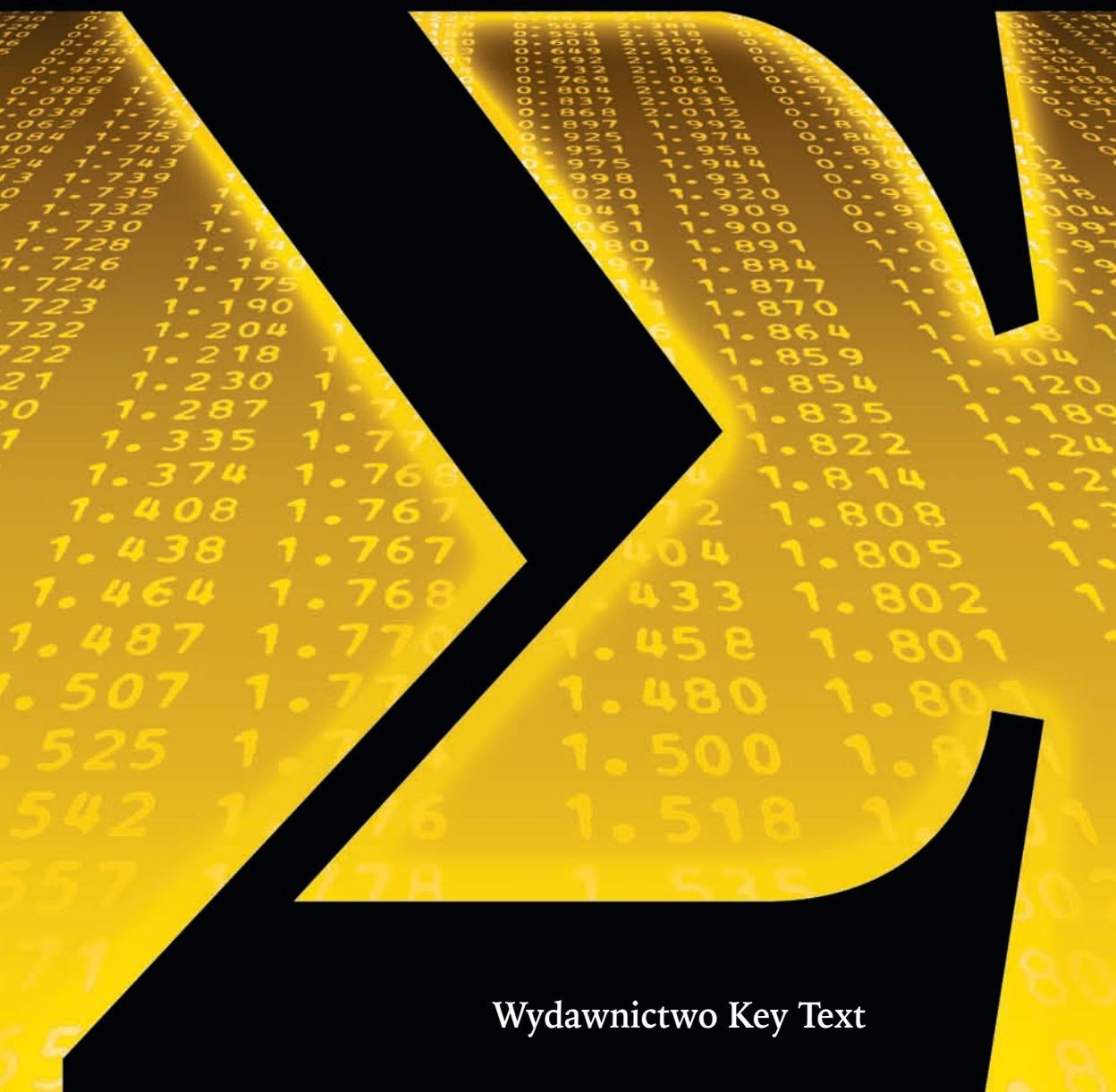


Brunon R. Górecki

Ekonometria

podstawy teorii i praktyki



Wydawnictwo Key Text

Ekonometria

podstawy teorii i praktyki

Brunon R. Górecki

Ekonometria

podstawy teorii i praktyki

Wydawnictwo Key Text

Recenzent

prof. dr hab. Jan B. Gajda

Opracowanie graficzne i typograficzne

Jacek Tarasiewicz

Redakcja

Zespół

© Copyright by Wydawnictwo Key Text

ISBN: 978-83-87251-13-0

Warszawa 2010

Wydawnictwo Key Text sp. z o.o.

ul. Sokołowska 9/410, 01-142 Warszawa

tel. 22 632 11 36, faks wew. 212, kom. 665 108 002

www.keytext.com.pl

wydawnictwo@keytext.com.pl

Spis treści

Wstęp	9
Część 1. Klasyczny model regresji liniowej	11
1. Wprowadzenie	13
1.1. Czym jest ekonometria?	13
1.2. Pojęcie modelu ekonometrycznego	14
1.3. Dane statystyczne	16
1.4. Metodologia ekonometrii	18
Podsumowanie	23
2. Podstawy klasycznego modelu regresji liniowej	25
2.1. Zapis macierzowy modelu	25
2.2. Od populacji do próby i od próby do populacji	26
2.3. Założenia klasycznego modelu regresji liniowej	29
Podsumowanie	34
3. Metoda najmniejszych kwadratów	35
3.1. Estymatory metody najmniejszych kwadratów	35
3.2. Własności algebraiczne rozwiązania MNK	38
Podsumowanie	44
4. Wnioskowanie o estymatorach metody najmniejszych kwadratów	45
4.1. Jeszcze o założeniu normalności zaburzeń losowych	45
4.2. Twierdzenie Gaussa-Markowa	46
4.3. Estymator wariancji zaburzenia losowego i błędy standardowe estymatorów	48
4.4. Rozkład <i>t-Studenta</i> , weryfikacja prostych hipotez i przedziały ufności ..	50
4.5. Istotność równania regresji	55
4.6. Asymptotyczne własności estymatorów MNK	56
Podsumowanie	58
5. Interpretacja równania regresji	59
5.1. Interpretacja współczynników regresji i założenie liniowości	59
5.2. Jakościowe zmienne objaśniające – regresory zerojedynkowe, oznaczane również jako zmienne 0–1 lub zmienne binarne	65
5.3. Restrykcje i modele zagnieżdżone. Łączna istotność zmiennych zerojedynkowych	71
5.4. Jakościowa zmienna objaśniana	73
5.5. Wybór regresorów zgodnie z zasadą „Od ogólnego do szczegółowego”. Skutki pominięcia w równaniu regresji istotnych zmiennych objaśniających; skutki dodania do równania regresji zmiennych nieistotnych	74
5.6. Testowanie łącznej istotności podzbioru regresorów	78
5.7. Testowanie hipotez złożonych	80
Podsumowanie	86

