

Modele równań strukturalnych w metodologii badań szarej strefy

Katarzyna Brzozowska-Rup, Mariola Chrzanowska

Konferencja naukowa MET2019
Metodologia Badań Statystycznych

Warszawa, 4 lipca 2019 r.



Urząd Statystyczny
w Kielcach

Problem badawczy

**SZARA
GOSPODARKA**

określa grupę aktywności gospodarczych, które są legalne (zarejestrowane) w danym kraju ale prowadzone są bez pełnej wiedzy państwa, co implikuje brak możliwości dokonania bezpośredniego pomiaru wielkości analizowanego zjawiska.

Problem badawczy

Gdyby wszystkie obecnie rejestrowane ale ukrywane działania oraz nierejestrowane działania były obserwowalne, powiększyłyby oficjalnie obliczany produkt narodowy brutto.

Problem badawczy

Badane zjawisko

- nie jest bezpośrednio obserwowalne - zmienna objaśniana jest zmienną ukrytą (latentną),
- jest na tyle złożone, że odtworzenie i weryfikacja jego rozmiaru i struktury wymaga ogromnego nakładu pracy i finansów,
- pojawia się problem jednoznacznej identyfikacji przyczyny, skutku i związku przyczynowego.

Wzrost szarej strefy



Zmniejszenie dochodów państwa



Wzrost stawek podatkowych dla firm i osób działających legalnie



Wzrost szarej strefy

Cel

- Konstrukcja teoretycznego modelu umożliwiającego szacowanie rozmiaru szarej strefy - metoda pośrednia (wskaźnikowa).
- Poszukiwanie modelu, który daje możliwość testowania hipotez badawczych o dużej złożoności relacji między zmiennymi, umożliwia uwzględnienie zarówno zmiennych obserwowanych jak i nie dających się zmierzyć bezpośrednio.
- (?) Próba identyfikacji czynników behawioralnych w kontekście złożonych współzależności zachodzących w gospodarce.

Zmienne latentne

Pojęcie struktury ukrytej („głębokiej”), będące podstawą wyjaśniania obserwowalnych systemów i struktur społecznych znane jest za sprawą przedstawicieli antropologii strukturalnej (Claude Lévi-Strauss).

Zmienne latentne

Ze względu na typ relacji pomiędzy zmiennymi obserwowalnymi i ukrytymi wyróżnia się modele:

- **refleksyjne** – zmienna ukryta rzeczywiście istnieje, lecz nie można jej bezpośrednio obserwować. Przyjmuje się, że zmienne obserwowalne mierzą (z pewnym błędem) wartości zmiennej ukrytej – są jej odbiciem (refleksem).
- **formacyjne** – przyjmuje się, że zmienna ukryta jest jedynie wynikiem oddziaływania pewnej liczby zmiennych, nie musi faktycznie istnieć.

Idea

„Tam, gdzie klasyczna ekonomia staje się bezradna, pomocna okazuje się metodologia badań psychologicznych. Psychologia jest nauką, która pomaga wyjaśnić oraz zrozumieć podłoża i motywy ludzkich zachowań. Zachowań, których katalizatorem są zawsze ludzkie emocje i wartości etyczne. Zachowań, których nie da się analizować bez zrozumienia czynników na nie wpływających i je kształtujących” (Zygan, 2013).

Idea

- Teoria procesu substancywnego jest bardzo ogólną i wysoce abstrakcyjną reprezentacją praw rządzących rozważanym procesem generującym dane DGP (*data generating process*).
- Celem modelu substancywnego jest wyjaśnienie procesu DGP leżącego u podstaw obserwowanego zbioru danych w postaci modelu strukturalnego.

Idea

Konstrukcja modelu (podejście kompleksowe), w którym zmiana oszacowanej wielkości szarej strefy jest spójna ze zmianami wskaźników makroekonomicznych,

takich jak:

- dynamika PKB
- stopa bezrobocia,
- obrót pieniądza w gospodarce (metoda Gutmanna (1977), Feige (2008)
 - ✓ proporcja pieniądza gotówkowego,
 - ✓ popyt (/podaż) na pieniądz (gotówkę)/ zasób pieniądza w gospodarce,
 - ✓ relacja gotówki i depozytów.

Idea

- udział płac w dochodzie narodowym,
- relacja zasobów gotówki w obiegu i miary podaży pieniądza,
- średnia, ważona stopa opodatkowania (obciążenia podatkowe).

Oczekuje się, że wzrost powyższych wartości przekłada się na większy popyt na gotówkę (większość transakcji w szarej strefie odbywa się gotówką). Wzrost stopy oprocentowania depozytów przekłada się na większy popyt na gotówkę.

