $\mathbf{W}$ jest standaryzowana wierszowo (co implikuje równość $W = N$) mamy (por. rozdział I)

$$\mathbb{E}_0\, I = -\frac{1}{N-1}.$$

Ponownie lematy III.2 i III.1 pozwalają wyliczyć żądaną wariancję. Uwzględniając symetryczność i idempotentność macierzy $\mathbf{D}$ możemy zapisać

$$\begin{aligned}\mathbb{E}_0\, I^2 &= \mathbb{E}\left(\frac{N}{W}\cdot\frac{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})^{\mathsf{T}}(\mathbf{y}-\bar{\mathbf{y}})}\right)^2\\ &= \frac{N^2}{W^2}\cdot\frac{\mathbb{E}\left(\mathbf{y}^{\mathsf{T}}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}\mathbf{W}(\mathbf{I}-\mathbf{D})\mathbf{y}\right)^2}{\mathbb{E}\left(\mathbf{y}^{\mathsf{T}}(\mathbf{I}-\mathbf{D})^{\mathsf{T}}(\mathbf{I}-\mathbf{D})\mathbf{y}\right)^2}\\ &= \frac{N^2}{W^2}\cdot\frac{\operatorname{tr}\left((\mathbf{I}-\mathbf{D})\mathbf{W}^{\mathsf{T}}\mathbf{R} + (\mathbf{I}-\mathbf{D})\mathbf{R}^2\right) + 1}{2\cdot\operatorname{tr}(\mathbf{I}-\mathbf{D}) + \left(\operatorname{tr}(\mathbf{I}-\mathbf{D})\right)^2},\end{aligned}$$

gdzie $\mathbf{R} = \mathbf{W}(\mathbf{I} - \mathbf{D})$, a zatem

$$\begin{aligned}\operatorname{Var}_0 I &= \mathbb{E}_0 I^2 - (\mathbb{E}_0 I)^2 \\ &= \frac{N^2}{W^2} \cdot \left(\frac{\operatorname{tr}\left(\mathbf{W}^{\mathsf{T}}(\mathbf{I}-\mathbf{D})\mathbf{R} + \mathbf{R}^2\right)}{N^2-1} + \frac{2N}{(N+1)(N-1)^2} \right).\end{aligned}$$

Gdy macierz $\mathbf{W}$ jest standaryzowana wierszowo, czyli $\mathbf{W} \cdot \mathbf{1} = \mathbf{1}$ oraz $W = N$, wówczas

$$\mathbf{WD} = \mathbf{W} \cdot \frac{1}{N}\mathbf{1} \cdot \mathbf{1}^{\mathsf{T}} = \frac{1}{N}(\mathbf{W} \cdot \mathbf{1}) \cdot \mathbf{1}^{\mathsf{T}} = \frac{1}{N}\mathbf{1} \cdot \mathbf{1}^{\mathsf{T}} = \mathbf{D}.$$

W takim wypadku wyrażenie określające wariancję statystyki I Morana można dalej uprościć do postaci

$$\operatorname{Var}_0 I = \frac{\operatorname{tr}\left(\mathbf{W}^{\mathsf{T}} - \mathbf{W}^{\mathsf{T}}\mathbf{D} + \mathbf{W}\right)\mathbf{W}}{N^2-1} + \frac{1}{(N-1)^2}.$$

Na koniec zauważmy, że przy założeniach

$$\sup_{N \in \mathbb{N}} \|\mathbf{W}\| < \infty$$

oraz

$$\lim_{N \to \infty} \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}} = \infty,$$

korzystając z ogólniejszego twierdzenia III.4, sformułowanego w następnym podrozdziale, możemy wnioskować o zbieżności statystyki normalizowanej

$$I_{\text{norm}} = \frac{I - \mathbb{E}_0 I}{\sqrt{\operatorname{Var}_0 I}} \xrightarrow{N \to \infty} \mathcal{N}(0,1).$$

Warto też zaobserwować, że warunek rozbieżności normy $\|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}}$ do nieskończoności, jest równoważny takiej rozbieżności dla samej macierzy $\mathbf{W} = [w_{ij}]_{1 \leqslant i,j \leqslant N}$, jeśli elementy w_{ij} są nieujemne.

1.3. Rozkład statystyki Morana dla procesów autoregresji w obecności zmiennych objaśniających

Rozważania z poprzedniego podrozdziału uogólnimy teraz na przypadek specyfikacji procesu generującego obserwacje, opartej na modelu SAR ze zmiennymi objaśniającymi, tj.

$$\begin{aligned}&\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I}),\end{aligned} \tag{3.3}$$

gdzie $\mathbf{X}$ jest macierzą o rozmiarze $N \times k$ obserwacji zmiennych objaśniających, a $\boldsymbol{\beta}$ wektorem odpowiadających im parametrów. Jak poprzednio, hipoteza zerowa o braku autokorelacji przestrzennej procesu $\mathbf{y}$ jest hipotezą prostą postaci

$$H_0: \quad \rho = 0.$$

Przyjmijmy następujące konwencjonalne oznaczenia

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}, \qquad \mathbf{M} = \mathbf{I} - \mathbf{P}.$$

Macierze $\mathbf{P}$ i $\mathbf{M}$ reprezentują rzuty ortogonalne, a więc w szczególności są symetryczne i idempotentne. Ponadto przyjmijmy, że $\hat{\boldsymbol{\beta}}$ jest oszacowaniem MNK parametru $\boldsymbol{\beta}$ przy prawdziwości hipotezy zerowej, tj. $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$. Wówczas odpowiednia w przypadku powyższej specyfikacji postać statystyki I Morana, uwzględniająca obecność zmiennych objaśniających, to

$$\begin{aligned} I &= \frac{N}{W} \cdot \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} = \frac{N}{W} \cdot \frac{\mathbf{y}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{W}\mathbf{M}\mathbf{y}}{\mathbf{y}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{M}\mathbf{y}} \\ &= \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})(y_j - \mathbf{x}_i\hat{\boldsymbol{\beta}})}{\sum_{i=1}^N (y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})^2}, \end{aligned}$$

gdzie $\mathbf{x}_i$, $1 \leqslant i \leqslant N$, to wiersze macierzy $\mathbf{X}$, a liczba W zdefiniowana jest równaniem (3.2).

Analogicznie jak w przypadku poprzedniego modelu, korzystając z lematów III.2 i III.1, możemy wyliczyć wartość oczekiwaną i wariancję statystyki I Morana przy założeniu prawdziwości hipotezy zerowej. I tak, ponieważ $\mathbf{M}\mathbf{y} = \mathbf{M}\boldsymbol{\varepsilon}$, mamy

$$\begin{aligned} \mathbb{E}_0\, I = \mathbb{E}\left(\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}\right) &= \frac{N}{W} \cdot \frac{\mathbb{E}\,\mathbf{y}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{W}\mathbf{M}\mathbf{y}}{\mathbb{E}\,\mathbf{y}^\mathsf{T}\mathbf{M}\mathbf{y}} \\ &= \frac{N}{W} \cdot \frac{\sigma^2 \operatorname{tr}\mathbf{M}\mathbf{W}\mathbf{M}}{\sigma^2 \operatorname{tr}\mathbf{M}} \frac{1}{N-k} \cdot \frac{N}{W} \cdot \operatorname{tr}\mathbf{M}\mathbf{W}. \end{aligned}$$

Lematy III.2 i III.1 pozwalają również wyliczyć żądaną wariancję statystyki I Morana. Najpierw uzyskujemy jej drugi moment:

$$\begin{aligned} \mathbb{E}_0\, I^2 = \mathbb{E}\left(\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\mathsf{T}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}\right)^2 &= \frac{N}{W} \cdot \frac{\mathbb{E}\,(\mathbf{y}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{W}\mathbf{M}\mathbf{y})^2}{\mathbb{E}\,(\mathbf{y}^\mathsf{T}\mathbf{M}\mathbf{y})^2} \\ &= \frac{N^2}{W^2} \cdot \frac{\operatorname{tr}\left(\mathbf{M}\mathbf{W}^\mathsf{T}\mathbf{M}\mathbf{W} + (\mathbf{M}\mathbf{W})^2\right) + (\operatorname{tr}\mathbf{M}\mathbf{W})^2}{2 \cdot \operatorname{tr}\mathbf{M} + (\operatorname{tr}\mathbf{M})^2}, \end{aligned}$$

a zatem

$$\begin{aligned}\operatorname{Var}_0 I &= \mathbb{E}_0 I^2 - (\mathbb{E}_0 I)^2 \\ &= \frac{N^2}{W^2}\left(\frac{\operatorname{tr}\left(\mathbf{MW}^{\mathsf{T}}\mathbf{MW} + (\mathbf{MW})^2\right) + (\operatorname{tr}\mathbf{MW})^2}{(N-k)(N-k+2)} + \frac{(\operatorname{tr}\mathbf{MW})^2}{(N-k)^2}\right).\end{aligned}$$

Do wyznaczenia przedziału krytycznego dla procedury testowej wykorzystuje się aproksymację rozkładu znormalizowanej statystyki I Morana, tj. statystyki $I_{\text{norm}} = \frac{I-\mathbb{E}_0 I}{\sqrt{\operatorname{Var}_0 I}}$ rozkładem normalnym $\mathcal{N}(0,1)$. Jak zaznaczyliśmy już w podrozdziale 1.3, fakt takiej zbieżności musi być jednak formalnie wykazany. Poniżej prezentujemy autorskie twierdzenie o takiej zbieżności.

Twierdzenie III.4

Niech $\mathbf{W}$ będzie macierzą wag przestrzennych spełniającą warunki

$$\sup_{N\in\mathbb{N}} \|\mathbf{W}\| < \infty, \qquad \lim_{N\to\infty} \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}} = \infty,$$

gdzie $\mathbf{W}_{\text{sym}} \coloneqq \frac{\mathbf{W}+\mathbf{W}^{\mathsf{T}}}{2}$ oraz niech $\mathbf{y}$ będzie procesem przestrzennym zgodnym ze specyfikacją SAR w równaniu (3.3). Wówczas normalizowana statystyka Morana $I_{\text{norm}} = \frac{I-\mathbb{E}_0 I}{\sqrt{\operatorname{Var}_0 I}}$, gdzie wartości momentów zostały wyznaczone powyżej, zbiega według rozkładu do $\mathcal{N}(0,1)$.

Dowód. Z uwagi na oszacowanie

$$\begin{aligned}\|\mathbf{W}_{\text{sym}}(\mathbf{I}-\mathbf{P})\|_{\mathsf{F}}^2 &= \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}}^2 - 2\operatorname{tr}\left(\mathbf{P}\mathbf{W}_{\text{sym}}^{\mathsf{T}}\mathbf{W}_{\text{sym}}\mathbf{P}\right) + \|\mathbf{W}_{\text{sym}}\mathbf{P}\|_{\mathsf{F}}^2 \\ &= \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}}^2 - \|\mathbf{W}_{\text{sym}}\mathbf{P}\|_{\mathsf{F}}^2 \\ &\geqslant \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}}^2 - \|\mathbf{W}\|^2 \cdot \|\mathbf{P}\|_{\mathsf{F}}^2 \\ &= \|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}}^2 - k\cdot\|\mathbf{W}\|^2,\end{aligned}$$

równość $\mathbf{My} = \mathbf{M}\boldsymbol{\varepsilon}$ oraz lemat III.1 mamy rozbieżność

$$V_0 \coloneqq \operatorname{Var}_0 \frac{1}{W}\mathbf{y}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{WMy} = \frac{2}{W}\cdot\|\mathbf{W}_{\text{sym}}\|_{\mathsf{F}} \xrightarrow{N\to\infty} \infty.$$

Oznaczmy

$$\xi = \frac{\frac{1}{W}\mathbf{y}^{\mathsf{T}}\mathbf{MWMy} - \frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{My}\cdot\mathbb{E}_0 I}{\sqrt{V_0}}.$$

Mamy wówczas

$$\xi = \frac{\frac{1}{W}\mathbf{y}^{\mathsf{T}}\mathbf{MWMy} - \frac{\sigma^2}{W}\operatorname{tr}\mathbf{MW}}{\sqrt{V_0}} + \frac{\frac{1}{W}\left(\sigma^2 - \frac{1}{N-k}\mathbf{y}^{\mathsf{T}}\mathbf{My}\right)}{\sqrt{V_0}},$$

przy czym drugi składnik w powyższej sumie zbiega do zera według prawdopodobieństwa. Istotnie, zgodnie z przyjętymi założeniami mianownik rozbieżny jest do nieskończoności, a z teorii estymacji MNK wynika, że zmienna

$$\zeta := \frac{1}{N-k}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathsf{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

zbiega według prawdopodobieństwa do σ_0^2. Zatem, korzystając z lematu III.3 wnioskujemy, że

$$\frac{I - \mathbb{E}_0\, I}{\sqrt{V_0}} = \frac{\xi}{\zeta} \xrightarrow{N\to\infty} \mathcal{N}\Big(0, \frac{1}{\sigma^2}\Big).$$

Powyższa zbieżność implikuje w szczególności, że

$$\lim_{N\to\infty} \frac{\operatorname{Var}_0 I}{\sqrt{V_0}} = \lim_{N\to\infty} \operatorname{Var}_0 \Big(\frac{I - \mathbb{E}_0\, I}{\sqrt{V_0}}\Big) = \frac{1}{\sigma^2},$$

a zatem

$$I_{\text{norm}} = \frac{I - \mathbb{E}_0\, I}{\sqrt{\operatorname{Var}_0 I}} = \frac{\sqrt{V_0}}{\operatorname{Var}_0 I} \cdot \frac{I - \mathbb{E}_0\, I}{\sqrt{V_0}} \xrightarrow{N\to\infty} \mathcal{N}(0,1).$$

□

Rozważmy teraz model mechanizmu generującego obserwacje procesu przestrzennego $\mathbf{y}$, oparty na specyfikacji SEM ze zmiennymi objaśniającymi

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2\mathbf{I}),\end{aligned}$$

gdzie $\mathbf{X}$ jest macierzą o rozmiarze $N \times k$ obserwacji zmiennych objaśniających, a $\boldsymbol{\beta}$ wektorem odpowiadających im parametrów. Hipotezą zerową jest stwierdzenie o braku autokorelacji przestrzennej procesu $\mathbf{y}$ wyrażone przez równość

$$H_0^{\text{SEM}}\colon \lambda = 0.$$

Tak jak w poprzednim podrozdziale, specyfikacja modelu dla procesu $\mathbf{y}$ w przypadku prawdziwości H_0^{SEM} jest tożsama w przypadku hipotezy H_0 dla omówionej już specyfikacji SAR. Wynika stąd, iż postać statystyki I Morana oraz procedura testowa są takie same w obu przypadkach. Naturalnie jednak, z uwagi na różne hipotezy alternatywne, można spodziewać się różnych funkcji mocy testu.

1.4. Uwagi praktyczne dotyczące statystyki I Morana

Analizując powyższe rozumowania, łatwo zauważyć, że czynnik $\frac{1}{W}$, patrz (3.2), występujący w tradycyjnej definicji statystyki I Morana nie spełnia roli normalizującej. Można nawet powiedzieć, że w pewnym sensie „przeszkadza" przy przekształcaniu wzorów. Istotnie, należałoby się spodziewać, że czynnikiem normalizującym będzie raczej wyrażenie związane z wariancją (a dokładniej odchyleniem standardowym) formy kwadratowej, występującej w liczniku formuły opisującej właściwą statystykę, tj. $\mathbf{y}^{\mathsf{T}}\widetilde{\mathbf{W}}\mathbf{y}$, gdzie $\widetilde{\mathbf{W}}$ jest macierzą w pewien sposób związaną z macierzą wag przestrzennych $\mathbf{W}$ i zależną od specyfikacji procesu przestrzennego. Taka wariancja jest proporcjonalna do wartości śladu $\operatorname{tr}(\widetilde{\mathbf{W}}^{\mathsf{T}} + \widetilde{\mathbf{W}})\widetilde{\mathbf{W}}$, która z kolei w ogólności nie jest równa i nie jest proporcjonalna do sumy $\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}$ (por. Kelejian, Prucha, 2001; Olejnik, 2013). Co ciekawe, w pierwszych pracach dotyczących statystyki I Morana rozważano symetryczne macierze zero-jedynkowe, dla których mamy równość

$$\frac{1}{2}\operatorname{tr}(\mathbf{W}^{\mathsf{T}} + \mathbf{W})\mathbf{W} = \frac{1}{2}\|\mathbf{W}\|_{\mathsf{F}}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2 = W.$$

Jest jednak faktem, że w badaniach empirycznych powszechnie stosuje się postać statystyki I Morana z czynnikiem „normalizującym" $\frac{1}{W}$. To w zasadzie nie jest błędem metodologicznym, jeśli statystyką testową jest postać już znormalizowana pierwiastkiem odpowiedniej wariancji. Chcielibyśmy jednak w tym miejscu z pełną mocą podkreślić, że należy zachować ostrożność przy porównywaniu wartości nieznormalizowanych odchyleniem standardowym wartości statystki I Morana, dla różnych macierzy wag, w szczególności o innym rozmiarze próby N. Tak uzyskane wartości mogą nie być porównywalne. W takim wypadku procedurą gwarantującą metodologiczną poprawność jest porównywanie albo znormalizowanych wartości statystyki tj. I_{norm}, albo też *explicite* wyznaczanych przez nie wartości p (ang. *p-values*).

Kolejną istotną kwestią aplikacyjną, na którą należy zwrócić szczególną uwagę jest stosowalność poszczególnych form statystyki I Morana w przypadku różnych specyfikacji *de facto* stosowanych do modelowania procesu przestrzennego. Niestety, często spotyka się w badaniach empirycznych stosowanie niewłaściwej postaci statystyki I Morana. Na przykład stosuje się wzór

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \cdot \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\,(y_i - \bar{y})\,(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

gdzie $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, czyli postać statystyki właściwą do badania autokorelacji procesu jedynie ze składnikiem stałym, w przypadku, gdy proces $\mathbf{y} =$

$(y_1, \ldots, y_N)$ jest *explicite* procesem regresyjno-autoregresyjnym z nietrywialną macierzą zmiennych objaśniających. Nierzadko też można przeczytać w publikacjach różnych autorów nieprawdziwe stwierdzenie, że $\mathbb{E}\, I = -\frac{1}{N-1}$, niezależnie od założeń dotyczących postaci losowości procesu przestrzennego (por. Suchecki [red.], 2010). Innym, często powtarzanym błędem jest formułowanie stwierdzeń dotyczących asymptotyki statystyki testowej I Morana bez jakichkolwiek założeń odnośnie asymptotycznego zachowania macierzy wag przestrzennych $\mathbf{W}$.

2. Testy oparte na mnożnikach Lagrange'a

W tym podrozdziale opiszemy metodę testowania obecności autokorelacji przestrzennej w obserwowanym przestrzennym procesie stochastycznym, opartą na tzw. mnożnikach Lagrange'a (ang. *LM*, *Lagrange/Lagrangian Multipliers*, patrz Anselin, 1988b). Jest to podejście alternatywne dla testów opartych na statystyce I Morana, ale — jak się często okazuje — asymptotycznie z nim zgodne. Przedstawioną tu procedurę, w dość ogólnej, choć nieprzestrzennej formie, po raz pierwszy zaprezentowano w pracy Silvey'a (1959). Nieco wcześniej Rao (1948) opisał właściwie równoważną procedurę testową, opartą na zachowaniu informanty (pochodnej logarytmu funkcji wiarogodności) dla parametru wyznaczającego hipotezę zerową. Stąd testy omawiane w tym podrozdziale są również nazywane testami informanty Rao (ang. *Rao's Score Tests*, patrz Anselin, 2001).

2.1. Test mnożników Lagrange'a dla procesu czystej autoregresji

Rozważmy następującą specyfikację procesu czystej autoregresji przestrzennej SAR/SEM:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2 \mathbf{I}), \end{aligned}$$

gdzie $\mathbf{y}$ jest modelowanym procesem przestrzennym. Rozważmy także procedurę estymacyjną największej wiarogodności parametru σ^2, stosowaną przy ograniczeniu wynikającym z hipotezy zerowej o braku autokorelacji przestrzennej, tj.

$$H_0 \colon \rho = 0.$$

W wyniku otrzymujemy program optymalizacyjny, który możemy rozwiązywać metodą mnożników Lagrange'a. Zatem, najpierw poszukujemy punktów zerowania się pochodnej funkcji

$$\mathcal{L}(\rho, \sigma^2, \alpha) = \ln L_{\mathbf{y}}(\rho, \sigma^2) - \alpha \cdot g(\rho),$$

gdzie

$$L_{\mathbf{y}}(\rho, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln\det \mathbf{\Delta}(\rho) - \frac{1}{2\sigma^2}(\mathbf{\Delta}(\rho)\cdot\mathbf{y})^{\mathsf{T}}(\mathbf{\Delta}(\rho)\cdot\mathbf{y}),$$

jest funkcją wiarogodności parametrów procesu, $\mathbf{\Delta}(\rho) = \mathbf{I} - \rho\mathbf{W}$, a $g(\rho) = \rho$ jest funkcją nakładanego ograniczenia (tj. $g(\rho) = 0$). Według Silvey'a (1959), statystykę testową dla H_0 można zbudować w oparciu o wartość kwadratu mnożnika $\alpha = \alpha(\mathbf{y})$ rozwiązującego równanie

$$\frac{\partial \mathcal{L}}{\partial(\rho, \sigma^2, \alpha)} = 0,$$

a dokładniej rozważając jego wartości normalizowaną wariancją, czyli

$$\mathrm{LM} = \frac{\alpha^2}{\operatorname{Var}\alpha}.$$

Takie podejście jest zgodne z konwencjonalną interpretacją mnożników Lagrange'a, jako efektu krańcowego ograniczenia (tutaj $\rho = 0$) nałożonego na wartość optymalną funkcji celu (w naszym przypadku funkcji log-wiarogodności).

Metoda poszukiwania ekstremum warunkowego Lagrange'a prowadzi do następującego układu równań

$$\begin{aligned}
\frac{\partial \ln L_{\mathbf{y}}}{\partial \rho} &= \operatorname{tr}\mathbf{W}\mathbf{\Delta}(\rho)^{-1} + \frac{1}{2\sigma^2}\cdot 2\cdot \mathbf{y}^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\mathbf{\Delta}(\rho)\mathbf{y} = \alpha \\
\frac{\partial \ln L_{\mathbf{y}}}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{\sigma^2}(\mathbf{\Delta}(\rho)\cdot\mathbf{y})^{\mathsf{T}}(\mathbf{\Delta}(\rho)\cdot\mathbf{y}) = 0 \\
\frac{\partial(\alpha\rho)}{\partial\alpha} &= \rho = 0,
\end{aligned}$$

z którego wyliczamy formułę

$$\alpha = \frac{N}{\mathbf{y}^{\mathsf{T}}\mathbf{y}}\cdot \mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{y}.$$

Wartości statystyki α bliskie zeru będziemy uważać za wspierające hipotezę zerową, a odległe od zera za wskazujące na jej odrzucenie.

Ostatecznie możemy zauważyć, że statystyka LM równa jest normalizowanej statystyce Morana I_{norm}, a więc (jak wynika z rozważań w poprzednim podrozdziale dotyczących asymptotycznych własności statystyki I) jej wartość zbiega według dystrybuanty do rozkładu χ^2 z jednym stopniem swobody, tj.

$$\mathrm{LM} \xrightarrow{N\to\infty} \chi^2(1).$$

Innym podejściem do uzyskania statystyki testowej jest normalizacja informanty za pomocą funkcji informacji Fishera. Istotnie, niech

$$\mathcal{S}(\rho,\sigma^2)=\frac{\partial \ln L_{\mathbf{y}}}{\partial \rho}=\operatorname{tr}\mathbf{W}\boldsymbol{\Delta}(\rho)^{-1}+\frac{1}{\sigma^2}\cdot\mathbf{y}^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Delta}(\rho)\mathbf{y}$$

będzie funkcją informanty (ang. *score*) oraz niech

$$\mathcal{I}(\rho,\sigma^2)=-\mathbb{E}\left(\frac{\partial^2 \ln L_{\mathbf{y}}}{(\partial\rho)^2}\mid\rho,\sigma^2\right)=\operatorname{tr}\mathbf{W}\boldsymbol{\Delta}(\rho)^{-1}\mathbf{W}\boldsymbol{\Delta}(\rho)^{-1}+\operatorname{tr}\mathbf{W}^{\mathsf{T}}\mathbf{W}.$$

będzie funkcją informacji Fishera dla parametru autoregresyjnego. Wówczas, jak wiadomo (patrz Lehmann, Casella, 1998), dla prawdziwych wartości parametrów mamy $\mathbb{E}\,\mathcal{S}(\rho,\sigma^2)=0$ oraz $\operatorname{Var}\mathcal{S}(\rho,\sigma^2)=\mathcal{I}(\rho,\sigma^2)$. Zatem test informanty Rao przyjmuje postać

$$\mathrm{RS}=\frac{\left(\mathcal{S}(\rho,\sigma^2)\right)^2}{\operatorname{Var}\mathcal{S}(\rho,\sigma^2)}=\frac{(\mathcal{S}(\rho,\sigma^2))^2}{\mathcal{I}(\rho,\sigma^2)},$$

co przy założeniu prawdziwości hipotezy zerowej daje

$$\mathrm{RS}=\frac{(\mathcal{S}(0,\sigma^2))^2}{\mathcal{I}(0,\sigma^2)}=\frac{\left(\frac{1}{\sigma^2}\mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{y}+\operatorname{tr}\mathbf{W}\right)^2}{\operatorname{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{W}+\mathbf{W}^2\right)}.$$

Co ciekawe, wykonując naturalne podstawienie $\sigma^2=\hat{\sigma}^2:=\frac{1}{n}\mathbf{y}^{\mathsf{T}}\mathbf{y}$ oraz uwzględniając równość $\operatorname{tr}\mathbf{W}=0$, uzyskujemy

$$\mathrm{LM}=\frac{(\mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{y})^2}{\left(\frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right)^2\cdot\operatorname{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{W}+\mathbf{W}^2\right)}=\mathrm{RS}.$$

2.2. Testy mnożników Lagrange'a dla procesów o specyfikacjach regresjno-autoregresyjnych SAR oraz SEM

Rozważmy następującą specyfikację procesu regresji z autoregresją składnika losowego SEM

$$\begin{aligned}\mathbf{y}&=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}\\ \mathbf{u}&=\lambda\mathbf{W}\mathbf{u}+\boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon}&\sim\mathcal{N}(0,\sigma^2\mathbf{I}),\end{aligned}$$

gdzie $\mathbf{y}$ jest modelowanym procesem przestrzennym, $\mathbf{X}$ jest macierzą wartości zmiennej objaśniającej, a $\boldsymbol{\beta}$ wektorem odpowiadających tym zmiennym parametrów nachylenia. Rozważmy też problem optymalizacyjny wynikający z procedury MNW dla parametrów $\boldsymbol{\beta}$ i σ^2, przy ograniczeniu wynikającym z hipotezy

zerowej o braku autokorelacji przestrzennej

$$H_0 \colon \lambda = 0.$$

Analogicznie jak w przypadku czystego modelu SEM, poszukujemy zatem punktów zerowania się pochodnej funkcji Lagrange'a

$$\mathcal{L}(\boldsymbol{\beta}, \lambda, \sigma^2, \alpha) = \ln L_{\mathbf{y}}(\boldsymbol{\beta}, \lambda, \sigma^2) - \alpha \cdot g(\alpha),$$

gdzie

$$L_{\mathbf{y}}(\boldsymbol{\beta}, \lambda, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln\det \boldsymbol{\Gamma}(\lambda) - \frac{1}{2\sigma^2}\|\boldsymbol{\Gamma}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2,$$

jest funkcją wiarogodności parametrów procesu, $\boldsymbol{\Gamma}(\lambda) = \mathbf{I} - \rho\mathbf{W}$, a funkcja $\lambda \mapsto g(\lambda)$, poprzez równanie $g(\lambda) = 0$, definiuje nakładane ograniczenie.

Metoda Lagrange'a poszukiwania ekstremum warunkowego prowadzi do następującego układu równań

$$\begin{aligned}
\frac{\partial \ln L_{\mathbf{y}}}{\partial \lambda} &= \operatorname{tr} \mathbf{W}\boldsymbol{\Gamma}(\lambda)^{-1} + \frac{1}{2\sigma^2} \cdot 2 \cdot (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Gamma}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \alpha \\
\frac{\partial \ln L_{\mathbf{y}}}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{\sigma^2}\|\boldsymbol{\Gamma}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2 = 0 \\
\frac{\partial \ln L_{\mathbf{y}}}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\lambda)^{\mathsf{T}}\boldsymbol{\Gamma}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \\
\frac{\partial(\alpha\lambda)}{\partial \alpha} &= \lambda = 0,
\end{aligned}$$

z czego uzyskujemy

$$\alpha(\mathbf{y}) = (\frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{y})^{-1}\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{y},$$

gdzie $\mathbf{M} = \mathbf{I} - \mathbf{X}^{\mathsf{T}}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$. Przeprowadzając odpowiednie wyliczenia można wykazać, iż $\operatorname{Var}\alpha$ jest asymptotycznie równoważna wartości $\operatorname{Var}\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{y}$. Stąd, przyjmuje się następującą definicję statystyki testowej

$$\mathrm{LM}_{\mathrm{SEM}} = \frac{(\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{y})^2}{\left(\frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{y}\right)^2 \cdot \operatorname{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{M}\mathbf{W}\mathbf{M} + (\mathbf{W}\mathbf{M})^2\right)}.$$

Jak wynika z rozważań dotyczących asymptotycznych własności statystyki I Morana (patrz twierdzenie III.4), statystyka testu mnożników Lagrange'a zbiega według dystrybuanty do rozkładu $\chi^2(1)$.

Podobnie rozważać można specyfikację SAR dla procesu $\mathbf{y}$, patrz specyfikacja (3.5). W takim przypadku analogiczne rozumowanie prowadzi do statystyki

$$\mathrm{LM}_{\mathrm{SAR}} = \frac{\left(\frac{1}{N}\mathbf{y}^{\mathsf{T}}\mathbf{M}\mathbf{y}\right)^{-2} \cdot (\mathbf{y}^{\mathsf{T}}\mathbf{W}\mathbf{M}\mathbf{y})^2}{\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{M}\mathbf{W}\mathbf{M} + (\mathbf{W}\mathbf{M})^2\right) + \|\mathbf{M}\mathbf{W}(\mathbf{I}-\mathbf{M})\mathbf{y}\|^2},$$

zbieżnej według dystrybuanty do rozkładu $\chi^2(1)$.

3. Test F dla modelu z krzyżowymi zależnościami przestrzennymi zmiennych objaśniających

Najprostszą formą zależności przestrzennych w modelu ekonometrycznym są przestrzenne zależności krzyżowe zmiennych objaśniających. Taki model ma wówczas postać (patrz również specyfikacja SADL w równaniu (2.3))

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2\mathbf{I}).\end{aligned}$$

Występujący powyższej element $\mathbf{WX}$ stanowi przestrzenną, średnią, ważoną wartości poszczególnych zmiennych objaśniających w sąsiednich lokalizacjach, a $\boldsymbol{\gamma}$ to wektor odpowiadających im parametrów. Ponadto przyjmijmy, że k jest liczbą kolumn w macierzy $\mathbf{X}$. Hipoteza zerowa o braku przestrzennych zależności krzyżowych zmiennych objaśniających to hipoteza prosta

$$H_0 \colon \boldsymbol{\gamma} = 0,$$

przy złożonej hipotezie alternatywnej

$$H_1 \colon \boldsymbol{\gamma} \neq 0.$$

Okazuje się, że obecność takiej formy zależności może być testowana, analogicznie jak dla modeli klasycznych, za pomocą testu F. Istotnie, przyjmijmy statystykę testową postaci

$$F = \frac{N-2k}{k} \cdot \frac{\hat{\varepsilon}_{H_0}^{\mathsf{T}}\hat{\varepsilon}_{H_0} - \hat{\varepsilon}_{H_1}^{\mathsf{T}}\hat{\varepsilon}_{H_1}}{\hat{\varepsilon}_{H_1}^{T}\hat{\varepsilon}_{H_1}},$$

gdzie $\hat{\varepsilon}_{H_0}$ jest wektorem reszt uzyskanym z estymacji modelu metodą MNK bez składnika $\mathbf{WX}\boldsymbol{\gamma}$, a $\hat{\varepsilon}_{H_1}$ jest wektorem reszt z estymacji tego modelu uwzględniającego składnik $\mathbf{WX}\boldsymbol{\gamma}$. Gdy hipoteza zerowa jest prawdziwa, wartość $\hat{\varepsilon}_{H_1}^{\mathsf{T}}\hat{\varepsilon}_{H_1}$ nie powinna być znacznie mniejsza niż $\hat{\varepsilon}_{H_0}^{\mathsf{T}}\hat{\varepsilon}_{H_0}$, relatywnie do wartości σ^2.

Przy oznaczeniu $\mathbf{M_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}$ oraz $\mathbf{M_Z} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^\mathsf{T}\mathbf{Z})^{-1}\mathbf{Z}^\mathsf{T}$, $\mathbf{Z} = \begin{bmatrix}\mathbf{X} & \mathbf{WX}\end{bmatrix}$ mamy

$$F = \frac{N-2k}{k} \cdot \frac{\mathbf{y}^T\mathbf{M_X}\mathbf{y} - \mathbf{y}^\mathsf{T}\mathbf{M_Z}\mathbf{y}}{\mathbf{y}^\mathsf{T}\mathbf{M_Z}\mathbf{y}} = \frac{\frac{1}{k}\boldsymbol{\varepsilon}^\mathsf{T}(\mathbf{M_X} - \mathbf{M_Z})\boldsymbol{\varepsilon}}{\frac{1}{N-2k}\boldsymbol{\varepsilon}^\mathsf{T}\mathbf{M_Z}\boldsymbol{\varepsilon}} \frac{\frac{1}{k}\|(\mathbf{M_X} - \mathbf{M_Z})\boldsymbol{\varepsilon}\|^2}{\frac{1}{N-2k}\|\mathbf{M_Z}\boldsymbol{\varepsilon}\|^2},$$

przy czym $\mathbf{M_X} - \mathbf{M_Z}$ jest macierzą rzutu ortogonalnego, gdyż zachodzi macierzowa nierówność $\mathbf{M_Z} \leqslant \mathbf{M_X}$, a więc $\mathbf{M_Z}\mathbf{M_X} = \mathbf{M_X}\mathbf{M_Z} = \mathbf{M_Z}$. Licznik i mianownik mają rozkłady chi-kwadrat z, odpowiednio, k i $N - 2k$ stopniami swobody. Ponadto mamy

$$\mathbb{E}\,(\mathbf{M_Z}\boldsymbol{\varepsilon})^\mathsf{T}(\mathbf{M_X} - \mathbf{M_Z})\boldsymbol{\varepsilon} = \operatorname{tr}\mathbf{M_Z}(\mathbf{M_X} - \mathbf{M_Z}) = 0,$$

zatem licznik i mianownik są niezależne. Ostatecznie, statystyka F ma rozkład Fishera-Snedecora z $(k, 2N - k)$ stopniami swobody.

4. Testowanie niestacjonarności przestrzennej

Problem niestacjonarności w przypadku procesów przestrzennych jako pierwszy rozważał Fingleton (1999). W swojej pracy zauważył, że niestacjonarność przestrzenna, analogicznie jak w przypadku szeregów czasowych, może prowadzić do wystąpienia regresji pozornej (por. Olejnik, 2008, 2013). W literaturze proponowane były różne procedury testowe, umożliwiające wykluczenie występowania problemu niestacjonarności przestrzennej. Podejścia te z reguły stanowiły analogi do testów znanych z ekonometrii klasycznej (tj. nieprzestrzennej). Na przykład Lauridsen (1999) zaproponował procedurę analogiczną do testu Dickey–Fullera, a późniejsza praca Kosfelda i Lauridsena (2004) jest poświęcona podejściu opartemu na teście Walda. My jednak przyjrzymy się w tym podrozdziale procedurze zaczerpniętej z pracy Kosfelda i Lauridsena (2006). Na jej przykładzie pokażemy, na czym polega problem poprawnej specyfikacji przestrzennej niestacjonarności. Ostatecznie prezentujemy tutaj oryginalną, poprawioną przez nas wersję owej procedury.

Jako punkt wyjścia przyjmijmy model SEM postaci

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2\mathbf{I}), \end{aligned} \tag{3.4}$$

gdzie ρ, $\boldsymbol{\beta}$ i σ^2 są nieznanymi parametrami. Jak wynika z rozważań w poprzednim rozdziale, przy założeniu zerowej przekątnej i standaryzacji wierszowej macierzy wag przestrzennych $\mathbf{W}$, naturalną przestrzenią dopuszczalnych wartości

parametru autoregresji składnika losowego ρ jest przedział $(-1, 1)$. Niestety, gdy wartość parametru ρ jest bliska jedności, wówczas oszacowania parametrów modelu przestają być stabilne. Przypadek graniczny, gdy $\rho = 1$, jest nazywany przestrzennym pierwiastkiem jednostkowym. Ponadto, przez analogię do szeregów czasowych, proces autoregresji, dla którego parametr autoregresyjny implikuje osobliwość macierzy $\mathbf{I} - \rho\mathbf{W}$ jest nazywany — zgodnie z powszechnie przyjętą w literaturze konwencją — procesem niestacjonarnym. Podobnie można zdefiniować niestacjonarność przestrzenną dla modelu o specyfikacji z autoregresją zmiennej objaśnianej SAR, inaczej — SARAR$(1, 0)$.

Zauważmy, że tak zdefiniowana niestacjonarność przestrzenna nie jest równoważna brakowi stacjonarności (rozumianej standardowo jako niezmienniczość charakterystyk rozkładów skończenie wymiarowych ze względu na przesunięcia) procesu przestrzennego, ani słabej, ani mocnej. Należy zauważyć, że niestacjonarność przestrzenna jest zjawiskiem określonym w kontekście specyficznego modelu dla obserwowanego procesu przestrzennego.