6.	Problemy wynikające z niedoskonałości danych statystycznych	89
6.1.	Współliniowość i jej konsekwencje. Wykrywanie współliniowości i środki zaradcze	89
6.2.	Obserwacje opuszczone	94
6.3.	Wykrywanie nietypowych wartości zmiennej objaśnianej i nietypowych wartości zmiennych objaśniających (obserwacje znaczące)	94
	Podsumowanie	96
7.	Prognozowanie na podstawie klasycznej metody regresji liniowej	99
7.1.	Prognoza i błąd standardowy prognozy	99
	Podsumowanie	104
	Literatura uzupełniająca do części I	105
	Część 2. Złagodzenie założeń modelu klasycznego	107
8.	Uogólniona metoda najmniejszych kwadratów	109
8.1.	Heteroskedastyczność i autokorelacja zaburzeń losowych w KMRL	109
8.2.	Estymatory uogólnionej metody najmniejszych kwadratów	111
8.3.	Testowanie heteroskedastyczności: testy Goldfelda-Quandt, Breuscha-Pagana oraz White'a	113
8.4.	Estymacja macierzy wariancji-kowariancji zaburzeń losowych w przypadku heteroskedastyczności. Stosowalna uogólniona metoda najmniejszych kwadratów	120
8.5.	Estymator White'a macierzy wariancji-kowariancji dla b wyznaczonego za pomocą MNK – odporny na heteroskedastyczność	122
8.6.	Testowanie autokorelacji: testy Durбина-Watsona i Breuscha-Godfrey'a	124
8.7.	Estymacja macierzy wariancji-kowariancji zaburzeń losowych w przypadku autokorelacji zaburzeń pierwszego rzędu	131
8.8.	Estymator Neweya-Westa macierzy wariancji-kowariancji dla b oszacowanego za pomocą MNK – odporny na heteroskedastyczność i odporny na autokorelację	133
	Podsumowanie	134
9.	Diagnostyka w klasycznej metodzie regresji liniowej	135
9.1.	Test White'a	135
9.2.	Test RESET błędu specyfikacji postaci funkcyjnej równania regresji Ramsey'a	136
9.3.	Test niezagnieźdzonych alternatyw	138
9.4.	Testy stabilności parametrów Chowa	142
9.5.	Test Jarque-Bera normalności zaburzeń	149
9.6.	Ocena wyników analizy regresji	151
	Podsumowanie	151
	Literatura uzupełniająca do części II	152
	Część 3. Szczególnie ważne modele ekonometryczne	153
10.	Ograniczona zmienna objaśniana	155
10.1.	Liniowa funkcja prawdopodobieństwa	155
10.2.	Metody logitowa i probitowa	156

10.3. Wielomianowa metoda logitowa, metoda tobitowa, modele samoselekcji próby	162
Podsumowanie	164
11. Modele jednowymiarowych szeregów czasowych	167
11.1. Analiza klasyczna	167
11.2. Szereg czasowy jako realizacja procesu stochastycznego	168
11.3. Procedura Boxa-Jenkinsa	172
11.4. Funkcja autokorelacji i cząstkowej autokorelacji szeregu Dow Jones ...	176
11.5. Procesy ARIMA dla danych sezonowych	180
Podsumowanie	183
12. Modele dynamiczne	185
12.1. Problemy ekonometryczne modeli dynamicznych	185
12.2. Modele o opóźnieniach rozłożonych (<i>Distributed Lag Models</i>)	186
12.3. Estymacja modeli DL i wybór rzędu opóźnienia	187
12.4. Modele autoregresyjne i modele autoregresyjne z opóźnieniami rozłożonymi (<i>AutoRegressive Distributed Lag Models</i> – Modele ADL lub ARDL)	188
12.5. Niestacjonarność i integracja szeregu	192
12.6. Test pierwiastka jednostkowego Dickeya-Fullera (test DF)	193
12.7. Rozszerzony test pierwiastka jednostkowego (test ADF)	196
12.8. Kointegracja szeregów czasowych	199
12.9. Przyczynowość w ekonometrii	201
Podsumowanie	203
13. Modele wektorowej autoregresji (VAR) i modele korekty błędem	205
13.1. Modele wielorównaniowe	205
13.2. Modele wektorowej autoregresji (<i>Vector AutoRegressive Models</i> – VAR)	206
13.3. Model korekty błędem (równowagi) (<i>Error Correction Model</i> – ECM)	219
Podsumowanie	221
14. Opracowanie projektów badawczych	223
Literatura uzupełniająca do części III	226
Aneksy	227
A. Elementy algebry macierzy	229
B. Wybrane zagadnienia rachunku prawdopodobieństwa	243
C. Bazy danych	253
Bibliografia	255
Indeks	257

Wstęp

Książka zawiera podstawowy kurs teorii i praktyki ekonometrii. Jest przeznaczona dla studentów różnych dyscyplin ekonomicznych poza specjalizacją ekonometrii. Będzie również użyteczna dla ekonomistów prowadzących analizy danych ekonomicznych, a jednocześnie niedysponujących solidnymi podstawami matematycznymi.

W podręczniku uwzględnione zostały najnowsze ujęcia ekonometrii, które rozumiemy dwojako. Po pierwsze – jest to nowe ujęcie problemów tradycyjnej ekonometrii. Spośród przykładów odmiennych ujęć tej teorii od spojrzenia tradycyjnego należy wymienić: podstawową myśl filozoficzną współczesnej ekonometrii – „Od populacji do próby i od próby do populacji” (podrozdział 2.2), fundamentalny problem wyboru zmiennych objaśniających (podrozdział 5.5.), zagadnienie błędu standardowego White’a, usuwającego komplikacje wywołane heteroskedastycznością zaburzeń losowych (podrozdział 8.5), czy też nadanie podstawowego znaczenia testom diagnostycznym przy ocenie poprawności szacowanego modelu (rozdział 9).

Po drugie – jest to szersze ujęcie modeli dynamicznych wykorzystujących szeregi czasowe, stanowiących dominujący zbiór danych używanych w ekonomii. Analizowane są modele oparte na szeregach stacjonarnych (łącznie z testem pierwiastka jednostkowego) i na szeregach niestacjonarnych (łącznie z problematyką kointegracji). Natomiast modele wielorównaniowe są rozważane jedynie w kontekście modeli wektorowej autoregresji (VAR) lub modeli korekty błędem równowagi (ECM), a nie w kontekście wielkich wielorównaniowych modeli gospodarki.

Książka, stawiając pytanie „Dlaczego tak, a nie inaczej estymujemy modele ekonometryczne?” nie zaniedbuje odpowiedzi na pytanie: „Jak je obliczamy?”. Zawiera liczne przykłady (dotyczące Polski, Unii Europejskiej i gospodarki światowej) zastosowania omawianych metod w różnych dziedzinach nauk ekonomicznych. Obliczenia przykładów prowadzono przy użyciu bezpłatnego pakietu ekonometrycznego GRET. Został on opublikowany przez Free Software Foundation i jest dostępny pod adresem internetowym <http://gretl.sourceforge.net>. Polskojęzyczna wersja tego pakietu, opracowana przez prof. Tadeusza Kufla, jest udostępniona na stronie internetowej <http://www.kufel.torun.pl>. Do stosowania tego pakietu pomocna

jest książka T. Kufla, *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, WN PWN, Warszawa 2004. Niech to będzie jednocześnie okazja dla złożenia najszczerzego podziękowania prof. Tadeuszowi Kuflowi za zgodę na wykorzystanie w niniejszej książce niepublikowanych wcześniej zbiorów danych.

Istotnym dopełnieniem podręcznika są trzy aneksy. Pierwszy aneks zawiera niezbędną, ograniczoną jedynie do wątków bezpośrednio stosowanych w trakcie wykładu, wiedzę z zakresu algebry macierzy. Drugi aneks jest poświęcony celowo wybranym fragmentom statystyki matematycznej, które są niezbędne do swobodnego korzystania z tekstu. Trzeci aneks jest zbiorem dostępnych w internecie baz danych ekonomicznych stanowiących materiał nie tylko do sensownego formułowania zadań ćwiczeniowych, lecz także do wykorzystywania w interesujących ekonomistę analizach.

Pragnę podziękować mgr. Dariuszowi Szymańskiemu za przygotowanie niektórych przykładów przedstawionych w niniejszym tekście oraz mgr. Tomaszowi Rybnikowi za opracowanie informacji o dostępnych w internecie ekonomicznych bazach danych, które Czytelnik może zastosować przy samodzielnym formułowaniu i rozwiązywaniu zagadnień ekonometrycznych.

Część I

Klasyczny model regresji liniowej

Wprowadzenie

1.1. Czym jest ekonometria?

Ekonometria jest nauką zajmującą się ilościowym (liczbowym) opisem, na podstawie danych statystycznych, prawidłowości ekonomicznych, postulowanych przez teorię ekonomii lub sugerowanych przez sensowne hipotezy ekonomiczne.

Hipoteza to przypuszczenie wymagające sprawdzenia. Na przykład: Czy wykształcenie przysparza tyle samo złotych miesięcznej płacy kobietom, co mężczyznom? Czy w każdym wieku awansuje się jednakowo szybko? Czy kobiety w młodym wieku awansują szybciej od mężczyzn? Czy krańcowa skłonność do konsumpcji w Polsce jest taka sama jak w Stanach Zjednoczonych?

W zbiorze metod ilościowych – obok matematyki, statystyki i badań operacyjnych – ekonometria jest ważnym i użytecznym narzędziem wspomagającym prowadzenie analiz ekonomicznych. Można powiedzieć, że ekonometria pomaga poznać przeszłość i teraźniejszość, a także określać przyszłość wydarzeń ekonomicznych.

Ważnym problemem staje się sposób rozpoznawania teraźniejszości i przyszłości.

Badania ekonometryczne są przeprowadzone na podstawie sformułowań teorii ekonomii lub wyraźnie określonych hipotez, dotyczących procesów lub zjawisk ekonomicznych, które mają być przedmiotem badania i weryfikacji empirycznej. Teoria jest modelem opisowym całej rzeczywistości gospodarczej lub jej części, zawiera również zbiór reguł wiążących podstawowe wielkości ekonomiczne. Model ekonometryczny jest liczbowym przedstawieniem opartym na zaobserwowanych danych sformułowanej w teorii prawidłowości.

Jednorównaniowym modelem ekonometrycznym nazywamy równanie, w którym występuje:

- a) zmienna objaśniana,
- b) zmienne objaśniające – kształtujące procesy ujęte w teorii lub w postawionych hipotezach,

- c) parametry – mierzące efekty oddziaływania zmiennych objaśniających na zmienną objaśnianą oraz
- d) zaburzenie losowe – oddające wpływ drugorzędnych, *explicite* niewyróżnionych czynników.

Zbiór metod, którymi posługuje się ekonometria, nazywa się **ekonometrią teoretyczną lub teorią ekonometrii** – w odróżnieniu od zastosowań tych metod zwanych **ekonometrią stosowaną**.

Rozważania będą poświęcone zarówno teorii ekonometrii, jak i jej zastosowaniom, dlatego wszystkie wątki teorii będą ilustrowane konkretnymi aplikacjami praktycznymi. W takim ujęciu poniżej prezentowany materiał staje się rodzajem przewodnika po metodach i jednocześnie po zastosowaniach tych metod. Nie będzie on miał charakteru wykładu matematycznego, obarczonego dowodami twierdzeń o wysokim stopniu trudności matematycznych.