Zaproponowana przez Kosfelda i Lauridsena (2006) dwustopniowa procedura testowa jest oparta na teście mnożników Lagrange'a, który jest stosowany na obu jej etapach. Najpierw testuje się hipotezę o braku korelacji przestrzennej, tj. hipotezę

$$H_0 \colon \rho = 0.$$

Gdy zostanie ona odrzucona, a więc zostanie stwierdzona obecność autokorelacji przestrzennej, wówczas Kosfeld i Lauridsen proponują ponowne wykonanie testu mnożników Lagrange'a dla modelu przekształconego operatorem $\boldsymbol{\Delta} = \mathbf{I} - \mathbf{W}$, a więc

$$\boldsymbol{\Delta}\mathbf{y} = \boldsymbol{\Delta}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{u} = \boldsymbol{\Delta}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Autorzy zdają się jednak nie zauważać, że w przypadku granicznym specyfikacja (3.4) jest w zasadzie wewnętrznie sprzeczna. Istotnie, gdy $\rho = 1$, zmienna $\boldsymbol{\varepsilon} = \boldsymbol{\Delta}\mathbf{u} = (\mathbf{I} - \mathbf{W})\mathbf{u}$ nie może mieć rozkładu normalnego $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ przy osobliwej macierzy $\boldsymbol{\Delta}$.

Alternatywnie, w drugim etapie, możemy zastosować podejście Kosfelda–Lauridsena–Olejnika, w którym hipoteza zerowa

$$H_0' \colon \rho = 1.$$

prowadzi do zmodyfikowanej specyfikacji

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2\mathbf{P}), \end{aligned} \tag{3.5}$$

gdzie $\mathbf{P}$ jest macierzą rzutu ortogonalnego na przestrzeń będącą zbiorem wartości operatora liniowego $\boldsymbol{\Delta}$, czyli

$$\mathbf{P} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^{\mathsf{T}}\boldsymbol{\Delta})^{+}\boldsymbol{\Delta}^{\mathsf{T}} = \boldsymbol{\Delta} \cdot \boldsymbol{\Delta}^{+}.$$

Przez $\boldsymbol{A}^{+}$, dla dowolnej macierzy $\boldsymbol{A}$, oznaczamy jej pseudoodwrotność Moore'a-Penrose'a. Jeśli dalej w konstrukcji testu Lagrange'a uwzględni się fakt, że macierz wariacji składnika losowego ε nie jest pełnego rzędu, otrzymuje się właściwy test fazy drugiej.

Analogiczna procedura może zostać zastosowana w celu badania przestrzennej niestacjonarności każdej ze stochastycznych zmiennych rozważanego modelu. Wtedy należy wykonać regresję danej zmiennej na czynnik stały (patrz Olejnik, 2013; Kosfeld, Lauridsen, 2006). Praca Olejnik (2008) podaje przykład badania empirycznego, w którym rozważana jest przestrzenna stacjonarność procesu konwergencji w Unii Europejskiej.

Rozdział IV

Zgodność oszacowań estymatorów QNW dla modeli przestrzennych

Wstęp

Procedury estymacji modeli przestrzennych, oparte na metodzie największej wiarogodności, były opisywane i stosowane w praktyce już od początków ekonometrii przestrzennej (patrz Anselin, 1988a). Jednak, jak zauważa Anselin w rozdziale 5.2 cytowanej publikacji, elementy próby przestrzennej z samej definicji nie są

niezależne, a więc nie mogą być w takim wypadku stosowane klasyczne twierdzenia i teorie dotyczące zachowania asymptotycznego estymatorów MNW. Wynika stąd potrzeba odrębnego formalnego dowodu twierdzeń opisujących właściwości takich estymatorów dla parametrów modeli przestrzennych. Rozumowania te pojawiły się dopiero ponad dekadę później, niemniej jednak, odpowiednia teoria była oczekiwana przez specjalistów, co częściowo i nieformalnie uzasadniało stosowanie tej metodologii przez praktyków.

Przełomem okazał się artykuł Lee (2004), w którym przedstawiono spójną matematycznie teorię właściwości asymptotycznych oszacowań quasi-największej wiarogodności (QNW) dla parametrów autoregresyjnego modelu przestrzennego. Wyprowadzone tam rozumowanie opiera się na wcześniejszych pracach Kelejiana i Pruchy, gdzie sformułowano centralne twierdzenie graniczne dla form kwadratowych pojawiających się w wyrażeniach opisujących asymptotyczne zachowanie odpowiednich estymatorów.

Praca Lee (2004) zapoczątkowała dynamiczny rozwój teorii estymatorów opartych na metodzie największej wiarogodności, sięgający nawet dalece rozbudowanych specyfikacji modeli autoregresji przestrzennej. Na przykład Lee i Yu (2010) proponują estymator największej wiarogodności dla przestrzennego modelu z indywidualnymi i czasowymi efektami stałymi (ang. *individual and time fixed effects*) dla danych panelowych. Dodatkowo, w swojej pracy sformułowali oni poprawkę usuwającą obciążenie dla estymatora wariancji. Z kolei w artykule Lee i inni (2010) rozważano dalsze rozszerzenie tych teorii na panele niezbalansowane oraz ich zastosowanie do modelowania sieci społecznych (ang. *social networking models*). Rozumowania Lee (2004) były również wykorzystywane do analizy specyfikacji modeli z dynamiczną (ze względu na opóźnienie czasowe) zależnością zmiennej objaśnianej (patrz Yu i inni, 2008). Shi i Lee (2017) opisywali podejście największej wiarogodności do estymacji dynamicznego modelu panelowego z efektami interaktywnymi. Z kolei Qu i Lee (2017) rozważali modele dynamiczne, w których dopuszcza się macierz wag zmienne w czasie.

W ostatnich latach obserwuje się wzrost popularności aplikacji modeli autoregresyjnych wyższych rzędów, a w szczególności specyfikacji SARAR rzędu $(r, 0)$, dla $r > 1$. W efekcie, zyskują one również zainteresowanie teoretyków. Dla przykładu, Gupta i Robinson (2015) rozważają estymator oparty na metodzie największej wiarogodności dla modelu o specyfikacji z rosnącą do nieskończoności (wraz ze wzrostem rozmiaru próby) liczbą parametrów autoregresyjnych $r = r(N)$. Innymi słowy, badają asymptotykę oszacowań modelu SARAR$(r, 0)$, gdzie $\lim_{N\to\infty} r(N) = \infty$. Z kolei praca Li (2017) przedstawia analizę impulsu-odpowiedzi (inaczej charakterystykę impulsową) dynamicznego modelu panelowego z autoregresją wyższego rzędu i efektami stałymi. Badane są również alternatywne metody estymacji. Na uwagę zasługuje również praca Han i inni (2017), gdzie do estymacji stosowane jest podejście bayesowskie, oraz opraco-

wanie Badingera i Eggera (2013), w którym zaproponowano ulepszony wariant estymatora uogólnionej metody momentów.

Istotną cechą teorii opracowanej przez Lee (2004), którą nazywać tu będziemy standardową teorią asymptotyczną, jest założenie nakładane na zachowanie asymptotyczne macierzy wag przestrzennych, wymagające tzw. sumowalności macierzy. Dokładniej, powiemy, że macierz wag przestrzennych $\mathbf{W} = \mathbf{W}_N = [w_{N,ij}]_{i,j\leqslant N}$ jest sumowalna, jeśli

$$\sup_{N\in\mathbb{N}} \left\{ \sum_{k=1}^{N} |w_{N,ij}| + \sum_{k=1}^{N} |w_{N,ij}| : 1 \leqslant i,j \leqslant N \right\} < \infty,$$

czyli gdy zostaje spełniony warunek wyrażony w założeniu II.B w rozdziale II. Jak zauważono (patrz Olejnik, Olejnik (2020), wymaganie to w sposób zauważalny ogranicza stosowalność metod estymacji. Po pierwsze, nie pozwala na modelowanie ekonometryczne zależności przestrzennych o większym niż sumowalny (w sensie macierzy wag) stopniu interakcji między jednostkami. Po drugie, w przypadkach, gdy oryginalna specyfikacja modelu jest dodatkowo przekształcana poprzez pewną transformację liniową, przeprowadzenie rozumowania z użyciem standardowej analizy asymptotycznej wymaga pewności, że użyta transformacja zachowuje sumowalność macierzy wag. Niestety, taka konieczność dodatkowo komplikuje ewentualne rozważania teoretyczne. Dla ilustracji załóżmy przez chwilę, że rozważamy pewną specyfikację modelu przestrzennego typu SAR

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Następnie przekształcamy wyjściową specyfikację, stosując pewną transformację liniową $\mathbb{T}$ (np. zamianę współrzędnych lub filtr), o której dla uproszczenia załóżmy, że jest odwracalna. Wówczas, otrzymujemy specyfikację przekształconą

$$\widetilde{\mathbf{y}} = \boldsymbol{\rho}\widetilde{\mathbf{W}}\widetilde{\boldsymbol{y}} + \widetilde{\mathbf{X}}\boldsymbol{\beta} + \mathbb{T}\boldsymbol{\varepsilon},$$

gdzie $\widetilde{\mathbf{y}} = \mathbb{T}\mathbf{y}$, $\widetilde{\mathbf{X}} = \mathbb{T}\mathbf{X}$ oraz $\widetilde{\mathbf{W}} = \mathbb{T}\mathbf{W}\mathbb{T}^{-1}$. Jeśli nawet oryginalna macierz $\mathbf{W}$ jest sumowalna, to nowa macierz $\widetilde{\mathbf{W}}$ nie musi być sumowalna nawet w prostym przypadku, gdy $\mathbb{T}$ jest izometrią. Jednocześnie, dla normy spektralnej problem nie występuje, gdyż zwyczajnie zachodzi równość $\|\mathbf{W}\| = \|\widetilde{\mathbf{W}}\|$.

W tym rozdziale prezentujemy spójną i matematycznie kompletną teorię estymacji quasi-największej wiarogodności dla modeli przestrzennej autoregresji wyższych rzędów, w ramach której uzyskujemy rozszerzenie zakresu jej stosowalności poprzez zastąpienie warunku sumowalności macierzy przez ograniczenie jej normy spektralnej. Twierdzenia dotyczące specyfikacji autoregresji zmiennej zależnej zostały pierwotnie opublikowane w pracy Olejnik, Olejnik (2020).

Elementem nowatorskim w obecnej monografii jest przeniesienie tej teorii na przypadek specyfikacji z przestrzenną zależnością autoregresyjną składnika losowego. Szczególną uwagę będziemy przykładać do zupełności rozważań matematycznych, prowadzących do kluczowych wyników.

1. Podstawowe definicje

Niech $d \geqslant 1$ będzie dowolną liczbą całkowitą. Będziemy rozważać przestrzenne modele autoregresyjne ustalonego rzędu d, mianowicie model SARAR$(d,0)$ oraz SARAR$(0,d)$. Dla uproszczenia notacji związanej z mnogością macierzy wag występującą w wyprowadzanych formułach, zastosujemy następujące oznaczenia. Niech $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_d$ będą macierzami kwadratowymi tego samego wymiaru oraz $\mathbf{A} = \langle \boldsymbol{A}_1, \ldots, \boldsymbol{A}_d \rangle^{\mathsf{T}}$ będzie (kolumnowym) wektorem złożonym z tychże macierzy. Niech $\boldsymbol{\rho} \in \mathbb{R}^d$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_d)^{\mathsf{T}}$, będzie wektorem liczbowym. Wówczas definiujemy macierz

$$\boldsymbol{\rho}^{\mathsf{T}}\mathbf{A} \coloneqq \sum_{r=1}^{d} \rho_r \boldsymbol{A}_r.$$

Można zauważyć, że dla zdefiniowanego w ten sposób iloczynu zachodzi następujący ciąg nierówności

$$\begin{aligned} \|\boldsymbol{\rho}^{\mathsf{T}}\mathbf{A}\| = \Big\| \sum_{r=1}^{d} \rho_r \boldsymbol{A}_r \Big\| &\leqslant \sum_{r=1}^{d} |\rho_r| \cdot \|\boldsymbol{A}_r\| \leqslant \|\boldsymbol{\rho}\| \cdot \sqrt{\sum_{r=1}^{d} \|\mathbf{A}_r\|} \\ &\leqslant d \cdot \|\boldsymbol{\rho}\| \cdot \max_{r \leqslant d} \|\boldsymbol{A}_r\|, \end{aligned}$$

gdzie symbol $\|\cdot\|$ oznacza tradycyjnie normę l_2 dla wektorów i operatorową normę spektralną dla macierzy, tj. normą spektralną macierzy.

1.1. Specyfikacje niegaussowskie modeli autoregresji przestrzennej

Przyjmijmy teraz, że $\mathbf{W}_1, \ldots, \mathbf{W}_d$ są ustalonymi macierzami. Chociaż będą one występować w roli przestrzennych macierzy wag, celowo, dla zachowania istotnej w dalszej części rozdziału ogólności, nie zakładamy zerowania się ich przekątnych ani standaryzacji wierszowej. Oznaczmy przez $\mathbf{W}$ wektor złożony z tych macierzy, tj. $\mathbf{W} = \langle \mathbf{W}_1, \ldots, \mathbf{W}_d \rangle^{\mathsf{T}}$. Rozważmy model ekonometryczny o specyfikacji typu SARAR$(d,0)$, z niekoniecznie gaussowskim składnikiem losowym.

Stosując taką notację możemy model zapisać w postaci

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \mathbb{E}\,\boldsymbol{\varepsilon} &= 0 \\ \operatorname{Var}(\boldsymbol{\varepsilon}) &= \sigma^2\mathbf{I}, \end{aligned} \tag{4.1}$$

gdzie $\mathbf{X}$ jest macierzą k zmiennych objaśniających, a $\boldsymbol{\rho} \in \mathcal{P} \subset \mathbb{R}^d$, $\boldsymbol{\beta} \in \mathbb{R}^k$ i $\sigma^2 > 0$ są nieznanymi parametrami. Nieznany jest również rozkład wektora zaburzeń modelu $\boldsymbol{\varepsilon}$, który jednak nie podlega estymacji, a traktowany jest raczej jako nieliczbowy parametr poboczny (ang. *non-numerical nuisance parameter*). Podobnie niegaussowski model typu SARAR$(0, d)$ przyjmie postać

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \\ \mathbb{E}\,\boldsymbol{\varepsilon} &= 0 \\ \operatorname{Var}\boldsymbol{\varepsilon} &= \sigma^2\mathbf{I}, \end{aligned} \tag{4.2}$$

gdzie $\boldsymbol{\lambda} \in \mathcal{L} \subset \mathbb{R}^d$ jest wektorem parametrów autoregresji, a elementy $\boldsymbol{\varepsilon}$, $\mathbf{X}$, $\boldsymbol{\beta}$ i σ^2 rozumiane są analogicznie.

1.2. Gaussowskie estymatory quasi-największej wiarogodności

W przypadku modeli o specyfikacji (4.1) oraz (4.2), gdy nieznany jest rozkład prawdopodobieństwa składnika losowego, nie jest również dostępna postać funkcji gęstości rozkładu obserwowanej zmiennej zależnej. Zatem, w odróżnieniu od teorii prezentowanej w rozdziale II, nie możemy bezpośrednio zastosować procedury estymacji największej wiarogodności. Jak się jednak okazuje, przy pewnych technicznych założeniach dotyczących między innymi przestrzennej macierzy wag i momentów rozkładu zaburzenia modelu, nieznana funkcja log-wiarogodności asymptotycznie upodabnia się do funkcji log-wiarogodności dla zaburzenia gaussowskiego. Ostatecznie, jak dowiedziemy w następnych podrozdziałach, ten fakt pozwala na stosowanie estymatora największej wiarogodności, wyznaczonego dla przypadku gaussowskiego składnika losowego, nawet w przypadku dużych odstępstw od normalności rozkładów. Wynikający stąd estymator nazwiemy gaussowskim estymatorem quasi-największej wiarogodności (QNW, ang. *Quasi-Maxmium Likelihood*, QML).

1.2.A. Estymator QNW dla specyfikacji SAR z zaburzeniem niegaussowskim

Analogicznie jak w rozdziale II, dla specyfikacji z równania (4.1), przy założeniu o gaussowskim błędzie, można wyprowadzić postać gęstości zmiennej zależnej

$$f_{\mathbf{y}}(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Omega}_{\mathbf{y}}}} \exp\Big\{ \|\widetilde{\boldsymbol{y}} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \Big\},$$

a następnie uzyskać funkcję log-wiarogodności

$$\begin{aligned} \ln L_{\mathbf{y}}(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln|\det \boldsymbol{\Delta}(\boldsymbol{\rho})| \\ - \frac{1}{2\sigma^2}(\boldsymbol{\Delta}(\boldsymbol{\rho})\cdot\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{\Delta}(\boldsymbol{\rho})\cdot\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \end{aligned} \tag{4.3}$$

gdzie $\boldsymbol{\Delta}(\boldsymbol{\rho}) = \mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}$. W przeciwieństwie do poprzednich rozważań, nie zakładamy dodatniości wyznacznika przekształcenia $\boldsymbol{\Delta}(\boldsymbol{\rho})$. Należy zauważyć, że gdyby przestrzeń dla parametru $\boldsymbol{\rho}$ była niespójna lub nie zawierała zera, dopuszczalna byłaby sytuacja, w której wyznacznik przekształcenia $\boldsymbol{\Delta}(\boldsymbol{\rho})$ uzyskiwałby wartości ujemne. Natomiast — podobnie jak w przypadku estymacji modelu pierwszego rzędu — z różniczkowych warunków koniecznych optymalizacji $\ln L_{\mathbf{y}}(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$ uzyskujemy

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}) &= \big(\mathbf{X}^{\mathsf{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}\big(\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\big)\mathbf{y}, \\ \hat{\sigma}^2(\boldsymbol{\rho}) &= \frac{1}{N}\big\|\mathbf{y} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\rho})\big\|^2. \end{aligned} \tag{4.4}$$

Ostatecznie, wartość estymatora $\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}$ uzyskujemy, znajdując argument maksymalizujący funkcję skoncentrowanej log-wiarogodności

$$\mathcal{P} \ni \boldsymbol{\varrho} \mapsto \ln L_{\mathbf{y}}\big(\boldsymbol{\varrho}, \hat{\boldsymbol{\beta}}(\boldsymbol{\varrho}), \hat{\sigma}^2(\boldsymbol{\varrho})\big) = -\frac{N}{2}\Big(\ln\big(2\pi\cdot\hat{\sigma}^2(\boldsymbol{\varrho})\big) + 1\Big) + \ln\big|\det\boldsymbol{\Delta}(\boldsymbol{\varrho})\big|, \tag{4.5}$$

gdzie $\mathcal{P}$ jest przestrzenią dopuszczalnych wartości wektorowego parametru $\boldsymbol{\rho}$. W obecnym opracowaniu nie będziemy rozważać możliwych metod określania zbioru $\mathcal{P}$, jednak temat ten jest szeroko dyskutowany w literaturze (np. Elhorst i inni (2012); Olejnik i inni, 2020). Definicję estymatora QNW dopełniają równości $\hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}})$ oraz $\hat{\sigma}^2_{\text{SAR\_QNW}} = \hat{\sigma}^2(\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}})$.

1.2.B. Estymator QNW dla specyfikacji SEM z zaburzeniem niegaussowskim

Analogicznie, do specyfikacji z równania (4.2), można wyprowadzić postać gęstości zmiennej zależnej przy założeniu o normalności rozkładu błędu losowego $\boldsymbol{\varepsilon}$, tj.

$$f_{\mathbf{y}}(\widetilde{\mathbf{y}}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Omega}_{\mathbf{y}}}} \exp\frac{1}{2}\big\|\widetilde{\boldsymbol{y}} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}\mathbf{X}\boldsymbol{\beta}\big\|.$$

W dalszej kolejności możemy uzyskać funkcję log-wiarogodności

$$\begin{aligned}\ln L_{\mathbf{y}}\big(\boldsymbol{\beta},\boldsymbol{\lambda},\sigma^2\big) = &-\frac{N}{2}\ln(2\pi\sigma^2)+\ln\left|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda})\right| \\ &-\frac{1}{2\sigma^2}\big(\boldsymbol{\Gamma}(\boldsymbol{\lambda})\cdot(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\big)^{\mathsf{T}}\big(\boldsymbol{\Gamma}(\boldsymbol{\lambda})\cdot(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\big),\end{aligned} \tag{4.6}$$

gdzie $\boldsymbol{\Gamma}(\boldsymbol{\lambda}) = \mathbf{I}-\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}$. Podobnie jak w przypadku estymacji modelu SAR, z koniecznych warunków różniczkowych optymalizacji $\ln L_{\mathbf{y}}(\boldsymbol{\beta},\boldsymbol{\lambda},\sigma^2)$ otrzymujemy

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) &= \big(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}\big)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{y}, \\ \hat{\sigma}^2(\boldsymbol{\lambda}) &= \frac{1}{N}\big(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\big)^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\big(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\big).\end{aligned} \tag{4.7}$$

Ostatecznie, wartość estymatora $\hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}}$ uzyskujemy, odnajdując argument maksymalizujący funkcję skoncentrowanej log-wiarogodności

$$\begin{aligned}\mathcal{L}\ni\boldsymbol{\lambda}\mapsto\ln L_{\mathbf{y}}\big(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}),\boldsymbol{\lambda},\hat{\sigma}^2(\boldsymbol{\lambda})\big) = \\ -\frac{N}{2}\left(\ln\left(2\pi\cdot\hat{\sigma}^2(\boldsymbol{\lambda})\right)+1\right)+\ln\left|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda})\right|,\end{aligned} \tag{4.8}$$

gdzie $\mathcal{L}$ jest przestrzenią dopuszczalnych wartości wektorowego parametrów $\boldsymbol{\lambda}$. Kwestię określenia zbioru $\mathcal{L}$ można rozstrzygać w taki sam sposób, jak problem identyfikacji przestrzeni $\mathcal{P}$. Definicję estymatora QNW dopełniają równości $\hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}})$ oraz $\hat{\sigma}^2_{\text{SEM\_QNW}} = \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}})$.

2. Zgodność estymatorów QNW

Poniżej przedstawiamy ścisłe dowody zgodności estymatorów quasi-największej wiarogodności zdefiniowanych w poprzednim podrozdziale. Nasze rozumowanie wykorzystuje elementy dość ogólnej teorii asymptotycznej, opisanej w monografii Pötschera i Pruchy (1997). Argumentację rozpoczynamy od przytoczenia założeń formalnych koniecznych dla ścisłości wywodu.

2.1. Założenia formalne

Aby przeprowadzić argumentację dowodu formalnego, musimy poczynić następujące założenia.

Założenie IV.A

Normy spektralne przestrzennych macierzy wag modelu są ograniczone, tj.

$$\sup_{N\in\mathbb{N}}\max_{r\leqslant d}\|\mathbf{W}_r\|<\infty.$$

Założenie IV.B$_{\text{SAR}}$
Zbiór $\mathcal{P}$ jest zwartym podzbiorem $\mathbb{R}^d$ i dla każdego $\boldsymbol{\rho} \in \mathcal{P}$ macierz $\Delta(\boldsymbol{\rho}) = \mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}$ jest nieosobliwa. Ponadto, dla każdego $\boldsymbol{\rho} \in \mathcal{P}$ zachodzi

$$\sup_{N\in\mathbb{N}} \left\|(\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W})^{-1}\right\| < \infty.$$

Założenie IV.C
Dla dowolnego $N \in \mathbb{N}$ istnieje liczba $\bar{N} \geqslant N$ i semi-ortogonalna macierz $\mathbf{E}$, tj. taka, że $\mathbf{E}\mathbf{E}^{\mathsf{T}} = \mathbf{I}$, oraz istnieje $\bar{N}$-elementowy wektor losowy $\bar{\boldsymbol{\varepsilon}}$, o własnościach

a) $\mathbb{E}\,\bar{\boldsymbol{\varepsilon}} = 0$ oraz $\operatorname{Var}(\bar{\boldsymbol{\varepsilon}}) = \sigma^2\mathbf{I}$,

b) elementy wektora $\bar{\boldsymbol{\varepsilon}}$ są czwórkami niezależne, a ich czwarte momenty są wspólnie ograniczone,

dla którego wektor zaburzeń modelu $\boldsymbol{\varepsilon}$ spełnia zależność $\boldsymbol{\varepsilon} = \mathbf{E}\cdot\bar{\boldsymbol{\varepsilon}}$.

Założenie IV.D
Macierz $\mathbf{X}$ jest deterministyczną macierzą obserwacji k zmiennych objaśniających. Ponadto, dla $\mathbf{X}_N^2 := \frac{1}{N}\mathbf{X}^{\mathsf{T}}\mathbf{X}$, mamy

a) $\sup_{N\in\mathbb{N}} \left\|\mathbf{X}_N^2\right\| < \infty$,

b) dla każdego $N \in \mathbb{N}$ macierz $\mathbf{X}_N^2$ jest nieosobliwa,

c) $\sup_{N\in\mathbb{N}} \left\|(\mathbf{X}_N^2)^{-1}\right\| < \infty$.

Założenie IV.E$_{\text{SAR}}$
Dla dowolnych dwóch różnych dopuszczalnych wartości $\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2 \in \mathcal{P}$ parametru $\boldsymbol{\rho}$, co najmniej jeden z poniższych warunków jest spełniony:

$$\liminf_{N\to\infty} \frac{\frac{1}{\sqrt{N}}\left\|(\mathbf{I} - \boldsymbol{\varrho}_1^{\mathsf{T}}\mathbf{W})(\mathbf{I} - \boldsymbol{\varrho}_2^{\mathsf{T}}\mathbf{W})^{-1}\right\|_{\mathsf{F}}}{\sqrt[N]{\left|\det(\mathbf{I} - \boldsymbol{\varrho}_1^{\mathsf{T}}\mathbf{W})(\mathbf{I} - \boldsymbol{\varrho}_2^{\mathsf{T}}\mathbf{W})^{-1}\right|}} > 1 \tag{4.9}$$

lub

$$\liminf_{N\to\infty} \left\|\frac{1}{\sqrt{N}}\mathbf{M}_{\mathbf{X}}(\mathbf{I} - \boldsymbol{\varrho}_1^{\mathsf{T}}\mathbf{W})(\mathbf{I} - \boldsymbol{\varrho}_2^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right\| > 0, \tag{4.10}$$

dla każdego $\boldsymbol{\beta} \in \mathbb{R}^k$.

Założenia IV.A i IV.B_{SAR} dotyczą kluczowych warunków ograniczoności, nakładanych na macierz (macierze — w przypadku modeli wyższych rzędów) wag oraz na implikowaną przez nią funkcję operatora opóźnienia przestrzennego

$$\mathcal{P} \ni \boldsymbol{\varrho} \mapsto \boldsymbol{\Delta}(\boldsymbol{\varrho}) = \mathbf{I} - \boldsymbol{\varrho}^{\mathsf{T}}\mathbf{W}.$$

Wymaganie jednostajnej ograniczoności norm macierzy $\mathbf{W}_N$ gwarantuje, że ilość interakcji przestrzennych, a tym samym stopień autozależności procesu generującego próbę, pozwala na zmniejszanie błędu estymatora wraz ze wzrostem N. Jak argumentują Olejnik i Olejnik (2020), użycie w tym celu normy spektralnej macierzy jest w pewnym sensie optymalne. Dokładniej, gdyby norma którejkolwiek z macierzy $\mathbf{W}_r$, $1 \leqslant r \leqslant d$, nie była ograniczona, czyli

$$\limsup_{N\to\infty} \|\mathbf{W}_r\| = \infty,$$

wówczas zbiór możliwych wartości parametru autoregresyjnego ρ_r, który odpowiada tej macierzy, mogłyby być ograniczony do zbioru jednoelementowego $\{0\}$, formalnie wykluczając autoregresję przestrzenną.

Zauważmy, że wymaganie odwracalności operatora opóźnienia przestrzennego $\boldsymbol{\Delta}(\boldsymbol{\rho})$, $\boldsymbol{\rho} \in \mathcal{P}$ jest w istocie dość naturalne. W przeciwnym wypadku, nie można by było uzyskać jednoznacznej postaci jawnej modelu, czyli rozwiązującej jego równanie ze względu na zmienną zależną. Warunek ograniczoności norm macierzy $\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}$, $\boldsymbol{\rho} \in \mathcal{P}$ gwarantuje taką odwracalność również w sensie asymptotycznym. Z kolei postulat zwartości przestrzeni parametrów $\mathcal{P}$ ma charakter czysto techniczny.

Warto zauważyć, że założenie IV.C dotyczące rozkładu składnika losowego $\boldsymbol{\varepsilon}$ nie wymaga, aby jego elementy były gaussowskie, niezależne lub posiadały ten sam rozkład. Zamiast tego postulowana jest postać zaburzenia losowego jako rezultat pewnego ortogonalnego przekształcenia zmiennych losowych, niezależnych jedynie czwórkami. Zauważmy, że w efekcie zakładamy jedynie nieskorelowanie elementów wektora $\boldsymbol{\varepsilon}$, gdyż mamy

$$\begin{aligned} \mathbb{E}\,\boldsymbol{\varepsilon} &= \mathbb{E}\,(\mathbf{E}\bar{\boldsymbol{\varepsilon}}) = \mathbf{E}\cdot\mathbb{E}\,\bar{\boldsymbol{\varepsilon}} = 0, \\ \operatorname{Var}\boldsymbol{\varepsilon} &= \mathbb{E}\,(\mathbf{E}\bar{\boldsymbol{\varepsilon}}\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}) = \sigma^2\mathbf{E}\mathbf{E}^{\mathsf{T}} = \sigma^2\mathbf{I}, \end{aligned} \tag{4.11}$$

zgodnie ze specyfikacją (4.1). Odróżnienie warunku niezależności od braku korelacji zmiennych ε_i, gdzie $1 \leqslant i \leqslant N$, jest istotne ze względu na brak założenia normalności ich rozkładu łącznego. Należy również rozgraniczyć brak założenia równości rozkładów dla elementów składnika losowego od ich heteroskedastyczności. Warto zauważyć, że w naszej teorii nie zakładamy tożsamości rozkładów, niemniej jednak utrzymujemy założenie równości odpowiednich wariancji, co

wynika z równania (4.11). Prostym sposobem poradzenia sobie z heteroskedastycznością zaburzenia losowego $\boldsymbol{\varepsilon}$ mogłoby być zastosowanie przekształcania normalizującego wariancję, czyli sprowadzającego model do postaci z losowością homoskedastyczną. Jednak, żeby uprościć rozumowanie, nie stosujemy takiego podejścia. Alternatywne podejście do problemu heteroskedastyczności zaproponowali Liu i Yang (2015). Ich unikalny pomysł pozwala na uwzględnienie w procedurze MNW heteroskedastyczności składnika losowego o nieznanym profilu wariancji.

Założenie IV.D przyjmuje strukturę wartości obserwacji zmiennych objaśniających, która zapewnia identyfikowalność parametru $\boldsymbol{\beta}$ w kontekście wybranej metody estymacji. Warunek jednostajnej ograniczoności norm macierzy $\mathbf{X}_N^2 = \frac{1}{N}\mathbf{X}^\mathsf{T}\mathbf{X}$ jednocześnie kontroluje wielkość wszystkich obserwacji w macierzy $\mathbf{X}$, w taki sposób, aby żadna z nich nie wywierała dominującego wpływu na oszacowanie parametru nachylenia. Założenie o odwracalności macierzy $\mathbf{X}_N^2$ jest zupełnie naturalne, ze względu na konieczność zapewnienia braku współliniowości między zmiennymi objaśniającymi. Jak zauważono w publikacji Olejnik i Olejnik (2020), warunek „odwracalności asymptotycznej", tj. $\sup_{N\in\mathbb{N}}\|(\mathbf{X}_N^2)^{-1}\| < \infty$, nie jest w istocie różny od klasycznego $\limsup_{N\in\mathbb{N}}\|(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\| = 0$, koniecznego nawet w przypadku zwykłej nieprzestrzennej metody najmniejszych kwadratów. Z kolei warunek ograniczoności norm macierzy $\mathbf{X}_N^2$ zabezpiecza nas przed sytuacją, w której niektóre obserwacje mają dominujący wpływ na wartości oszacowań parametrów nachylenia, co mogłoby wykluczać gaussowski rozkład asymptotyczny. Zauważmy też, że klasyczna analiza asymptotyczna, oparta na warunku sumowalności macierzy $\mathbf{W}$, wymaga wprost ograniczoności elementów macierzy $\mathbf{X}$, a więc jest w tym względzie bardziej restrykcyjna.

W naszych rozumowaniach zakładamy, że obserwacje w macierzy $\mathbf{X}$ mają charakter niedeterministyczny (założenie IV.D), niemniej jednak możliwe są rozszerzenia tej teorii, w których $\mathbf{X}$ ma charakter losowy. Na przykład, można by przyjąć, że założenia dotyczące $\mathbf{X}$ są spełnione na pewnym zbiorze $A = A(N)$ o malejącym do zera prawdopodobieństwie, tj. takim, że $\lim_{N\to\infty}\mathbb{P}(\Omega \setminus A) = 0$. Co więcej, gdy założenia dotyczące składnika losowego modelu $\boldsymbol{\varepsilon}$ są spełnione warunkowo względem stochastycznego procesu wektorowego $\mathbf{X}$, wówczas teoria przedstawiona w tym rozdziale prowadzi do zgodności estymatorów quasi-największej wiarogodności. Istotnie, dla dowolnego z omawianych tu estymatorów (oznaczmy go $\hat{\theta}$) i odpowiadającej mu wartości prawdziwej θ_0, twierdzenia IV.10 lub IV.11 implikują

$$\lim_{N\to\infty}\mathbb{P}\left(\|\hat{\theta} - \theta_0\| \mid \mathbf{X}\right) = 0$$

prawie pewnie, dla dowolnej liczby $\delta > 0$. Zatem z twierdzenia o zbieżności

zmajoryzowanej Lebesgue'a mamy

$$\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\theta} - \theta_0\| > \delta\right) = \lim_{N\to\infty} \mathbb{E}\left[\mathbb{P}\left(\|\hat{\theta} - \theta_0\| > \delta \mid \mathbf{X}\right)\right] = 0,$$

czyli estymator $\hat{\theta}$ jest zgodny. Zauważmy, że warunkowanie założeń względem $\mathbf{X}$ pociąga za sobą, m.in. $\mathbb{E}\left[\boldsymbol{\varepsilon} \mid \mathbf{X}\right] = 0$ oraz $\mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}} \mid \mathbf{X}\right] = \sigma_0^2\mathbf{I}$. Zmienna $\mathbf{X}$ nie musi być niezależna (według prawdopodobieństwa) od reszt $\boldsymbol{\varepsilon}$ (por. założenie E2 w pracy Shi, Lee, 2017). Należy jednak pamiętać, że w przypadku takiej teorii wyprowadzenie rozkładu asymptotycznego oszacowań może być możliwe w pełnej ogólności jedynie warunkowo ze względu na wartości zmiennych w macierzy $\mathbf{X}$.

Założenie IV.$\mathrm{E}_{\mathrm{SAR}}$ ma charakter techniczny. Zapewnia ono taką strukturę zależności w macierzy wag przestrzennych, która pozwala na asymptotyczną identyfikację prawdziwej wartości parametru $\boldsymbol{\rho}$. Dokładniej, gwarantuje ono, że dowolne dwie wartości $\boldsymbol{\varrho}_1$, $\boldsymbol{\varrho}_2 \in \mathcal{P}$ parametru $\boldsymbol{\rho}$ implikują wystarczająco rozbieżne ciągi wartości gaussowskiej funkcji wiarogodności. Wtedy, obserwowane dane dostarczają wystarczającą ilość informacji, aby zidentyfikować parametr autoregresyjny. Te informacje mogą pochodzić z samej struktury zależności przestrzennych, przy spełnionym warunku (4.9) albo z zależności zmiennej $\mathbf{y}$ od wartości opóźnień przestrzennych zmiennych $\mathbf{X}$, gdy działającym warunkiem w założeniu IV.$\mathrm{E}_{\mathrm{SAR}}$ jest warunek (4.10). Dokładniej, jeśli zgodnie ze specyfikacją (4.1) zachodzi

$$\mathbf{y} \approx \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta},$$

wówczas

$$\mathbf{y} \approx \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\mathbf{X}\boldsymbol{\beta},$$

a więc, można by powiedzieć, że $\mathbf{W}\Delta(\boldsymbol{\rho}_0)^{-1}\mathbf{X}$ stanowi niejawną zmienną objaśniającą.

Do rozważań dotyczących specyfikacji modelu z autoregresją składnika losowego, zawierającego niesumowalną macierz wag, wprowadzamy również założenia IV.$\mathrm{B}_{\mathrm{SEM}}$ oraz IV.$\mathrm{E}_{\mathrm{SEM}}$, które są odpowiednikami założeń IV.$\mathrm{B}_{\mathrm{SAR}}$ oraz IV.$\mathrm{E}_{\mathrm{SAR}}$. Warto zwrócić uwagę na fakt, iż identyfikowalność problemu estymacji w przypadku modelu SEM jest niezależna od właściwości występujących w specyfikacji zmiennych egzogenicznych.

Założenie IV.$\mathrm{B}_{\mathrm{SEM}}$

Zbiór $\mathcal{L}$ jest zwartym podzbiorem $\mathbb{R}^d$ i dla każdego $\boldsymbol{\lambda} \in \mathcal{L}$ macierz $\Delta(\boldsymbol{\lambda}) = \mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}$ jest nieosobliwa. Ponadto, dla każdego $\boldsymbol{\lambda} \in \mathcal{L}$ zachodzi

$$\sup_{N\in\mathbb{N}} \left\|(\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W})^{-1}\right\| < \infty.$$

Założenie IV.E$_{\text{SEM}}$

Dla dowolnych dwóch różnych dopuszczalnych wartości $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathcal{L}$ parametru $\boldsymbol{\lambda}$ mamy

$$\liminf_{N\to\infty} \frac{\frac{1}{\sqrt{N}}\left\|(\mathbf{I}-\boldsymbol{\lambda}_1^{\mathsf{T}}\mathbf{W})(\mathbf{I}-\boldsymbol{\lambda}_2^{\mathsf{T}}\mathbf{W})^{-1}\right\|}{\sqrt[N]{\left|\det\left(\mathbf{I}-\boldsymbol{\lambda}_1^{\mathsf{T}}\mathbf{W}\right)(\mathbf{I}-\boldsymbol{\lambda}_2^{\mathsf{T}}\mathbf{W})^{-1}\right|}} > 1. \tag{4.12}$$

2.2. Stwierdzenia pomocnicze

Poniższa definicja i następujący po niej lemat zostały zaadaptowane z opracowania Pötschera i Pruchy (1997).

Definicja

Niech $Q = Q(N)$ będzie funkcją rzeczywistą, określoną na pewnym zbiorze $\Theta \subset \mathbb{R}^d$ i niech $\theta_0 \in \Theta$ będzie ustalonym elementem tej przestrzeni. Powiemy, że θ_0 jest **identyfikowalnie jedynym** (ang. *identifiably unique*) **argumentem maksymalizującym** funkcję Q, jeśli, dla każdego $\delta > 0$ mamy

$$\liminf_{N\to\infty}\left(\inf_{\theta\in\Theta:\ \|\theta-\theta_0\|>\delta} Q(\theta_0) - Q(\theta)\right) > 0. \tag{4.13}$$

Lemat IV.1

Niech $R = R(N, \boldsymbol{\lambda})$, dla $\boldsymbol{\lambda} \in \mathcal{L}$, będzie zmienną losową, a $\bar{R} = \bar{R}(N)$ funkcją określoną na przestrzeni $\mathcal{L}$, taką, że

$$\sup_{\boldsymbol{\lambda}\in\mathcal{L}} |R(\boldsymbol{\lambda}) - \bar{R}(\boldsymbol{\lambda})| \to 0$$

według prawdopodobieństwa, przy $N \to \infty$. Jeśli $\boldsymbol{\lambda}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$ (patrz definicja powyżej) oraz zmienna losowa $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(N)$, o wartościach w zbiorze $\mathcal{L}$, maksymalizuje R, tzn. spełnia równość

$$R(\hat{\boldsymbol{\lambda}}) = \sup_{\boldsymbol{\lambda}\in\mathcal{L}} R(\boldsymbol{\lambda}) \tag{4.14}$$

prawie pewnie, wówczas $\hat{\boldsymbol{\lambda}}$ jest zgodnym estymatorem $\boldsymbol{\lambda}_0$, czyli $\|\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}}\|$ zbiega do zera według prawdopodobieństwa.

Dowód. Oznaczmy $\mathcal{E} = \mathcal{E}(\delta) = \left\{\boldsymbol{\lambda} \in \mathcal{L}\colon \|\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}}\| > \delta\right\}$. Dla dowolnego podciągu ciągu wszystkich liczb naturalnych możemy znaleźć dalszy podciąg $(N_n)_{n\in\mathbb{N}}$, dla którego

$$\lim_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\mathcal{L}} \left|R(N_n, \boldsymbol{\lambda}) - \bar{R}(N_n, \boldsymbol{\lambda})\right| = 0$$

prawie pewnie. Używając warunku (4.13) dla $Q = \bar{R}$ mamy

$$\liminf_{N\to\infty}\left(\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(\bar{R}(N_n,\boldsymbol{\lambda}_0)-\bar{R}(N_n,\boldsymbol{\lambda})\right)\right)>\eta>0,$$

dla pewnej liczby $\eta > 0$. Zatem wnioskujemy, że

$$\begin{aligned}\liminf_{n\to\infty}\left[\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(R(N_n,\boldsymbol{\lambda}_0)-R(N_n,\boldsymbol{\lambda})\right)\right]\geqslant\\ \liminf_{n\to\infty}\left(R(N_n,\boldsymbol{\lambda}_0)-\bar{R}(N_n,\boldsymbol{\lambda}_0)\right)+\liminf_{n\to\infty}\left[\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(\bar{R}(N_n,\boldsymbol{\lambda}_0)-\bar{R}(N_n,\boldsymbol{\lambda})\right)\right]\\ +\liminf_{n\to\infty}\left[\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(\bar{R}(N_n,\boldsymbol{\lambda})-\bar{R}(N_n,\boldsymbol{\lambda}_0)\right)\right]\\ \geqslant\eta-2\cdot\liminf_{n\to\infty}\left[\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(R(N_n,\boldsymbol{\lambda})-\bar{R}(N_n,\boldsymbol{\lambda})\right)\right]=\eta>0\end{aligned}$$

prawie pewnie, a więc

$$\inf_{\boldsymbol{\lambda}\in\mathcal{E}}\left(R(N_n,\boldsymbol{\lambda}_0)-R(N_n,\boldsymbol{\lambda})\right)\geqslant\frac{\eta}{2}>0,$$

dla wystarczająco dużych $n \in \mathbb{N}$. Skoro, według założenia (4.14), $R(N_n, \boldsymbol{\lambda}_0)$ nie przekracza $R(N_n, \hat{\boldsymbol{\lambda}}(N_n))$, wynika stąd, że $\|\hat{\boldsymbol{\lambda}}(N_n) - \boldsymbol{\lambda}_0\| < \delta$. Z dowolności wybranego pierwotnie podciągu liczb naturalnych uzyskujemy ostatecznie żądaną zbieżność według prawdopodobieństwa: $\hat{\boldsymbol{\lambda}} \to \boldsymbol{\lambda}_0$, przy $N \to \infty$. □

Lemat IV.2

Niech $R = R(N, \boldsymbol{\rho})$, dla $\boldsymbol{\rho} \in \mathcal{R}$, będzie zmienną losową, a $\bar{R} = \bar{R}(N)$ funkcją określoną na przestrzeni $\mathcal{R}$, taką, że

$$\sup_{\boldsymbol{\rho}\in\mathcal{R}}|R(\boldsymbol{\rho})-\bar{R}(\boldsymbol{\rho})|\to 0$$

według prawdopodobieństwa, przy $N \to \infty$. Jeśli $\boldsymbol{\rho}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$ (patrz definicja powyżej lematu IV.1) oraz zmienna losowa $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\rho}}(N)$, o wartościach w zbiorze $\mathcal{R}$, maksymalizuje R, tzn. spełnia równość

$$R(\hat{\boldsymbol{\rho}})=\sup_{\boldsymbol{\rho}\in\mathcal{R}}R(\boldsymbol{\rho}) \tag{4.15}$$

prawie pewnie, wówczas $\hat{\boldsymbol{\rho}}$ jest zgodnym estymatorem $\boldsymbol{\rho}_0$, czyli $\|\boldsymbol{\rho}_0 - \hat{\boldsymbol{\rho}}\|$ zbiega do zera według prawdopodobieństwa.

Do tezy lematu IV.2 prowadzi rozumowanie analogiczne do dowodu lematu IV.1 z $\boldsymbol{\rho}$ w miejscu $\boldsymbol{\lambda}$ i z $\mathcal{R}$ w miejscu $\mathcal{L}$.

LEMAT IV.3

Przy złożeniu IV.B$_{\text{SEM}}$ istnieje otwarty ograniczony nadzbiór $U_{\mathcal{L}} \subset \mathbb{R}^d$ przestrzeni parametrów $\mathcal{L}$, niezależny od $N \in \mathbb{N}$, taki, że operator $\Gamma(\boldsymbol{\lambda}) = \mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}$ jest odwracalny dla każdego $\boldsymbol{\lambda} \in U_{\mathcal{L}}$. Ponadto mamy

$$\sup_{\boldsymbol{\lambda} \in U_{\mathcal{L}}} \sup_{N \in \mathbb{N}} \|\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W}\| < \infty$$

oraz

$$\sup_{\boldsymbol{\lambda} \in U_{\mathcal{L}}} \sup_{N \in \mathbb{N}} \|(\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W})^{-1}\| < \infty.$$

DOWÓD. Użyjemy rozumowania przedstawionego oryginalnie w pracy Olejnik i Olejnik (2020). Dla dowolnej macierzy (ciągu macierzy) $\boldsymbol{A}$ definiujemy wielkość $\|\boldsymbol{A}\|_{\mathcal{A}} := \sup_{N \in \mathbb{N}} \|\boldsymbol{A}\|$. Rozważmy zbiór $\mathcal{A} = \{\boldsymbol{A} \colon \|A\|_{\mathcal{A}} < \infty\}$. Jak można wykazać, stosując klasyczną argumentację, funkcja $\boldsymbol{A} \mapsto \|A\|_{\mathcal{A}}$ jest (nieujemnie) jednorodna i podaddytywna, zaś zbiór $\boldsymbol{A}$ w nią wyposażony stanowi algebrę Banacha z jednością. Istotnie, mnożeniem w tej algebrze jest mnożenie odpowiadających sobie wyrazów ciągu macierzowego, a jego element neutralny stanowi macierz (ciąg macierzy) $\mathbf{I}$. Korzystając ze stwierdzenia 1.7 w monografii Takesakiego (1979), zbiór

$$G(\mathcal{A}) = \{\boldsymbol{A} : \text{ istnieje } \boldsymbol{B} \in \mathcal{A} \text{ takie, że } \boldsymbol{AB} = \mathbf{I}\}$$

jest otwarty w $\mathcal{A}$. Odwzorowanie Γ przekształcające $\mathbb{R}^d$ w $\mathcal{A}$ jest ciągłe, a zatem przeciwobraz V względem Γ zbioru $G(\mathcal{A})$ jest również otwarty w $\mathbb{R}^d$.

Można zauważyć, że funkcja $\gamma \colon V \to \mathbb{R}$ określona wzorem $\gamma(\boldsymbol{\lambda}) = \|(\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{W})^{-1}\|_{\mathcal{A}}$ jest ciągła. Istotnie, z wniosku 1.8 (Takesaki, 1979) wynika, że funkcja przypisująca elementowi algebry jego element odwrotny $G(\mathcal{A}) \ni \boldsymbol{A} \mapsto \boldsymbol{A}^{-1}$ jest ciągła względem normy w $G(\mathcal{A})$. Z samej definicji również norma $\|\cdot\|_{\mathcal{A}}$ jest funkcją ciągłą na $\mathcal{A}$, a w rezultacie ciągła jest funkcja γ, będąc ich złożeniem. Ostatecznie, zbiór

$$U = \left\{\boldsymbol{\lambda} \in V : \gamma(\boldsymbol{\lambda}) < 2 \sup_{\boldsymbol{\lambda}' \in \mathcal{L}} \gamma(\boldsymbol{\lambda}') < \infty\right\}$$

jest otwarty w V, a tym samym otwarty w $\mathbb{R}^d$. Korzystając z założenia IV.B$_{\text{SEM}}$ mamy $\mathcal{L} \subset V$, gdyż $\Gamma(\mathcal{L}) \subset G(\mathcal{A})$. Co więcej, dla każdego $\boldsymbol{\lambda} \in \mathcal{L}$ zachodzi

$$\gamma(\boldsymbol{\lambda}) \leqslant \sup_{\boldsymbol{\lambda}' \in \mathcal{L}} \gamma(\boldsymbol{\lambda}') < 2 \cdot \sup_{\boldsymbol{\lambda}' \in \mathcal{L}} \gamma(\boldsymbol{\lambda}'),$$

a więc $\mathcal{L} \subset U$ z samego określenia zbioru U. Dodatkowo, uwzględniając zwartość przestrzeni $\mathcal{L}$, zbiór $U_{\mathcal{L}}$ można również wybrać jako ograniczony i wciąż zawierający $\mathcal{L}$. Wystarczy przyjąć

$$U_{\mathcal{L}} \coloneqq U \cap \Big\{ \boldsymbol{\lambda} \in \mathbb{R}^d : \|\boldsymbol{\lambda}\| < \max_{\boldsymbol{\lambda}' \in \mathcal{L}} \|\boldsymbol{\lambda}'\| \Big\}.$$

W ten sposób uzyskujemy oczywistą nierówność

$$\sup_{\boldsymbol{\lambda} \in U_{\mathcal{L}}} \sup_{N \in \mathbb{N}} \|\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{W}\| < 2 + \max_{\boldsymbol{\lambda}' \in \mathcal{L}} \cdot \sup_{N \in \mathbb{N}} \max_{r \leqslant d} \|\mathbf{W}_r\| < \infty.$$

□

Lemat IV.4

Przy złożeniu IV.B_{SAR} istnieje otwarty ograniczony nadzbiór $U_{\mathcal{R}} \subset \mathbb{R}^d$ przestrzeni parametrów $\mathcal{R}$, niezależny od $N \in \mathbb{N}$, taki, że operator $\boldsymbol{\Delta}(\boldsymbol{\rho}) = \mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}} \mathbf{W}$ jest odwracalny dla każdego $\boldsymbol{\rho} \in U_{\mathcal{P}}$ oraz mamy

$$\sup_{\boldsymbol{\rho} \in U_{\mathcal{P}}} \sup_{N \in \mathbb{N}} \|\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}} \mathbf{W}\| < \infty,$$
$$\sup_{\boldsymbol{\rho} \in U_{\mathcal{P}}} \sup_{N \in \mathbb{N}} \|(\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}} \mathbf{W})^{-1}\| < \infty.$$

Rozumowanie przebiega analogicznie do dowodu lematu IV.3, z zastąpieniem symboli $\boldsymbol{\lambda}$ i $\mathcal{L}$ przez $\boldsymbol{\rho}$ oraz $\mathcal{P}$.

Lemat IV.5

Rozważmy funkcję $\log L_{\mathbf{y}}$ daną wzorem (4.6), określoną na dziedzinie $\mathbb{R}^k \times U \times (0, \infty)$, gdzie zbiór $U \subset \mathbb{R}^d$ dany jest w lemacie IV.3, parametryzowaną wartością $\mathbf{y} \in \mathbb{R}$. Dla uproszczenia zapisu przyjmijmy następujące oznaczenia

$$\mathbf{u}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}, \qquad \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta}) = \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{u}(\boldsymbol{\beta}),$$

$$\widetilde{\mathbf{W}}_r^{\boldsymbol{\lambda}} = \mathbf{W}_r \cdot \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1} = \mathbf{W}_r \cdot (\mathbf{I} - \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{W})^{-1}, \qquad \text{dla } 1 \leqslant r \leqslant d.$$

Pierwsze pochodne cząstkowe funkcji $\log L$ są dane wzorami

$$\frac{\partial \log L_{\mathbf{y}}}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta}),$$
$$\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{X},$$
$$\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda}} = \left[-\operatorname{tr} \widetilde{\mathbf{W}}_r^{\boldsymbol{\lambda}} + \frac{1}{\sigma^2} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \mathbf{W}_r \mathbf{u}(\boldsymbol{\beta}) \right]_{1 \leqslant r \leqslant d}.$$

Pochodne cząstkowe drugiego rzędu to

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}} = \left[-\operatorname{tr} \widetilde{\mathbf{W}}_{r_1}^{\boldsymbol{\lambda}} \widetilde{\mathbf{W}}_{r_2}^{\boldsymbol{\lambda}} - \frac{1}{\sigma^2} \mathbf{u}(\boldsymbol{\beta})^{\mathsf{T}} \mathbf{W}_{r_1}^{\mathsf{T}} \mathbf{W}_{r_2} \mathbf{u}(\boldsymbol{\beta}) \right]_{1 \leqslant r_1, r_2 \leqslant d},$$

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}} = \left[-\frac{1}{\sigma^2} \mathbf{u}(\boldsymbol{\beta})^{\mathsf{T}} (\mathbf{W}_r^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) - \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}} \mathbf{W}_r) \mathbf{X} \right]_{1 \leqslant r \leqslant d},$$

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda} \partial \sigma^2} = \left[-\frac{1}{\sigma^4} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \beta)^{\mathsf{T}} \mathbf{W}_r \mathbf{u}(\beta) \right]_{1 \leqslant r \leqslant d},$$

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2} \mathbf{X}^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{X},$$

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \boldsymbol{\beta}} = -\frac{1}{\sigma^4} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{X},$$

$$\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^4} + \frac{1}{\sigma^6} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta}).$$

Z kolei pochodne cząstkowe trzeciego rzędu są dane wzorami:

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}} = \left[-\operatorname{tr} \left(\widetilde{\mathbf{W}}_{r_1}^{\boldsymbol{\lambda}} \widetilde{\mathbf{W}}_{r_2}^{\boldsymbol{\lambda}} \widetilde{\mathbf{W}}_{r_3}^{\boldsymbol{\lambda}} \right) - \operatorname{tr} \left(\widetilde{\mathbf{W}}_{r_1}^{\boldsymbol{\lambda}} \widetilde{\mathbf{W}}_{r_3}^{\boldsymbol{\lambda}} \widetilde{\mathbf{W}}_{r_2}^{\boldsymbol{\lambda}} \right) \right]_{1 \leqslant r_1, r_2, r_3 \leqslant d},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}} = \left[\frac{1}{\sigma^4} \mathbf{u}(\boldsymbol{\beta})^{\mathsf{T}} \mathbf{W}_{r_1}^{\mathsf{T}} \mathbf{W}_{r_2} \mathbf{u}(\boldsymbol{\beta}) \right]_{1 \leqslant r_1, r_2 \leqslant d},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta} \partial \sigma^2} = \left[\frac{1}{\sigma^4} \mathbf{u}(\boldsymbol{\beta})^{\mathsf{T}} (\mathbf{W}_r^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) - \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}} \mathbf{W}_r) \mathbf{X} \right]_{1 \leqslant r \leqslant d},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta} \partial \sigma^2} = \frac{1}{\sigma^4} \mathbf{X}^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{X},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \sigma^2 \partial \boldsymbol{\beta}} = \frac{2}{\sigma^6} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{X},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \sigma^2 \partial \sigma^2} = -\frac{n}{\sigma^6} + \frac{3}{\sigma^8} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\varepsilon}(\boldsymbol{\lambda}, \boldsymbol{\beta}),$$

$$\frac{\partial^3 \log L_y}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}} = \left[\frac{1}{\sigma^2} \mathbf{X}^{\mathsf{T}} \left(\mathbf{W}_r^{\mathsf{T}} \boldsymbol{\Gamma}(\boldsymbol{\lambda}) + \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}} \mathbf{W}_r \right) \mathbf{X} \right]_{1 \leqslant r \leqslant d},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}} = \left[-\frac{1}{\sigma^2} \mathbf{u}(\boldsymbol{\beta})^{\mathsf{T}} \left(\mathbf{W}_{r_1}^{\mathsf{T}} \mathbf{W}_{r_2} + \mathbf{W}_{r_2}^{\mathsf{T}} \mathbf{W}_{r_1} \right) \mathbf{X} \right]_{1 \leqslant r_1, r_2 \leqslant d},$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = 0.$$

Powyższe formuły uzyskuje się przez bezpośrednie wyliczenie pochodnych, zgodnie z zasadami macierzowego rachunku różniczkowego. Dla wygody czytelnika podajemy najważniejsze własności pozwalające na samodzielne wyprowadzenie kolejnych pochodnych.

Jeśli t jest parametrem skalarnym, a $\boldsymbol{x}$ parametrem wektorowym, wówczas dla dowolnej macierzy $\boldsymbol{A}$ oraz funkcji macierzowych $\boldsymbol{B}(t)$, $\boldsymbol{C}_1(\boldsymbol{x})$, $\boldsymbol{C}_2(\boldsymbol{x})$, przy odpowiednich założeniach różniczkowalności i wykonywalności działań, prawdziwe są następujące stwierdzenia. Po pierwsze, własności liniowości pociągają za sobą wzory:

$$\mathrm{d}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\mathrm{d}\boldsymbol{x}, \quad \mathrm{d}\boldsymbol{A} = 0\mathrm{d}\boldsymbol{x}, \quad \mathrm{d}(\boldsymbol{C}_1(\boldsymbol{x})^{\mathsf{T}}) = (\mathrm{d}\boldsymbol{C}_1(\boldsymbol{x}))^{\mathsf{T}},$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\boldsymbol{B}(t) = \operatorname{tr}\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)\right).$$

Ponadto, z formuły Jacobiego dla macierzy nieosobliwych, tj.

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\boldsymbol{B}(t) = \operatorname{tr}\left(\left(\frac{\boldsymbol{B}(t)}{\det\boldsymbol{B}(t)}\right)^{-1}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)\right),$$

można wyprowadzić równość

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\left(\det\mathbf{B}(t)\right) = \operatorname{tr}\left(\mathbf{B}(t)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\right).$$

Następujący przepis na różniczkę iloczynu:

$$\mathrm{d}(\boldsymbol{C}_1(\boldsymbol{x})\boldsymbol{C}_2(\boldsymbol{x})) = \mathrm{d}(\boldsymbol{C}_1(\boldsymbol{x}))\cdot\boldsymbol{C}_2(\boldsymbol{x}) + \boldsymbol{C}_1(\boldsymbol{x})\cdot\mathrm{d}\boldsymbol{C}_2(\boldsymbol{x})$$

jest analogiem wzoru znanego z analizy funkcji rzeczywistych. Ostatecznie, za sprawą tożsamości $\boldsymbol{B}(t)^{-1}\boldsymbol{B}(t) = \boldsymbol{I}$, z powyższego wynika wzór opisujący pochodną macierzy odwrotnej

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{B}(t)^{-1}\right) = -\boldsymbol{B}(t)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)\right)\boldsymbol{B}(t)^{-1}.$$

LEMAT IV.6

Niech $\boldsymbol{\varepsilon}$ będzie składnikiem losowym modelu, spełniającym założenie IV.C dla pewnej semi-ortogonalnej macierzy $\mathbf{E}$ o wymiarach $N\times\bar{N}$ oraz dla pewnego N-elementowego wektora zaburzeń losowych $\bar{\boldsymbol{\varepsilon}}$. Wówczas, dla dowolnej macierzy $\boldsymbol{A}$ mamy

$$\mathbb{E}\left(\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon}\right) = \sigma^2\cdot\|\boldsymbol{A}\|_{\mathsf{F}}^2$$

oraz

$$\operatorname{Var}\left(\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon}\right) \leqslant 3N\cdot\|\boldsymbol{A}\|^4\cdot\sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4,$$

gdzie zmienne losowe $\bar{\varepsilon}_i$, $1\leqslant i\leqslant\bar{N}$ są elementami wektora $\bar{\boldsymbol{\varepsilon}}$.

Dowód. Wyliczając wartość oczekiwaną formy kwadratowej $\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\bar{\boldsymbol{\varepsilon}}$, otrzymujemy

$$\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon} = \mathbb{E}\,\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\mathbf{E}\bar{\boldsymbol{\varepsilon}} = \operatorname{tr}\mathbf{E}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\mathbf{E} = \operatorname{tr}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} = \|\boldsymbol{A}\|_{\mathsf{F}}^2,$$

gdyż z założenia IV.C wynika, że $\|\mathbf{E}\| = 1$. Aby dowieść zapowiedzianej nierówności dla wariancji, oznaczmy $\boldsymbol{\Omega} = \mathbf{E}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\mathbf{E}$ oraz przyjmijmy, że liczby ω_{ij}, $1 \leqslant i, j \leqslant \bar{N}$ będą elementami macierzy $\boldsymbol{\Omega}$. Wtedy uzyskujemy równość

$$\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon} = \operatorname{tr}\mathbf{E}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\mathbf{E} = \sigma^2 \cdot \sum_{i=1}^{\bar{N}} \omega_{ii},$$

a zatem

$$(\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon})^2 = \sigma^2 \cdot \sum_{i=1}^{\bar{N}} \omega_{ii} \cdot \sigma^2 \cdot \sum_{j=1}^{\bar{N}} \omega_{jj} = \sigma^4 \cdot \sum_{i,j=1}^{\bar{N}} \omega_{ii}\omega_{jj}.$$

Dalej wyliczamy drugi moment formy kwadratowej $\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon} = \bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\boldsymbol{\Omega}\bar{\boldsymbol{\varepsilon}}$ w następujący sposób

$$\begin{aligned}\mathbb{E}\,(\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon})^2 = \mathbb{E}\,(\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\boldsymbol{\Omega}\bar{\boldsymbol{\varepsilon}})^2 &= \sum_{i=1}^{\bar{N}}\sum_{j=1}^{\bar{N}}\sum_{i'=1}^{\bar{N}}\sum_{j'=1}^{\bar{N}} \omega_{ij}\omega_{i'j'}\,\mathbb{E}\,\bar{\varepsilon}_i\bar{\varepsilon}_j\bar{\varepsilon}_{i'}\bar{\varepsilon}_{j'} \\ &= \sum_{i=1}^{\bar{N}} \omega_{ii}\,\mathbb{E}\,\bar{\varepsilon}_i^4 + \sigma^4 \cdot \sum_{i=1}^{\bar{N}}\sum_{\substack{j=1 \\ i\neq j}}^{\bar{N}} (\omega_{ii}\omega_{jj} + \omega_{ij}\omega_{ij} + \omega_{ij}\omega_{ji}).\end{aligned}$$

Ostatecznie, uwzględniając fakt, że $\|\boldsymbol{\Omega}\| = \|\mathbf{E}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\mathbf{E}\| \leqslant 1^2 \cdot \|\boldsymbol{A}\|^2$, mamy

$$\begin{aligned}\operatorname{Var}(\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon}) &= \mathbb{E}\,(\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon})^2 - (\mathbb{E}\,\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\varepsilon})^2 \\ &= 2\sigma^4 \cdot \|\boldsymbol{\Omega}\|_{\mathsf{F}}^2 + \sum_{i=1}^{\bar{N}} \left(\mathbb{E}\,\bar{\varepsilon}_i^4 - 3\sigma^4\right) \cdot \omega_{ii}^2 \\ &\leqslant 3N \cdot \|\boldsymbol{\Omega}\|^2 \cdot \sup_{N'\in\mathbb{N}} \sup_{i\leqslant\bar{N}(N')} \mathbb{E}\,\bar{\varepsilon}_i^4 \\ &\leqslant 3N \cdot \|\boldsymbol{A}\|^4 \cdot \sup_{N'\in\mathbb{N}} \sup_{i\leqslant\bar{N}(N')} \mathbb{E}\,\bar{\varepsilon}_i^4.\end{aligned}$$

□

LEMAT IV.7

Niech $U \subset \mathbb{R}^d$ będzie zbiorem otwartym i $L \subset U$ jego zwartym podzbiorem. Jeśli $\boldsymbol{Q}\colon U \to \mathbb{R}^m$ jest funkcją różniczkowalną oraz dla pewnej stałej $M < \infty$ zachodzi

$$\sup_{\boldsymbol{x}\in L}\|\boldsymbol{Q}(\boldsymbol{x})\| + \sup_{\boldsymbol{x}\in U}\left\|\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}\boldsymbol{x}}\right\| < M,$$

wówczas $\boldsymbol{Q}$ ma własność Lipschitza na zbiorze L z pewną stałą $K_L = K_L(M)$.

DOWÓD. Zauważmy, że istotnie, $\boldsymbol{Q}$, jako funkcja różniczkowalna, jest ograniczona na zbiorze zwartym. Niech δ będzie odległością zbioru L od dopełnienia zbioru U, czyli

$$\delta = \inf\big\{\|\boldsymbol{x} - \boldsymbol{\xi}\| : \boldsymbol{x} \in L \wedge \boldsymbol{\xi} \in \mathbb{R}^d \setminus U\big\}.$$

Ze zwartości zbioru L możemy wnioskować, że ta odległość jest różna od zera. Istotnie, zdefiniujmy funkcję

$$L \ni \boldsymbol{x} \mapsto \delta_{\boldsymbol{x}} = \inf\big\{\|\boldsymbol{x} - \boldsymbol{\xi}\| : \boldsymbol{\xi} \in \mathbb{R}^d \setminus U\big\},$$

która jest półciągła z góry jako infimum funkcji ciągłych (a więc też półciągłych z góry). Zgodnie z twierdzeniem Bolzano-Weierstrassa, $L \ni \boldsymbol{x} \mapsto \delta_{\boldsymbol{x}}$ osiąga swoje maksimum, a więc $\max_{\boldsymbol{x}\in L}\delta_{\boldsymbol{x}} = \delta$. Zatem, gdyby δ było równe zero, mielibyśmy element $\boldsymbol{x}^*$ w $\mathcal{L}$, dla którego $\delta_{\boldsymbol{x}^*} = 0$, i w konsekwencji istniałby ciąg (ξ_n) elementów spoza U zbieżny do $\boldsymbol{x}^* \in L \subset U$. Ostatni wniosek jest jednak sprzeczny z założeniem, że U jest zbiorem otwartym.

Niech $\boldsymbol{x}$ i $\boldsymbol{y}$ będą dowolnymi elementami zbioru L. Jeśli $\|\boldsymbol{x} - \boldsymbol{y}\| < \delta$, wówczas odcinek

$$\overline{\boldsymbol{x},\boldsymbol{y}} \coloneqq \{\alpha\boldsymbol{x} + \beta\boldsymbol{y} : \alpha, \beta \geqslant 0 \wedge \alpha + \beta = 1\}$$

zawiera się w całości w U, jak wynika z określenia liczby δ. Wtedy, przy oznaczeniu $\bar{\boldsymbol{Q}}(t) = \boldsymbol{Q}(\boldsymbol{x} + t\cdot(\boldsymbol{y} - \boldsymbol{x}))$, dla $t \in [0,1]$, również mamy

$$\begin{aligned}\|\boldsymbol{Q}(\boldsymbol{y}) - \boldsymbol{Q}(\boldsymbol{x})\| &= \left\|\int_0^1 \frac{\mathrm{d}\bar{\boldsymbol{Q}}}{\mathrm{d}t}(t)\,\mathrm{d}t\right\| \\ &= \left\|\int_0^1 (\boldsymbol{y} - \boldsymbol{x})^{\mathsf{T}}\cdot\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x} + t\cdot(\boldsymbol{y} - \boldsymbol{x}))\,\mathrm{d}t\right\| \leqslant M\cdot\|\boldsymbol{y} - \boldsymbol{x}\|.\end{aligned}$$

Jeśli przeciwnie, $\|\boldsymbol{y} - \boldsymbol{x}\| \geqslant \delta$, wówczas

$$\|\boldsymbol{Q}(\boldsymbol{x}) - \boldsymbol{Q}(\boldsymbol{y})\| \leqslant \|\boldsymbol{Q}(\boldsymbol{x})\| + \|\boldsymbol{Q}(\boldsymbol{y})\| \leqslant \frac{2M}{\delta}\cdot\|\boldsymbol{x} - \boldsymbol{y}\|.$$

□

LEMAT IV.8

Rozważmy funkcję $\log L_{\mathbf{y}}$ daną wzorem (4.3), określoną na dziedzinie $U_{\mathcal{P}} \times \mathbb{R}^k \times (0,\infty)$, gdzie zbiór $U_{\mathcal{P}} \subset \mathbb{R}^d$ dany jest w lemacie IV.4 parametryzowaną wartością $\mathbf{y} \in \mathbb{R}$. Aby uprościć zapis, przyjmijmy następujące oznaczenia

$$\widetilde{\mathbf{W}}_r^{\rho} = \mathbf{W}_r \cdot \boldsymbol{\Delta}(\boldsymbol{\rho})^{-1} = \mathbf{W}_r \cdot \left(\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\right)^{-1},$$

dla $1 \leqslant r \leqslant d$, oraz

$$\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta}) = \mathbf{y} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}.$$

Pierwsze pochodne cząstkowe funkcji $\log L_{\mathbf{y}}$ dane są wzorami

$$\begin{aligned}
\frac{\partial \log L_{\mathbf{y}}}{\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta}),\\
\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\mathbf{X},\\
\frac{\partial \log L_{y}}{\partial \boldsymbol{\rho}} &= \left[-\operatorname{tr}\widetilde{\mathbf{W}}_r^{\rho} + \frac{1}{\sigma^2}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\mathbf{W}_r\mathbf{y}\right]_{1\leqslant r\leqslant d}.
\end{aligned}$$

Pochodne cząstkowe drugiego rzędu to

$$\begin{aligned}
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}} &= \left[-\operatorname{tr}\widetilde{\mathbf{W}}_{r_1}^{\rho}\widetilde{\mathbf{W}}_{r_2}^{\rho} - \frac{1}{\sigma^2}\mathbf{y}^{\mathsf{T}}\mathbf{W}_{r_1}^{\mathsf{T}}\mathbf{W}_{r_2}\mathbf{y}\right]_{1\leqslant r_1,r_2\leqslant d},\\
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho}\partial \boldsymbol{\beta}} &= \left[-\frac{1}{\sigma^2}\mathbf{y}^{\mathsf{T}}\mathbf{W}_r\mathbf{X}\right]_{1\leqslant r\leqslant d},\\
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho}\partial \sigma^2} &= \left[-\frac{1}{\sigma^4}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\mathbf{W}_r\mathbf{y}\right]_{1\leqslant r\leqslant d},\\
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^2}\mathbf{X}^{\mathsf{T}}\mathbf{X},\\
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \sigma^2\partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\mathbf{X},\\
\frac{\partial^2 \log L_{\mathbf{y}}}{\partial \sigma^2\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^4} + \frac{1}{\sigma^6}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta})^{\mathsf{T}}\boldsymbol{\varepsilon}(\boldsymbol{\rho},\boldsymbol{\beta}).
\end{aligned}$$

Z kolei pochodne cząstkowe trzeciego rzędu dane są wzorami

$$\begin{aligned}
\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho}} &= \left[-\operatorname{tr}\left(\widetilde{\mathbf{W}}_{r_1}^{\rho}\widetilde{\mathbf{W}}_{r_2}^{\rho}\widetilde{\mathbf{W}}_{r_3}^{\rho} + \widetilde{\mathbf{W}}_{r_1}^{\rho}\widetilde{\mathbf{W}}_{r_3}^{\rho}\widetilde{\mathbf{W}}_{r_2}^{\rho}\right)\right]_{1\leqslant r_1,r_2,r_3\leqslant d},\\
\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}} &= \left[\frac{1}{\sigma^4}\mathbf{y}^{\mathsf{T}}\mathbf{W}_{r_1}^{\mathsf{T}}\mathbf{W}_{r_2}\mathbf{y}\right]_{1\leqslant r_1,r_2\leqslant d},
\end{aligned}$$

$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho} \partial \boldsymbol{\beta} \partial \sigma^2} = \left[\frac{1}{\sigma^4} \mathbf{y}^{\mathsf{T}} \mathbf{W}_r \mathbf{X} \right]_{1 \leqslant r \leqslant d},$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \sigma^2} = \frac{1}{\sigma^4} \mathbf{X}^{\mathsf{T}} \mathbf{X},$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \sigma^2 \partial \boldsymbol{\beta}} = \frac{2}{\sigma^6} \boldsymbol{\varepsilon}(\boldsymbol{\rho}, \boldsymbol{\beta})^{\mathsf{T}} \mathbf{X},$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \sigma^2 \partial \sigma^2 \partial \sigma^2} = -\frac{n}{\sigma^6} + \frac{3}{\sigma^8} \boldsymbol{\varepsilon}(\boldsymbol{\rho}, \boldsymbol{\beta})^{\mathsf{T}} \boldsymbol{\varepsilon}(\boldsymbol{\rho}, \boldsymbol{\beta}),$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta} \partial \boldsymbol{\rho}} = 0,$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} = 0,$$
$$\frac{\partial^3 \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = 0.$$

Formuły wymienione w powyższym lemacie uzyskuje się przez bezpośrednie wyliczenie pochodnych zgodnie z zasadami macierzowego rachunku różniczkowego (por. komentarz po lemacie IV.5).

Lemat IV.9

Niech Y_0 będzie zmienną losową i niech $X = X_N$, $Y = Y_N$ będą ciągami zmiennych losowych, takimi, że $Y \xrightarrow{N\to\infty} Y_0$ według rozkładu oraz $X \xrightarrow{N\to\infty} 0$ według prawdopodobieństwa. Wówczas ciąg zmiennych losowych o elementach będących iloczynem $X \cdot Y$ dąży według prawdopodobieństwa do zera.

Dowód. Niech $\delta > 0$ będzie dowolne. Możemy wybrać liczbę $K \in \mathbb{R}$, dla której $\mathbb{P}(|Y_0| > K) < \delta$. Dodatkowo można założyć, że $\mathbb{P}(|Y_0| = K) = 0$, gdyż dystrybuanta zmiennej Y_0 może mieć co najwyżej przeliczalną liczbę skoków. Dla dowolnego $\epsilon > 0$ mamy oszacowanie

$$\begin{aligned}\mathbb{P}(|X \cdot Y| > \epsilon) &= \mathbb{P}(|X| \cdot |Y| > \epsilon) \\ &\leqslant \mathbb{P}(|Y| > K) + \mathbb{P}(|X| \cdot K > \epsilon) \xrightarrow{N\to\infty} \mathbb{P}(|Y_0| > K) < \delta.\end{aligned}$$

Z dowolności $\delta > 0$ mamy żądaną zbieżność. □

2.3. Twierdzenia o zgodności, dowód dla modelu SEM

Zagadnienie zgodności gaussowskich estymatorów quasi-największej wiarogodności dla modeli autoregresyjnych typu SAR zostało rozstrzygnięte w pracy Olejnik i Olejnik (2020), gdzie zaprezentowano formalny dowód poniższego

twierdzenia IV.10. W tym podrozdziale skupiamy się jednak na sformułowaniu i dowodzie nowego wyniku, dotyczącego modeli autoregresyjnych typu SEM (twierdzenie IV.11).

Twierdzenie IV.10

Przy założeniach IV.A, IV.B$_{\text{SAR}}$, IV.C, IV.D i IV.E$_{\text{SAR}}$ estymatory $\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}$, $\hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}}$ oraz $\hat{\sigma}^2_{\text{SAR\_QNW}}$ parametrów specyfikacji (4.1) są zgodne, tj. dla dowolnej liczby $\delta > 0$ mamy

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\boldsymbol{\rho}}_{\text{SAR_QNW}} - \boldsymbol{\rho}_0\| \geqslant \delta\right) &= 0,\\
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\boldsymbol{\beta}}_{\text{SAR_QNW}} - \boldsymbol{\beta}_0\| \geqslant \delta\right) &= 0,\\
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\sigma}^2_{\text{SAR_QNW}} - \sigma_0^2\| \geqslant \delta\right) &= 0,
\end{aligned}$$

gdzie $\boldsymbol{\rho}_0$, $\boldsymbol{\beta}_0$, σ_0^2 są prawdziwymi wartościami odpowiednich parametrów, a rozkład zaburzenia modelu $\boldsymbol{\varepsilon}$ jest dowolny.