Znacznie silniej niż w innych polskojęzycznych podręcznikach będą akcentowane problemy testowania poprawności doboru modelu, weryfikacji stawianych hipotez, dyskusji nad konsekwencjami wykrywania różnego rodzaju błędów specyfikacji i dyskusji nad kryteriami wyboru modelu.

Przyjmuję, że studenci korzystający z tego opracowania mają możliwości używania pakietów ekonometrycznych, takich jak LIMPED, TSP, EVIEWS, SAS, SPSS, STATA, PcGive i szeregu innych (w tym pakietów nieodpłatnie udostępnianych w internecie), których listę można znaleźć pod adresem: <http://www.oswego.edu/economic/econsoftware.htm>

Za ich pomocą mogą być wykonywane prawie wszystkie obliczenia dla diskutowanych w niniejszym opracowaniu technik i testów. W obliczeniach zawartych w tekście jest wykorzystywany darmowy, nieodpłatnie udostępniony pakiet, publikowany przez Free Software Foundation pod nazwą **GRET**. Jego adres internetowy: <http://gretl.sourceforge.net> oraz <http://www.kufel.torun.pl>. Dla wykorzystania tego pakietu pomocne będzie opracowanie Tadeusza Kufla, *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*, WN PWN, Warszawa 2004.

1.2. Pojęcie modelu ekonometrycznego

Model ekonometryczny, jak każdy model, jest uproszczoną wizją rzeczywistości. Szczególną cechą modelu ekonometrycznego jest przedstawienie zależności zachodzącej między zjawiskiem objaśnianym i najważniejszymi zjawiskami objaśniającymi za pomocą równania zwanego **równaniem regresji**¹. Równanie

¹ Nazwa „regresja” została użyta po raz pierwszy przez Francis Galtona w końcu XIX wieku w badaniu nad wzrostem potomstwa w zależności od wzrostu rodziców, w którym to badaniu Galton sformułował tezę, że wzrost potomstwa w całej populacji zmierza do średniego wzrostu w populacji. Owo zmierzanie do średniej Galton określił właśnie słowem „regresja”.

to ma konkretną postać matematyczną, w której pojedyncza zmienna, zwana zmienną objaśnianą, przedstawiana jest jako funkcja deterministyczna (najczęściej liniowa) najważniejszych (w świetle teorii ekonomicznej) zmiennych, zwanych zmiennymi objaśniającymi. Do takiego równania dodane jest tak zwane zaburzenie losowe zwane również zaburzeniem stochastycznym², składnikiem losowym lub błędem losowym, którego głównym celem jest przedstawienie sumarycznego oddziaływania na zmienną objaśnianą wszystkich innych czynników, pominiętych w równaniu, ze względu na ich drugorzędne znaczenie dla opisu badanego związku. Błąd ten może również wynikać z nieadekwatności teorii ekonomii lub niepoprawności postawionych hipotez, które ma weryfikować model, a także z niewłaściwej postaci równania matematycznego przyjętego w modelu. Zaburzenie losowe zawiera także błędy pomiaru zmiennych, wynikające z niedokładności mierzenia wartości, jakie przyjmują zmienne, jak również z uchybień zawinionych zarówno przez ankieterów i osoby gromadzące dane, jak i udzielające odpowiedzi na pytania ankieterów. Wreszcie, w zaburzeniu losowym są zawarte wpływy indywidualnych cech jednostek, które obok wyróżnionych zmiennych objaśniających mogą oddziaływać na zmienną objaśnianą.

Zmienna objaśniana zwana jest również:

- zmienną zależną,
- zmienną endogeniczną,
- regresantem,
- zmienną kontrolowaną.

Zmienne objaśniające w różnych kontekstach mogą być nazywane:

- zmiennymi niezależnymi,
- zmiennymi egzogenicznymi,
- regresorami,
- zmiennymi kontrolującymi.

Zaburzenie losowe zwane jest również:

- błędem losowym,
- składnikiem stochastycznym lub składnikiem losowym,
- zakłóceniem,
- innowacją (niekiedy).

W związku z tym określenia te będziemy używać zamiennie.

Dodanie do równania regresji składnika losowego powoduje, że **równanie regresji nie ma charakteru deterministycznego, a stochastyczny** (co znaczy tyle samo co losowy), gdyż zmienna objaśniana będąc funkcją składnika losowego staje się sama zmienną losową. Związek regresyjny (jaki zachodzi między zmienną objaśnianą a zbiorem zmiennych objaśniających) jest związkiem statystycznym, który jednak nie implikuje charakteru przy-

² Stochastyczny to znaczy losowy lub przypadkowy (od greckiego słowa *stochasis* – domysł).

czynowego tej zależności. Treść ekonomiczna zmiennej objaśnianej i zmiennych objaśniających może wskazywać, że relacja wyznaczona przez równanie regresji jest relacją przyczynową.

Analiza regresyjna jest koncepcyjnie odmienna od analizy korelacyjnej, często stosowanej w statystyce w celu opisanie związku między dwiema zmiennymi. Na podstawie analizy regresyjnej szacujemy wartość oczekiwaną zmiennej objaśnianej za pomocą konkretnych wartości zmiennych objaśniających. W tym sensie analiza regresyjna pozwala prognozować zmienną objaśnianą na podstawie znanych, konkretnych wartości zmiennych objaśniających. Nie jest to więc relacja symetryczna, jaką jest współczynnik korelacji.

1.3. Dane statystyczne

Zbiory gromadzonych informacji o zjawiskach ekonomicznych noszą ogólną nazwę danych statystycznych. Odgrywają one podstawową rolę w modelowaniu ekonometrycznym. Potwierdzają poprawność specyfikacji funkcji regresji, poprawność postawionych hipotez lub – ogólniej rzecz ujmując – poprawność modelowania ekonometrycznego i wyprowadzanych na podstawie tego modelowania wniosków.

Możemy wyróżnić trzy rodzaje danych statystycznych wykorzystywanych w modelowaniu ekonometrycznym.