Twierdzenie IV.11

Przy założeniach IV.A, IV.B$_{\text{SEM}}$, IV.C, IV.D i IV.E$_{\text{SEM}}$ estymatory $\hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}}$, $\hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}}$ oraz $\hat{\sigma}^2_{\text{SEM\_QNW}}$ parametrów specyfikacji (4.2) są zgodne, tj. dla dowolnej liczby $\delta > 0$ mamy

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\boldsymbol{\lambda}}_{\text{SEM_QNW}} - \boldsymbol{\lambda}_0\| \geqslant \delta\right) &= 0,\\
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\boldsymbol{\beta}}_{\text{SEM_QNW}} - \boldsymbol{\beta}_0\| \geqslant \delta\right) &= 0,\\
\lim_{N\to\infty} \mathbb{P}\left(\|\hat{\sigma}^2_{\text{SEM_QNW}} - \sigma_0^2\| \geqslant \delta\right) &= 0,
\end{aligned}$$

gdzie $\boldsymbol{\lambda}_0$, $\boldsymbol{\beta}_0$, σ_0^2 są prawdziwymi wartościami odpowiednich parametrów, a rozkład zaburzenia modelu $\boldsymbol{\varepsilon}$ jest dowolny.

Dowód. Dla uproszczenia zapisu, pominiemy indeks dolny w nazwach estymatorów, ustalając $\hat{\boldsymbol{\lambda}} \coloneqq \hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}}$, $\hat{\boldsymbol{\beta}} \coloneqq \hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}}$ oraz $\hat{\sigma}^2 \coloneqq \hat{\sigma}^2_{\text{SEM\_QNW}}$. Przypomnijmy również popularne oznaczenie $\mathbf{P}_{\boldsymbol{A}}$ dla operatora projekcji na podprzestrzeń rozpiętą przez kolumny dowolnie wybranej macierzy $\boldsymbol{A}$, tj. $\mathbf{P}_{\boldsymbol{A}} = \boldsymbol{A}\cdot(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}$. Dla dopełnienia przyjmujemy również $\mathbf{M}_{\boldsymbol{A}} = \mathbf{I} - \mathbf{P}_{\boldsymbol{A}}$.

Korzystając bezpośrednio z określenia estymatora $\hat{\boldsymbol{\lambda}}$ wnioskujemy, że jego wartość maksymalizuje na zbiorze $\mathcal{L}$ funkcję skoncentrowanej wiarogodności $\mathcal{L} \ni \boldsymbol{\lambda} \mapsto \ln L_{\mathbf{y}}(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}, \hat{\sigma}^2(\boldsymbol{\lambda}))$ daną w równaniu (4.8). Zatem, maksymalizuje ona również funkcję losową $\boldsymbol{\lambda} \mapsto R(\boldsymbol{\lambda})$, określoną wzorem

$$\begin{aligned}
R(\boldsymbol{\lambda}) &= \frac{1}{N}\ln L_{\mathbf{y}}\left(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}, \hat{\sigma}^2(\boldsymbol{\lambda})\right) + \frac{1}{2}\ln(2\pi) + \frac{1}{2}\\
&= -\frac{1}{2}\ln\hat{\sigma}^2(\boldsymbol{\lambda}) + \frac{1}{N}\ln|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda})|,
\end{aligned} \tag{4.16}$$

przy czym $\hat{\sigma}^2(\boldsymbol{\lambda})$ zdefiniowane jest w równaniu (4.7). Niech $U_{\mathcal{L}}$ będzie otwartym nadzbiorem przestrzeni $\mathcal{L}$ danym w lemacie IV.3. Łatwo zauważyć, że funkcja R jest poprawnie określona na również na zbiorze $U_{\mathcal{L}}$.

Zdefiniujmy funkcję deterministyczną

$$U_{\mathcal{L}} \ni \boldsymbol{\lambda} \mapsto \bar{R}(\boldsymbol{\lambda}) = \frac{1}{N} \ln |\det \boldsymbol{\Gamma}(\boldsymbol{\lambda})| - \frac{1}{2} \ln \Big(\frac{\sigma_0^2}{N} \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}} \Big).$$

Stosując standardowe zasady różniczkowania funkcji macierzowych można pokazać, że pochodne cząstkowe funkcji $\bar{R}$ są określone w całym zbiorze $U_{\mathcal{L}}$ następującym wzorem

$$\frac{\partial \bar{R}(\boldsymbol{\lambda})}{\partial \lambda_r} = -\frac{\operatorname{tr} \mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1}}{N} + \frac{\operatorname{tr} \left(\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-\mathsf{T}} (\mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda}) + \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{W}_r) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right)}{2 \cdot \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2},$$

dla $\boldsymbol{\lambda} \in U_{\mathcal{L}}$ oraz $1 \leqslant r \leqslant d$. Z kolei, na podstawie lematu IV.3 stwierdzamy, że następujące wartości są skończone:

$$\begin{aligned} \mathfrak{B}_{\mathbf{W}} &\coloneqq \sup_{N \in \mathbb{N}} \max_{r' \leqslant d} \|\mathbf{W}_{r'}\|, \\ \mathfrak{B}_{\mathcal{L}} &\coloneqq \sup_{\boldsymbol{\lambda}' \in \mathcal{L}} \|\boldsymbol{\lambda}'\|, \\ \mathfrak{B}_{\text{inv}} &\coloneqq \sup_{\boldsymbol{\lambda}' \in \mathcal{L}} \sup_{N \in \mathbb{N}} \|(\mathbf{I} - \boldsymbol{\lambda}'^{\mathsf{T}} \mathbf{W})^{-1}\|. \end{aligned} \tag{4.17}$$

Można zatem wnioskować, że pochodna funkcji $\bar{R}(\boldsymbol{\lambda})$ jest ograniczona na zbiorze $U_{\mathcal{L}}$. Istotnie, dla pierwszego składnika mamy nierówność

$$\begin{aligned} \Big| \frac{1}{N} \operatorname{tr} \mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1} \Big| &\leqslant \frac{1}{N} \sum \big\{ e \in \mathbb{R} : \det (\mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1} - e \cdot \mathbf{I}) = 0 \big\} \\ &\leqslant \|\mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1}\| \leqslant \mathfrak{B}_{\mathbf{W}} \cdot \mathfrak{B}_{\text{inv}} < \infty. \end{aligned}$$

Podobnie możemy otrzymać ograniczenie wyrażenia w liczniku drugiego składnika:

$$\begin{aligned} \Big| \frac{1}{N} \operatorname{tr} \Big(\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-\mathsf{T}} (\mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda}) + \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{W}_r) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \Big) \Big| \\ \leqslant \big\| \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-\mathsf{T}} (\mathbf{W}_r \boldsymbol{\Gamma}(\boldsymbol{\lambda}) + \boldsymbol{\Gamma}(\boldsymbol{\lambda}) \mathbf{W}_r) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \big\| \\ \leqslant 2 \mathfrak{B}_{\mathbf{W}} \cdot (1 + d \cdot \mathcal{B}_{\mathcal{L}} \cdot \mathfrak{B}_{\mathbf{W}}) \cdot \mathfrak{B}_{\text{inv}}^2. \end{aligned}$$

Z kolei ograniczenie (z dołu) wyrażenia w mianowniku wynika z nierówności

$$\begin{aligned} N = \|\mathbf{I}\|_{\mathsf{F}}^2 &= \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \cdot \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0) \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1}\|_{\mathsf{F}}^2 \\ &\leqslant \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}) \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2 \cdot \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0) \boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1}\|_{\mathsf{F}}^2, \end{aligned}$$

która prowadzi do oszacowania

$$\frac{1}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2 \geqslant \frac{1}{\|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{-1}\|^2} \geqslant (1+d\mathcal{B}_{\mathcal{L}}\mathfrak{B}_{\mathbf{W}})^{-2}\,\mathfrak{B}_{\text{inv}}^{-2}. \quad (4.18)$$

Aby wykazać zgodność estymatora $\hat{\boldsymbol{\lambda}}$, wykorzystamy udowodniony wcześniej lemat IV.1. Pokażemy, że różnica między wartościami funkcji R i $\bar{R}$, tj.

$$\big|\bar{R}(\boldsymbol{\lambda}) - R(\boldsymbol{\lambda})\big| = \left|\ln \frac{\hat{\sigma}^2(\boldsymbol{\lambda})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2}\right|,$$

maleje do zera według prawdopodobieństwa wraz ze wzrostem rozmiaru próby, jednostajnie względem $\boldsymbol{\lambda} \in \mathcal{L}$. Uwzględniając równość

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}$$

oraz wykorzystując równości (4.7), dla dowolnego $\boldsymbol{\lambda} \in \mathcal{L}$, uzyskujemy

$$\begin{aligned}
\hat{\sigma}^2(\boldsymbol{\lambda}) &= \frac{1}{N}\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\big)^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\big(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\big) \\
&= \frac{1}{N}\big\|\mathbf{M}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{y}\big\|^2 \\
&= \frac{1}{N}\big\|\mathbf{M}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{M}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\| \\
&= \frac{1}{N}\big\|\mathbf{M}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\| \\
&= \frac{\sigma_0^2}{N}\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\big\|_{\mathsf{F}}^2 - \xi_1(\boldsymbol{\lambda}) + \xi_2(\boldsymbol{\lambda}),
\end{aligned} \quad (4.19)$$

gdzie wyrazy resztowe $\xi_1(\boldsymbol{\lambda})$ oraz $\xi_2(\boldsymbol{\lambda})$ określone są wzorami

$$\begin{aligned}
\xi_1(\boldsymbol{\lambda}) &= \frac{1}{N}\big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\| \\
\xi_2(\boldsymbol{\lambda}) &= \frac{1}{N}\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 - \frac{\sigma_0^2}{N}\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\big\|_{\mathsf{F}}^2.
\end{aligned} \quad (4.20)$$

Istotnie, ostatnią równość w (4.19) uzyskujemy z twierdzenia Pitagorasa, przez wzajemną ortogonalność operatorów $\mathbf{M}_{\boldsymbol{A}}$ i $\mathbf{P}_{\boldsymbol{A}}$, dla $\boldsymbol{A} = \boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}$, gdyż

$$\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 = \big\|\mathbf{M}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 + \big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2.$$

Zauważmy teraz, że reszta $\xi_1(\boldsymbol{\lambda})$ dąży według prawdopodobieństwa do zera, jednostajnie względem $\boldsymbol{\lambda}$, z uwagi na oszacowania, wynikające z lematu IV.6 i z własności $\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\| \leqslant 1$ oraz

$$\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\|_{\mathsf{F}} = \operatorname{rank}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X} = \operatorname{rank}\mathbf{X} = k.$$

Po pierwsze, dla dowolnego $\boldsymbol{\lambda} \in \mathcal{L}$ mamy

$$\begin{aligned}
|\mathbb{E}\,\xi_1(\boldsymbol{\lambda})| &= \frac{1}{N}\,\mathbb{E}\,\big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 \\
&= \frac{1}{N}\big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\big\|_{\mathsf{F}}^2 \\
&\leqslant \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|^2 \cdot \|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\|^2 \cdot \frac{1}{N}\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\|_{\mathsf{F}}^2 \\
&\leqslant \mathfrak{B}_{\text{inv}}^2\,(1 + d \cdot \mathcal{B}_{\mathcal{L}} \cdot \mathfrak{B}_{\mathbf{W}})^2\,\frac{k^2}{N},
\end{aligned}$$

zgodnie z lematem IV.6, patrz również (4.17). Co za tym idzie, zachodzi zbieżność $\lim_{N\to\infty} \mathbb{E}\,\xi_1(\boldsymbol{\lambda}) = 0$, jednostajnie względem $\boldsymbol{\lambda} \in \mathcal{L}$. Po drugie,

$$\begin{aligned}
\operatorname{Var}\xi_1(\boldsymbol{\lambda}) &= \frac{1}{N^2}\operatorname{Var}\Big(\big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2\Big) \\
&\leqslant \frac{3}{N}\big\|\mathbf{P}_{\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\big\|^4 \sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4 \leqslant \frac{C}{N},
\end{aligned}$$

dla pewnej stałej $0 < C < \infty$, a to z kolei, z uwagi na założenie IV.C, implikuje zbieżność $\lim_{N\to\infty}\operatorname{Var}\xi_1(\boldsymbol{\lambda}) = 0$, jednostajnie względem $\boldsymbol{\lambda} \in \mathcal{L}$.

Wykażemy też, że reszta $\xi_2(\boldsymbol{\lambda})$ dąży według prawdopodobieństwa do zera, jednostajnie względem $\boldsymbol{\lambda} \in \mathcal{L}$. Zgodnie z lematem IV.6 mamy

$$\mathbb{E}\,\Big(\frac{1}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\|^2\Big) = \frac{\sigma_0^2}{N}\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\big\|_{\mathsf{F}}^2,$$

a więc $\mathbb{E}\,\xi_2(\boldsymbol{\lambda}) = 0$. Wariancję ξ_2 szacujemy podobnie jak w przypadku składnika ξ_1, mianowicie

$$\begin{aligned}
\operatorname{Var}\xi_2(\boldsymbol{\lambda}) &= \frac{1}{N^2}\operatorname{Var}\big\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 \\
&\leqslant \frac{3}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|^4 \cdot \sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4 \\
&\leqslant \frac{3\mathfrak{B}_{\text{inv}}^4\,(1 + d \cdot \mathcal{B}_{\mathcal{L}} \cdot \mathfrak{B}_{\mathbf{W}})^4}{N}\sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4,
\end{aligned}$$

a w konsekwencji $\operatorname{Var}\xi_2(\boldsymbol{\lambda})$ zbiega jednostajnie do zera na zbiorze $U_{\mathcal{L}}$.

Z powyższych obserwacji oraz z faktu, że wartości $\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2$ są jednostajnie odsunięte od zera, patrz (4.18), możemy wywnioskować jednostajną zbieżność według prawdopodobieństwa

$$\frac{\xi_1(\boldsymbol{\lambda}) + \xi_2(\boldsymbol{\lambda})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2} \xrightarrow{N\to\infty} 0,$$

a w konsekwencji również stwierdzić, że wyrażenie

$$2\big(\bar{R}(\boldsymbol{\lambda}) - R(\boldsymbol{\lambda})\big) = \ln \frac{\hat{\sigma}^2(\boldsymbol{\lambda})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2} = \ln\left(1 + \frac{\xi_1(\boldsymbol{\lambda}) + \xi_2(\boldsymbol{\lambda})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2}\right)$$

dąży do zera według prawdopodobieństwa, jednostajnie względem $\boldsymbol{\lambda} \in \mathcal{L}$. Zatem pokazaliśmy zbieżność $\sup_{\boldsymbol{\lambda}\in\mathcal{L}} |R(\boldsymbol{\lambda}) - \bar{R}(\boldsymbol{\lambda})| \to 0$ i aby skorzystać z lematu IV.1 wykażemy, że $\boldsymbol{\lambda}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$. Rozumowanie przeprowadzimy nie wprost.

Zauważmy najpierw, że $\bar{R}(\boldsymbol{\lambda}_0) \geqslant \bar{R}(\boldsymbol{\lambda})$ dla każdego $\boldsymbol{\lambda} \in \mathcal{L}$. Istotnie, ponieważ zachodzi równość

$$\bar{R}(\boldsymbol{\lambda}_0) = \frac{1}{N}\ln\det\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0) - \frac{1}{2}\ln(\sigma_0^2),$$

z elementarnej nierówności pomiędzy średnią arytmetyczną a średnią geometryczną mamy

$$\begin{aligned} 2\left(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\boldsymbol{\lambda})\right) &= \frac{2}{N}\ln|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)| - \ln(\sigma_0^2) - \frac{2}{N}\ln|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda})| \\ &\qquad + \ln\left(\frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2\right) \\ &= \ln\frac{\frac{1}{N}\|\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_{\mathsf{F}}^2}{|\det\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}|^{2/N}} \geqslant 0, \end{aligned}$$

gdzie licznik ułamka pod logarytmem można interpretować jako średnią arytmetyczną kwadratów wartości własnych macierzy $\boldsymbol{\Gamma}(\boldsymbol{\lambda})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}$, a mianownik — jako średnią geometryczną kwadratów tych wartości. Co więcej, na podstawie założenia IV.$\mathrm{E}_{\mathrm{SEM}}$ możemy stwierdzić, że

$$\liminf_{N\to\infty}\left(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\boldsymbol{\lambda})\right) > 0, \qquad \boldsymbol{\lambda} \in \mathcal{L}. \tag{4.21}$$

Załóżmy, że $\boldsymbol{\lambda}_0$ nie jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$, jak określono w nierówności (4.13), czyli

$$0 \leqslant \liminf_{N\to\infty}\left(\inf_{\boldsymbol{\lambda}\in\mathcal{L}:\ \|\boldsymbol{\lambda}-\boldsymbol{\lambda}_0\|>\delta}\left(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\boldsymbol{\lambda})\right)\right) \leqslant 0.$$

Istnieje więc liczba $\delta > 0$ oraz ściśle rosnący ciąg liczb naturalnych $(N_n)_{n\in\mathbb{N}}$, dla których wskazać można ciąg $(\tilde{\boldsymbol{\lambda}}_n)_{n\in\mathbb{N}}$ elementów przestrzeni $\mathcal{L}$ spełniający

$$\left\{\tilde{\boldsymbol{\lambda}}_n\right\}_{n\in\mathbb{N}} \subset \mathcal{E}(\delta) \coloneqq \{\boldsymbol{\lambda} \in \mathcal{L} : \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\| \geqslant \delta\}$$

oraz

$$0 \leqslant \lim_{N_n \to \infty} \big(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}}_n)\big) \leqslant 0.$$

Zbiór $\mathcal{E}(\delta)$ jest zwarty, jako domknięty podzbiór zwartej przestrzeni $\mathcal{L}$. Możemy więc wybrać dalszy podciąg $(\widetilde{\boldsymbol{\lambda}}_{m_n})_{n\in\mathbb{N}}$ ciągu $(\widetilde{\boldsymbol{\lambda}}_n)_{n\in\mathbb{N}}$, zbieżny w $\mathcal{L}$. Oznaczając $\widetilde{\boldsymbol{\lambda}} = \lim_{n\to\infty} \widetilde{\boldsymbol{\lambda}}_{m_n}$, na podstawie własności (4.21) wnioskujemy, że

$$\epsilon := \liminf_{N\to\infty} \big(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}})\big) > 0.$$

Z zaobserwowanej wcześniej ograniczoności pochodnej funkcji $\bar{R}$ oraz lematu IV.7 wynika, że $\bar{R}$ ma własność Lipschitza na przestrzeni $\mathcal{L}$ z pewną stałą $K_{\mathcal{L}}$, niezależną od rozmiaru próby. Dla dostatecznie dużych wartości indeksu n, tzn. dla wartości n przekraczających pewien poziom $n_0 \in \mathbb{N}$, mamy $\|\tilde{\boldsymbol{\lambda}}_{m_n} - \tilde{\boldsymbol{\lambda}}\| \leqslant \frac{\epsilon}{3K_{\mathcal{L}}}$. Ostatecznie, pożądana sprzeczność wynika z nierówności trójkąta poprzez następujące oszacowanie

$$\begin{aligned}
\epsilon = \liminf_{N\to\infty} \big(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}})\big) &\leqslant \liminf_{N_{m_n}\to\infty} \big(\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}})\big) \\
&\leqslant \liminf_{N_{m_n}\to\infty} \big(|\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}}_{m_n})| + |\bar{R}(\tilde{\boldsymbol{\lambda}}_{m_n}) - \bar{R}(\tilde{\boldsymbol{\lambda}})|\big) \\
&\leqslant \liminf_{N_{m_n}\to\infty} |\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}}_{m_n})| + \liminf_{N_{m_n}\to\infty} |\bar{R}(\tilde{\boldsymbol{\lambda}}_{m_n}) - \bar{R}(\tilde{\boldsymbol{\lambda}})| \\
&\leqslant \liminf_{N_{m_n}\to\infty} |\bar{R}(\boldsymbol{\lambda}_0) - \bar{R}(\tilde{\boldsymbol{\lambda}}_{m_n})| + K_{\mathcal{L}} \cdot \frac{\epsilon}{3K_{\mathcal{L}}} = \frac{\epsilon}{3}.
\end{aligned}$$

Wynika stąd, iż $\boldsymbol{\lambda}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$. Ponieważ estymator $\hat{\boldsymbol{\lambda}}$ jest — wprost ze swojej definicji — argumentem maksymalizującym funkcję R, określoną formułą (4.16), udowodniony fakt zbieżności jednostajnej według prawdopodobieństwa

$$\sup_{\boldsymbol{\lambda}\in\mathcal{L}} |R(\boldsymbol{\lambda}) - \bar{R}(\boldsymbol{\lambda})| \xrightarrow{N\to\infty} 0$$

pozwala użyć lematu IV.1 do ustalenia zgodności estymatora $\hat{\boldsymbol{\lambda}}$.

Pozostało nam zatem uzasadnić zgodność estymatorów $\hat{\boldsymbol{\lambda}}$ oraz $\hat{\sigma}^2$. Uwzględniając równania (4.7) oraz pamiętając o zależności $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}})$, otrzymujemy

$$\hat{\boldsymbol{\beta}} = \mathbf{P}_{\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}} \boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{y} = \mathbf{P}_{\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}} \boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})(\mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}) = \mathbf{I}\cdot\boldsymbol{\beta}_0 + \zeta_N,$$

gdzie $\zeta_N = \mathbf{P}_{\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}} \boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}$. Pokażemy teraz, że zaburzenie ζ_N zbiega do zera według prawdopodobieństwa. Najpierw zauważmy, iż na mocy założenia IV.D mamy

$$\frac{1}{\sqrt{N}}\|\mathbf{X}^{\mathsf{T}}\| \leqslant \frac{1}{\sqrt{N}}\|\mathbf{X}^{\mathsf{T}}\|_{\mathsf{F}} \leqslant \frac{1}{\sqrt{N}}\sqrt{k\cdot\|\mathbf{X}^{\mathsf{T}}\mathbf{X}\|} \leqslant C\sqrt{k}, \tag{4.22}$$

dla pewnej stałej $0 < C < \infty$. Ponadto, dla dowolnego $\boldsymbol{\lambda} \in \mathcal{L}$, prawdziwa jest nierówność

$$\begin{aligned}\frac{1}{\left\|\left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}\right)^{-1}\right\|} &= e_{\min}\left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X}\right)\\ &\leqslant e_{\min}(\mathbf{X}^{\mathsf{T}}\mathbf{X}) \cdot e_{\min}\left(\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\right)\\ &\leqslant \frac{1}{\left\|\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right\|} \cdot \frac{1}{\left\|\left(\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\right)^{-1}\right\|},\end{aligned}$$

więc w konsekwencji otrzymujemy

$$\left\|(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda})\mathbf{X})^{-1}\right\| \leqslant \left\|(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\right\| \cdot \|(\boldsymbol{\Gamma}(\boldsymbol{\lambda})^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda}))^{-1}\|.$$

Z powyższych obserwacji wynikają oszacowania spełnione prawie pewnie:

$$\begin{aligned}\left\|(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X})^{-1}\right\| &\leqslant \frac{1}{n} \cdot \|\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{-1}\|^2 \cdot \left\|\left(\frac{1}{N}\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right\|\\ &\leqslant \frac{\mathfrak{B}_{\text{inv}}^2}{N} \sup_{N' \in \mathbb{N}} \left\|\left(\frac{1}{N'}\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right\| \leqslant \frac{C'}{N}\end{aligned}$$

oraz

$$\begin{aligned}\left\|\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\right\| &\leqslant \|\mathbf{X}^{\mathsf{T}}\| \cdot \|\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\|^2 \cdot \|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|\\ &\leqslant C \cdot \sqrt{kN} \cdot (1 + d\mathfrak{B}_{\mathcal{L}}\mathfrak{B}_{\mathbf{W}})^2\, \mathfrak{B}_{\text{inv}} \leqslant C'\sqrt{kN},\end{aligned}$$

dla pewnej stałej $C' < \infty$.

Z uwagi na zależność

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}}) = \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0) + (\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}$$

mamy

$$\begin{aligned}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}}) = G(\boldsymbol{\lambda}_0)^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0) + G(\boldsymbol{\lambda}_0)^{\mathsf{T}}(\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}\\ + \boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)(\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}^{\mathsf{T}} + (\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}(\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}.\end{aligned}$$

Zatem zaburzenie ζ_N możemy dalej rozłożyć na sumę zaburzeń składowych $\zeta_N = \chi_N + \zeta'_N + \zeta''_N + \zeta'''_N$, gdzie

$$\begin{aligned}\chi_N &= \left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\varepsilon},\\ \zeta'_N &= \left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^{\mathsf{T}}(\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}})^{\mathsf{T}}\mathbf{W}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon},\end{aligned}$$

$$\zeta_N'' = \left(\mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})(\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}})^\mathsf{T}\mathbf{W}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon},$$
$$\zeta_N''' = \left(\mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}(\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}})^\mathsf{T}\mathbf{W}^\mathsf{T}(\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}})^\mathsf{T}\mathbf{W}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}.$$

Teraz wykażemy zbieżność ζ_N do zera według prawdopodobieństwa, wskazując ograniczenia i zbieżności kolejnych składników. Mamy, patrz (4.22),

$$\begin{aligned}\mathbb{E}\|\chi_N\|^2 &= \mathbb{E}\left(\left\|(\mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X})^{-1}\right\|^2\cdot\|\mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^\mathsf{T}\boldsymbol{\varepsilon}\|^2\right)\\ &\leqslant \left(\frac{C'}{N}\right)^2\cdot\sigma_0^2\cdot\|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)\mathbf{X}\|_\mathsf{F}^2 \leqslant \left(\frac{C'}{N}\right)^2\cdot k\cdot\|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)\mathbf{X}\|^2\\ &\leqslant \frac{(C'\cdot Ck)^2}{N}\left(1+d\mathfrak{B}_{\mathcal{L}}\mathfrak{B}_{\mathbf{W}}\right)^2,\end{aligned}$$

czyli χ_N zbiega do zera w $\mathbb{R}^k$, a z uwagi na nierówność Czebyszewa zbiega również według prawdopodobieństwa. Kolejny składnik rozważymy, ograniczając jego pierwszy moment. Mianowicie, oznaczając $\mathbf{U} = \mathbf{X}^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})^\mathsf{T}\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\mathbf{X}$, mamy

$$\begin{aligned}\mathbb{E}\|\xi_N'\| &\leqslant \mathbb{E}\left\|\left(\frac{1}{N}\mathbf{U}\right)^{-1}\right\|\left\|\frac{1}{\sqrt{N}}\mathbf{X}^\mathsf{T}\right\|\|\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)\|\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|\left\|\mathbf{W}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\right\|\|\boldsymbol{\varepsilon}\|\\ &\leqslant C''\cdot\frac{1}{\sqrt{N}}\,\mathbb{E}\left(\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|\cdot\|\boldsymbol{\varepsilon}\|\right),\end{aligned}$$

dla odpowiednio dobranej stałej $C'' < \infty$. Z nierówności Schwarza wnioskujemy, że

$$\frac{1}{\sqrt{N}}\,\mathbb{E}\left(\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|\cdot\|\boldsymbol{\varepsilon}\|\right) \leqslant \sqrt{\sigma_0^2\cdot\mathbb{E}\,\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|^2}.$$

Dla dowolnej liczby $\delta > 0$ zachodzi

$$\mathbb{E}\left(\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|^2\right) \leqslant \delta + \sup_{\boldsymbol{\lambda},\boldsymbol{\lambda}'\in\mathcal{L}}\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|\cdot\mathbb{P}\left(\|\boldsymbol{\lambda}_0-\hat{\boldsymbol{\lambda}}\|>\delta\right)\xrightarrow{N\to\infty}\delta,$$

zatem $\hat{\boldsymbol{\lambda}}$ zbiega do $\boldsymbol{\lambda}_0$ również w L_2, co daje zbieżność $\mathbb{E}\|\zeta_N'\|$ do zera według prawdopodobieństwa. Dla składników ζ_N'' oraz ζ_N''' można przeprowadzić analogiczne rozważania.

Aby wykazać zgodność estymatora $\hat{\sigma}^2$, przypomnijmy, że $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$. Zgodnie z reprezentacją opisaną przez równości (4.19) i (4.20) mamy

$$\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \frac{\sigma_0^2}{N}\|\boldsymbol{\Gamma}(\hat{\boldsymbol{\lambda}})\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|_\mathsf{F}^2 - \xi_1(\hat{\boldsymbol{\lambda}}) + \xi_2(\hat{\boldsymbol{\lambda}}).$$

Ponadto, ze względu na wykazaną, jednostajną względem $\boldsymbol{\lambda}$ zbieżność elementów $\xi_1(\boldsymbol{\lambda})$ i $\xi_2(\boldsymbol{\lambda})$ do zera, możemy wywnioskować, że

$$0 \leqslant \left|-\xi_1(\hat{\boldsymbol{\lambda}})+\xi_2(\hat{\boldsymbol{\lambda}})\right| \leqslant \sup_{\boldsymbol{\lambda}\in\mathcal{L}}|\xi_1(\hat{\boldsymbol{\lambda}})| + \sup_{\boldsymbol{\lambda}\in\mathcal{L}}|\xi_2(\hat{\boldsymbol{\lambda}})| \xrightarrow{N\to\infty} 0$$

według prawdopodobieństwa. Pozostaje zatem wykazać, iż

$$\lim_{N\to\infty} \frac{\sigma_0^2}{N} \left\| \mathbf{\Gamma}(\hat{\boldsymbol{\lambda}}) \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right\|_{\mathsf{F}}^2 = \sigma_0^2$$

według prawdopodobieństwa. Zauważmy, że funkcja

$$U_{\mathcal{L}} \ni \boldsymbol{\lambda} \mapsto \theta(\boldsymbol{\lambda}) \coloneqq \left\| \mathbf{\Gamma}(\boldsymbol{\lambda}) \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right\|_{\mathsf{F}}^2$$

jest różniczkowalna, a jej pochodne cząstkowe, dla $1 \leqslant r \leqslant d$,

$$U_{\mathcal{L}} \ni \boldsymbol{\lambda} \mapsto \frac{\partial \theta}{\partial \lambda_r}(\boldsymbol{\lambda}) = \frac{\sigma_0^2}{N} \operatorname{tr} \left(\mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-\mathsf{T}} (\mathbf{W}_r^{\mathsf{T}} + \mathbf{W}_r) \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right) + \frac{\sigma_0^2}{N} \sum_{r'=1}^{d} \lambda_{r'} \left\| \mathbf{W}_{r'} \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right\|_{\mathsf{F}}^2$$

są — ze względu na argument $\boldsymbol{\lambda}$ i rozmiar próby N — jednostajnie ograniczone przez pewną wartość, wynikającą z założeń IV.A i IV.$\mathrm{B}_{\mathrm{SEM}}$. Istotnie, mamy

$$\left| \frac{\sigma_0^2}{N} \operatorname{tr} \left(\mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-\mathsf{T}} (\mathbf{W}_r^{\mathsf{T}} + \mathbf{W}_r) \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \right) \right| \leqslant 2\sigma_0^2 \mathfrak{B}_{\mathbf{W}} \mathfrak{B}_{\mathrm{inv}}^2$$

oraz

$$\left| \frac{\sigma_0^2}{N} \sum_{r'=1}^{d} \lambda_{r'} \| \mathbf{W}_{r'} \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \|_{\mathsf{F}}^2 \right| \leqslant d\sigma_0^2 \mathfrak{B}_{\mathcal{L}} \mathfrak{B}_{\mathbf{W}}^2 \mathfrak{B}_{\mathrm{inv}}^2.$$

Na mocy lematu IV.7 wnioskujemy zatem, że na przestrzeni $\mathcal{L}$ funkcja θ spełnia warunek Lipschitza, z pewną stałą $K_{\mathcal{L}}^{\theta}$. Otrzymujemy ostatecznie

$$\left| \| \mathbf{\Gamma}(\hat{\boldsymbol{\lambda}}) \mathbf{\Gamma}(\boldsymbol{\lambda}_0)^{-1} \|_{\mathsf{F}}^2 - \sigma_0^2 \right| = |\theta(\hat{\boldsymbol{\lambda}}) - \theta(\boldsymbol{\lambda}_0)| \leqslant K_{\mathcal{L}}^{\theta} \cdot \| \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 \| \xrightarrow{N\to\infty} 0$$

według prawdopodobieństwa. □

2.4. Dowód twierdzenia o zgodności dla modelu SAR

W tym podrozdziale przedstawiamy dowód twierdzenia IV.10 o zgodności estymatora QNW dla dla modelu, którego specyfikacja (patrz równanie (4.1)) uwzględnia przestrzenną autokorelację zmiennej zależnej. Schemat rozumowania argumentującego tezę twierdzenia prowadzona jest w sposób analogiczny do dowodu twierdzenia IV.11. Mianowicie wykorzystywane jest pojęcie identyfikowalnie jedynego argumentu maksymalizującego, zbieżność jednostajna ciągu funkcji wiarogodności oraz, prowadzący do zgodności, lemat IV.2. Niemniej jednak, różne specyfikacje modelu procesu generującego obserwacje wymagają szczególnego w obu przypadkach szacowania elementów resztowych i zaburzeń losowych estymatorów.

Dowód. Ponownie, dla uproszczenia zapisu w nazwach estymatorów pominiemy indeks dolny ustalając: $\hat{\boldsymbol{\rho}} \coloneqq \hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}$, $\hat{\boldsymbol{\beta}} \coloneqq \hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}}$ oraz $\hat{\sigma}^2 \coloneqq \hat{\sigma}^2_{\text{SAR\_QNW}}$. Przyjmujemy też popularne oznaczenie macierzy rzutowych $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$ oraz $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$.

Przypomnijmy, że wartość estymatora $\hat{\boldsymbol{\rho}}$ została określona jako ta maksymalizująca z prawdopodobieństwem 1 funkcję skoncentrowanej wiarogodności $\mathcal{P} \ni \boldsymbol{\rho} \mapsto \ln L_{\mathbf{y}}\Big(\boldsymbol{\rho}, \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}), \hat{\sigma}^2(\boldsymbol{\rho})\Big)$ daną w równaniu (4.5). Co za tym idzie, maksymalizuje ona również na zbiorze $\mathcal{P}$ funkcję losową

$$\begin{aligned} U_{\mathcal{P}} \ni \boldsymbol{\rho} \mapsto R(\boldsymbol{\rho}) &= \frac{1}{N} \ln L_{\mathbf{y}}\left(\boldsymbol{\rho}, \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}), \hat{\sigma}^2(\boldsymbol{\rho})\right) + \frac{1}{2}\ln(2\pi) + \frac{1}{2} \\ &= -\frac{1}{2}\ln\left(\hat{\sigma}^2(\boldsymbol{\rho})\right) + \frac{1}{N}\ln|\det \boldsymbol{\Delta}(\boldsymbol{\rho})|, \end{aligned} \tag{4.23}$$

gdzie zbiór $U_{\mathcal{P}}$ jest otwartym nadzbiorem przestrzeni $\mathcal{P}$ danym w lemacie IV.4, natomiast $\hat{\sigma}^2(\boldsymbol{\rho})$, dla $\boldsymbol{\rho} \in U_{\mathcal{P}}$, dane jest w równaniu (4.4). Dodatkowo zdefiniujmy funkcję deterministyczną $U_{\mathcal{P}} \ni \boldsymbol{\rho} \mapsto \bar{R}(\boldsymbol{\rho})$ określoną formułą

$$\begin{aligned} \bar{R}(\boldsymbol{\rho}) = -\frac{1}{2}\ln\Big(\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}} + \|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\|^2\Big) \\ + \frac{1}{N}\ln|\det \boldsymbol{\Delta}(\boldsymbol{\rho})|. \end{aligned} \tag{4.24}$$

Jak można wyliczyć, stosując standardowe zasady różniczkowania funkcji macierzowych, pochodne cząstkowe funkcji $\bar{R}$ określone są w całym zbiorze $U_{\mathcal{P}}$ i są równe

$$\begin{aligned} \frac{\partial \bar{R}(\boldsymbol{\rho})}{\partial \rho_r} = \frac{\frac{\sigma_0^2}{2N}\chi_1(\boldsymbol{\rho}) + \frac{1}{2N}\chi_2(\boldsymbol{\rho})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2 + \frac{1}{N}\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\|^2} \\ - \frac{1}{N}\operatorname{tr}\left(\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\right), \end{aligned}$$

dla $\boldsymbol{\rho} \in U_{\mathcal{P}}$, $1 \leqslant r \leqslant d$, gdzie

$$\begin{aligned} \chi_1(\boldsymbol{\rho}) &= \frac{\partial\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2}{\partial\rho_r} \\ &= \operatorname{tr}\Big(\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}\big(\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}) + \boldsymbol{\Delta}(\boldsymbol{\rho})\mathbf{W}_r\big)\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\Big), \end{aligned}$$

oraz

$$\chi_2(\boldsymbol{\rho}) = \frac{\partial\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\|^2}{\partial\rho_r}$$

$$= -\frac{1}{N}\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho})^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}\mathbf{M}_{\mathbf{X}}\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0\mathbf{X}^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}\mathbf{W}_r^{\mathsf{T}}\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0.$$

Na podstawie lematu IV.4 można wnioskować, że pochodne te są ograniczone na całym zbiorze $U_{\mathcal{P}}$. Istotnie, przy oznaczeniach

$$\begin{aligned}
\mathfrak{B}_{\mathbf{W}} &= \sup_{N\in\mathbb{N}} \max_{r\leqslant d} \|\mathbf{W}_r\| < \infty,\\
\mathfrak{B}_{\mathcal{P}} &= \sup_{\boldsymbol{\rho}\in\mathcal{P}} \|\boldsymbol{\rho}\| < \infty,\\
\mathfrak{B}_{\boldsymbol{\Delta}} &= \sup_{\boldsymbol{\rho}\in\mathcal{P}} \sup_{N\in\mathbb{N}} \|\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W}\| < \infty,\\
\mathfrak{B}_{\text{inv}} &= \sup_{\boldsymbol{\rho}\in\mathcal{P}} \sup_{N\in\mathbb{N}} \|(\mathbf{I} - \boldsymbol{\rho}^{\mathsf{T}}\mathbf{W})^{-1}\| < \infty,
\end{aligned} \tag{4.25}$$

ostatni składnik wzoru opisującego pochodne cząstkowe $\bar{R}$ spełnia nierówność

$$\begin{aligned}
\left|\frac{1}{N}\operatorname{tr}\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\right| &\leqslant \frac{1}{N}\sum\left\{e\in\mathbb{R} : \det\left(\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1} - e\cdot\mathbf{I}\right) = 0\right\}\\
&\leqslant \|\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\| \leqslant \mathfrak{B}_{\mathbf{W}}\cdot\mathfrak{B}_{\text{inv}} < \infty,
\end{aligned}$$

dla $\boldsymbol{\rho}\in U_{\mathcal{P}}$. Podobnie szacujemy wartość elementu $\frac{1}{N}\chi_1$:

$$\begin{aligned}
\left|\frac{1}{N}\chi_1(\boldsymbol{\rho})\right| &\leqslant \left\|\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}(\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}) + \boldsymbol{\Delta}(\boldsymbol{\rho})\mathbf{W}_r)\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\right\|\\
&\leqslant 2\mathfrak{B}_{\mathbf{W}}\cdot\mathfrak{B}_{\boldsymbol{\Delta}}\cdot\mathfrak{B}_{\text{inv}}^2 < \infty.
\end{aligned}$$

Ograniczenie wartości elementu $\frac{1}{N}\chi_2$ możemy z kolei skonstruować odwołując się do submultiplikatywności normy i tożsamości $\|\mathbf{X}\|^2 = \|\mathbf{X}^{\mathsf{T}}\mathbf{X}\|$, przez co otrzymujemy

$$\left|\frac{1}{N}\chi_2(\boldsymbol{\rho})\right| \leqslant 2\|\boldsymbol{\beta}_0\|^2 \sup_{N'\in\mathbb{N}}\left\|\frac{1}{N'}\mathbf{X}^{\mathsf{T}}\mathbf{X}\right\|\cdot\mathfrak{B}_{\mathbf{W}}\mathfrak{B}_{\boldsymbol{\Delta}}\mathfrak{B}_{\text{inv}}^2 < \infty.$$

Zauważmy też, że wartości wyrażenia $\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2$ są odcięte od zera jednostajnie względem $\boldsymbol{\rho}\in U_{\mathcal{P}}$ (wartość $\boldsymbol{\rho}_0$ traktujemy jako ustaloną). Wniosek ten wypływa z następującego oszacowania

$$\begin{aligned}
N = \|\mathbf{I}\|_{\mathsf{F}}^2 = \left\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\cdot\boldsymbol{\Delta}(\boldsymbol{\rho}_0)\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\right\|_{\mathsf{F}}^2\\
\leqslant \left\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\right\|_{\mathsf{F}}^2\cdot\left\|\boldsymbol{\Delta}(\boldsymbol{\rho}_0)\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\right\|_{\mathsf{F}}^2,
\end{aligned}$$

które prowadzi do

$$\frac{\sigma_0^2}{N}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|_{\mathsf{F}}^2 \geqslant \frac{1}{\|\boldsymbol{\Delta}(\boldsymbol{\rho}_0)\boldsymbol{\Delta}(\boldsymbol{\rho})^{-1}\|^2} \geqslant (\mathcal{B}_{\boldsymbol{\Delta}} \cdot \mathfrak{B}_{\text{inv}})^{-2} > 0.$$

Aby wykazać zgodność estymatora $\hat{\boldsymbol{\rho}}$, wykorzystamy lemat IV.2. Pokażemy, że bezwzględna różnica między wartościami powyższych dwóch funkcji, a więc $|R(\boldsymbol{\rho}) - \bar{R}(\boldsymbol{\rho})|$ maleje do zera według prawdopodobieństwa wraz ze wzrostem wielkości próby, jednostajnie względem $\boldsymbol{\rho} \in \mathcal{P}$. Biorąc pod uwagę równość

$$\mathbf{y} = \boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}$$

oraz (4.4), dla dowolnego $\boldsymbol{\rho} \in \mathcal{R}$, uzyskujemy

$$\begin{aligned}\hat{\sigma}^2(\boldsymbol{\rho}) &= \frac{1}{N}\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\mathbf{y}\|^2 \\ &= \frac{1}{N}\big\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\| \\ &= \frac{1}{N}\big\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\big\|^2 + \frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2 \\ &\qquad + \chi(\boldsymbol{\rho}) - \xi_1(\boldsymbol{\rho}) + \xi_2(\boldsymbol{\rho}),\end{aligned} \tag{4.26}$$

gdzie wyrazy resztowe $\xi_1(\boldsymbol{\rho})$, $\xi_2(\boldsymbol{\rho})$ oraz $\chi(\boldsymbol{\rho})$ są określone wzorami

$$\begin{aligned}\chi(\boldsymbol{\rho}) &= \frac{2}{N}\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho})^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}, \\ \xi_1(\boldsymbol{\rho}) &= \frac{1}{N}\big\|\mathbf{P}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2, \\ \xi_2(\boldsymbol{\rho}) &= \frac{1}{N}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 - \frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2.\end{aligned} \tag{4.27}$$

Istotnie, powyższe równości wynikają ze wzoru skróconego mnożenia dla kwadratu sumy oraz z twierdzenia Pitagorasa mamy poprzez wzajemną ortogonalność operatorów $\mathbf{M}_{\mathbf{X}}$ i $\mathbf{P}_{\mathbf{X}}$.

Zauważmy teraz, że wartość oczekiwana składnika $\chi(\boldsymbol{\rho})$ wynosi zero, dla wszystkich $\boldsymbol{\rho} \in U_{\mathcal{P}}$ oraz jego wariancja jest jednostajnie ograniczona, tj.

$$\begin{aligned}\mathbb{E}\left(\chi(\boldsymbol{\rho})\right)^2 &= \frac{4}{N^2}\big\|\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho})^{\mathsf{T}}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-\mathsf{T}}\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|^2 \\ &\leqslant \frac{4}{N}\|\boldsymbol{\beta}_0\|^2 \sup_{N' \in \mathbb{N}}\Big\|\frac{1}{N'}\mathbf{X}^{\mathsf{T}}\mathbf{X}\Big\|\mathfrak{B}_{\boldsymbol{\Delta}}^4\mathfrak{B}_{\text{inv}}^4 < \infty.\end{aligned}$$

Wykażemy też, że reszta $\xi_1(\boldsymbol{\rho})$ dąży według prawdopodobieństwa do 0, jednostajnie względem $\boldsymbol{\rho}$, z uwagi na następujące oszacowania, wynikające z lematu IV.6

oraz z własności: $\|\mathbf{P}_{\mathbf{X}}\|$ i $\|\mathbf{P}_{\mathbf{X}}\|_{\mathsf{F}} = \operatorname{rank}(\mathbf{X}) = k$. Po pierwsze, dla dowolnego $\boldsymbol{\rho} \in \mathcal{P}$ mamy

$$\begin{aligned}\mathbb{E}\,\xi_1(\boldsymbol{\rho}) &= \frac{1}{N}\,\mathbb{E}\,\big\|\mathbf{P}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 = \frac{1}{N}\big\|\mathbf{P}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|_{\mathsf{F}}^2 \\ &\leqslant \|\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|^2 \cdot \|\boldsymbol{\Delta}(\boldsymbol{\rho})\|^2 \cdot \frac{1}{N}\|\mathbf{P}_{\mathbf{X}}\|_{\mathsf{F}}^2 \leqslant \frac{k^2}{N}\mathfrak{B}_{\boldsymbol{\Delta}}^2\mathfrak{B}_{\text{inv}}^2,\end{aligned}$$

czyli $\lim_{N\to\infty}\mathbb{E}\,\xi_1(\boldsymbol{\rho}) = 0$, jednostajnie względem $\boldsymbol{\rho} \in \mathcal{L}$. Po drugie,

$$\begin{aligned}\operatorname{Var}\xi_1(\boldsymbol{\rho}) &= \frac{1}{N^2}\operatorname{Var}\big\|\mathbf{P}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 \\ &= \frac{3}{N}\big\|\mathbf{P}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|^4 \sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4 \\ &\leqslant \frac{3}{N}\mathfrak{B}_{\boldsymbol{\Delta}}^4\mathfrak{B}_{\text{inv}}^4 \leqslant \frac{C}{N},\end{aligned}$$

dla pewnej stałej $C < \infty$, co implikuje zbieżność $\operatorname{Var}\xi_1(\boldsymbol{\rho})$ do zera, jednostajnie względem $\boldsymbol{\rho} \in \mathcal{L}$, z uwagi na założenia IV.A, IV.B$_{\text{SAR}}$ i IV.C, patrz (4.25).

Podobnie, reszta $\xi_2(\boldsymbol{\rho})$ dąży według prawdopodobieństwa do zera, jednostajnie względem $\boldsymbol{\rho} \in \mathcal{P}$. Istotnie, zgodnie z lematem IV.6 otrzymujemy

$$\mathbb{E}\left(\frac{1}{N}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2\right) = \frac{\sigma_0^2}{N}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|_{\mathsf{F}}^2,$$

a więc $\mathbb{E}\,\xi_2(\boldsymbol{\rho}) = 0$. Następnie uzyskujemy oszacowanie

$$\begin{aligned}\operatorname{Var}\xi_2(\boldsymbol{\rho}) &= \frac{1}{N^2}\operatorname{Var}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\big\|^2 \\ &= \frac{3}{N}\big\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\big\|^4 \sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4 \leqslant \frac{3}{N}\mathfrak{B}_{\boldsymbol{\Delta}}^4\mathfrak{B}_{\text{inv}}^4 \leqslant \frac{C}{N},\end{aligned}$$

dla pewnej stałej $C < \infty$, co implikuje zbieżność $\lim_{N\to\infty}\operatorname{Var}\xi_2(\boldsymbol{\rho}) = 0$.

Z przedstawionych obserwacji oraz z faktu, że wartości $\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2$ są jednostajnie odsunięte od zera, możemy wywnioskować jednostajną zbieżność według prawdopodobieństwa

$$\frac{\chi(\boldsymbol{\rho}) - \xi_1(\boldsymbol{\rho}) + \xi_2(\boldsymbol{\rho})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2 + \frac{1}{N}\|\mathbf{M}_{\mathbf{X}}\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\|^2} \xrightarrow{N\to\infty} 0,$$

a w konsekwencji również zbieżność różnicy $R(\boldsymbol{\rho}) - \bar{R}(\boldsymbol{\rho})$ do zera jednostajnie względem $\boldsymbol{\rho} \in \mathcal{P}$, z uwagi na oszacowanie

$$\big(R(\boldsymbol{\rho}) - \bar{R}(\boldsymbol{\rho})\big)^2 \leqslant \frac{1}{4}\ln^2\left(1 + \frac{\chi(\boldsymbol{\rho}) - \xi_1(\boldsymbol{\rho}) + \xi_2(\boldsymbol{\rho})}{\frac{\sigma_0^2}{N}\|\boldsymbol{\Delta}(\boldsymbol{\rho})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2}\right).$$

Zatem, aby skorzystać z lematu IV.2 pozostaje wykazać, że $\boldsymbol{\rho}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$. Rozumowanie przeprowadzimy nie wprost.

Zauważmy najpierw, że $\bar{R}(\boldsymbol{\rho}_0) \geqslant \bar{R}(\boldsymbol{\rho})$ dla każdego $\boldsymbol{\rho} \in \mathcal{P}$. Istotnie, przypominając (4.24) możemy wyliczyć, że

$$\bar{R}(\boldsymbol{\rho}_0) = \frac{1}{N} \ln |\det \boldsymbol{\Delta}(\boldsymbol{\rho}_0)| - \frac{1}{2} \ln(\sigma_0^2).$$

Zatem, na podstawie elementarnej nierówności między średnią arytmetyczną a średnią geometryczną mamy

$$\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\boldsymbol{\rho}) = \frac{1}{2} \ln \frac{\frac{1}{N} \|\boldsymbol{\Delta}(\boldsymbol{\rho}) \boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|_{\mathsf{F}}^2}{|\det \boldsymbol{\Delta}(\boldsymbol{\rho}) \boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}|^{\frac{2}{N}}} \geqslant 0,$$

gdzie licznik ułamka pod logarytmem można interpretować jako średnią arytmetyczną kwadratów wartości własnych macierzy $\boldsymbol{\Delta}(\boldsymbol{\rho}) \boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}$, a mianownik jako ich średnią geometryczną. Co więcej, na podstawie założenia IV.$\mathrm{E}_{\mathrm{SAR}}$ możemy stwierdzić, że

$$\liminf_{N \to \infty} \big(\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\boldsymbol{\rho}) \big) > 0, \qquad \boldsymbol{\rho} \in \mathcal{P}. \tag{4.28}$$

Załóżmy teraz przeciwnie, że $\boldsymbol{\rho}_0$ nie jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$, jak określono w nierówności (4.13), czyli

$$0 \leqslant \liminf_{N \to \infty} \Big(\inf_{\boldsymbol{\rho} \in \mathcal{P} \colon \|\boldsymbol{\rho} - \boldsymbol{\rho}_0\| > \delta} \bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\boldsymbol{\rho}) \Big) \leqslant 0.$$

Istnieje więc liczba $\delta > 0$ oraz ściśle rosnący ciąg liczb naturalnych $(N_n)_{n \in \mathbb{N}}$, tj. podciąg ciągu rozmiarów prób $(N)_{N \in \mathbb{N}}$, dla których można wskazać ciąg $(\widetilde{\boldsymbol{\rho}}_n)_{n \in \mathbb{N}}$ elementów przestrzeni $\mathcal{P}$ spełniający

$$\{\tilde{\boldsymbol{\rho}}_n\}_{n \in \mathbb{N}} \subset \mathcal{E}(\delta) \coloneqq \{\boldsymbol{\rho} \in \mathcal{P} : \|\boldsymbol{\rho} - \boldsymbol{\rho}_0\| \geqslant \delta\}$$

oraz

$$0 \leqslant \liminf_{N_n \to \infty} \big(\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\boldsymbol{\rho}) \big) \leqslant 0.$$

Zbiór $\mathcal{E}(\delta)$ jest domknięty w zwartej przestrzeni $\mathcal{P}$, a zatem jest sam zwarty. Możemy więc wybrać podciąg $(\widetilde{\boldsymbol{\rho}}_{m_n})_{n \in \mathbb{N}}$ ciągu $(\widetilde{\boldsymbol{\rho}}_n)_{n \in \mathbb{N}}$, zbieżny w $\mathcal{P}$. Oznaczając $\widetilde{\boldsymbol{\rho}} = \lim_{n \to \infty} \widetilde{\boldsymbol{\rho}}_{m_n}$, na podstawie własności (4.28), wnioskujemy, że liczba

$$\epsilon \coloneqq \liminf_{N \to \infty} \big(\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\boldsymbol{\rho}) \big)$$

jest dodatnia.

Z zaobserwowanej wcześniej ograniczoności pochodnej funkcji $\bar{R}$ oraz lematu IV.7 wynika, że $\bar{R}$ ma własność Lipschitza na przestrzeni $\mathcal{L}$ z pewną stałą $K_{\mathcal{P}}$, niezależną od wielkości próby. Dla dostatecznie dużych wartości indeksu n, tzn. dla wartości n przekraczających pewien poziom $n_0 \in \mathbb{N}$, mamy $\|\tilde{\boldsymbol{\rho}}_{m_n} - \tilde{\boldsymbol{\rho}}\| < \frac{\varepsilon}{3K_{\mathcal{P}}}$. Ostatecznie, oczekiwana sprzeczność wynika z nierówności trójkąta poprzez następujące oszacowanie

$$\begin{aligned}
\epsilon &= \liminf_{N\to\infty} \big(\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\tilde{\boldsymbol{\rho}})\big) \leqslant \liminf_{N_{m_n}\to\infty} \big(\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\tilde{\boldsymbol{\rho}})\big) \\
&\leqslant \liminf_{N_{m_n}\to\infty} \big(|\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\tilde{\boldsymbol{\rho}}_{m_n})| + |\bar{R}(\tilde{\boldsymbol{\rho}}_{m_n}) - \bar{R}(\tilde{\boldsymbol{\rho}})|\big) \\
&\leqslant \liminf_{N_{m_n}\to\infty} |\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\tilde{\boldsymbol{\rho}}_{m_n})| + \liminf_{N_{m_n}\to\infty} |\bar{R}(\tilde{\boldsymbol{\rho}}_{m_n}) - \bar{R}(\tilde{\boldsymbol{\rho}})| \\
&\leqslant \liminf_{N_{m_n}\to\infty} |\bar{R}(\boldsymbol{\rho}_0) - \bar{R}(\tilde{\boldsymbol{\rho}}_{m_n})| + K_{\mathcal{P}} \cdot \frac{\epsilon}{3K_{\mathcal{P}}} = \frac{\epsilon}{3}.
\end{aligned}$$

Zatem, $\boldsymbol{\rho}_0$ jest identyfikowalnie jedynym argumentem maksymalizującym funkcję $\bar{R}$. Ponieważ estymator $\hat{\boldsymbol{\rho}}$ jest — wprost ze swojej definicji — argumentem maksymalizującym funkcję R, określoną formułą (4.23), udowodniony fakt zbieżności jednostajnej według prawdopodobieństwa

$$\sup_{\boldsymbol{\rho}\in\mathcal{P}} \big|R(\boldsymbol{\rho}) - \bar{R}(\boldsymbol{\rho})\big| \xrightarrow{N\to\infty} 0$$

pozwala użyć lematu IV.1 do ustalenia zgodności estymatora $\hat{\boldsymbol{\rho}}$.

Pozostało zatem uzasadnić zgodność estymatorów $\hat{\boldsymbol{\beta}}$ oraz $\hat{\sigma}^2$. Uwzględniając (4.4) oraz pamiętając, że $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\rho}})$, możemy uzyskać

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Delta}(\hat{\boldsymbol{\rho}})\mathbf{y} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Delta}(\hat{\boldsymbol{\rho}})\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}(\mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) \\
&= (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\big((\mathbf{I} - \boldsymbol{\rho}_0^\mathsf{T}\mathbf{W}) + (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)^\mathsf{T}\mathbf{W}\big)\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}(\mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) \\
&= \mathbf{I}\cdot\boldsymbol{\beta}_0 + \boldsymbol{\zeta}_N + \boldsymbol{\zeta}'_N,
\end{aligned}$$

gdzie

$$\begin{aligned}
\boldsymbol{\zeta}_N &= (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)^\mathsf{T}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0, \\
\boldsymbol{\zeta}'_N &= (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Delta}(\hat{\boldsymbol{\rho}})\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}.
\end{aligned}$$

Pokażemy, że oba te składniki zaburzenia losowego estymatora $\hat{\boldsymbol{\beta}}$ zbiegają do zera według prawdopodobieństwa. Najpierw zauważmy, iż na mocy założenia IV.D mamy

$$\frac{1}{\sqrt{N}}\|\mathbf{X}^\mathsf{T}\| \leqslant \frac{1}{\sqrt{N}}\|\mathbf{X}^\mathsf{T}\|_\mathsf{F} \leqslant \frac{1}{\sqrt{N}}\sqrt{k\cdot\|\mathbf{X}^\mathsf{T}\mathbf{X}\|} \leqslant C\sqrt{k}, \tag{4.29}$$

dla pewnej stałej $C < \infty$. Otrzymujemy zatem nierówność

$$\mathbb{E}\,\|\boldsymbol{\zeta}_N\| = \mathbb{E}\,\Big\|\Big(\frac{1}{N}\mathbf{X}^\mathsf{T}\mathbf{X}\Big)^{-1}\frac{1}{\sqrt{N}}\mathbf{X}^\mathsf{T}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)^\mathsf{T}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\frac{1}{\sqrt{N}}\mathbf{X}\boldsymbol{\beta}_0\Big\| \leqslant C_{\boldsymbol{\zeta}}\cdot kd\cdot\|\boldsymbol{\beta}_0\|\,\mathbb{E}\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|,$$

gdzie

$$\infty > C_{\boldsymbol{\zeta}} \geqslant C^2 \sup_{N'\in\mathbb{N}}\Big\|\frac{1}{N'}\mathbf{X}^\mathsf{T}\mathbf{X}\Big\|\mathfrak{B}_{\mathbf{W}}\mathfrak{B}_{\text{inv}}.$$

Korzystając z nierówności Czebyszewa, dla dowolnej liczby $\epsilon > 0$ i $n \in \mathbb{N}$ otrzymujemy nierówność

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\zeta}_N\| \geqslant \epsilon\right) &\leqslant \frac{1}{\epsilon}\,\mathbb{E}\left(\|\boldsymbol{\zeta}_N\|\mathbf{I}_{\|\boldsymbol{\zeta}_N\|\geqslant\epsilon}\right) \leqslant \frac{C_{\boldsymbol{\zeta}}kd\|\boldsymbol{\beta}_0\|}{\epsilon}\,\mathbb{E}\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\| \\ &\leqslant \frac{C_{\boldsymbol{\zeta}}kd\|\boldsymbol{\beta}_0\|}{\epsilon}\,\mathbb{E}\left(\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|\mathbf{I}_{\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|>\frac{1}{n}} + \|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|\mathbf{I}_{\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|\leqslant\frac{1}{n}}\right) \\ &\leqslant \frac{C_{\boldsymbol{\zeta}}kd\|\boldsymbol{\beta}_0\|}{\epsilon}\mathfrak{B}_{\mathcal{P}}\,\mathbb{P}\left(\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\| > \frac{1}{n}\right) + \frac{C_{\boldsymbol{\zeta}}kd\|\boldsymbol{\beta}_0\|}{n\epsilon}.\end{aligned}$$

Wyliczając odpowiednie granice otrzymujemy zatem

$$0 \leqslant \lim_{n\to\infty}\left[\lim_{N\to\infty}\mathbb{P}\left(\|\boldsymbol{\zeta}_N\| \geqslant \epsilon\right)\right] \leqslant \lim_{n\to\infty}\frac{C_{\boldsymbol{\zeta}}kd\|\boldsymbol{\beta}_0\|}{n\epsilon} = 0,$$

a ponieważ wartość w nawiasie kwadratowym nie zależy od n, mamy zbieżność $\lim_{N\to\infty}\mathbb{P}\left(\|\boldsymbol{\zeta}_N\| \geqslant \epsilon\right) = 0$. Podobnie, używając submultiplikatywności normy macierzowej, wnioskujemy o zbieżności

$$\big\|\operatorname{Var}\boldsymbol{\zeta}'_N\big\| \leqslant C_{\boldsymbol{\zeta}'}\cdot k\cdot\frac{\|\operatorname{Var}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\mathsf{T}\|}{N} = \frac{C_{\boldsymbol{\zeta}'}\cdot k\sigma_0^2}{N} \xrightarrow{N\to\infty} 0,$$

gdzie stała $C_{\boldsymbol{\zeta}'}$ spełnia

$$\infty > C_{\boldsymbol{\zeta}'} \geqslant C^2 \sup_{N'\in\mathbb{N}}\Big\|\frac{1}{N'}\mathbf{X}^\mathsf{T}\mathbf{X}\Big\|^2\mathfrak{B}^2_{\text{inv}}\mathfrak{B}^2_{\boldsymbol{\Delta}}.$$

Oznacza to, że również każda z wariancji k współrzędnych $\zeta'_{N,m}$, $1 \leqslant m \leqslant k$ wektora $\boldsymbol{\zeta}'_N$ zbiega wraz ze wzrostem rozmiaru próby do zera, a co za tym idzie, dla każdego $m \leqslant k$ oraz $\epsilon > 0$, mamy

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\zeta}_N\| \geqslant \epsilon\right) &\leqslant \sum_{m=1}^{k}\mathbb{P}\left(|\zeta_{N,m}| \geqslant \epsilon\right) \leqslant \sum_{m=1}^{k}\frac{\operatorname{Var}\zeta'_{N,m}}{\epsilon} \\ &\leqslant \frac{k}{\epsilon}\|\operatorname{Var}\zeta'_{N,m}\| \xrightarrow{N\to\infty} 0.\end{aligned}$$

Aby wykazać zgodność estymatora $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\boldsymbol{\rho}})$, rozwijamy równanie (4.4) otrzymując

$$\begin{aligned}\hat{\sigma}^2(\hat{\boldsymbol{\rho}}) &= \frac{1}{N}\left\|\mathbf{y} - \hat{\boldsymbol{\rho}}^{\mathsf{T}}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\rho}})\right\|^2 \\ &= \frac{1}{N}(\boldsymbol{\varepsilon} - \boldsymbol{\vartheta}_N - \boldsymbol{\varphi}_N)^{\mathsf{T}}(\boldsymbol{\varepsilon} - \boldsymbol{\vartheta}_N - \boldsymbol{\varphi}_N) \\ &= \frac{\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon}}{N} + \frac{\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N} + \frac{\boldsymbol{\varphi}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N} + \frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N} + \frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varphi}_N}{N} + \frac{2\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N},\end{aligned} \tag{4.30}$$

gdzie

$$\begin{aligned}\boldsymbol{\vartheta}_N &= (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\mathbf{y}, \\ \boldsymbol{\varphi}_N &= \mathbf{X}\cdot(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).\end{aligned} \tag{4.31}$$

Zauważmy, że w konsekwencji słabego prawa wielkich liczb mamy zbieżność $\frac{1}{N}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ do σ_0^2 według prawdopodobieństwa. Istotnie, na mocy założenia IV.C zachodzi

$$\frac{1}{N}\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon} = \frac{1}{N}(\mathbf{E}\hat{\boldsymbol{\varepsilon}})^{\mathsf{T}}\mathbf{E}\hat{\boldsymbol{\varepsilon}} = \frac{1}{N}\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\bar{\boldsymbol{\varepsilon}} \xrightarrow{N\to\infty} \sigma_0^2,$$

gdyż dla dowolnego $\epsilon > 0$, uwzględniając lemat IV.6 mamy

$$\mathbb{P}\left(\left|\frac{1}{N}\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\bar{\boldsymbol{\varepsilon}} - \sigma_0^2\right| \geqslant \epsilon\right) \leqslant \frac{\operatorname{Var}(\bar{\boldsymbol{\varepsilon}}^{\mathsf{T}}\bar{\boldsymbol{\varepsilon}})}{N^2\epsilon} = \frac{3\sup_{N'\in\mathbb{N}}\sup_{i\leqslant\bar{N}(N')}\mathbb{E}\,\bar{\varepsilon}_i^4}{N\epsilon} \xrightarrow{N\to\infty} 0.$$

Wskażemy teraz ograniczenie dla pozostałych składników w równaniu (4.30). Łatwo też zauważyć, że $\left\|\frac{1}{\sqrt{N}}\boldsymbol{\varphi}_N\right\|$ dąży według prawdopodobieństwa do zera. Wynika to z obserwacji, iż przy dowolnym $\epsilon > 0$ zgodność estymatora $\hat{\boldsymbol{\beta}}$ implikuje ograniczenie

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{N}}\boldsymbol{\varphi}_N\right\| \geqslant \epsilon\right) \leqslant \mathbb{P}\left(\left\|\frac{1}{\sqrt{N}}\mathbf{X}\right\|\cdot\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \geqslant \epsilon\right) \xrightarrow{N\to\infty} 0.$$

Element $\boldsymbol{\vartheta}_N$ spełnia z kolei nierówność

$$\begin{aligned}\left\|\frac{1}{\sqrt{N}}\boldsymbol{\vartheta}_N\right\| \leqslant{}& \left\|\frac{1}{\sqrt{N}}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\right\| \\ &+ \left\|\frac{1}{\sqrt{N}}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\right\|.\end{aligned}$$

Z faktu zgodności estymatora $\hat{\boldsymbol{\rho}}$ wynika zbieżność

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{N}}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\right\|\geqslant\epsilon\right)$$
$$\leqslant\mathbb{P}\left(d\cdot\mathfrak{B}_{\mathbf{W}}\sup_{N'\in\mathbb{N}}\left\|\frac{1}{\sqrt{N'}}\mathbf{X}\right\|\cdot\|\boldsymbol{\beta}_0\|\cdot\mathfrak{B}_{\text{inv}}\cdot\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|\geqslant\epsilon\right)\xrightarrow{N\to\infty}0$$

i podobnie

$$\mathbb{P}\left(\left\|(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\right\|\geqslant\epsilon\right)$$
$$\leqslant\mathbb{P}\left(d\cdot\mathfrak{B}_{\mathbf{W}}\|\boldsymbol{\beta}_0\|\mathfrak{B}_{\text{inv}}\cdot\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0\|\geqslant\epsilon\right)\xrightarrow{N\to\infty}0.$$

Ta druga pozwala, poprzez lemat IV.9, wnioskować o zbieżności normy

$$\left\|\frac{1}{\sqrt{N}}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\right\|\leqslant\left\|(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0)^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\right\|\cdot\left\|\frac{1}{\sqrt{N}}\boldsymbol{\varepsilon}\right\|$$

do zera według prawdopodobieństwa, gdyż $\left\|\frac{1}{\sqrt{N}}\boldsymbol{\varepsilon}\right\|$ zbiega do wartości $\sqrt{\sigma_0^2}$, na mocy prawa wielkich liczb.

Ostatecznie, zgodność estymatora $\hat{\sigma}^2$ otrzymujemy, używając lematu IV.9 do oszacowania normy kolejnych składników w równości (4.30). Mianowicie mamy nierówność

$$\left|\frac{\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N}+\frac{\boldsymbol{\varphi}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}+\frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N}+\frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}+\frac{2\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}\right|$$
$$\leqslant\left\|\frac{\boldsymbol{\vartheta}_N}{\sqrt{N}}\right\|^2+\left\|\frac{\boldsymbol{\varphi}_N}{\sqrt{N}}\right\|^2+2\left\|\frac{\boldsymbol{\varepsilon}}{\sqrt{N}}\right\|\left\|\frac{\boldsymbol{\vartheta}_N}{\sqrt{N}}\right\|+2\left\|\frac{\boldsymbol{\varepsilon}}{\sqrt{N}}\right\|\left\|\frac{\boldsymbol{\varphi}_N}{\sqrt{N}}\right\|+2\left\|\frac{\boldsymbol{\vartheta}_N}{\sqrt{N}}\right\|\left\|\frac{\boldsymbol{\varphi}_N}{\sqrt{N}}\right\|,$$

a więc różnica

$$\left|\hat{\sigma}^2-\frac{1}{N}\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon}\right|=\left|\frac{\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N}+\frac{\boldsymbol{\varphi}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}+\frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\vartheta}_N}{N}+\frac{2\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}+\frac{2\boldsymbol{\vartheta}_N^{\mathsf{T}}\boldsymbol{\varphi}_N}{N}\right|$$

dąży do zera według prawdopodobieństwa. □

Rozdział V

Rozkład asymptotyczny estymatorów QNW dla modeli przestrzennych

Wstęp

Istotnym elementem każdej procedury estymacyjnej jest identyfikacja wynikowego rozkładu prawdopodobieństwa oszacowań. Taka wiedza pozwala na konstrukcję przedziałów ufności dla parametrów oraz testów statystycznych dla hipotez ich dotyczących. Cechą charakterystyczną metod estymacji opartych na zasadzie największej wiarogodności jest — poza pewnymi trywialnymi przypadkami — duża trudność, bądź wręcz niemożliwość, wyprowadzenia dokładnego rozkładu tychże oszacowań. W efekcie, wnioskowanie statystyczne musi być oparte na asymptotycznym zachowaniu estymatorów. Okazuje się, że pod pewnymi

warunkami odpowiedni rozkład graniczny istnieje i jest nim rozkład normalny. Odpowiednie twierdzenia o zbieżności rozkładów, choć nie dla procesów przestrzennych, są znane od dekad (por. Lehmann, Casella, 1998).

W tym rozdziale przedstawimy formalną teorię asymptotycznego rozkładu prawdopodobieństwa oszacowań, uzyskiwanych metodą quasi-największej wiarogodności. Wprawdzie zgodność tych estymatorów została wykazana w poprzednim podrozdziale, jednak jakiekolwiek wnioskowanie statystyczne na podstawie szacowanego modelu, w tym określenie istotności statystycznej jego składników, wymaga identyfikacji rozkładu użytych estymatorów. Podobnie wyznaczenie przedziałów ufności dla estymowanych parametrów jest możliwe tylko wtedy, gdy znamy co najmniej przybliżone wartości dystrybuanty, właściwej dla uzyskiwanych oszacowań.

Poniżej prezentujemy i formalnie udowadniamy autorskie centralne twierdzenie graniczne dla form liniowo-kwadratowych. Na tym wyniku opiera się argumentacja omawianych dalej w rozdziale własności estymatorów QNW. Twierdzenie to wykorzystaliśmy również w rozdziale III do uzyskania rozkładów asymptotycznych statystyk testowych autokorelacji przestrzennej. Podrozdział drugi zawiera serię lematów i stwierdzeń pomocniczych, które okażą się przydatne w kolejnych rozumowaniach. W podrozdziale trzecim przedstawiamy i udowadniamy autorskie, nigdy wcześniej niepublikowane twierdzenie o rozkładzie asymptotycznym estymatora quasi-największej wiarogodności dla modeli wyższych rzędów z autoskorelowanym składnikiem losowym. Ostatni podrozdział jest poświęcony autorskiemu opracowaniu analogicznego twierdzenia dla modeli z autoregresją zmiennej objaśnianej, które zostało pierwotnie opublikowane w pracy Olejnik i Olejnik (2020).

1. Nowe centralne twierdzenie graniczne dla form liniowo-kwadratowych

Narzędziem matematycznym używanym w teorii asymptotycznej estymacji parametrów modeli ekonometrycznych jest pojęcie zbieżności ciągu zmiennych losowych. Gdy mówimy o zgodności estymatora, rozważamy jego zbieżność według prawdopodobieństwa do wartości prawdziwej. W przypadku asymptotycznego rozkładu oszacowań właściwym pojęciem jest zbieżność ciągu zmiennych losowych według dystrybuanty. Intuicyjnie, pojęcie takiej zbieżności można rozumieć jako właściwość upodabniania się dystrybuanty zmiennej losowej do dystrybuanty rozkładu granicznego, wraz ze wzrostem rozmiaru próby. W praktyce oznacza to, że jesteśmy w stanie uniknąć konieczności wyprowadzania dokładnego rozkładu rozważanych statystyk. Zamiast tego możemy (przy pewnej ostrożności w przypadku rozkładów nieciągłych) przybliżać wartości prawdopodobieństw

zdarzeń związanych z elementami ciągu wartościami wyliczonymi z użyciem dystrybuanty granicznej. Jest to o tyle istotne, że wyprowadzenie rozkładu dokładnego może być trudne lub wręcz niemożliwe. W szczególności tak jest właśnie w przypadku estymacji metodą największej wiarogodności, gdy prawdziwy rozkład prawdopodobieństwa, z którego pochodzą dane nie jest znany (por. Założenia IV.E_{SAR} i IV.E_{SEM}).

W teorii estymacji parametrów modeli ekonometrycznych najczęściej pojawiającym się rozkładem granicznym jest rozkład normalny, a twierdzenia o zbieżności według dystrybuanty o gaussowskiej granicy nazywa się zwyczajowo centralnymi twierdzeniami granicznymi. Do rozwoju teorii ekonometrii nie wystarczają jednak standardowe, kursowe twierdzenia. Poza prostymi przypadkami, takimi jak wyliczanie średniej arytmetycznej z próby o niezależnych obserwacjach, jest wręcz przeciwnie. Zależności między obserwacjami w próbie wymagają użycia specjalistycznego twierdzenia granicznego, uwzględniającego specyficzną ich naturę. W szczególności, takiego, które pozwoli uwzględnić reprezentację zależności w formie macierzy wag, użytej w specyfikacji autoregresyjnej procesu generującego obserwacje.

W przypadku rozważań dotyczących ekonometrii przestrzennej, szczególną postacią wyrażenia, które decyduje o rozkładzie badanego estymatora (patrz dowody twierdzeń V.8 oraz V.9) jest forma kwadratowo-liniowa zaburzenia losowego modelu. Podobnie jest w przypadku statystyk przestrzennych typu I Morana (patrz rozdział III), których postać jest *explicite* ilorazem form kwadratowych bądź liniowo-kwadratowych pewnej zmiennej losowej, związanej z obserwowanym zjawiskiem. Przyjmijmy, że mamy do czynienia z funkcjami postaci $\mathbb{R}^N \ni \xi \mapsto \xi^{\mathsf{T}} \boldsymbol{A} \xi$ o niediagonalnej macierzy $\boldsymbol{A} = [a_{ij}]_{1 \leqslant i,j \leqslant N}$, której wartość obliczono dla wektorowej zmiennej losowej $\xi = (\xi_i)_{i=1}^N$, o elementach niezależnych według prawdopodobieństwa. Wówczas, w sumie $\sum_{i=1}^N \sum_{j=1}^N a_{ij} \xi_i \xi_j$, poza wciąż niezależnymi składnikami kwadratowymi ξ_i^2, gdzie $1 \leqslant i \leqslant N$, pojawiają się składniki iloczynów krzyżowych $\xi_i \xi_j$, dla $i \neq j$. W sposób oczywisty nie są one już niezależne, gdyż, dla każdego $1 \leqslant i \leqslant N$, wartość zmiennej losowej $\xi_i \xi_j$ będzie związana (poza przypadkami trywialnymi) z wartościami zmiennej $\xi_i \xi_k$, dla $1 \leqslant j, k \leqslant N$.

Warto tu wyjaśnić, że pojawienie się form liniowo-kwadratowych jest naturalnie związane z zastosowaniem metody największej wiarogodności. Wynika to z zastosowania rozwinięcia w szereg Taylora pochodnej funkcji wiarogodności do składnika rzędu drugiego, tak jak ma to miejsce w prezentowanych dowodach twierdzeń V.8 i V.9, patrz również praca Feng i inni (2014).

Literatura przedmiotu oferuje cały wachlarz twierdzeń granicznych dla rozkładów form kwadratowych przy różnych założeniach. Temat ten był podejmowany między innymi w pracach Whittle (1964), Beran (1972), Sen (1976), de Jong

(1987), Giraitis i Taqqu (1998). W kontekście teorii ekonometrii przestrzennej najczęściej wykorzystuje się twierdzenie z pracy Kelejiana i Pruchy (2001) dla form liniowo-kwadratowych o macierzach jednostajnie sumowalnych w sensie modułów wierszy i kolumn. Na szczególną uwagę zasługuje opracowanie Bhansali i inni (2007), w którym sformułowano centralne twierdzenie graniczne dla form czysto kwadratowych o macierzy z dowolną przekątną, oraz z warunkami wyrażonymi w języku norm spektralnych. Bazując na rozumowaniu zaczerpniętym z tej pracy, prezentujemy jednak wynik rozszerzony na użytek naszej teorii — twierdzenie centralne o zbieżności rozkładów form liniowo-kwadratowych. Wynik ten pierwotni został przedstawiony w materiałach dodatkowych do pracy Olejnik i Olejnik (2020).