Dane szeregów czasowych. Są to najbardziej popularne zbiory informacji, w których kolejne obserwacje na różnych zmiennych rejestrują badane zjawiska ekonomiczne w następujących po sobie momentach lub przedziałach czasu. Takimi zmiennymi są np. PKB, zatrudnienie, stopa inflacji, liczba ludności, zestawiane jako dane roczne, kwartalne, miesięczne a nawet dzienne (jak np. wartość jednostek uczestnictwa funduszy inwestycyjnych), czy dane godzinowe (kursy walutowe, stopy zwrotu papierów wartościowych). Dane roczne, kwartalne lub miesięczne, z którymi najczęściej spotykamy się w modelowaniu ekonometrycznym są z reguły danymi zagregowanymi, wyrażającymi przeciętną wartość badanego zjawiska rejestrowanego w określonym przedziale czasu lub na określony moment (np. stan ludności na 31 grudnia czy liczba bezrobotnych na 30 czerwca).

Dane szeregów czasowych, choć powszechnie wykorzystywane w ekonometrii mogą powodować poważne problemy modelowania, związane z tak zwaną niestacjonarnością szeregów czasowych. Sygnalizując jedynie problem powiemy, że jeżeli szeregi, wraz z upływem czasu, wykazują trendy lub wzrastającą wariancję obserwowanego zjawiska, co wskazuje na ich niestacjonarność, to fakt ten rodzi szczególne komplikacje budowy i weryfikacji modeli. Do problemu tego wrócimy w rozdziale 11.

Dane przekrojowe powstają jako obserwacje dokonywane w tym samym czasie na wielu jednostkach. Typowymi danymi przekrojowymi są obser-

wacje budżetów gospodarstw domowych. Główny Urząd Statystyczny bada każdego miesiąca około trzech tysięcy gospodarstw z terenu całego kraju, dostarczając informacji o dochodach i wydatkach badanych rodzin, o ich składzie demograficznym, o ich mieszkaniu i jego wyposażeniu w dobra trwałego użytku itd. Są to dane bardzo szczegółowe, uzewnętrzniające różnorodność badanych gospodarstw domowych. Różnorodność ta wywołuje problemy zwane heteroskedastycznością, co niekiedy komplikuje modelowanie ekonometryczne.

Dane panelowe (połączone, longitudinalne) łączą cechy danych szeregów czasowych i danych przekrojowych. Na przykład dane PKB dla poszczególnego kraju są pojedynczym szeregiem czasowym, ale zestawienie PKB dla np. krajów OECD tworzy dane połączone. Typowymi danymi panelowymi są panele gospodarstw domowych. Na przykład dla Polski zostały zestawione panele z lat 1993–1997 i 1997–2000, zawierające około trzech tysięcy tych samych gospodarstw badanych przez okresy czteroletnie. W Luksemburgu znajduje się baza danych panelowych gospodarstw domowych dla wszystkich krajów Unii i Stanów Zjednoczonych. Dane te są dostępne również na Wydziale Nauk Ekonomicznych Uniwersytetu Warszawskiego. Zalety danych panelowych ilustruje następujący przykład. Na podstawie danych szeregów czasowych możemy ustalić, jak zmienia się z czasem procent pracujących kobiet. Na podstawie danych przekrojowych możemy sprawdzić, jak procent ten zależy od np. wieku kobiet, ich poziomu wykształcenia czy fazy cyklu rozwojowego rodziny. Ale dopiero dane panelowe umożliwiają ustalenie, czy w kolejnych latach kobietami pracującymi są te same kobiety, a więc umożliwiają zbadanie, jak zmienia się struktura pracujących kobiet: kiedy kobiety rozpoczynają, przerywają i wznowiają pracę.

Niedoskonałość danych statystycznych, mająca swoje różnorodne źródła, może mieć niekiedy decydujące znaczenie dla oszacowanego modelu. Możemy się dopatrywać złych wyników modelu w jego niepoprawnej specyfikacji i starać się go udoskonalać, gdy w rzeczywistości błąd tkwi w niedoskonałości danych statystycznych, której powody mogą być bardzo różnorodne. Ze względu na nieeksperymentalny charakter większości danych ekonomicznych mogą być one obciążone znacznymi błędami obserwacji lub pomiaru. W danych ankietowych braki odpowiedzi mogą prowadzić do tak zwanego obciążenia doboru próby, gdy odmawiający uczestnictwa zachowują się zdecydowanie odmiennie od poddających się badaniu.

Niekiedy wśród obserwacji pojawiają się tzw. **obserwacje znacząco wpływające** (*influential observations*), których wpływ na oszacowany model jest niezwykle silny, znacznie odbiegający od przeciętnego. Wyłączenie takich kilku lub nawet jednej obserwacji z próby może niekiedy istotnie zmienić szacowane parametry. Dlatego umiejętność ich zidentyfikowania staje się ważnym zabiegiem w przygotowaniu danych. Problemom związanym z obserwacjami wpływowymi (zwanymi w przypadku zmiennej objaśnianej obserwacjami nietypowymi, a w przypadku zmiennych objaśniających – ob-

serwacjami dźwigniowymi) na oszacowania parametrów modelu i ogólnie na jego jakość, ze względu na szczególną wagę tego zagadnienia poświęcimy specjalny fragment rozważań w rozdziale 6.

1.4. Metodologia ekonometrii

Przez metodologię ekonometrii rozumiemy sposób postępowania w trakcie budowy modelu ekonometrycznego. W tradycyjnym i do dzisiaj powszechnie zalecanym sposobie można wyróżnić następujące elementy:

1. Ustalenie teorii ekonomicznej lub zbioru hipotez, które model ekonometryczny ma potwierdzić lub odrzucić. W praktyce modelowania zwykle na tle teorii stawiamy hipotezy o tym, czy w konkretnych warunkach ekonomicznych teoria funkcjonuje w swojej niezmiennionej postaci, czy może dotąd rozpoznane czynniki, w świetle postawionych hipotez, zaczynają wywierać silniejszy lub słabszy wpływ niż poprzednio, a może ujawniają swoje działanie nowe, nieznane w teorii mechanizmy, które modyfikują istniejącą dotąd wizję rzeczywistości. Może modelowane procesy przebiegają szybciej lub wolniej niż w dotychczas rozpoznanych przypadkach. Teoria ekonomii i w jej świetle postawione hipotezy badawcze stanowią pierwszy krok w procedurze modelowania ekonometrycznego.
2. Określenie postaci matematycznej modelu ekonometrycznego polega na specyfikacji funkcji matematycznej, wiążącej zmienną objaśnianą ze zdefiniowanymi zmiennymi objaśniającymi. Metodologia ekonometrii zaleca (co znajduje swoje uzasadnienie ekonometryczne, patrz rozdz. 5), aby w początkowym modelu znalazł się możliwie najszerszy zbiór zmiennych ekonomicznych, które mają wyjaśnić zachowanie się zmiennej objaśnianej. W kolejnych krokach poprawy modelu zbiór ten może być redukowany. Jest to postępowanie zwane w ekonometrii „Od ogólnego do szczegółowego” (o czym będzie mowa w rozdz. 5).
3. Dołączenie zaburzenia losowego, dzięki czemu wyspecyfikowane równanie staje się modelem ekonometrycznym.
4. Ustalenie zbioru danych statystycznych użytych dla oszacowania modelu. Jest to szalenie istotny etap modelowania. Zaleca się wstępne sporządzenie wykresów zmiennych, a w dalszej procedurze wykrywanie obserwacji wpływowych (rozdz. 5). Często są bowiem przypadki, gdy mało satysfakcjonujący model ekonometryczny ma swoją przyczynę nie w wadliwej procedurze modelowania, a w istotnych niedomaganiach danych statystycznych użytych do jego budowy.
5. Estymacja parametrów modelu ekonometrycznego. W niniejszym tekście zajmujemy się jedynie dwiema metodami estymacji o powszechnym zastosowaniu, a mianowicie metodą najmniejszych kwadratów i uogólnioną metodą najmniejszych kwadratów. Zasygnalizujemy również za-

lety metody największej wiarygodności, stosowanej w przypadkach dużych prób.

6. Po estymacji modelu następuje jego diagnostyka, która pomaga ustalić, czy model nie zawiera istotnych wad wymagających poprawek. W tym celu oszacowany model jest poddawany szczegółowej weryfikacji za pomocą całego szeregu testów diagnostycznych, których przeprowadzenie pozwala na uzyskanie odpowiedzi, czy równanie modelu jest poprawnie wyspecyfikowane, czy zawiera on wszystkie ekonomicznie ważne zmienne, opisujące badane zjawisko, czy poprawne są założenia dotyczące specyfikacji zaburzenia losowego, czy uzyskane estymatory mają pożądane własności, wreszcie czy można znaleźć model lepszy od wyestymowanego.
7. Satysfakcjonujący model może służyć do sprawdzenia teorii ekonomii, inicjującej jego powstanie lub do testowania postawionych na wstępie hipotez. Należy zauważyć, że ważność teorii lub testowanych hipotez ma charakter warunkowy, to znaczy, że zależy od konkretnego zbioru danych statystycznych, wykorzystanych w procedurze estymacyjnej. Można by oczekiwać, że dla innego zbioru danych, obejmującego np. inny przedział czasu, wyniki estymacji mogłyby być odmienne.
8. Oszacowany model może być następnie wykorzystany dla celów progностycznych. Zakładając, że prawidłowość z okresu próby obowiązuje również w okresie prognozowanym, możemy wyznaczyć prognozę ekonometryczną i błąd *ex ante* dla tej prognozy. Po zrealizowaniu się prognozy możemy wyznaczyć jej błąd *ex post*.
9. Oszacowany model może być również wykorzystany dla obliczeń oczekiwanych efektów polityki ekonomicznej, fiskalnej, społecznej itp. Znając np. wymodelowane zachowanie się płatników podatku od dochodów osobistych ludności przy obowiązujących przepisach podatkowych, możemy wykorzystać model dla wyznaczenia spodziewanych efektów nowej formy polityki podatkowej.

Powyższe elementy metodologii przedyskutujemy na klasycznym przykładzie ekonomii, jakim jest Keynesowska teoria konsumpcji.

Zgodnie z teorią Keynesa „podstawowe prawo psychologiczne głosi, że ludzie są skłonni do zwiększania konsumpcji wraz ze wzrostem dochodów, ale w mniejszym stopniu niż wzrasta dochód”. Oznacza to, że krańcowa skłonność do konsumpcji MPC (*Marginal Propensity to Consume*) jest stopą zmiany konsumpcji. Oczekujemy, że jest ona większa od zera, ale mniejsza od jedności.

Matematyczny model dla tej teorii może przyjąć postać:

$$y = \beta_1 + \beta_2 x \quad 0 < \beta_2 < 1, \quad (1.1)$$

gdzie:

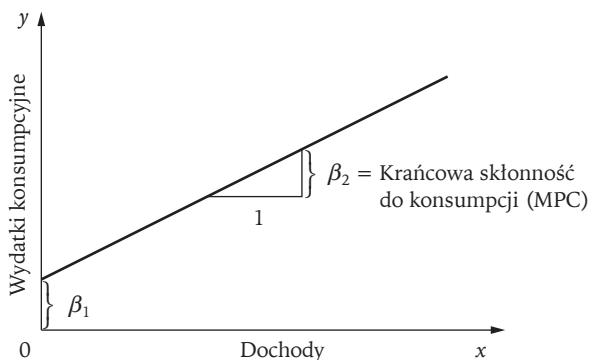
y – wydatki konsumpcyjne,

x – dochód,

β_1 i β_2 – parametry równania.

β_1 zwane jest **stałą równania**, β_2 jest **parametrem nachylenia** lub **współczynnikiem kierunkowym**. Ze wzoru (1.1) widzimy, że parametr kierunkowy β_2 mierzy krańcową skłonność do konsumpcji (MPC). Możemy to przedstawić graficznie, jak na rys. 1.1.