Twierdzenie V.1

Niech $\boldsymbol{\varepsilon}_N = (\varepsilon_i)_{i=1}^N$ będzie wektorem niezależnych zmiennych losowych (por. założenie V.C' na s. 143) o zerowej wartości oczekiwanej i stałej wariancji $\sigma_0^2 > 0$. Ponadto załóżmy, że rodzina ich czwartych potęg, tj. ε_i^4, dla $1 \leqslant i \leqslant N$, jest jednostajnie całkowalna ze względu na indeks i oraz rozmiar próby N (patrz definicja na s. 143). Dla dowolnej liczby naturalnej N niech $\boldsymbol{x}_N = (x_i)_{i=1}^N \in \mathbb{R}^N$. Dodatkowo, niech $\boldsymbol{A}_N = (a_{ij})_{1 \leqslant i,j \leqslant N}$ będą macierzami o wymiarach $N \times N$. Oznaczmy $Q_N \coloneqq \boldsymbol{\varepsilon}_N^{\mathsf{T}} \boldsymbol{A}_N \boldsymbol{\varepsilon}_N + \boldsymbol{\varepsilon}^{\mathsf{T}} \boldsymbol{x}_N$ oraz załóżmy, że $\operatorname{Var} Q_N > 0$, dla dostatecznie dużych N. Jeśli spełnione są warunki

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{x}_N\|^2}{\operatorname{Var} Q_N} < \infty, \tag{5.1}$$

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{x}_N\|_\infty^2}{\operatorname{Var} Q_N} = 0, \tag{5.2}$$

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{A}_N\|^2}{\operatorname{Var} Q_N} = 0, \tag{5.3}$$

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{A}_N\|_{\mathsf{F}}^2}{\operatorname{Var} Q_N} < \infty, \tag{5.4}$$

gdzie $\|\boldsymbol{x}_N\|_\infty^2 = \max_{1 \leqslant i \leqslant N} |x_i|$, wówczas forma standaryzowana $\frac{Q_N - \mathbb{E}\, Q_N}{\sqrt{\operatorname{Var} Q_N}}$ zbiega według dystrybuanty do zmiennej o rozkładzie $\mathcal{N}(0,1)$.

Dowód. Bez straty ogólności możemy założyć, że $\sigma_0^2 = 1$. Zauważmy, że macierz $\boldsymbol{A}'_N = \frac{1}{2}\boldsymbol{A}_N + \frac{1}{2}\boldsymbol{A}_N^{\mathsf{T}}$ jest symetryczna i, użyta w miejscu A_n, wyznacza identyczną część kwadratową formy Q_N. Ze względu na nierówności $\|\boldsymbol{A}'_N\|^2 \leqslant \|\boldsymbol{A}_N\|^2$ oraz $\|\boldsymbol{A}'_N\|_{\mathsf{F}}^2 \leqslant \|\boldsymbol{A}_N\|_{\mathsf{F}}^2$, warunki (5.3) i (5.4) są spełnione dla macierzy $\boldsymbol{A}'_N$ (w miejscu $\boldsymbol{A}_N$) wtedy, gdy są one spełnione dla $\boldsymbol{A}_N$. Zatem, na użytek dowodu możemy założyć, że rozważana macierz $\boldsymbol{A}_N$ jest symetryczna.

Mając na uwadze zwartość zapisu, wprowadzimy następujące oznaczenia. Dla dowolnego $N \in \mathbb{N}$ oraz $1 \leqslant i \leqslant N$ przyjmujemy

$$\begin{aligned} \mathcal{V}_N &\coloneqq \operatorname{Var} Q_N, \\ \varsigma_i^2 = \varsigma_i^2(N) &\coloneqq \operatorname{Var} \varepsilon_i^2, \\ \eta_i^2 = \eta_i^2(N) &\coloneqq \mathbb{E}\, \varepsilon_i^3, \end{aligned}$$

a następnie

$$U_i = U_i(N) \coloneqq a_{ii} \cdot (\varepsilon_i^2 - 1) + x_i \varepsilon_i + 2\varepsilon_i \sum_{j=1}^{i-1} a_{ij} \varepsilon_j.$$

Jak łatwo sprawdzić, dla każdego ustalonego $N = 1, 2, \ldots$ zachodzi równość

$$\sum_{i=1}^{N} U_i = Q_N - \mathbb{E}\, Q_N.$$

Co więcej, można zaobserwować, że każdy skończony ciąg $(U_i)_{i=1}^N$ jest martyngałem (patrz Jakubowski, Sztencel, 2001). Istotnie, niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie odpowiednią przestrzenią probabilistyczną oraz niech $\mathcal{F}_i = \mathcal{F}_i(N)$, dla $1 \leqslant i \leqslant N$, będą σ-ciałami generowanymi przez zmienne poprzedzające U_i z U_i włącznie, tj. $\mathcal{F}_i \coloneqq \sigma(U_1, \ldots, U_i)$. Załóżmy też, że σ-ciało $\mathcal{F}_0 = \mathcal{F}_0(N) \subset \mathcal{F}$ jest niezależne (według prawdopodobieństwa) od wszystkich zmiennych ε_i, $1 \leqslant i \leqslant N$, np. $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Wtedy, używając założenia o łącznej niezależności zmiennych $(\varepsilon_i)_{i=1}^N$ mamy

$$\begin{aligned} \mathbb{E}\,[U_i \mid \mathcal{F}_{i-1}] &= \mathbb{E}\left(a_{ii}(\varepsilon_i^2 - 1) \mid \mathcal{F}_{i-1}\right) + \mathbb{E}\,(x_i \varepsilon_i \mid \mathcal{F}_{i-1}) \\ &\qquad + 2\,\mathbb{E}\left(\varepsilon_i \sum_{j=1}^{i-1} a_{ij} \varepsilon_j \mid \mathcal{F}_{i-1}\right) \\ &= a_{ii}\,\mathbb{E}\left(\varepsilon_i^2 - 1\right) + x_i\,\mathbb{E}\,(\varepsilon_i) + 2\,\mathbb{E}\,(\varepsilon_i) \sum_{j=1}^{i-1} a_{ij} \varepsilon_j = 0. \end{aligned}$$

Dla dowolnej wartości $N = 1, 2, \ldots$ oraz dowolnej liczby $\delta > 0$ dodatkowo oznaczmy

$$\begin{aligned} S_N &= \frac{1}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\left[U_i^2 \mid \mathcal{F}_{i-1}\right], \\ q_N(\delta) &= \frac{1}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\left[U_i^2 \mathbb{I}_{\{U_i^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} \mid \mathcal{F}_{i-1}\right]. \end{aligned}$$

Przedstawiane poniżej rozumowanie opiera się na centralnym twierdzeniu granicznym dla trójkątnej tablicy różnic martyngałowych, które można znaleźć we wniosku 3.1 w monografii Hall i Hyde (1980). Zgodnie z tym twierdzeniem, zbieżność według prawdopodobieństwa

$$\lim_{N\to\infty} S_N = 1 \tag{5.5}$$

oraz zbieżność według prawdopodobieństwa dla dowolnego $\delta > 0$

$$\lim_{N\to\infty} q_N(\delta) = 1 \tag{5.6}$$

implikują asymptotyczną normalność sumy $\frac{1}{\mathcal{V}_N}\sum_{i=1}^{N} U_i$. Ta z kolei sprowadza się do tezy dowodzonego twierdzenia.

Przy oznaczeniu $T_i = 2\sum_{j=1}^{i-1} a_{ij}\varepsilon_j$ możemy zapisać

$$\begin{aligned} U_i^2 &= \left(a_{ii}(\varepsilon_i^2 - 1) + x_i\varepsilon_i + \varepsilon_i T_i\right)^2 \\ &= 2a_{ii}x_i\varepsilon_i(\varepsilon_i^2 - 1) + 2a_{ii}\varepsilon_i(\varepsilon_i^2 - 1)T_i + 2x_i\varepsilon_i^2 T_i \\ &\qquad + \varepsilon_i^2 T_i^2 + a_{ii}^2(\varepsilon_i^2 - 1)^2 + x_i^2\varepsilon_i^2, \end{aligned} \tag{5.7}$$

a następnie

$$S_N = \frac{1}{\mathcal{V}_N}\sum_{i=1}^{N} \mathbb{E}\left[U_i^2 \mid \mathcal{F}_{i-1}\right] = 2\eta_i^3 a_{ii}x_i + 2\eta_i^3 a_{ii}T_i + 2x_iT_i + T_i^2 + \varsigma_i^2 a_{ii}^2 + x_i^2.$$

Z własności martyngału dla ciągu zmiennych losowych U_i, $1 \leqslant i \leqslant N$, mamy

$$\begin{aligned} \mathcal{V}_N &= \mathbb{E}\left(Q_N - \mathbb{E}\,Q_N\right)^2 = \mathbb{E}\left(\sum_{i=1}^{N} U_i\right)^2 = \sum_{i=1}^{N}\mathbb{E}\left(U_i^2 + 2\sum_{j=1}^{i-1} U_iU_j\right) \\ &= \sum_{i=1}^{N}\mathbb{E}\,U_i^2 + 2\sum_{i=1}^{N}\mathbb{E}\left(\sum_{j=1}^{i-1}\mathbb{E}\left(U_iU_j \mid \mathcal{F}_j\right)\right) \\ &= \sum_{i=1}^{N}\mathbb{E}\,U_i^2 + 2\sum_{i=1}^{N}\mathbb{E}\left(\sum_{j=1}^{i-1} U_j\cdot\mathbb{E}\left(U_i \mid \mathcal{F}_j\right)\right) = \sum_{i=1}^{N}\mathbb{E}\,U_i^2 + 0. \end{aligned}$$

Zatem, podstawiając tożsamość (5.7) oraz uwzględniając niezależność zmiennych ε_i i T_i (gdzie $T_1 := 0$) według prawdopodobieństwa oraz fakt zerowania się

wartości oczekiwanej $\mathbb{E}\,T_i$, dla każdego $1 \leqslant i \leqslant N$, uzyskujemy równość

$$\begin{aligned}\mathcal{V}_N &= 2\sum_{i=1}^{N} a_{ii}x_i\,\mathbb{E}\,\varepsilon_i\left(\varepsilon_i^2-1\right) + 2\sum_{i=1}^{N} a_{ii}\,\mathbb{E}\,\varepsilon_i\left(\varepsilon_i^2-1\right)T_i + 2\sum_{i=1}^{N} x_i\,\mathbb{E}\,\varepsilon_i^2 T_i \\ &\qquad + \sum_{i=1}^{N}\mathbb{E}\,\varepsilon_i^2 T_i^2 + \sum_{i=1}^{N} a_{ii}^2\,\mathbb{E}\,(\varepsilon_i^2-1)^2 + \sum_{i=1}^{N} x_i^2\,\mathbb{E}\,\varepsilon_i^2 \\ &= 2\sum_{i=1}^{N}\eta_{n,i}^3 x_{n,i}a_{n,ii} + \sum_{i=1}^{N}\mathbb{E}\,T_i^2 + \sum_{i=1}^{N}\varsigma_i^2 a_{ii}^2 + \sum_{i=1}^{N} x_i^2.\end{aligned}$$

Poniżej wykażemy, że zbieżność wyrażona w równaniu (5.6) zachodzi średniokwadratowo w przestrzeni $L^2(\Omega,\mathcal{F},\mathbb{P})$, co implikuje żądaną zbieżność według prawdopodobieństwa. Dla dowolnego $N = 1, 2, \ldots$ mamy

$$(S_N - 1)^2 = \frac{1}{\mathcal{V}_N^2}\left(\mathcal{V}_N S_N - \mathcal{V}_N\right)^2.$$

Uwzględniając wyliczone wcześniej formuły na S_N i $\mathcal{V}_N$ otrzymujemy

$$\begin{aligned}(S_N-1)^2 &= \mathcal{V}_N^{-2}\cdot\left(\sum_{i=1}^{N}\left(2\eta_i^3 a_{ii}T_i + 2x_iT_i + (T_i^2 - \mathbb{E}\,T_i^2)\right)\right)^2 \\ &\leqslant \frac{6}{\mathcal{V}_N^2}\left(\sum_{i=1}^{N}\eta_i^3 a_{ii}T_i\right)^2 + \frac{6}{\mathcal{V}_N^2}\left(\sum_{i=1}^{N} x_iT_i\right)^2 + \frac{3}{\mathcal{V}_N^2}\left(\sum_{i=1}^{N}\left(T_i^2-\mathbb{E}\,T_i^2\right)\right)^2.\end{aligned}$$

Zatem, przykładając operator wartości oczekiwanej uzyskujemy oszacowanie

$$\|S_N - 1\|_{L^2} = \mathbb{E}\,(S_N-1)^2 \leqslant \frac{6}{\mathcal{V}_N^2}\xi_N^{(1)} + \frac{6}{\mathcal{V}_N^2}\xi_N^{2} + \frac{3}{\mathcal{V}_N^2}\xi_N^{(3)}, \tag{5.8}$$

gdzie

$$\begin{aligned}\xi_N^{(1)} &= \mathbb{E}\left(\sum_{i=1}^{N} a_{ii}T_i\right)^2, \\ \xi_N^{(2)} &= \mathbb{E}\left(\sum_{i=1}^{N} x_iT_i\right)^2, \\ \xi_N^{(3)} &= \mathbb{E}\left(\sum_{i=1}^{N}\left(T_i^2-\mathbb{E}\,T_i^2\right)\right)^2.\end{aligned}$$

Dalej wskażemy ograniczenie każdego z powyższych elementów oddzielnie, zaczynając od wartości $\xi_N^{(3)}$. Uwzględniając tożsamości

$$T_i^2 = 4\sum_{k=1}^{i-1}\sum_{l=1}^{i-1} a_{ik}a_{il}\varepsilon_j\varepsilon_k,$$
$$\mathbb{E}\,T_i^2 = 4\sum_{k=1}^{i-1} a_{ik}^2,$$

dla $1 \leqslant i \leqslant N$, otrzymujemy oszacowanie

$$\begin{aligned}\xi_N^{(3)} &= 16\cdot\mathbb{E}\left(\sum_{i=1}^{N}\sum_{l=1}^{i-1}\sum_{k=1}^{i-1} a_{il}a_{ik}\varepsilon_l\varepsilon_k - \sum_{i=1}^{N}\sum_{l=1}^{i-1} a_{il}^2\right)^2\\ &= 16\cdot\mathbb{E}\left(\sum_{i=1}^{N}\sum_{l=1}^{i-1}\sum_{\substack{k=1\\k\neq l}}^{i-1} a_{il}a_{ik}\varepsilon_l\varepsilon_k + \sum_{i=1}^{N}\sum_{l=1}^{i-1} a_{il}^2\varepsilon_l^2 - \sum_{i=1}^{N}\sum_{l=1}^{i-1} a_{il}^2\right)^2\\ &\leqslant 32\cdot\mathbb{E}\left(\sum_{i=1}^{N}\sum_{l=1}^{i-1}\sum_{\substack{k=1\\k\neq l}}^{i-1} a_{il}a_{ik}\varepsilon_l\varepsilon_k\right)^2 + 32\cdot\mathbb{E}\left(\sum_{i=1}^{N}\sum_{k=1}^{i-1} a_{ik}^2(\varepsilon_k^2-1)\right)^2.\end{aligned}$$

Rozwijając kwadrat wyrażenia w nawiasie możemy zauważyć, że pierwszy składnik po lewej stronie powyższej nierówności można zapisać jako

$$\begin{aligned}&\mathbb{E}\left(\sum_{i=1}^{N}\sum_{l=1}^{i-1}\sum_{\substack{k=1\\k\neq l}}^{i-1} a_{il}a_{ik}\varepsilon_l\varepsilon_k\right)^2\\ &\qquad= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{i-1}\sum_{l'=1}^{j-1}\sum_{\substack{k=1\\k\neq l}}^{i-1}\sum_{\substack{k'=1\\k'\neq l'}}^{j-1} a_{il}a_{ik}a_{i'l'}a_{i'k'}\,\mathbb{E}\,\varepsilon_l\varepsilon_k\varepsilon_{l'}\varepsilon_{k'}.\end{aligned}$$

Zauważmy, że czynnik $\mathbb{E}\,\varepsilon_j\varepsilon_k\varepsilon_{j'}\varepsilon_{k'}$ jest różny od zera tylko wtedy, gdy $l = l'$ i $k = k'$, lub gdy $j = k'$ i $k = j'$. Oba te przypadki, uwzględnione w odpowiednich sumach, prowadzą do takiego samego wyrażenia. Zatem, badany składnik jest dalej ograniczony przez

$$2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{\min\{i,j\}-1}\sum_{\substack{k=1\\k\neq l}}^{\min\{i,j\}-1} a_{il}a_{ik}a_{jl}a_{jk}\,\mathbb{E}\,\varepsilon_l^2\varepsilon_k^2$$

$$\leqslant 2\left(\sup_{i\leqslant N}\mathbb{E}\,\varepsilon_i^4\right)\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{\min\{i,j\}-1}\sum_{\substack{k=1\\k\neq l}}^{\min\{i,j\}-1}a_{il}a_{ik}a_{jl}a_{jk}.$$

Drugi składnik sumy szacującej element $\xi_N^{(3)}$ spełnia nierówność

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^{N}\sum_{k=1}^{i-1}a_{ik}^2(\varepsilon_k^2-1)\right)^2 &= \mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{i-1}\sum_{l=1}^{j-1}a_{ik}^2a_{jl}^2\big(\varepsilon_k^2-1\big)\big(\varepsilon_l^2-1\big)\\ &\leqslant\left(\sup_{i\leqslant N}\varsigma_i^2\right)\cdot\mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{\min\{i,j\}-1}a_{ik}^2a_{jk}^2\\ &=\left(\sup_{i\leqslant N}\varsigma_i^2\right)\cdot\mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{\min\{i,j\}-1}\sum_{\substack{l=1\\l=k}}^{\min\{i,j\}-1}a_{ik}a_{il}a_{jk}a_{jl}.\end{aligned}$$

Łącząc powyższe oszacowania wnioskujemy, że dla pewnej stałej

$$\infty > C \geqslant 32\cdot\sup_{i\leqslant N}(\varsigma_i^2+2\,\mathbb{E}\,\varepsilon_i^4)$$

mamy

$$\xi_N^{(3)}\leqslant C\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{\min\{i,j\}-1}\sum_{k=1}^{\min\{i,j\}-1}a_{ik}a_{il}a_{jk}a_{jl}=C\cdot B_N,$$

przy czym B_N zdefiniowane jest jako

$$\begin{aligned}B_N &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{k=1}^{\min\{i,j\}-1}a_{ik}a_{jk}\right)^2\\ &=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{\min\{i,j\}-1}\sum_{k=1}^{\min\{i,j\}-1}a_{ik}a_{il}a_{jk}a_{jl}.\end{aligned}\tag{5.9}$$

Wartość B_N jest z kolei ograniczona przez $2\|\boldsymbol{A}_N\|^2\cdot\|\boldsymbol{A}_N\|_{\mathsf{F}}^2$, co wykazujemy

w następujący sposób:

$$\begin{aligned} B_N &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\Bigg(\sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk}\Bigg)^2 + \sum_{i=1}^{N}\sum_{j=i}^{N}\Bigg(\sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk}\Bigg)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\Bigg(\sum_{k=1}^{j-1} a_{ik}a_{jk}\Bigg)^2 + \sum_{i=1}^{N}\sum_{j=i}^{N}\Bigg(\sum_{k=1}^{i-1} a_{ik}a_{jk}\Bigg)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\Bigg(\sum_{k=1}^{j-1} a_{ik}a_{jk}\Bigg)^2 + \sum_{j=1}^{N}\sum_{i=j}^{N}\Bigg(\sum_{k=1}^{j-1} a_{ik}a_{jk}\Bigg)^2 \\ &\leqslant 2\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg(\sum_{k=1}^{j-1} a_{ik}a_{jk}\Bigg)^2 = 2\Big\|\boldsymbol{A}_N\cdot\big[a_{ij}\mathbb{I}_{\{i<j\}}\big]_{i,j=1}^{N}\Big\|_{\mathsf{F}}^2 \\ &\leqslant 2\|\boldsymbol{A}_N\|^2\cdot\|\boldsymbol{A}_N\|_{\mathsf{F}}^2, \end{aligned} \tag{5.10}$$

gdzie przechodząc pomiędzy drugą a trzecią linijką zamieniamy rolami symbole i i j w drugim składniku sumy. Użyty powyżej symbol $\mathbb{I}_{\{i<j\}} = \mathbb{I}_{\{i<j\}}(i,j)$, dla $1 \leqslant i,j \leqslant N$, jest indykatorem relacji w indeksie dolnym — przyjmujemy, że jest równy jeden, gdy $i<j$, a zero, gdy $i \geqslant j$.

Aby wskazać ograniczenie dla elementu $\xi_N^{(2)}$, zaobserwujmy równość

$$\begin{aligned} \xi_N^{(2)} &= \mathbb{E}\Bigg(\sum_{i=1}^{N} x_i T_i\Bigg)^2 = 4\,\mathbb{E}\Bigg(\sum_{i=1}^{N} x_i \sum_{k=1}^{i-1} a_{ik}\varepsilon_k\Bigg)^2 \\ &= 4\sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sum_{k=1}^{i-1}\sum_{l=1}^{j-1} a_{ik}a_{jk}\,\mathbb{E}\,\varepsilon_k\varepsilon_l = 4\sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk}. \end{aligned}$$

Następnie, stosując nierówność Schwartza uzyskujemy

$$\begin{aligned} \xi_N^{(2)} &\leqslant 4\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg(\sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk}\Bigg)^2} \\ &\leqslant 4\sum_{j=1}^{N} x_i^2\cdot\sqrt{B_N} \leqslant 4\sqrt{2}\cdot\|\boldsymbol{x}_N\|^2\cdot\|\boldsymbol{A}_N\|\cdot\|\boldsymbol{A}_N\|_{\mathsf{F}}. \end{aligned} \tag{5.11}$$

Składnik $\xi_N^{(1)} = \mathbb{E}\left(\sum_{i=1}^N a_{ii}T_i\right)^2$ szacujemy podobnie. Najpierw zauważamy, że

$$\begin{aligned}\xi_N^{(1)} &= \mathbb{E}\left(\sum_{i=1}^N a_{ii}T_i\right)^2 = 4\,\mathbb{E}\left(\sum_{i=1}^N a_{ii}\sum_{k=1}^{i-1} a_{ik}\varepsilon_k\right)^2\\ &= 4\sum_{i=1}^N\sum_{j=1}^N a_{ii}a_{jj}\sum_{k=1}^{i-1}\sum_{l=1}^{j-1} a_{ik}a_{jk}\,\mathbb{E}\,\varepsilon_k\varepsilon_l\\ &= 4\sum_{i=1}^N\sum_{j=1}^N a_{ii}a_{jj}\sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk},\end{aligned}$$

a następnie nierówność Schwartza implikuje

$$\begin{aligned}\xi_N^{(1)} &\leqslant 4\sqrt{\sum_{i=1}^N\sum_{j=1}^N x_ix_j}\sqrt{\sum_{i=1}^N\sum_{j=1}^N\left(\sum_{k=1}^{\min\{i,j\}-1} a_{ik}a_{jk}\right)^2} \leqslant 4\sum_{j=1}^N x_i^2\cdot\sqrt{B_N}\\ &\leqslant 4\sqrt{2}\cdot\|\boldsymbol{A}_N\|\cdot\|\boldsymbol{A}_N\|_{\mathsf{F}}^3.\end{aligned}$$

Ostatecznie, sięgając do nierówności (5.8), wnioskujemy, że

$$\begin{aligned}\|S_N-1\|_{L^2} &\leqslant \frac{6}{\mathcal{V}_N^2}\xi_N^{(1)} + \frac{6}{\mathcal{V}_N^2}\xi_N^{(2)} + \frac{3}{\mathcal{V}_N^2}\xi_N^{(3)}\\ &\leqslant \frac{C'}{\mathcal{V}_N^2}\left(\|\boldsymbol{x}_N\|^2\|\boldsymbol{A}_N\|\|\boldsymbol{A}_N\|_{\mathsf{F}} + \|\boldsymbol{A}_N\|\|\boldsymbol{A}_N\|_{\mathsf{F}}^3 + \|\boldsymbol{A}_N\|^2\|\boldsymbol{A}_N\|_{\mathsf{F}}^2\right).\end{aligned}$$

Uwzględniając założenia (5.1), (5.3) i (5.4), uzyskujemy zatem zbieżność (5.5).

Aby zakończyć dowód, musimy jeszcze wykazać zależność (5.6). Na mocy nierówności Czebyszewa, dla dowolnych $\delta,\tau>0$ mamy

$$\mathbb{P}\left(|q_N(\delta)-0|>\tau\right) \leqslant \frac{\mathbb{E}\,|q_N(\delta)|}{\tau}.$$

Ponieważ z definicji $q_N(\delta)\geqslant 0$, wystarczy wykazać, że $\mathbb{E}\,q_N(\delta)$ zbiega do zera, dla każdej wartości $\delta>0$.

Niech $K>1$ będzie dowolnie ustaloną liczbą. Dla $N\in\mathbb{N}$ oraz $1\leqslant i\leqslant N$, zdefiniujmy

$$\begin{aligned}Z_i = Z_i(N) &\coloneqq \varepsilon_i\mathbb{I}_{\{\varepsilon_i<K\}} - \mathbb{E}\,\varepsilon_i\mathbb{I}_{\{\varepsilon_i<K\}},\\ H_i = H_i(N) &\coloneqq \varepsilon_i - Z_i,\\ u_i = u_i(N) &\coloneqq 2Z_i\sum_{j=1}^{i-1} a_{ij}Z_j + a_{ii}(Z_i^2 - \mathbb{E}\,Z_i^2) + x_iZ_i,\\ h_i = h_i(N) &\coloneqq U_i - u_i.\end{aligned} \tag{5.12}$$

Zauważmy, że $\mathbb{E}\, Z_i = EH_i = 0$, dla wszystkich $1 \leqslant i \leqslant N$. Ponadto, elementy zarówno ciągu $Z_1, \ldots, Z_N$, jak i ciągu $H_1, \ldots, H_N$ są niezależne, gdyż są funkcjami zmiennych losowych ε_i, o odpowiadających indeksach $1 \leqslant i \leqslant N$.

Z definicji wynika, że dla dowolnego $1 \leqslant i \leqslant N$ mamy $U_i^2 \leqslant 2u_i^2 + 2h_i^2$ prawie pewnie. Otrzymujemy więc oszacowanie

$$\begin{aligned} q_n(\delta) &= \frac{1}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, U_i^2 \mathbb{I}_{\{U_i^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} \\ &\leqslant \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, U_i^2 \mathbb{I}_{\{U_i^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} + \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, h_i^2. \end{aligned}$$

Dla dowolnego zdarzenia elementarnego w przestrzeni Ω oraz $1 \leqslant i \leqslant N$ spełniony jest jeden z dwóch warunków: albo $|u_i| \geqslant \frac{1}{2}|U_i|$, albo $|u_i| < \frac{1}{2}|U_i| < |h_i|$. Istotnie, w przypadku, gdy $|u_i| < \frac{1}{2}|U_i|$ mamy

$$|U_i| = |u_i + h_i| \leqslant |u_i| + |h_i| < \frac{1}{2}|U_i| + |h_i|,$$

a więc $\frac{1}{2}|U_i| < |h_i|$. Możemy zatem wnioskować, że

$$\mathbb{I}_{\{U_i^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} \leqslant \mathbb{I}_{\{(2u_i)^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} + \mathbb{I}_{\{h_i > u_i\}},$$

dla dowolnego $1 \leqslant i \leqslant N$, czyli w konsekwencji

$$\begin{aligned} \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, U_i^2 \mathbb{I}_{\{U_i^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} &\leqslant \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, u_i^2 \mathbb{I}_{\{(2u_i)^2 \geqslant \delta^2 \cdot \mathcal{V}_N\}} + \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, h_i^2 \\ &\leqslant \frac{2}{\mathcal{V}_N^2 \cdot \delta} \sum_{i=1}^{N} \mathbb{E}\, u_i^4 + \frac{2}{\mathcal{V}_N} \sum_{i=1}^{N} \mathbb{E}\, h_i^2. \end{aligned} \tag{5.13}$$

Jako pierwszą rozważymy sumę czwartych momentów zmiennych u_i, $1 \leqslant i \leqslant N$. Ponieważ zachodzi nierówność

$$\frac{1}{27} u_i^4 \leqslant 2^4 Z_i^4 \left(\sum_{j=1}^{i-1} a_{ij} Z_j \right)^4 + a_{ii}^4 \left(Z_i^2 - \mathbb{E}\, Z_i^2 \right)^4 + x_i^4 Z_i^4,$$

element $\sum_{i=1}^{N} \mathbb{E}\, u_i^4$, występujący w (5.13), możemy ograniczyć za pomocą sumy

$$432 \left(\sup_{i \leqslant N} \mathbb{E}\, Z_i^4 \right) \kappa_N^{(1)} + 27 K^8 \kappa_N^{(2)} + 27 \left(\sup_{i \leqslant N} \mathbb{E}\, Z_i^4 \right) \kappa_N^{(3)},$$

gdzie

$$\kappa_N^{(1)} = \mathbb{E}\sum_{i=1}^{N}\left(\sum_{j=1}^{i-1} a_{ij}Z_j\right)^4,$$
$$\kappa_N^{(2)} = \sum_{i=1}^{N} a_{ii}^4,$$
$$\kappa_N^{(3)} = \sum_{i=1}^{N} x_i^4.$$

Rozwijając potęgę w wyrażeniu $\kappa_N^{(1)}$, przy użyciu zaobserwowanej wcześniej niezależności zmiennych Z_i, przy $1 \leqslant i \leqslant N$ otrzymujemy

$$\begin{aligned}\kappa_N^{(1)} &= \mathbb{E}\sum_{i=1}^{N}\left(\sum_{j=1}^{i-1} a_{ij}Z_j\right)^4 = \sum_{i=1}^{N}\mathbb{E}\left(\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} a_{ij}a_{ik}Z_jZ_k\right)^2\\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\sum_{j'=1}^{i-1}\sum_{k'=1}^{i-1} a_{ij}a_{ik}a_{ij'}a_{ik'}\,\mathbb{E}\,Z_jZ_kZ_{j'}Z_{k'}\\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\sum_{j'=1}^{i-1}\sum_{k'=1}^{i-1} a_{ij}a_{ik}a_{ij'}a_{ik'}\,\mathbb{E}\,Z_jZ_kZ_{j'}Z_{k'}\\ &\leqslant C\cdot\sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} a_{ij}^2a_{ik}^2,\end{aligned}$$

dla pewnej stałej $C > 0$. To z kolei implikuje

$$\frac{\kappa_N^{(1)}}{C} + \kappa_N^{(2)} \leqslant 2\sum_{i=1}^{N}\left(\sum_{j=1}^{i-1} a_{ij}^2\right)^2 \leqslant 2\left\|\boldsymbol{A}_N\cdot\left[a_{ij}\mathbb{I}_{\{i\leqslant j\}}\right]_{ij\leqslant N}\right\|_{\mathsf{F}}^2 \leqslant \|\boldsymbol{A}_N\|^2\|\boldsymbol{A}_N\|_{\mathsf{F}}^2.$$

Ponieważ $\kappa_N^{(1)}, \kappa_N^{(2)} \geqslant 0$, na postawie założeń (5.3) i (5.4) otrzymujemy

$$\lim_{N\to\infty}\frac{1}{\mathcal{V}_N^2}\kappa_N^{(1)} = \lim_{N\to\infty}\frac{1}{\mathcal{V}_N^2}\kappa_N^{(2)} = 0.$$

Co więcej, założenia (5.1) i (5.2) implikują nierówność

$$0 \leqslant \lim_{N\to\infty}\frac{1}{\mathcal{V}_N^2}\kappa_N^{(3)} \leqslant \lim_{N\to\infty}\frac{1}{\mathcal{V}_N^2}\left(\max_{1\leqslant i\leqslant N} x_i^2\right)\|\boldsymbol{x}_N\|^2 = 0.$$

Dowód będzie można uznać za ukończony, gdy wykażemy jednostajną względem $N \in \mathbb{N}$ zbieżność wyrażenia $\mathcal{V}_N^{-1} \sum_{i=1}^{N} \mathbb{E}\left[h_i^2\right]$ do zera przy $K \to \infty$. Najpierw zaobserwujmy nierówność

$$\mathbb{E}\, h_i^2 \leqslant E\Bigg(\sum_{j=1}^{i-1} a_{ij}(\varepsilon_i\varepsilon_j - Z_iZ_j)\Bigg)^2 + 3a_{ii}^2\, \mathbb{E}\left(\varepsilon_i^2 - Z_i^2 - (1 - \mathbb{E}\, Z_i^2)\right)^2 + 3x_i^2\, \mathbb{E}\left(\varepsilon_i - Z_i\right)^2.$$

Ponieważ $\boldsymbol{\varepsilon} = (Z_i)_{i\leqslant N} + (H_i)_{i\leqslant N}$, z użyciem prostego rachunku można pokazać, że dla dowolnych indeksów $1 \leqslant j < i \leqslant N$ mamy

$$\varepsilon_i\varepsilon_j - Z_iZ_j = H_iH_j + Z_iH_j + H_iZ_j,$$
$$\varepsilon_i^2 - Z_i^2 = 2Z_iH_i + H_i^2.$$

Zatem, dla pewnej stałej $C > 0$ prawdziwe jest ograniczenie

$$\mathcal{V}_N^{-1} \sum_{i=1}^{N} \mathbb{E}\, h_i^2 \leqslant C\mathcal{V}_N^{-1} \|\boldsymbol{A}_N\|_{\mathsf{F}}^2 \sqrt{\sup_{1\leqslant i,j\leqslant N} \mathbb{E}\, H_i^4\, \mathbb{E}\, Z_j^4} + C\mathcal{V}_N^{-1} \|\boldsymbol{A}_N\|_{\mathsf{F}}^2 \sup_{1\leqslant i\leqslant N} \mathbb{E}\, H_i^4 + C\mathcal{V}_N^{-1} \|\boldsymbol{x}_N\|^2 \sup_{1\leqslant i\leqslant N} \mathbb{E}\, H_i^2.$$

Ostatecznie, na mocy założeń (5.1) i (5.4), wyrażenia $\mathcal{V}_N^{-1}\|\boldsymbol{A}_N\|_{\mathsf{F}}^2$ i $\mathcal{V}_N^{-1}\|\boldsymbol{x}_N\|^2$ są ograniczone jednostajnie względem $N \in \mathbb{N}$. Ponieważ, zgodnie z założeniem, czwarte potęgi ε_i, $1 \leqslant i \leqslant N$, są jednostajnie całkowalne mamy

$$\sup_{K>0} \sup_{N\in\mathbb{N}} \sup_{1\leqslant i\leqslant N} \mathbb{E}\, Z_i^4 < \infty$$

oraz

$$0 \leqslant \sqrt{\sup_{N\in\mathbb{N}} \sup_{i\leqslant N} \mathbb{E}\, H_i^2} \leqslant \sup_{N\in\mathbb{N}} \sup_{i\leqslant N} \mathbb{E}\, H_i^4 \leqslant 16 \sup_{N\in\mathbb{N}} \sup_{i\leqslant N} \mathbb{E}\, \varepsilon_i^4 \mathbb{I}_{\{\varepsilon_i \geqslant K\}} \xrightarrow{K\to\infty} 0,$$

co wynika z określenia zmiennych Z_i i H_i w równaniu (5.12). □

2. Twierdzenia o rozkładzie granicznym

2.1. Założenia formalne

Aby uzyskać asymptotyczną normalność rozważanych estymatorów QNW, konieczne jest przyjęcie dodatkowych założeń oraz wzmocnienie tych wprowadzonych w poprzednim podrozdziale. W szczególności modyfikacja założeń dotyczących natury stochastycznej składnika losowego wymaga przytoczenia następującej definicji.

Definicja

Powiemy, że rodzina zmiennych losowych $\{V_z\}_{z\in Z}$, gdzie Z jest pewnym zbiorem indeksów, jest **jednostajnie całkowalna**, jeśli dla dowolnej liczby $\epsilon > 0$ istnieje liczba K_ϵ, taka, że dla dowolnego $z \in Z$ mamy

$$\mathbb{E}\left(|V_z| \cdot \mathbb{I}_{\{|V_z|>K_\epsilon\}}\right) \leqslant \epsilon,$$

gdzie $\mathbb{I}$ jest indykatorem zdarzenia opisanego w indeksie dolnym, a więc zmienną losową postaci

$$\mathbb{I}_{\{|V_z|>K_\epsilon\}} = \begin{cases} 1 & \text{gdy } |V_z| > K_\epsilon \\ 0 & \text{gdy } |V_z| \leqslant K_\epsilon \end{cases}.$$

Zauważmy, że pojęcie jednostajnej całkowalności opiera się wyłącznie na własnościach (jednowymiarowych) rozkładów poszczególnych zmiennych. Zatem ignoruje ono potencjalne zależności między zmiennymi. Intuicyjnie, założenie jednostajnej całkowalności ogranicza „grubość" ogonów rozkładów — jednocześnie dla wszystkich elementów rodziny rodziny. Jak wynika z twierdzenia de la Vallée Poussina (patrz la Vallée Poussin, 1915, lub Meyer, 1966, twierdzenie T22, tamże), warunek ten jest nieznacznie mocniejszy niż warunek wspólnej ograniczoności całek, tj. $\sup_{z\in Z} \mathbb{E}\,|V_z| < \infty$.

Założenie V.C'

Elementy wektora zaburzeń $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^N$ są zmiennymi losowymi o zerowej wartości oczekiwanej i stałej nieznanej wariancji $\sigma_0^2 > 0$. Ponadto, rodzina ich czwartych potęg, tj. ε_i^4, dla $1 \leqslant i \leqslant N$, jest jednostajnie całkowalna ze względu na indeks i oraz rozmiar próby N.

Definicja

Niech $\boldsymbol{x} = \boldsymbol{x}(N)$ będzie wektorem w $\mathbb{R}^N$ oraz $\boldsymbol{x} = (x_i)_{i=1}^N$. Elementy wektora $\boldsymbol{x}$ wykazują **równomierny wzrost**, jeśli zachodzi zbieżność

$$\lim_{N\to\infty} \frac{1}{N}\|\boldsymbol{x}_N\|_\infty^2 = \lim_{N\to\infty} \max_{1\leqslant i\leqslant N} \frac{x_i^2}{N} = 0.$$

Założenie V.D'$_{\text{SAR}}$

Macierz zmiennych objaśniających $\mathbf{X}$ spełnia założenie IV.D. Ponadto, elementy kolumn macierzy $\mathbf{X}$ oraz $\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\lambda}_0)\mathbf{X}$, dla $1 \leqslant r \leqslant d$ wykazują równomierny wzrost.

Założenie V.F$_{\text{SAR}}$

Niech $\mathcal{P} \times \mathbb{R}^k \times (0,\infty) \ni (\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) \mapsto \ln L_{\boldsymbol{y}}(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$ będzie funkcją parametryzowaną wartością $\boldsymbol{y} \in \mathbb{R}^N$, określoną formułą (4.3) na s. 94. Rozważmy zmienną losową

$$\boldsymbol{\mathcal{G}} = \big(\mathrm{D} \ln L_{\mathbf{y}}\big)(\boldsymbol{\rho}_0, \boldsymbol{\beta}_0, \sigma_0^2), \tag{5.14}$$

gdzie D, zgodnie z notacją Eulera, oznacza operator funkcji pochodnej (po argumentach $\boldsymbol{\rho}$, $\boldsymbol{\beta}$, σ^2, przy ustalonej wartości parametru $\boldsymbol{y}$), a jej wartość obliczona jest dla argumentów $\boldsymbol{\rho}_0$, $\boldsymbol{\beta}_0$, σ_0^2 i wektora zmiennej objaśnianej $\mathbf{y}$ jako parametru, patrz specyfikacja (4.1). Przy analogicznie rozumianej drugiej pochodnej oznaczmy

$$\boldsymbol{\mathcal{J}} = -\frac{1}{N}\,\mathbb{E}\left(\mathrm{D}^2 \ln L_{\mathbf{y}}\right)(\boldsymbol{\rho}_0, \boldsymbol{\beta}_0, \sigma_0^2).$$

Istnieją granice

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}} \coloneqq \lim_{N\to\infty} \frac{1}{N}\,\mathbb{E}\,\boldsymbol{\mathcal{G}}^{\mathsf{T}}\boldsymbol{\mathcal{G}},$$
$$\boldsymbol{\mathcal{J}}_0 \coloneqq \lim_{N\to\infty} \boldsymbol{\mathcal{J}},$$

przy czym $\boldsymbol{\mathcal{J}}_0$ jest macierzą nieosobliwą.

Założenie V.G$_{\text{SAR}}$
Prawdziwa wartość $\boldsymbol{\rho}_0$ parametru autoregresyjnego $\boldsymbol{\rho}$, patrz specyfikacja (4.1), jest elementem topologicznego wnętrza zbioru $\mathcal{P} \subset \mathbb{R}^d$.

Założenie V.D'$_{\text{SEM}}$
Macierz zmiennych objaśniających $\mathbf{X}$ spełnia założenie IV.D. Co więcej, elementy kolumn macierzy $\mathbf{X}$ wykazują równomierny wzrost (patrz definicja na s. 143).

Założenie V.F$_{\text{SEM}}$
Niech $\mathbb{R}^k \times \mathcal{L} \times (0,\infty) \ni (\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) \mapsto \ln L_{\boldsymbol{y}}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma^2)$ będzie funkcją parametryzowaną wartością $\boldsymbol{y} \in \mathbb{R}$, określoną formułą (4.6) na s. 89. Przy oznaczeniach, por. założenie V.F$_{\text{SAR}}$,

$$\boldsymbol{\mathcal{S}} = \left(\mathrm{D} \ln L_{\mathbf{y}}\right)(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2) \tag{5.15}$$

oraz

$$\boldsymbol{\mathcal{I}} = -\frac{1}{N}\,\mathbb{E}\left(\mathrm{D}^2 \ln L_{\mathbf{y}}\right)(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2),$$

gdzie $\mathbf{y}$ jest zmienną zależną w specyfikacji (4.2), istnieją granice

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} \coloneqq \lim_{N\to\infty} \frac{1}{N}\,\mathbb{E}\,\boldsymbol{\mathcal{S}}^{\mathsf{T}}\boldsymbol{\mathcal{S}},$$
$$\boldsymbol{\mathcal{I}}_0 \coloneqq \lim_{N\to\infty} \boldsymbol{\mathcal{I}},$$

przy czym $\boldsymbol{\mathcal{I}}_0$ jest macierzą nieosobliwą.

Założenie V.G$_{\text{SEM}}$
Prawdziwa wartość $\boldsymbol{\lambda}_0$ parametru autoregresyjnego $\boldsymbol{\lambda}$ (patrz specyfikacja (4.2)) jest elementem topologicznego wnętrza zbioru $\mathcal{L} \subset \mathbb{R}^d$.

Uwagi dotyczące wzmocnionych założeń

Łatwo zauważyć, że założenie V.C' jest wzmocnieniem założenia IV.C wprowadzonego na użytek dowodu zgodności gaussowskich estymatorów quasi-największej wiarogodności. Jak się okazuje, por. Pruss (1998), warunki wyrażone w założeniu IV.C nie są wystarczające do wyprowadzenia ich asymptotycznego rozkładu. W naszym opracowaniu przyjmujemy zatem standardowy postulat ekonometryczny o niezależności według prawdopodobieństwa elementów wektora zaburzeń losowych w ramach jednej próby. Zaznaczmy, że wciąż nie wymagamy równości dystrybuant tychże elementów, a więc, w szczególności, ich rozkłady nie muszą być gaussowskie. Co istotne, założenie jednorodności rozkładu, przy utrzymanym żądaniu homoskedastyczności, zastępujemy założeniem jednostajnej całkowalności rodziny czwartych momentów elementów zaburzenia losowego (patrz definicja na s. 143).

Założenia V.F$_{\text{SAR}}$ i V.F$_{\text{SEM}}$ uwzględniają warunki konieczne istnienia granicznej wariancji rozkładu oszacowań. W szczególności, z każdego z nich wynika, że macierz (dokładniej ciąg macierzy) $\frac{1}{N}\mathbf{X}^{\mathsf{T}}\mathbf{X}$ ma nieosobliwą granicę. Można się spodziewać, że ta granica odegra kluczową rolę w ustaleniu wariancji estymatorów parametrów nachylenia $\hat{\boldsymbol{\beta}}$, choć — w przeciwieństwie do klasycznego modelu liniowego estymowanego metodą najmniejszych kwadratów — rolę nie wyłączną. Dodatkowo, założenia V.F$_{\text{SAR}}$, V.F$_{\text{SEM}}$ kontrolują warunki istnienia granicznej wariancji estymatorów parametrów autoregresyjnych, poprzez nietrywialność wariancji formy kwadratowej składnika losowego modelu o macierzy $\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\lambda}_0)^{-1}$. Dodatkowo, założenie V.F$_{\text{SAR}}$ implikuje istnienie granicznej wariancji estymatora parametru autoregresyjnego $\hat{\boldsymbol{\rho}}$, także poprzez ograniczenie korelacji między jawnymi (kolumny w $\mathbf{X}$) a niejawnymi ($\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\lambda}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0$, dla $1 \leqslant r \leqslant d$) zmiennymi objaśniającymi modelu autoregresji zmiennej zależnej.

Warto tutaj przypomnieć, że funkcje $\ln L_{\mathbf{y}}$, zarówno w założeniu V.F$_{\text{SAR}}$, jak i w V.F$_{\text{SEM}}$, nie są prawdziwymi funkcjami log-wiarogodności, gdyż prawdziwy rozkład składnika losowego modelu jest nieznany. Zatem nie możemy się spodziewać równości pomiędzy drugim momentem informanty a informacją Fishera, znanej z klasycznej metody największej wiarogodności. W kontekście przedstawionych założeń, taka równość implikowałaby tożsamość $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}} = \boldsymbol{\mathcal{J}}$ dla modelu autoregresji przestrzennej zmiennej objaśnianej i odpowiednio $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} = \boldsymbol{\mathcal{I}}$ dla modelu autoregresji przestrzennej składnika losowego.

2.2. Stwierdzenia pomocnicze

W tym podrozdziale wprowadzamy serię lematów wykorzystywanych w dowodach twierdzeń o rozkładach asymptotycznych estymatorów (twierdzenia V.8 i V.9). Prezentowane wyniki dotyczą regularności funkcji pseudowiarogodności

dla odpowiedniego modelu i są konsekwencją założeń przyjętych w poprzednim podrozdziale.

LEMAT V.2

Niech $U_{\mathcal{L}} \subset \mathbb{R}^d$ będzie otwartym, ograniczonym zbiorem danym w lemacie IV.3. Każdy element $e_N = e_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ którejkolwiek z macierzy reprezentujących pierwszą, drugą lub trzecią pochodną funkcji pseudowiarogodności $\log L_{\mathbf{y}}$ (patrz równanie (4.6), s. 95), jest zmienną losową postaci

$$e_N = \boldsymbol{\varepsilon}^{\mathsf{T}} \boldsymbol{A}_N \boldsymbol{\varepsilon} + \boldsymbol{x}_N^{\mathsf{T}} \boldsymbol{\varepsilon} + \boldsymbol{z}_N,$$

gdzie $\boldsymbol{\varepsilon}$ jest składnikiem losowym modelu (patrz specyfikacja (4.2)), a $\boldsymbol{A}_N = \boldsymbol{A}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$, $\boldsymbol{x}_N = \boldsymbol{x}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ i $\boldsymbol{z}_N = \boldsymbol{z}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ są ciągłymi nielosowymi funkcjami parametrów modelu. Co więcej, przy założeniach IV.A i IV.$\mathrm{B}_{\mathrm{SEM}}$ istnieje uniwersalna ciągła funkcja

$$\mathbb{R}^k \times (0, \infty) \ni (\boldsymbol{\beta}, \sigma^2) \mapsto K(\boldsymbol{\beta}, \sigma^2),$$

niezależna od rozmiaru próby N, dla której

$$\begin{aligned} \|\boldsymbol{A}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)\|^2 &\leqslant K(\boldsymbol{\beta}, \sigma^2), \\ \|\boldsymbol{x}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)\|^2 &\leqslant N \cdot K(\boldsymbol{\beta}, \sigma^2), \\ \|\boldsymbol{z}_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)\|^2 &\leqslant N \cdot K(\boldsymbol{\beta}, \sigma^2), \end{aligned}$$

dla wszystkich $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{\lambda} \in U_{\mathcal{L}}$, $\sigma^2 > 0$.

Fakt wyrażony w tezie lematu V.2 można sprawdzić, analizując bezpośrednio formuły opisujące pochodne przedstawione w lemacie IV.5.

LEMAT V.3

Niech $U_{\mathcal{L}} \subset \mathbb{R}^d$ będzie otwartym, ograniczonym zbiorem, danym w lemacie IV.3 oraz niech $U_{\boldsymbol{\beta}} \subset \mathbb{R}^k$ i $U_{\sigma^2} \subset (\varsigma, \infty)$, dla pewnego $\varsigma > 0$, będą zbiorami otwartymi i ograniczonymi w swoich przestrzeniach. Niech $\log L_{\mathbf{y}}$ będzie funkcją pseudowiarogodności, daną w równaniu (4.6) na s. 95. Przy założeniach IV.A, IV.$\mathrm{B}_{\mathrm{SEM}}$ i IV.D, wielkość

$$M(U_{\boldsymbol{\beta}}, U_{\sigma^2}) \coloneqq \sup_{\boldsymbol{\beta} \in U_{\boldsymbol{\beta}}} \sup_{\boldsymbol{\lambda} \in U_{\boldsymbol{\lambda}}} \sup_{\sigma^2 \in U_{\sigma^2}} \left\| \frac{1}{N} \big(\mathrm{D}^3 \ln L_{\mathbf{y}}\big)(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) \right\| \tag{5.16}$$

jest, przy zmiennym $N = 1, 2, \ldots$, stochastycznie ograniczoną zmienną losową.

Dowód. Zauważmy najpierw, że mierzalność $M(U_{\boldsymbol{\beta}}, U_{\sigma^2})$ wynika z ciągłości normy w wyrażeniu (5.16), względem parametrów modelu. Istotnie, mamy równość

$$M(U_{\boldsymbol{\beta}}, U_{\sigma^2}) = \sup_{\boldsymbol{\beta}\in U_{\boldsymbol{\beta}}\cap\mathbb{Q}^k} \sup_{\boldsymbol{\lambda}\in U_{\boldsymbol{\lambda}}\cap\mathbb{Q}^d} \sup_{\sigma^2\in U_{\sigma^2}\cap\mathbb{Q}} \Big\| \frac{1}{N}\big(\mathrm{D}^3 \ln L_{\mathbf{y}}\big)(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)\Big\|,$$

gdzie symbol $\mathbb{Q}$ oznacza zbiór liczb wymiernych.

Niech $e_N = e_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ będzie dowolnym elementem reprezentującym macierz $\mathrm{D}^3 \log L_{\mathbf{y}}$. Z lematu V.2 wnioskujemy, że

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in U_{\boldsymbol{\beta}}} \sup_{\boldsymbol{\lambda}\in U_{\boldsymbol{\lambda}}} \sup_{\sigma^2\in U_{\sigma^2}} |e_N| \\ \leqslant \frac{1}{N}\,\mathbb{E} \sup_{\boldsymbol{\beta}\in U_{\boldsymbol{\beta}}} \sup_{\boldsymbol{\lambda}\in U_{\boldsymbol{\lambda}}} \sup_{\sigma^2\in U_{\sigma^2}} \big(\|\boldsymbol{A}_N\|\|\boldsymbol{\varepsilon}\|^2 + \|\boldsymbol{x}_N\|\|\boldsymbol{\varepsilon}\| + \|\boldsymbol{z}_N\|\big) \\ \leqslant 3(\sigma_0^2+1) \max_{\boldsymbol{\beta}\in\overline{U_{\boldsymbol{\beta}}}} \max_{\sigma^2\in\overline{U_{\sigma^2}}} \sqrt{K(\boldsymbol{\beta}, \sigma^2)} < \infty,\end{aligned}$$

gdzie $\overline{U_{\boldsymbol{\beta}}}$ oraz $\overline{U_{\sigma^2}}$ są domknięciami topologicznymi odpowiednich zbiorów. □

Lemat V.4

Niech $\ln L_{\mathbf{y}}$ będzie funkcją pseudowiarogodności daną w równaniu (4.6) w rozdziale IV. Przy założeniach IV.A, IV.B$_{\mathrm{SEM}}$, V.C', IV.D i V.F$_{\mathrm{SEM}}$, zmienna losowa

$$\boldsymbol{\mathcal{I}}_* = -\frac{1}{N}\big(\mathrm{D}^3 \ln L_{\mathbf{y}}\big)(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2)$$

zbiega według prawdopodobieństwa do macierzy $\boldsymbol{\mathcal{I}}_0$ zdefiniowanej w założeniu V.F$_{\mathrm{SEM}}$.

Dowód. Niech $e_N = e_N(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ będzie dowolnym elementem macierzy $\boldsymbol{\mathcal{I}}_*$. Na mocy lematu V.2 mamy oszacowanie

$$\operatorname{Var} e_N \leqslant \frac{2}{N^2} \operatorname{Var}(\boldsymbol{\varepsilon}^{\mathsf{T}} \boldsymbol{A}_N \boldsymbol{\varepsilon}) + \frac{2}{N^2} \operatorname{Var}(\boldsymbol{x}_N^{\mathsf{T}} \boldsymbol{\varepsilon}).$$

Następnie, używając lematu IV.6, wnioskujemy, że

$$\operatorname{Var} e_N \leqslant \frac{6}{N} K(\boldsymbol{\beta}_0, \sigma_0^2) \cdot \sup_{N'\in\mathbb{N}} \sup_{i\leqslant \bar{N}(N')} \mathbb{E}\,\bar{\varepsilon}_i^4 + \frac{2\sigma_0^2}{N} K(\boldsymbol{\beta}_0, \sigma_0^2),$$

przy czym powyższe ograniczenie dąży do zera wraz z N zbiegającym do nieskończoności. Zatem, na mocy nierówności Jensena

$$\mathbb{E}\,\|\boldsymbol{\mathcal{I}}_* - \boldsymbol{\mathcal{I}}\| \leqslant \sqrt{\sum_{e_N} \mathbb{E}\,(e_N - \mathbb{E}\,e_N)^2} \xrightarrow{N\to\infty} 0,$$

gdzie sumowanie następuje po wszystkich $(k+d+1)^2$ elementach e_N macierzy $\boldsymbol{\mathcal{I}}_*$. Na mocy założenia V.F$_{\text{SEM}}$ normy różnic $\|\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{I}}_0\|$ również zbiegają do zera, więc

$$\mathbb{E}\,\|\boldsymbol{\mathcal{I}}_* - \boldsymbol{\mathcal{I}}_0\| \leqslant \mathbb{E}\,\|\boldsymbol{\mathcal{I}}_* - \boldsymbol{\mathcal{I}}\| + \|\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{I}}_0\| \xrightarrow{N\to\infty} 0.$$

Ostatecznie, tezę lematu uzyskujemy na mocy nierówności Czebyszewa. □

LEMAT V.5

Niech $U_{\mathcal{P}} \subset \mathbb{R}^d$ będzie zbiorem otwartym, ograniczonym, danym w lemacie IV.4. Każdy element $e_N = e_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$ którejkolwiek z macierzy, reprezentującej pierwszą, drugą lub trzecią pochodną funkcji pseudowiarogodności $\log L_{\mathbf{y}}$, patrz równanie (4.3) rozdziału IV, jest zmienną losową postaci

$$e_N = \boldsymbol{\varepsilon}^{\mathsf{T}} \boldsymbol{A}_N \boldsymbol{\varepsilon} + \boldsymbol{x}_N^{\mathsf{T}} \boldsymbol{\varepsilon} + \boldsymbol{z}_N,$$

gdzie $\boldsymbol{\varepsilon}$ jest składnikiem losowym modelu (patrz specyfikacja (4.1) w rozdziale IV), a $\boldsymbol{A}_N = \boldsymbol{A}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$, $\boldsymbol{x}_N = \boldsymbol{x}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$ i $\boldsymbol{z}_N = \boldsymbol{z}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$ są ciągłymi nielosowymi funkcjami parametrów modelu. Co więcej, przy założeniach IV.A i IV.B$_{\text{SAR}}$ istnieje uniwersalna ciągła funkcja

$$\mathbb{R}^k \times (0, \infty) \ni (\boldsymbol{\beta}, \sigma^2) \mapsto K(\boldsymbol{\beta}, \sigma^2),$$

niezależna od rozmiaru próby N, dla której

$$\begin{aligned}
\|\boldsymbol{A}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)\|^2 &\leqslant K(\boldsymbol{\beta}, \sigma^2),\\
\|\boldsymbol{x}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)\|^2 &\leqslant N \cdot K(\boldsymbol{\beta}, \sigma^2),\\
\|\boldsymbol{z}_N(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)\|^2 &\leqslant N \cdot K(\boldsymbol{\beta}, \sigma^2),
\end{aligned}$$

dla wszystkich $\boldsymbol{\beta} \in \mathbb{R}^k$, $\boldsymbol{\rho} \in U_{\mathcal{P}}$, $\sigma^2 > 0$.

Fakt wyrażony w tezie lematu V.5 można sprawdzić, analizując bezpośrednio formuły opisujące pochodne przedstawione w lemacie IV.8.

LEMAT V.6

Niech $U_{\mathcal{P}} \subset \mathbb{R}^d$ będzie zbiorem otwartym, ograniczonym, danym w lemacie IV.4 oraz niech $U_{\boldsymbol{\beta}} \subset \mathbb{R}^k$ i $U_{\sigma^2} \subset (\varsigma, \infty)$, dla pewnego $\varsigma > 0$, będą zbiorami otwartymi i ograniczonymi w swoich przestrzeniach. Niech $\log L_{\mathbf{y}}$ będzie funkcją pseudowiarogodności, daną w równaniu (4.3) rozdziału IV. Wówczas, przy założeniach IV.A, IV.B$_{\text{SAR}}$ i IV.D, wielkość

$$M(U_{\boldsymbol{\beta}}, U_{\sigma^2}) \coloneqq \sup_{\boldsymbol{\rho}\in U_\rho} \sup_{\boldsymbol{\beta}\in U_{\boldsymbol{\beta}}} \sup_{\sigma^2 \in U_{\sigma^2}} \left\| \frac{1}{N} \big(\mathrm{D}^3 \ln L_{\mathbf{y}}\big)(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) \right\| \tag{5.17}$$

jest, przy zmiennym $N = 1, 2, \ldots$, stochastycznie ograniczoną zmienną losową.

Dowód lematu V.6 jest analogiczny do dowody lematu V.3. Należy zamienić parametr $\boldsymbol{\lambda}$, parametrem $\boldsymbol{\rho}$ oraz skorzystać z nierówności opisanych w lemacie V.5.

LEMAT V.7

Niech $\ln L_{\mathbf{y}}$ będzie funkcją pseudowiarogodności daną w równaniu (4.3) w rozdziale IV. Przy założeniach IV.A, IV.B_{SAR}, V.C', IV.D i V.F_{SAR}, zmienna losowa

$$\boldsymbol{\mathcal{J}}_* = -\frac{1}{N}\big(\mathrm{D}^3 \ln L_{\mathbf{y}}\big)(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$$

zbiega według prawdopodobieństwa do macierzy $\boldsymbol{\mathcal{J}}_0$ zdefiniowanej w założeniu V.F_{SAR}.

Dowód powyższego lematu przebiega w sposób analogiczny do rozumowania argumentującego tezę lematu V.4, zamieniając parametr $\boldsymbol{\lambda}$ na ρ oraz zastępując symbole $\boldsymbol{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}_0$, $\boldsymbol{\mathcal{I}}_*$ przez odpowiednio $\boldsymbol{\mathcal{J}}$, $\boldsymbol{\mathcal{J}}_0$, $\boldsymbol{\mathcal{J}}_*$.

2.3. Asymptotyczna normalność estymatora dla modelu SEM

W tym podrozdziale prezentujemy autorskie twierdzenie dotyczące zachowania asymptotycznego estymatora quasi-największej wiarogodności dla modelu z przestrzennie skorelowanym składnikiem losowym. Przedstawiane rozumowanie wykorzystuje centralne twierdzenie graniczne (twierdznie V.1) sformułowane w podrozdziale pierwszym, s. 132.

TWIERDZENIE V.8

Gdy spełnione są założenia IV.A, IV.B_{SEM} i IV.E_{SEM} oraz założenia V.C', V.D'$_{SEM}$, V.F_{SEM} i V.G_{SEM}, wówczas łączny rozkład estymatorów $\hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}}$, $\hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}}$ oraz $\hat{\sigma}^2_{\text{SEM\_QNW}}$ jest $\sqrt{N}$-asymptotycznie normalny, a dokładniej, zmienna losowa

$$\sqrt{N} \cdot \left(\begin{bmatrix} \hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}} \\ \hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}} \\ \hat{\sigma}^2_{\text{SEM\_QNW}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\lambda}_0 \\ \sigma_0^2 \end{bmatrix} \right)$$

zbiega według dystrybuanty do rozkładu normalnego $\mathcal{N}(0, \boldsymbol{\mathcal{I}}_0^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{I}}_0^{-1})$.

DOWÓD. Dla uproszczenia zapisu oznaczmy

$$\hat{\boldsymbol{\beta}} \coloneqq \hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}}, \qquad \hat{\boldsymbol{\lambda}} \coloneqq \hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}}, \qquad \hat{\sigma}^2 \coloneqq \hat{\sigma}^2_{\text{SEM\_QNW}}.$$

Niech $\boldsymbol{\mathcal{S}}$ będzie pseudoinformantą dla parametrów modelu, zdefiniowaną równaniem (5.15). Używając lematu IV.5, łatwo obliczyć kolejne elementy wektora

losowego $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{S}}$. Wówczas mamy

$$\begin{aligned}
\frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2) &= \frac{1}{\sqrt{N}\sigma_0^2}\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)\mathbf{X},\\
\frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2) &= \frac{1}{\sqrt{N}\sigma_0^2}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}_r\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\boldsymbol{\varepsilon}\right]_{r\leqslant d}\\
&\quad - \frac{1}{\sqrt{N}\sigma_0^2}\left[\sigma_0^2 \operatorname{tr}\left(\mathbf{W}_r\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\right)\right]_{r\leqslant d},\\
\frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \sigma^2}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \sigma_0^2) &= \frac{1}{2\sqrt{N}\sigma_0^4}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon} - N\sigma_0^2\right]_{r\leqslant d}.
\end{aligned}$$

Pokażemy, że wektor losowy $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}}$ zbiega według rozkładu do $\mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}})$. Przypomnijmy, że w ogólności, suma dwóch ciągów, z których każdy jest zbieżny do rozkładu normalnego, sama nie musi być asymptotycznie gaussowska. Zatem, dla uzyskania tezy nie wystarczy wykazać asymptotyczną normalność elementów wektora $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}}$, jak to zostało zrobione w kontekście specyfikacji SAR w pracy Lee (2004). W naszym rozumowaniu użyjemy argumentu opartego na twierdzeniu Craméra–Wolda (patrz Billingsley, 2009).

Na początek zauważmy, że mamy $\mathbb{E}\,\boldsymbol{\mathcal{S}} = 0$, chociaż $\boldsymbol{\mathcal{S}}$ nie jest prawdziwą informantą, a więc równość ta nie jest wnioskiem z klasycznego kursu teorii estymacji. Niech $\boldsymbol{a} \in \mathbb{R}^k$, $\boldsymbol{b} \in \mathbb{R}^d$ i $c \in \mathbb{R}$ będą dowolne. Jeśli

$$\mathcal{V}(\boldsymbol{a}, \boldsymbol{b}, c) \coloneqq \begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix} \cdot \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} \cdot \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ c\end{bmatrix} = 0, \tag{5.18}$$

wówczas, korzystając z założenia $\mathrm{V.F_{SEM}}$, mamy

$$\begin{aligned}
\lim_{N\to\infty} \operatorname{Var}\left(\frac{\begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}}\right) &= \begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix} \cdot \lim_{N\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}}\right) \cdot \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ c\end{bmatrix}\\
= \begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix} \cdot \frac{1}{N}\left(\mathbb{E}\,\boldsymbol{\mathcal{S}}^{\mathsf{T}}\boldsymbol{\mathcal{S}} - 0\right) \cdot \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ c\end{bmatrix} &= \begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix} \cdot \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} \cdot \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ c\end{bmatrix} = 0,
\end{aligned}$$

a zatem $\frac{\begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}}$ zbiega według prawdopodobieństwa do zera — inaczej do rozkładu $\mathcal{N}(0, 0 \cdot \mathbf{I})$ — w $\mathbb{R}^{k+d+1}$. Założenie braku równości w (5.18) implikuje ciekawszy przypadek, w którym $\mathcal{V}(\boldsymbol{a}, \boldsymbol{b}, c) > 0$, jako, że $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}}$, będąca granicą

macierzy nieujemnie określonych (patrz założenie V.F$_{\text{SEM}}$) jest nieujemnie określona. Zauważmy, że wyrażenie $\frac{\begin{bmatrix}\boldsymbol{a}^\mathsf{T} & \boldsymbol{b}^\mathsf{T} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^\mathsf{T}$ ma postać

$$\frac{\begin{bmatrix}\boldsymbol{a}^\mathsf{T} & \boldsymbol{b}^\mathsf{T} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^\mathsf{T} = Q_N - \mathbb{E}\, Q_N,$$

gdzie $Q_N = \boldsymbol{\varepsilon}_N^\mathsf{T}\boldsymbol{A}_N\boldsymbol{\varepsilon}_N + \boldsymbol{\varepsilon}_N^\mathsf{T}\boldsymbol{x}_N$ jest formą liniowo-kwadratową o macierzy

$$\boldsymbol{A}_N = \frac{1}{\sqrt{N}\sigma_0^2}\Big(\boldsymbol{b}^\mathsf{T}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\lambda}_0)^{-1} + \frac{1}{2\sigma_0^2}c\mathbf{I}\Big)$$

oraz z wektorem współczynników części liniowej

$$\boldsymbol{x}_N = \frac{1}{\sqrt{N}\sigma_0^2}\mathbf{X}\cdot\boldsymbol{a}.$$

Pokażemy, że dla formy liniowo-kwadratowej Q_N możemy skorzystać z twierdzenia V.1. Skoro $\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c) > 0$, zbieżność wariancji pseudoinformanty, wyrażona w założeniu V.F$_{\text{SEM}}$, implikuje, że dla wszystkich dostatecznie dużych rozmiarów próby N zachodzi

$$\operatorname{Var} Q_N = \operatorname{Var}\left(\frac{\begin{bmatrix}\boldsymbol{a}^\mathsf{T} & \boldsymbol{b}^\mathsf{T} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^\mathsf{T}\right) > \frac{\begin{bmatrix}\boldsymbol{a}^\mathsf{T} & \boldsymbol{b}^\mathsf{T} & c\end{bmatrix}}{2}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ c\end{bmatrix} = \frac{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)}{2} > 0.$$

Następnie, z założenia V.D'$_{\text{SEM}}$, a dokładniej z zawartego w nim założenia IV.D, mamy

$$\limsup_{N\to\infty}\frac{\|\boldsymbol{x}_N\|^2}{\operatorname{Var} Q_N} \leqslant \frac{2\|\boldsymbol{a}\|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\limsup_{N\to\infty}\left\|\frac{1}{N}\mathbf{X}^\mathsf{T}\mathbf{X}\right\| < \infty.$$

Ponadto, zachodzi

$$\lim_{N\to\infty}\frac{\|\boldsymbol{x}_N\|_\infty^2}{\operatorname{Var} Q_N} \leqslant \frac{2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\lim_{N\to\infty}\frac{1}{N}\,\|\mathbf{X}\cdot\boldsymbol{a}\|_\infty^2 < \infty,$$

gdyż kombinacja liniowa wektorów, których elementy wykazują równomierny wzrost (por. definicja na s. 143) też jest takim wektorem. Z kolei założenia IV.A i IV.B$_{\text{SEM}}$ pozwalają stwierdzić, że

$$\mathfrak{B} \coloneqq \max_{r\leqslant d}\sup_{N\in\mathbb{N}}\|\mathbf{W}_r\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)^{-1}\|^2 < \infty,$$

a więc

$$0 \leqslant \frac{\|\boldsymbol{A}_N\|^2}{\operatorname{Var} Q_N} < \frac{2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\frac{1}{N}\left\|\boldsymbol{b}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)+\frac{1}{2\sigma_0^2}c\mathbf{I}\right\|^2$$
$$\leqslant \frac{4d\|\boldsymbol{b}\|^2\cdot\mathfrak{B}}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4\cdot N}+\frac{|c|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^8\cdot N}\xrightarrow{N\to\infty}0.$$

Podobnie, mamy

$$\frac{\|\boldsymbol{A}_N\|_{\mathsf{F}}^2}{\operatorname{Var} Q_N} < \frac{2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\frac{1}{N}\left\|\boldsymbol{b}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Gamma}(\boldsymbol{\lambda}_0)+\frac{1}{2\sigma_0^2}c\mathbf{I}\right\|_{\mathsf{F}}^2$$
$$\leqslant \frac{4d\|\boldsymbol{b}\|^2\cdot\mathfrak{B}}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}+\frac{|c|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}<\infty,$$

czyli $\limsup_{N\to\infty}\frac{\|\boldsymbol{A}_N\|_{\mathsf{F}}^2}{\operatorname{Var} Q_N}<\infty$. Zatem, uwzględniając założenie V.C' dla wektora zaburzeń modelu z twierdzenia V.1, wnioskujemy, że zachodzi zbieżność według rozkładu

$$\frac{\begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix}}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}} = Q_N - \mathbb{E}\,Q_N \xrightarrow{N\to\infty} \mathcal{N}\left(0,\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\right).$$

Ostatecznie z dowolności $\boldsymbol{a}$, $\boldsymbol{b}$ i c, na podstawie twierdzenia Craméra–Wolda, otrzymujemy zbieżność

$$\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{S}}^{\mathsf{T}} \xrightarrow{N\to\infty} \mathcal{N}(0,\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}}).$$

Zgodnie z tezą lematu V.4, zmienna losowa

$$\boldsymbol{\mathcal{I}}_* = -\frac{1}{N}\big(\mathrm{D}^2\ln L_{\mathbf{y}}\big)(\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\sigma_0^2)$$

zbiega według prawdopodobieństwa do $\boldsymbol{\mathcal{I}}_0$, przy $N\to\infty$. Z ciągłości wyznacznika jako funkcji macierzy ustalonego wymiaru wnioskujemy, że

$$\lim_{N\to\infty}\mathbb{P}\left(\{\det\boldsymbol{\mathcal{I}}_*=0\}\right)=0.$$

Przyjmijmy zatem $\|\boldsymbol{\mathcal{I}}_*^{-1}\|:=1$, tam, gdzie $\det\boldsymbol{\mathcal{I}}_*=0$. W efekcie, tak rozumiany ciąg norm $\|\boldsymbol{\mathcal{I}}_*^{-1}\|$ zbiega do $\|\boldsymbol{\mathcal{I}}_0^{-1}\|$ według prawdopodobieństwa.

Niech $(\Omega,\mathcal{F},\mathbb{P})$ będzie przestrzenią probabilistyczną, na której zdefiniowany jest proces stochastyczny $\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}(N)$, $N=1,2,\ldots$ Niech $O_{\boldsymbol{\beta}}\subset\mathbb{R}^k$, $O_{\boldsymbol{\lambda}}\subset\mathbb{R}^d$,

$O_{\sigma^2} \subset (\varsigma, \infty)$, dla pewnego $\varsigma > 0$, będą zbiorami otwartymi, ograniczonymi i wypukłymi, spełniającymi warunki

$$\boldsymbol{\beta}_0 \in O_{\boldsymbol{\beta}}, \qquad \boldsymbol{\lambda}_0 \in O_{\boldsymbol{\lambda}}, \qquad \sigma_0^2 \in O_{\sigma^2},$$

a ponadto, $O_{\boldsymbol{\lambda}} \subset \mathcal{L}$ (patrz założenie V.G$_{\text{SEM}}$). Zdefiniujmy

$$\boldsymbol{d}_N = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\lambda}} \\ \hat{\sigma}^2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\lambda}_0 \\ \sigma_0^2 \end{bmatrix}.$$

Pamiętając o zgodności estymatorów $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ i $\hat{\sigma}^2$ (patrz twierdzenie IV.11), możemy stwierdzić, że dla ciągu zdarzeń postaci

$$\Omega_N = \Omega_N^1 \cap \Omega_N^2 \cap \Omega_N^3,$$

gdzie

$$\begin{aligned} \Omega_N^1 &\coloneqq \{\det \boldsymbol{\mathcal{I}}_* \neq 0\}, \\ \Omega_N^2 &\coloneqq \{\hat{\boldsymbol{\beta}} \in O_{\boldsymbol{\beta}}\} \cap \{\hat{\boldsymbol{\lambda}} \in O_{\boldsymbol{\lambda}}\} \cap \{\hat{\sigma}^2 \in O_{\sigma^2}\}, \\ \Omega_N^3 &\coloneqq \big\{\|\boldsymbol{\mathcal{I}}_*^{-1}\| \cdot \|\boldsymbol{d}_N\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2})\big\}, \end{aligned}$$

a M dane jest w wyrażeniu (5.16), w lemacie V.3, mamy

$$\lim_{N\to\infty} \mathbb{P}\left(\Omega \setminus \Omega_N\right) = 0.$$

Niech $\omega \in \Omega_N$. Dla funkcji

$$O_{\boldsymbol{\beta}} \times O_{\boldsymbol{\lambda}} \times O_{\sigma^2} \ni \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \\ \sigma^2 \end{bmatrix} \mapsto f_\omega(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = \frac{1}{\sqrt{N}} \big(\mathrm{D} \ln L_{\mathbf{y}(\omega)}\big)(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)$$

możemy zastosować wielowymiarowe rozwinięcie Taylora w punkcie odpowiadającym wartościom parametrów $\boldsymbol{\beta}_0$, $\boldsymbol{\lambda}_0$, σ_0^2 (patrz np. twierdzenie 107 w monografii Hájek, Johanis, 2014). Zatem, dla pewnej funkcji resztowej $\mathcal{R}$, spełniającej nierówność

$$\|\mathcal{R}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)\| \leqslant \frac{1}{2} \sup_{O_{\boldsymbol{\beta}} \times O_{\boldsymbol{\lambda}} \times O_{\sigma^2}} \|\mathrm{D}\, f_\omega\| \cdot \|\boldsymbol{d}_N\|^2,$$

mamy

$$f_\omega(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{N}} \boldsymbol{\mathcal{S}} - \sqrt{N} \cdot \boldsymbol{d}_N^{\mathsf{T}} \cdot \boldsymbol{\mathcal{I}}_* + \mathcal{R}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2).$$

Dokonując w tym równaniu podstawień $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(\omega)$, $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}(\omega)$ oraz $\sigma^2 = \hat{\sigma}^2(\omega)$, otrzymujemy

$$\sqrt{N} \cdot \boldsymbol{d}_N^{\mathsf{T}} \cdot \boldsymbol{\mathcal{I}}_* = \frac{1}{\sqrt{N}} \boldsymbol{\mathcal{S}} + \mathcal{R}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\sigma}^2), \tag{5.19}$$

gdyż z samej definicji estymatorów, poprzez różniczkowy warunek optymalności, wynika, że $f_\omega(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = 0$.