Rysunek 1.1. Funkcja konsumpcji Keynesa

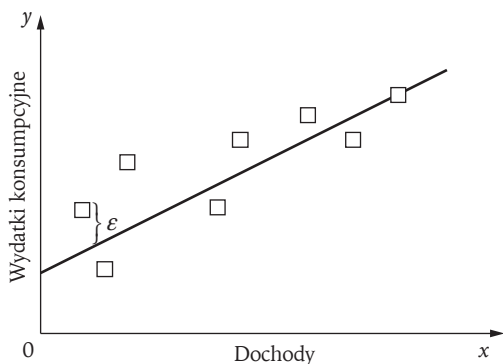


Możemy oczekiwać, że krańcową skłonność do konsumpcji nie we wszystkich rodzinach jest dokładnie taka sama. Wpływają na nią inne zmienne oprócz dochodu, takie jak wiek rodziny, liczba osób w rodzinie, miejsce zamieszkania, nawyki konsumpcyjne itd. Dlatego też modyfikujemy funkcję konsumpcji, dodając zaburzenie losowe, dzięki któremu funkcja deterministyczna konsumpcji staje się funkcją stochastyczną (losową). Taka postać funkcji jest modelem ekonometrycznym. Zapiszmy ją:

$$y = \beta_1 + \beta_2 x + \varepsilon. \quad (1.2)$$

Oznacza to że w rzeczywistości krańcową skłonność do konsumpcji różnych rodzin nie spełnia dokładnie relacji liniowej jak na rys. 1.1, a losowo się od niej odchyła na skutek działania efektu losowego ε . Możemy ją przedstawić za pomocą rys. 1.2.

Rysunek 1.2. Ekonometryczny model konsumpcji Keynesa



Dla estymacji modelu potrzebne są nam dane statystyczne, które umożliwią oszacowanie nieznanych parametrów β_1 oraz β_2 .

Przykład 1.1

Wykorzystamy dane o produkcie krajowym brutto (x) i spożyciu gospodarstw domowych, zarejestrowanym jako wydatki konsumpcyjne (y). Dane te są agregatami rocznymi wyrażonymi w mld złotych, liczonych w cenach stałych roku 2001, a więc w wielkościach realnych, a nie nominalnych. Zawiera je tabela 1.1.

Tabela 1.1. PKB i spożycie indywidualne w sektorze gospodarstw domowych w latach 1991–2004 w cenach roku 2001 (w mld zł)

Rok	PKB	SPO
1991	483,21	308,54
1992	495,68	315,51
1993	523,74	332,05
1994	545,56	346,27
1995	578,82	357,30
1996	613,63	388,36
1997	655,20	415,06
1998	686,89	435,09
1999	714,95	457,73
2000	743,53	470,21
2001	750,80	479,60
2002	761,18	490,50
2003	790,78	503,15
2004	832,84	523,17

Źródło: Roczniki Statystyczne GUS z kolejnych lat.

Stosując metodę najmniejszych kwadratów (którą omówimy w rozdz. 3), oszacowaliśmy funkcję regresji:

$$\hat{y}_i = -2,92 + 0,64x_i. \quad (1.3)$$

Daszek nad y_i oznacza, że jest to **wielkość oszacowana** dla i -tej obserwacji (zwana również **wielkością wyliczoną** lub **teoretyczną**), a nie rzeczywiście zaobserwowana wartość y_i .

Z równania (1.3) dowiadujemy się, że w latach 1990–2002, licząc w wielkościach realnych, średnio każda złotówka wzrostu PKB pociąga wzrost wydatków konsumpcyjnych gospodarstw domowych o 64 grosze, a więc $MPC = 0,64$. Stała równania wynosząca $-2,92$ nie ma sensownej interpretacji ekonomicznej. Oznaczałaby bowiem ona, ile wynosiłyby wydatki konsumpcyjne, gdyby PKB był równy zero. Sytuacja taka nigdy nie występuje w realnej gospodarce. Można przyjąć jako regułę, że z wyjątkiem szczególnych modeli

nie interpretujemy stałej równania regresji, a jedynie współczynniki kierunkowe, a więc wielkości przy zmiennych objaśniających.

Obliczenia zostały wykonane przy użyciu programu GRET. Wydruk komputerowy wygląda następująco:

Estymacja KMNK z wykorzystaniem 14 obserwacji 1991–2004

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p
const	-2,92007	6,79024	-0,4300	0,674788
PKB	0,63894	0,010212	62,5678	<0,00001*** ³

Zmienna zależna: SPO

Średnia arytmetyczna zmiennej zależnej = 415,897

Odchylenie standardowe zmiennej zależnej = 74,155

Suma kwadratów reszt = 218,461

Błąd standardowy reszt = 4,26674

Wsp. determinacji $R^2 = 0,996944$

Skorygowany $R^2 = 0,996689$

Stopnie swobody = 12

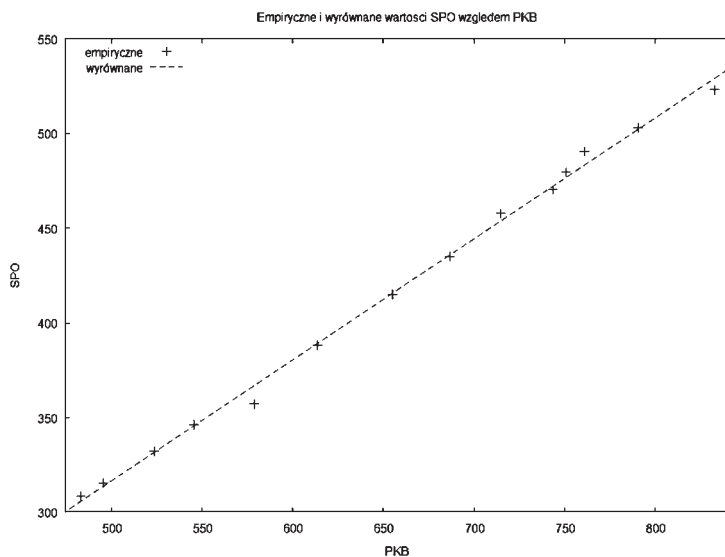
Statystyka testu Durбина-Watsona = 1,6828

Autokorelacja reszt rzędu pierwszego = 0,0697639

Logarytm wiarygodności = -39,098

Kryterium informacyjne Akaike'a (AIC) = 82,196

Kryterium bayesowskie Schwarza (BIC) = 83,4741



W podrozdziale 3.3 poznamy specjalną miarę dobroci dopasowania funkcji regresji do danych empirycznych, zwaną **współczynnikiem determinacji**, oraz na dalszych stronach zapoznamy się z pozostałymi statystykami występującymi na tym wydruku.

³ W całej książce jedną, dwoma lub trzema gwiazdkami oznaczono różne poziomy istotności (zob. podrozdział 4.4, str. 52).