Zauważmy, że gdyby na zbiorach Ω_N prawdziwa była nierówność

$$\left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* \right\| > 2 \left\| \frac{1}{N} \boldsymbol{\mathcal{S}} \right\|, \tag{5.20}$$

wówczas, uwzględniając (5.19), mielibyśmy ciąg oszacowań

$$\begin{aligned} \left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* \right\| &< 2 \left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* - \frac{1}{N} \boldsymbol{\mathcal{S}} \right\| = 2 \left\| \mathcal{R}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\sigma}^2) \right\| \\ &\leqslant \left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* \right\| \cdot \left\| \boldsymbol{\mathcal{I}}_*^{-1} \right\| \cdot \| \boldsymbol{d}_N \| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}). \end{aligned}$$

Stąd otrzymujemy zaprzeczenie nierówności (5.20). Ostatecznie uzyskujemy

$$\begin{aligned} \left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* - \frac{1}{N} \boldsymbol{\mathcal{S}} \right\| &= \left\| \mathcal{R}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\sigma}^2) \right\| \leqslant \frac{1}{2} M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \cdot \| \boldsymbol{d}_N \|^2 \\ &\leqslant \frac{1}{2} \left\| \sqrt{N} \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{I}}_* \right\| \cdot \left\| \boldsymbol{\mathcal{I}}_*^{-1} \right\| \cdot \| \boldsymbol{d}_N \| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \\ &\leqslant \left\| \frac{1}{N} \boldsymbol{\mathcal{S}} \right\| \cdot \left\| \boldsymbol{\mathcal{I}}_*^{-1} \right\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \cdot \| \boldsymbol{d}_N \|. \end{aligned}$$

Prawa strona powyższej nierówności zbiega według prawdopodobieństwa do zera, jako, że stanowi iloczyn zmiennych stochastycznie ograniczonych i zbieżnego do zera czynnika $\| \boldsymbol{d}_N \|$. Wynika stąd, że $\sqrt{N} \boldsymbol{\mathcal{I}}_* \boldsymbol{d}_N$, tak jak $\frac{1}{N} \boldsymbol{\mathcal{S}}^{\mathsf{T}}$, zbiega według dystrybuanty do $\mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{S}})$. Mamy

$$\sqrt{N} \boldsymbol{d}_N = \boldsymbol{\mathcal{I}}_0^{-1} \cdot \boldsymbol{\mathcal{I}}_0 \boldsymbol{\mathcal{I}}_*^{-1} \cdot \sqrt{N} \boldsymbol{\mathcal{I}}_* \boldsymbol{d}_N,$$

przy czym $\boldsymbol{\mathcal{I}}_0 \boldsymbol{\mathcal{I}}_*$ zbiega według prawdopodobieństwa do macierzy jednostkowej, ustalonego wymiaru $k+d+1$. Ostatecznie, otrzymujemy tezę o zbieżności według rozkładu

$$\sqrt{N} \cdot \left(\begin{bmatrix} \hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}} \\ \hat{\boldsymbol{\lambda}}_{\text{SEM\_QNW}} \\ \hat{\sigma}^2_{\text{SEM\_QNW}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\lambda}_0 \\ \sigma_0^2 \end{bmatrix} \right) \xrightarrow{N \to \infty} \mathcal{N}(0, \boldsymbol{\mathcal{I}}_0^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{I}}_0^{-1}).$$

□

2.4. Asymptotyczna normalność estymatora dla modelu SAR

W tym podrozdziale prezentujemy twierdzenie dotyczące zachowania asymptotycznego estymatora quasi-największej wiarogodności dla modelu z przestrzennie skorelowanym składnikiem losowym. Wynik został pierwotnie opublikowany w pracy Olejnik i Olejnik (2020). Schemat rozumowania argumentującego tezę twierdzenia prowadzona jest w sposób analogiczny do dowodu twierdzenia V.8. Mianowicie, za pomocą naszego centralnego twierdzenia granicznego V.1 (patrz s. 132) wykazywana jest asymptotyczna normalność wektora pseudoinformanty, po czym na tej podstawie, poprzez rozwinięcie jej pochodnej w szereg Taylora, uzyskiwany jest rozkład graniczny estymatorów QNW. Niemniej jednak, różne specyfikacje modelu procesu generującego obserwacje prowadzą do różnych postaci formy liniowo-kwadratowej i w obu przypadkach wymagają odrębnego szacowania.

Twierdzenie V.9

Gdy spełnione są założenia IV.A, IV.B_{SAR} i IV.E_{SAR} oraz założenia V.C', V.D'_{SAR}, V.F_{SAR} i V.G_{SAR}, wówczas łączny rozkład estymatorów $\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}$, $\hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}}$ oraz $\hat{\sigma}^2_{\text{SAR\_QNW}}$ jest $\sqrt{N}$-asymptotycznie normalny, a dokładniej, zmienna losowa

$$\sqrt{N}\cdot\left(\begin{bmatrix}\hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}\\ \hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}}\\ \hat{\sigma}^2_{\text{SAR\_QNW}}\end{bmatrix}-\begin{bmatrix}\boldsymbol{\rho}_0\\ \boldsymbol{\beta}_0\\ \sigma_0^2\end{bmatrix}\right)$$

zbiega według dystrybuanty do rozkładu normalnego $\mathcal{N}(0, \boldsymbol{\mathcal{J}}_0^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\mathcal{J}}_0^{-1})$.

Dowód. Dla uproszczenia zapisu oznaczmy

$$\hat{\boldsymbol{\rho}} \coloneqq \hat{\boldsymbol{\rho}}_{\text{SAR\_QNW}}, \qquad \hat{\boldsymbol{\beta}} \coloneqq \hat{\boldsymbol{\beta}}_{\text{SAR\_QNW}}, \qquad \hat{\sigma}^2 \coloneqq \hat{\sigma}^2_{\text{SAR\_QNW}}.$$

Niech $\boldsymbol{\mathcal{G}}$ będzie pseudoinformantą dla parametrów modelu, zdefiniowaną równaniem (5.14). Stosując lematy IV.8 i IV.5, łatwo obliczyć kolejne elementy wektora losowego $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}$. Wówczas mamy

$$\begin{aligned}\frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\rho}}(\boldsymbol{\rho}_0,\boldsymbol{\beta}_0,\sigma_0^2) &= \frac{1}{\sqrt{N}\sigma_0^2}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\varepsilon}\right]_{r\leqslant d}\\ &\quad-\frac{1}{\sqrt{N}\sigma_0^2}\left[\sigma_0^2\,\mathrm{tr}\,(\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1})\right]_{r\leqslant d}\\ &\quad+\frac{1}{\sqrt{N}\sigma_0^2}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\right]_{r\leqslant d},\\ \frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \boldsymbol{\beta}}(\boldsymbol{\rho}_0,\boldsymbol{\beta}_0,\sigma_0^2) &= \frac{1}{\sqrt{N}\sigma_0^2}\boldsymbol{\varepsilon}^{\mathsf{T}}\mathbf{X},\end{aligned}$$

$$\frac{1}{\sqrt{N}}\frac{\partial \log L_{\mathbf{y}}}{\partial \sigma^2}(\boldsymbol{\rho}_0, \boldsymbol{\beta}_0, \sigma_0^2) = \frac{1}{2\sqrt{N}\sigma_0^4}\left[\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon} - N\sigma_0^2\right]_{r\leqslant d}.$$

Pokażemy, że $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}}$ zbiega według rozkładu do $\mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}})$.

Aby skorzystać z twierdzenia Craméra-Wolda (patrz Cramér, Wold, 1936), ustalmy dowolny wektor

$$\boldsymbol{\alpha} = \begin{bmatrix}\boldsymbol{a}^{\mathsf{T}} & \boldsymbol{b}^{\mathsf{T}} & c\end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}.$$

Przyjmując oznaczenie $\mathcal{V}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\alpha}$, możemy zauważyć, że $\mathcal{V}(\boldsymbol{\alpha}) \geqslant 0$, gdyż $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}$ jest macierzą dodatnio określoną. Rozważymy zatem dwa przypadki. Jeśli $\mathcal{V}(\boldsymbol{\alpha}) = 0$, wówczas, korzystając z założenia V.F$_{\text{SAR}}$, zachodzi zbieżność

$$\begin{aligned}\lim_{N\to\infty} \operatorname{Var}\left(\boldsymbol{\alpha}\cdot\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}}\right) &= \boldsymbol{\alpha}\cdot\lim_{N\to\infty}\operatorname{Var}\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}}\cdot\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}\cdot\frac{1}{N}\left(\mathbb{E}\,\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}^{\mathsf{T}} - \mathbb{E}\,\boldsymbol{\mathcal{G}}\,\mathbb{E}\,\boldsymbol{\mathcal{G}}^{\mathsf{T}}\right)\cdot\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\alpha} = 0,\end{aligned}$$

gdyż, jak łatwo obliczyć, korzystając z wyprowadzonych pochodnych funkcji pseudowiarogodności, mamy $\mathbb{E}\,\boldsymbol{\mathcal{G}} = 0$. Wynika stąd, ze zmienna losowa $\frac{1}{\sqrt{N}}\boldsymbol{\alpha}\boldsymbol{\mathcal{G}}^{\mathsf{T}}$ zbiega prawdopodobieństwa do zera, co jest tożsame ze zbieżnością według rozkładu do rozkładu normalnego o wariancji $\boldsymbol{\alpha}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\alpha} = 0$.

Rozważmy teraz drugi przypadek, w którym $\mathcal{V}(\boldsymbol{\alpha}) > 0$ Zauważmy, że wyrażenie $\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}}$ ma postać centrowanej formy liniowo-kwadratowej zaburzenia losowego $\boldsymbol{\varepsilon}$, tj.

$$\frac{1}{\sqrt{N}}\boldsymbol{\alpha}\boldsymbol{\mathcal{G}}^{\mathsf{T}} = Q_N - \mathbb{E}\,Q_N,$$

gdzie $Q_N = \boldsymbol{\varepsilon}_N^{\mathsf{T}}\boldsymbol{A}_N\boldsymbol{\varepsilon}_N + \boldsymbol{\varepsilon}_N^{\mathsf{T}}\boldsymbol{x}_N$ oraz

$$\begin{aligned}\boldsymbol{A}_N &= \frac{1}{\sqrt{N}\sigma_0^2}\left(\boldsymbol{a}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1} + \frac{1}{2\sigma_0^2}c\mathbf{I}\right),\\ \boldsymbol{x}_N &= \frac{1}{\sqrt{N}\sigma_0^2}\left(\mathbf{X}\cdot\boldsymbol{a} + \boldsymbol{b}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\mathbf{X}\boldsymbol{\beta}_0\right).\end{aligned}$$

Pokażemy, że dla formy liniowo-kwadratowej Q_N możemy skorzystać z twierdzenia V.1. Zbieżność wariancji pseudoinformanty, wyrażona w założeniu V.F$_{\text{SAR}}$ implikuje, że dla wszystkich dostatecznie dużych rozmiarów próby N zachodzi

$$\operatorname{Var} Q_N = \operatorname{Var}\frac{1}{\sqrt{N}}\boldsymbol{\alpha}\boldsymbol{\mathcal{G}}^{\mathsf{T}} > \frac{1}{2}\boldsymbol{\alpha}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\alpha} = \frac{\mathcal{V}(\boldsymbol{\alpha})}{2} > 0.$$

Następnie, uwzględniając założenia IV.A, IV.B$_{\mathrm{SAR}}$, IV.D, przy oznaczeniach

$$\mathfrak{B}_{\boldsymbol{\Delta}} \coloneqq \max_{r\leqslant d} \sup_{N\in\mathbb{N}} \|\mathbf{W}_r\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\|^2 < \infty,$$
$$\mathfrak{B}_{\mathbf{X}} \coloneqq \sup_{N\in\mathbb{N}} \Big\|\frac{1}{N}\mathbf{X}^{\mathsf{T}}\mathbf{X}\Big\| < \infty,$$

mamy

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{x}_N\|^2}{\operatorname{Var} Q_N} \leqslant \frac{4\|\boldsymbol{b}\|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\mathfrak{B}_{\mathbf{X}} + \frac{4d\|\boldsymbol{a}\|^2\|\boldsymbol{\beta}_0\|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\mathfrak{B}_{\boldsymbol{\Delta}}\mathfrak{B}_{\mathbf{X}} < \infty.$$

Ponadto, własność równomiernego wzrostu, o której mowa w założeniu V.D'$_{\mathrm{SAR}}$ implikuje oszacowanie

$$\limsup_{N\to\infty} \frac{\|\boldsymbol{x}_N\|_\infty^2}{\operatorname{Var} Q_N} \leqslant \frac{4}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\sigma_0^4}\cdot\lim_{N\to\infty}\frac{1}{N}\|\mathbf{X}\cdot\boldsymbol{a}\|_\infty^2 + \frac{4}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\sigma_0^4}\cdot\lim_{N\to\infty}\frac{1}{N}\big\|\boldsymbol{a}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0)^{-1}\boldsymbol{\beta}_0\big\|_\infty^2 = 0.$$

Spełnione są zatem warunki (5.1) i (5.2). Z kolei (5.3) i (5.4) wnioskujemy z oszacowań

$$0 \leqslant \frac{\|\boldsymbol{A}_N\|^2}{\operatorname{Var} Q_N} < \frac{4d\|\boldsymbol{a}\|^2\cdot\mathfrak{B}_{\boldsymbol{\Delta}}}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4\cdot N} + \frac{|c|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^8\cdot N} \xrightarrow{N\to\infty} 0$$

oraz

$$\frac{\|\boldsymbol{A}_N\|_{\mathsf{F}}^2}{\operatorname{Var} Q_N} < \frac{2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4}\cdot\frac{1}{N}\Big\|\boldsymbol{a}^{\mathsf{T}}\mathbf{W}\boldsymbol{\Delta}(\boldsymbol{\rho}_0) + \frac{1}{2\sigma_0^2}c\mathbf{I}\Big\|_{\mathsf{F}}^2 \leqslant \frac{4d\|\boldsymbol{a}\|^2\cdot\mathfrak{B}}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4} + \frac{|c|^2}{\mathcal{V}(\boldsymbol{a},\boldsymbol{b},c)\cdot\sigma_0^4} < \infty.$$

Zatem, uwzględniając założenie V.C' dla wektora zaburzeń modelu, z twierdzenia V.1 wnioskujemy, że zachodzi zbieżność według rozkładu

$$\frac{\mathcal{V}(\boldsymbol{\alpha})}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}} = Q_N - \mathbb{E}\,Q_N \xrightarrow{N\to\infty} \mathcal{N}(0,\mathcal{V}(\boldsymbol{\alpha})),$$

a więc z dowolności wektora $\boldsymbol{\alpha}$ wynika zbieżność

$$\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}}^{\mathsf{T}} \xrightarrow{N\to\infty} \mathcal{N}(0,\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}).$$

Ze względu na zgodności estymatorów $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\beta}}$ i $\hat{\sigma}^2$ (patrz twierdzenie IV.10), mamy $\mathbb{P}(\Omega_N) \to 1$, dla $N \to \infty$, gdzie

$$\Omega_N = \Omega_N^1 \cap \Omega_N^2 \cap \Omega_N^3,$$

dla

$$\begin{aligned}
\Omega_N^1 &\coloneqq \{\det \boldsymbol{\mathcal{J}}_* \neq 0\},\\
\Omega_N^2 &\coloneqq \{\hat{\boldsymbol{\rho}} \in O_{\boldsymbol{\rho}}\} \cap \{\hat{\boldsymbol{\beta}} \in O_{\boldsymbol{\beta}}\} \cap \{\hat{\sigma}^2 \in O_{\sigma^2}\},\\
\Omega_N^3 &\coloneqq \left\{\|\boldsymbol{\mathcal{J}}_*^{-1}\| \cdot \|\boldsymbol{d}_N\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2})\right\},
\end{aligned}$$

przy czym

$$\boldsymbol{d}_N = \begin{bmatrix} \hat{\boldsymbol{\rho}} \\ \hat{\boldsymbol{\beta}} \\ \hat{\sigma}^2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\rho}_0 \\ \boldsymbol{\beta}_0 \\ \sigma_0^2 \end{bmatrix},$$

a M dane jest w wyrażeniu (5.17), w lemacie V.6.

Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną, na której zdefiniowany jest proces stochastyczny $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(N)$, $N = 1, 2, \ldots$ Niech $O_{\boldsymbol{\rho}} \subset \mathcal{P} \subset \mathbb{R}^d$, $O_{\boldsymbol{\beta}} \subset \mathbb{R}^k$, $O_{\sigma^2} \subset (\varsigma, \infty)$, dla pewnego $\varsigma > 0$, będą zbiorami otwartymi, ograniczonymi i wypukłymi, spełniającymi warunki

$$\boldsymbol{\rho}_0 \in O_{\boldsymbol{\rho}}, \qquad \boldsymbol{\beta}_0 \in O_{\boldsymbol{\beta}}, \qquad \sigma_0^2 \in O_{\sigma^2},$$

patrz założenie V.$\mathrm{G}_{\mathrm{SEM}}$.

Zgodnie z lematem V.4, dla zmiennej losowej $\boldsymbol{\mathcal{J}}_*$ mamy zbieżność według prawdopodobieństwa do $\boldsymbol{\mathcal{J}}_0$, przy $N \to \infty$. Wynika stąd, że macierz $\boldsymbol{\mathcal{J}}_*$ jest odwracalna na zbiorze $\Omega_N \subset \Omega$. W efekcie, norma macierzy $\boldsymbol{\mathcal{J}}_*^{-1}$ zbiega do $\|\boldsymbol{\mathcal{J}}_0^{-1}\|$ według prawdopodobieństwa, a co za tym idzie jest stochastycznie ograniczona.

Niech $\omega \in \Omega_N$. Dla funkcji

$$O_{\boldsymbol{\rho}} \times O_{\boldsymbol{\beta}} \times O_{\sigma^2} \ni \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\rho} \\ \sigma^2 \end{bmatrix} \mapsto f_\omega(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{N}} \left[\left(\mathrm{D} \ln L_{\mathbf{y}}\right)(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)\right]_{\mathbf{y}=\mathbf{y}(\omega)}$$

możemy zastosować wielowymiarowe rozwinięcie Taylora w punkcie odpowiadającym wartościom parametrów $\boldsymbol{\beta}_0$, $\boldsymbol{\rho}_0$, σ_0^2 (patrz np. twierdzenie 107 w monografii Hájek, Johanis, 2014). Zatem, dla pewnej funkcji resztowej $\mathcal{R}$, spełniającej

$$\|\mathcal{R}(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2)\| \leqslant \frac{1}{2} \sup_{O_{\boldsymbol{\rho}} \times O_{\boldsymbol{\beta}} \times O_{\sigma^2}} \|\mathrm{D}\, f_\omega\| \cdot \|\boldsymbol{d}_N\|^2,$$

mamy

$$f_\omega(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}} - \sqrt{N} \cdot \boldsymbol{d}_N^{\mathsf{T}} \boldsymbol{\mathcal{J}}_* + \mathcal{R}(\boldsymbol{\rho}, \boldsymbol{\beta}, \sigma^2).$$

Dokonując podstawień $\boldsymbol{\rho} = \hat{\boldsymbol{\rho}}(\omega)$, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(\omega)$ oraz $\sigma^2 = \hat{\sigma}^2(\omega)$, otrzymujemy

$$\sqrt{N} \cdot \boldsymbol{d}_N^{\mathsf{T}} \cdot \boldsymbol{\mathcal{J}}_* = \frac{1}{\sqrt{N}}\boldsymbol{\mathcal{G}} + \mathcal{R}(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2), \tag{5.21}$$

gdyż z samej definicji estymatorów wynika, poprzez różniczkowy warunek optymalności, że $f_\omega(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = 0$.

Zauważmy, że gdyby na zbiorach Ω_N prawdziwa była nierówność

$$\left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_*\right\| > 2\left\|\frac{1}{N}\boldsymbol{\mathcal{G}}\right\|, \tag{5.22}$$

wówczas, uwzględniając (5.21), mielibyśmy ciąg oszacowań

$$\begin{aligned}\left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_*\right\| &< 2\left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_* - \frac{1}{N}\boldsymbol{\mathcal{G}}\right\| = 2\left\|\mathcal{R}(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2)\right\| \\ &\leqslant \left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_*\right\| \cdot \left\|\boldsymbol{\mathcal{J}}_*^{-1}\right\| \cdot \|\boldsymbol{d}_N\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}).\end{aligned}$$

Zatem, w rzeczywistości zachodzi zaprzeczenie nierówności (5.22). Ostatecznie uzyskujemy

$$\begin{aligned}\left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_* - \frac{1}{N}\boldsymbol{\mathcal{G}}\right\| &= \left\|\mathcal{R}(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2)\right\| \leqslant \frac{1}{2}M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \cdot \|\boldsymbol{d}_N\|^2 \\ &\leqslant \frac{1}{2}\left\|\sqrt{N}\boldsymbol{d}_N^{\mathsf{T}}\boldsymbol{\mathcal{J}}_*\right\| \cdot \left\|\boldsymbol{\mathcal{J}}_*^{-1}\right\| \cdot \|\boldsymbol{d}_N\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \\ &\leqslant \left\|\frac{1}{N}\boldsymbol{\mathcal{G}}\right\| \cdot \left\|\boldsymbol{\mathcal{J}}_*^{-1}\right\| \cdot M(O_{\boldsymbol{\beta}}, O_{\sigma^2}) \cdot \|\boldsymbol{d}_N\|,\end{aligned}$$

przy czym prawa strona nierówności zbiega według prawdopodobieństwa do zera, jako że stanowi iloczyn zmiennych stochastycznie ograniczonych i czynnika $\|\boldsymbol{d}_N\|$. Ostatecznie, $\sqrt{N}\boldsymbol{\mathcal{J}}_*\boldsymbol{d}_N$ również zbiega według dystrybuanty do $\mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{S}})$, a więc

$$\sqrt{N} \cdot \left(\begin{bmatrix} \hat{\boldsymbol{\rho}}_{\text{SEM\_QNW}} \\ \hat{\boldsymbol{\beta}}_{\text{SEM\_QNW}} \\ \hat{\sigma}^2_{\text{SEM\_QNW}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\rho}_0 \\ \boldsymbol{\beta}_0 \\ \sigma_0^2 \end{bmatrix} \right) \xrightarrow{N\to\infty} \mathcal{N}(0, \boldsymbol{\mathcal{J}}_0^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{G}}}\boldsymbol{\mathcal{J}}_0^{-1}).$$

□

Zakończenie

W niniejszej monografii przybliżyliśmy wybrane zagadnienia z zakresu ekonometrii przestrzennej. Prezentując matematyczne podstawy metod stochastycznych stosowanych w tej dziedzinie, szczególny nacisk położyliśmy na formalną argumentację stwierdzeń o własnościach asymptotycznych. W szczególności, elementem nowatorskim było użycie szerokiej klasy przestrzennych macierzy wag. Poprzez dopuszczenie w naszych rozważaniach macierzy niekoniecznie sumowalnych, opracowane w pracy narzędzia pozwalają w znacznym stopniu rozszerzyć wachlarz zastosowań modeli, dając tym samym badaczom większe możliwości konstrukcji specyfikacji z zależnościami przestrzennymi. Dodatkowo, dowodząc rezultatu o rozkładzie asymptotycznym gaussowskiego estymatora quasi-największej wiarogodności dla modeli autoregresji składnika losowego wyższych rzędów, rozszerzyliśmy istotną dla rozwoju dziedziny teorię zapoczątkowaną w pracach Gupta i Robinson (2018) oraz Olejnik i Olejnik (2020).

Istotnym elementem opracowania jest uzyskane przez nas centralne twierdzenie graniczne dla form liniowo-kwadratowych, które wykorzystywaliśmy w pracy do argumentacji własności asymptotycznych rozważanych statystyk i estymatorów. Twierdzenie to celowo sformułowaliśmy w sposób możliwie najbardziej ogólny. Kierowaliśmy się przekonaniem, że wynik ten znajdzie zastosowanie w opracowaniu nowych twierdzeń dotyczących również innych metod ekonometrii przestrzennej, a tym samym przyczyni się do rozwoju dziedziny. Wierzymy, iż dociekliwi czytelnicy, w swoich własnych badaniach poświęconych podstawom formalnym ekonometrii, niekoniecznie przestrzennej, mogą napotkać problemy, w których rozwiązaniu pomoże teoria asymptotyczna, zapoczątkowana w tej książce.

Bibliografia

Anselin L. (1988a), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publications, Dordrecht.

Anselin L. (1988b), *Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity*, „Geographical Analysis" 20: 1–17.

Anselin L. (1996), *The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association*, [w:] M. Fischer, H. Scholten, D. Unwin (eds.), *Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences*, Taylor and Francis, London, s. 111–125.

Anselin L. (2001), *Rao's score test in spatial econometrics*, „Journal of Statistical Planning and Inference" 97: 113–139.

Anselin L. (2002), *Under the hood: issues in the specification and interpretation of spatial regression models*, „Agricultural Economics" 27: 247–267.

Anselin L., Bera A. K. (1998), *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*, [w:] A. Ullah, D. E. A. Giles (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, Inc., New York, s. 237–289.

Anselin L., Rey S. J. (2014), *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*, GeoDa Press LLC, Chicago.

Arbia G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Publishers, Boston.

Arbia G. (2006), *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Advances in Spatial Science, Springer, Berlin.

Arbia G., Baltagi B. H. (eds.) (2009), *Spatial Econometrics*. Methods and Applications, Springer, Berlin.

Badinger H., Egger P. (2013), *Estimation and testing of higher-order spatial autoregressive panel data error component models*, „Journal of Geographical Systems" 15 (4): 453–489.

Baltagi B. H., Lesage J. P., Pace R. K. (2016), *Spatial Econometrics: Qualitative and Limited Dependent Variables*, „Advances in Econometrics" 37, Emerald.

Beran J. (1972), *Rank spectral processes and tests for serial dependence*, „Annals of Mathematical Statistics" 43: 1749–1766.

Besner C. (2002), *A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme*, „Journal of Real Estate Research" 24: 193–212.

Bhansali R. J., Giraitis L., Kokoszka P. S. (2007), *Convergence of quadratic forms with non-vanishing diagonal*, „Statistics and Probability Letters" 77: 726–734.

Billingsley P. (2009), *Prawdopodobieństwo i miara*, przeł. K. Kizeweter, J. E. Roguski, wyd. 2, Wydawnictwo Naukowe PWN, Warszawa.

Bivand R., Hauke J., Kossowski T. (2013), *Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods*, „Geographical Analysis" 45 (2): 150–179.

Bodson P., Peeters D. (1975), *Estimations of the coefficients in a linear regression in the presence of spatial autocorrelation: an application to a Belgian labour–demand function*. „Environment and Planning" 7 (4): 455–472.

Born B., Breitung J. (2011), *Simple regression-based tests for spatial dependence*, „The Econometrics Journal" 14 (2): 330–342.

Cliff A. D., Ord J. K. (1972), *Testing for spatial autocorrelation among regression residuals*, „Geographical Analysis" 4: 267–284.

Cliff A. D., Ord J. K. (1973), *Spatial Autocorrelation*, Pion, London.

Cliff A. D., Ord J. K. (1981), *Spatial Processes: Models and Applications*, Pion, London.

Corrado L., Fingleton B. (2011), *Where is the economics in spatial econometrics?*, „Journal of Regional Science" 52 (2): 210–239.

Cramér H., Wold H. (1936), *Some Theorems on Distribution Functions*, „Journal of the London Mathematical Society" 11 (4): 290–294.

Dacey M. F. (1968), *A review on measures of contiguity for two and k–color maps*. Technical Report No. 2, „Spatial Diffusion Study", Department of Geography, Evanston, Northwestern University.

de Jong P. (1987), *A central limit theorem for generalized quadratic forms*, „Probability Theory and Related Fields" 75: 261–277.

de Jong P., Sprenger C., van Veen F. (1984), *On Extreme Values of Moran's I and Geary's c*, „Geographical Analysis" 16 (1): 17–24.

Deng M. (2008), *An anisotropic model for spatial processes*, Geographical Analysis 40 (1): 26–51.

Durbin J., Watson G. S. (1950), T*esting for serial correlation in least-squares regression I*, „Biometrika" 37: 159–178.

Durbin J., Watson G. S. (1951), *Testing for serial correlation in least-squares regression II*, „Biometrika" 38: 409–428.

Elhorst J. P. (2001), *Dynamic models in space and time*, „Geographical Analysis" 33 (2): 119–140.

Elhorst J. P., Halleck S. (2013), *On spatial econometric models, spillover effects, and W*, 53rd Congress of the European Regional Science Association: „Regional Integration: Europe, the Mediterranean and the World Economy", 27–31 August 2013, Palermo, Italy, http://hdl.handle.net/10419/123888 (dostęp 27.11.2020).

Elhorst J. P., Lacombe D. J., Piras G. (2012), *On model specification and parameter space definitions in higher order spatial econometric models*, „Regional Science and Urban Economics" 42 (1–2): 211–220.

Feng C., Wang H., Han Y., Xia Y., Tu, X. M. (2014), *The Mean Value Theorem and Taylor's Expansion in Statistics*. „The American Statistician" 67: 245–248.

Fingleton B. (1999), *Spurious spatial regression: some Monte Carlo results with spatial unit Root and Spatial Co-integration*, „Journal of Regional Science" 39: 1–19.

Fisher W. (1971), *Econometric estimation with spatial dependence*, „Regional and Urban Economics" 1: 19–40.

Florax R. J. G. M., Anselin L. (1995), *New Directions in Spatial Econometrics*, Springer, Berlin.

Florax R. J. G. M., Anselin L. (2004), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, Berlin.

Fujita M., Krugman P., Mori T. (1999a), *On the evolution of hierarchical urban systems*, „European Economic Review" 43 (2): 209–251.

Fujita M., Krugman P., Venables A. J. (1999b), *The spatial economy: Cities, regions and international trade*. MIT Press, Cambridge MA.

Getis A., Aldstadt J. (2004), *Constructing the Spatial Weights Matrix Using A Local Statistic*, „Geographical Analysis" 36: 90–114.

Getis A., Ord J. K. (1992), *The analysis of spatial association by distance statistics*, „Geographical Analysis" 24: 189–206.

Giraitis L., Taqqu M. (1998), *Central limit theorems for quadratic forms with time-domain conditions*, „Annals of Probability" 26: 377–398.

Griffith D. A. (2003), *Spatial Autocorrelation and Spatial Filtering. Gaining Understanding Through Theory and Scientific Visualization*, Springer.

Gupta A., Robinson P. (2015), *Inference on higher-order spatial autoregressive models with increasingly many parameters*, „Journal of Econometrics" 186: 19–31.

Gupta A., Robinson P. (2018), *Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension*, „Journal of Econometrics" 202: 92–107.

Hájek P., Johanis M. (2014), *Smooth analysis in Banach spaces*, De Gruyter Series in Nonlinear Analysis and Applications 19, De Gruyter, Berlin.

Hall P., Hyde C. C. (1980), *Martingale limit theory and its application*, Academic Press, Inc., New York.

Han X., Hsieh C., Lee L. F. (2017), *Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach*, „Regional Science and Urban Economics" 63: 97–120.

Hordijk L. (1974), *Spatial correlation in the disturbances of a linear interregional model*, „Regional Science and Urban Economics" 4 (3): 117–140.

Horn R. A., Johnson C. R. (2013), *Matrix Analysis*, Cambridge University Press, New York.

Jakubowski J., Sztencel R. (2001), *Wstęp do teorii prawdopodobieństwa*, SCRIPT, Warszawa.

Kelejian H. H., Piras G. (2014), *Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes*, „Regional Science and Urban Economics" 46: 140–149.

Kelejian H. H., Piras G. (2017), *Spatial Econometrics*, Academic Press, London.

Kelejian H. H., Prucha I. R. (1998), *A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances*, „Journal of Real Estate Finance and Economics" 17: 99–121.

Kelejian H. H., Prucha I. R. (2001), *On the Asymptotic Distribution of the Moran I Test Statistic with Applications*, „Journal of Econometrics" 104: 219–257.

Kelejian H. H., Prucha I. R. (2010), *Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances*, „Journal of Econometrics" 157 (1): 53–67.

Klaassen L. H., Paelinck J. H. P., Wagenaar S. (1979) *Spatial Systems*, Saxon House, Farnborough.

Kopczewska K. (2007), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa.

Kosfeld R., Lauridsen J. (2004), *Dynamic spatial modeling of regional convergence processes*, „Empirical Economics" 29: 705–722.

Kosfeld R., Lauridsen J. (2006), *A test strategy for spurious regression, spatial non-stationarity, and spatial co-integration*, „Papers in Regional Science" 85 (3): 363–377.

Kossowski T. (2010), *Teoretyczne aspekty modelowania przestrzennego w badaniach regionalnych*, „Rozwój Regionalny i Polityka Regionalna" 12: 9–26.

Kossowski T., Hauke J. (2011), *The method of computing the Log-Jacobian of the variable transformation for spatial models – test and comments*, „Acta Universitatis Lodziensis". Folia Oeconomica 252: 161–173.

Krugman P. (1991a), *Increasing returns and economic geography*. „Journal of Political Economy" 99 (3): 483–499.

Krugman P. (1991b), *Geography and Trade*. MIT Press.

Lauridsen J. (1999), *Spatial Co-integration Analysis in Econometric Modelling*, „ERSA Conference Papers" ersa99pa181, European Regional Science Association.

Lauridsen J. (2006), *Spatial autoregressively distributed lag models: equivalent forms, estimation and an illustrative commuting model*, „The Annals of Regional Science" 40: 297–311.

La Vallée Poussin C. de (1915), *Sur L'Integrale de Lebesgue*, „Transactions of the American Mathematical Society" 16 (4): 435–501.

Lee L. F. (2002), *Consistency and efficiency of least-squares estimation for mixed regressive spatial autoregressive models*, „Econometric Theory" 18: 252–277.

Lee L. F. (2004), *Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models*, „Econometrica" 72: 1899–1925.

Lee L. F., Yu J. (2010), *Estimation of spatial autoregressive panel data models with fixed effects*, „Journal of Econometrics" 154 (2): 165–168.

Lee L. F., Liu X., Lin X. (2010), *Specification and estimation of social interaction models with network structures*. „The Econometrics Journal" 13 (2): 145–176.

Lehmann E. L., Casella G. (1998), *Theory of Point Estimation*, 2nd ed., Springer, New York.

LeSage J. (1999), *Spatial Econometrics: The Web Book of Regional Science*, Regional Research Institute, West Virginia University, Morgantown.

LeSage J., Pace, R. K. (2009), *Introduction to Spatial Econometrics*, Statistics: Textbooks and Monographs, Chapman and Hall, Boca Raton, Florida.

Li K. (2017), *Fixed-effects dynamic spatial panel data models and impulse response analysis*, „Journal of Econometrics" 198 (1): 102–121.

Liu S. F., Yang Z. (2015), *Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and non-normality*, „Regional Science and Urban Economics" 52: 50–70.

Łaszkiewicz E. (2016), *Ekonometria przestrzenna III. Modele wielopoziomowe – teoria i zastosowania*, C.H. Beck, Warszawa.

Meyer P. A. (1966), *Probability and potentials*, Blaisdell Publishing Co., Waltham, Mass.

Moran P. (1950), *Notes on Continuous Stochastic Phenomena*, „Biometrika" 37: 17–23.

Mynbaev K. T. (2010), *Asymptotic distribution of the OLS estimator for a mixed spatial model*, „Journal of Multivariate Analysis" 101 (3): 733–748.

Mynbaev K. T. (2011), *Short-Memory Linear Processes and Econometric Applications*, John Wiley and Sons, Hoboken, NJ.

Mynbaev K. T., Ullah A. (2008), *Asymptotic distribution of the OLS estimator for a purely autoregressive spatial model*, „Journal of Multivariate Analysis" 99 (2): 245–277.

Nijkamp P., Fischer M. M. (eds.) (2014), *Handbook of Regional Science*, Springer, Heidelberg.

Olejnik A. (2008), U*sing the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU25*, „Papers in Regional Science" 87 (3): 371–384.

Olejnik A. (2013), *Wybrane metody testowania modeli regresji przestrzennej*, „Przegląd Statystyczny" 60 (3): 381–393.

Olejnik A., Olejnik J. (2019), *Increasing returns to scale, productivity and economic growth – a spatial analysis of the contemporary EU economy*, „Argumenta Oeconomica" 42 (1): 273–293.

Olejnik J., Olejnik A. (2020), *QML estimation with non-summable weight matrices*, „Journal of Geographical Systems" 22 (4): 469–495.

Olejnik A., Özyurt S., Olejnik J. (2020), *Introducing multi-dimensional weighting factors into spatial econometric models*, „Ekonomika Regiona / Economy of Region" [w recenzji].

Ord J. K., Getis A. (1995), *Local spatial autocorrelation statistics: distributional issues and an application*, „Geographical Analysis" 27 (4): 286–306.

Paelinck J. H. P., Klaassen L. H. (1979), Spatial Econometrics, Saxon House, Farnborough.

Panak [Olejnik] A. (2006), *Autokorelacja przestrzenna i kriging – metodologia i wybrane zastosowania*, „Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu", Taksonomia 13: 483–491.

Pinkse J. (1999), *Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models*, Department of Economics, University of British Columbia and University College, London.

Pötscher M. B., Prucha I. R. (1997), *Dynamic Non-linear Models: Asymptotic Theory*, Springer-Verlag, Berlin, Heidelberg.

Pruss A. R. (1998), *A bounded N-tuplewise independent and identically distributed counterexample to the CLT*, „Probability Theory and Related Fields" 111: 323–332.

Qu X., Lee L. F. (2017), *QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices*, „Journal of Econometrics" 197: 173–201.

Rao C. R. (1948), *Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation*, „Proceeding of the Cambridge Philosophical Society" 40: 50–57.

Sen A. (1976), *Large sample-size distribution of statistics used in testing for spatial correlation*, „Geographical Analysis" 9: 175–184.

Shi W., Lee L. F. (2017), *Spatial dynamic panel data models with interactive fixed effects*, „Journal of Econometrics" 197: 173–201.

Silvey S. D. (1959), *The Lagrangian multiplier test*, „Annals of Mathematical Statistics" 30: 389–407.

Suchecki B. (red.) (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H. Beck, Warszawa.

Suchecki B. (red.) (2012), *Ekonometria przestrzenna II. Modele zaawansowane*, C.H. Beck, Warszawa.

Szulc E. (2007), *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Takesaki M. (1979), *Theory of operator algebras I*, Springer-Verlag, New York.

Tobler W. (1970), *A computer movie simulating urban growth in the Detroit region*, „Economic Geography Supplement" 46: 234–240.

Whittle P. (1964), *On the convergence to normality of quadratic forms of independent variables*, „Theory of Probability and Its Applications" 9, 103–108.

Vega S. H., Elhorst J. P. (2015), *The SLX model*, „Journal of Regional Science" 55 (3): 339–363.

Yu J., de Jong R., Lee L. F. (2008), *Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects When Both n and T Are Large*. „Journal of Econometrics" 146 (1): 118–134.

Zeliaś A., Grabiński T., Ludwiczak B., Malina A. (1991), *Ekonometria przestrzenna*, PWE, Warszawa.