Wyżej oszacowany model wymaga dalszej weryfikacji. Powinniśmy sprawdzić, czy krańcowa skłonność do konsumpcji oszacowana jako 0,64 jest statystycznie mniejsza od 1, co zakłada teoria Keynesa. Problem ten przedyskutujemy w rozdziale 4, po zapoznaniu się z koncepcją błędów standardowych oszacowanych współczynników regresji.

Model może być następnie wykorzystany dla celów predykcji (prognozowania). Przypuśćmy, że chcemy wyznaczyć oczekiwaną średnią wielkość wydatków konsumpcyjnych gospodarstw domowych w 2005 roku. Jeśli założymy (lub przyjmijemy za prognozę makroekonomiczną, że PKB wzrośnie w 2005 roku o 4,5% w stosunku do roku poprzedniego, to wielkość jego wyniesie $(1 + 0,045) \times 832,84 = 870,32$ mld zł. Stąd z równania (1.3) $\hat{y}_i = -2,92 + 0,64x_i$ po podstawieniu danych otrzymamy:

$$\hat{y}_{2005} = -2,92 + 0,64 \times 870,32 = 559,92.$$

Po zapoznaniu się z ekonometrycznymi problemami prognozowania (rozdz. 7) będziemy mogli wyznaczyć błąd dla tak wyliczonej prognozy.

Podsumowanie

1. Ekonometria jest nauką zajmującą się ilościowym opisem, na podstawie danych statystycznych, prawidłowości ekonomicznych, postulowanych przez teorię ekonomii lub sugerowanych przez hipotezy ekonomiczne.
2. Hipoteza naukowa to przypuszczenie wymagające sprawdzenia.
3. Jednorównaniowym modelem ekonometrycznym nazywamy równanie wyjaśniające związek między zjawiskiem objaśnianym i najważniejszymi zjawiskami objaśniającymi. W modelu występują: zmienna objaśniana, zmienne objaśniające, zaburzenie losowe, oddające sumaryczny wpływ drugorzędnych, *explicite* niewyróżnionych czynników.
4. Rodzaje danych statystycznych
 1. Dane szeregów czasowych. Są to najbardziej popularne zbiory danych, gdzie kolejne obserwacje rejestrują badane zjawisko ekonomiczne w następujących po sobie momentach lub przedziałach czasu.
 2. Dane przekrojowe powstają jako zbiory obserwacji wielu jednostek w tym samym czasie.
 3. Dane panelowe (połączone, longitudinalne) łączą cechy danych szeregów czasowych i danych przekrojowych.
5. Metodologia ekonometrii to sposób postępowania w trakcie budowy modelu ekonometrycznego.

Kroki:

 1. Ustalenie teorii ekonomicznej lub zbioru hipotez, które model ekonometryczny ma potwierdzić lub odrzucić.

2. Określenie postaci matematycznej modelu ekonometrycznego polega na specyfikacji funkcji matematycznej, wiążącej zmienną objaśnianą ze zmiennymi objaśniającymi.
3. Dołączenie zaburzenia losowego, dzięki czemu wyspecyfikowane równanie staje się modelem ekonometrycznym.
4. Ustalenie zbioru danych statystycznych, wykorzystywanych do oszacowania modelu.
5. Estymacja parametrów modelu dokonywana jest poprzez zastosowanie właściwych metod szacowania.
6. Diagnostyka pomaga ustalić, czy model nie zawiera istotnych wad, wymagających poprawek. Diagnostykę przeprowadza się za pomocą różnorodnych testów statystycznych.
7. Sprawdzenie poprawności opracowanego modelu dla teorii ekonomii, która była przyczyną rozpoczęcia prac nad tym modelem. Przetestowanie postawionych na wstępie hipotez z użyciem opracowanego modelu.

Oszacowany i zweryfikowany model może być wykorzystywany dla celów prognostycznych i symulacyjnych.

gdzie $\mathbf{x}'_i = [1 \ x_{2i} \dots x_{Ki}]$ jest wektorem wierszowym zmiennych objaśniających dla i -tej obserwacji.

Dla zapisu macierzowego przyjmujemy oznaczenia:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{21} & x_{31} & \cdots & x_{K1} \\ 1 & x_{22} & x_{32} & \cdots & x_{K2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n} & x_{3n} & \cdots & x_{Kn} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}; \quad (2.4)$$

powyższy układ równań możemy zapisać jako jedno równanie macierzowe:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \cdots & x_{K1} \\ 1 & x_{22} & x_{32} & \cdots & x_{K2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n} & x_{3n} & \cdots & x_{Kn} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.5)$$

lub krócej:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.6)$$

Często korzystamy z modelu regresji, w którym występuje tylko jedna zmienna objaśniająca. Model taki możemy zapisać:

$$y_i = \beta_1 + \beta_2 x_{2i} + \varepsilon_i \quad i = 1, 2, 3, \dots, n \quad (2.7)$$

i będziemy go nazywać **modelem regresji prostej** lub krócej – **regresją prostą** w odróżnieniu od modelu z wieloma zmiennymi objaśniającymi, który nazywamy **regresją wieloraką**.

2.2. Od populacji do próby i od próby do populacji

Rozumowanie, które prowadzi nas od populacji do próby przebiega następująco. Równanie (2.5) przedstawia związek między zmienną objaśnianą a zmiennymi objaśniającymi w populacji. To, co jest przedmiotem naszego zainteresowania, to oczekiwane zachowanie się zmiennej y , pod warunkiem że zmienne objaśniające przyjmą zaobserwowane wartości, opisane macierzą $\mathbf{X}$.

To, co wyżej powiedzieliśmy możemy zapisać formalnie:

$$E(\mathbf{y} | \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \quad (2.8)$$

i co czytamy: warunkowa wartość oczekiwana zmiennej objaśnianej y przy danej macierzy obserwacji na zmiennych objaśniających $\mathbf{X}$ jest równa iloczynowi $\mathbf{X}\boldsymbol{\beta}$.

Dla i -tej obserwacji mamy podobnie:

$$E(y_i | \mathbf{x}'_i) = \beta_1 + \beta_2 x_{2i} + \dots + \beta_K x_{Ki} \quad i = 1, 2, \dots, n. \quad (2.9)$$