Idea

W przyszłości uzupełnione o wymiar:

- **mikroekonomiczny oraz**
- **moralny.**

(?) poziom gospodarstw domowych:

- **audyt podatkowy,**
- **analiza rozbieżności pomiędzy dochodami i wydatkami (w szczególności na żywność i dobra luksusowe)– wymaga badania ankietowego.**

Modele wielorównaniowe

Przyjmujemy następujące oznaczenia

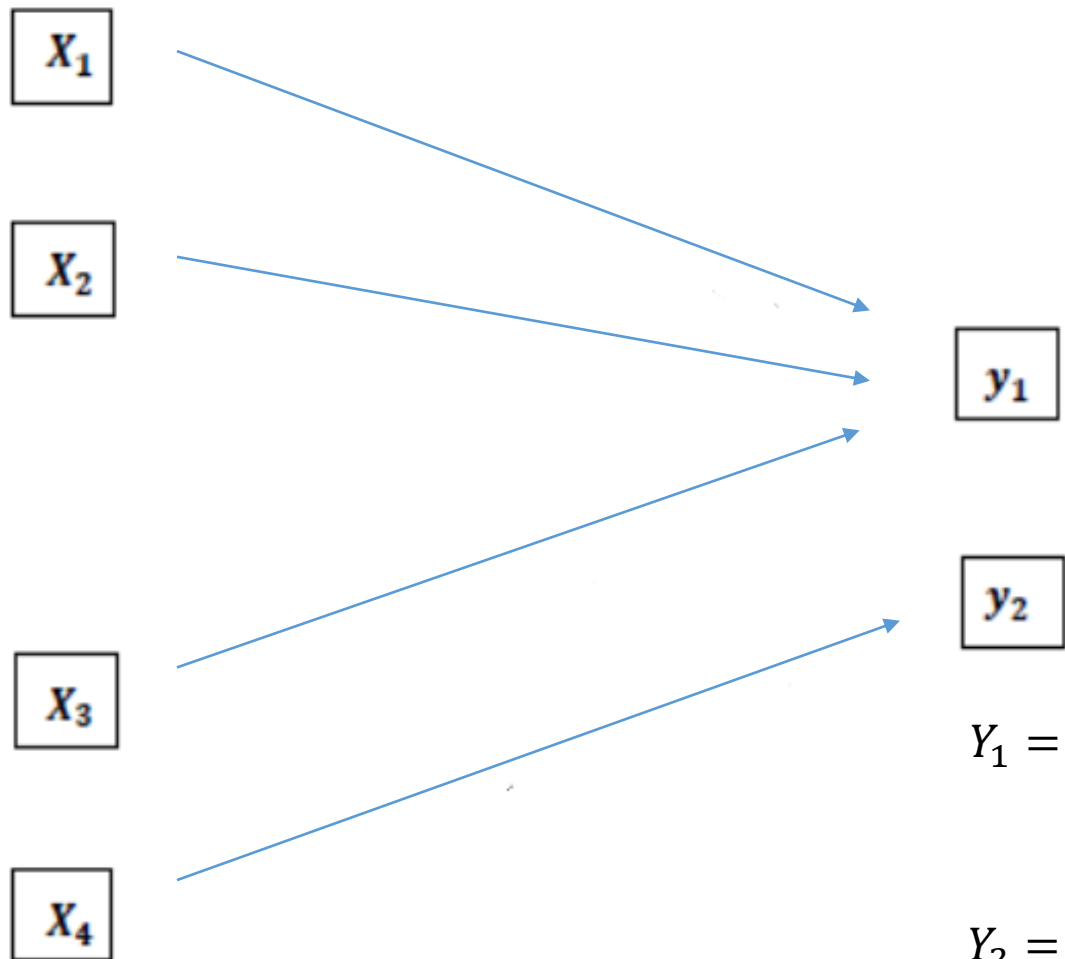
$Y_1, Y_2, \dots, Y_m$ zmienne endogeniczne bez opóźnień czasowych,
nazywane również zmiennymi łącznie
współzależnymi,

$X_1, X_2, \dots, X_k$ zmienne endogeniczne z opóźnieniami czasowymi
oraz zmienne egzogeniczne (bez i z opóźnieniami),
tzw. zmienne z góry ustalone.

Propozycja

Podjęte badania można sprowadzić do zweryfikowania modelu pewnego procesu substantywnego, w którym mnogość elementów połączonych wzajemnymi zależnościami, zostaje zredukowana poprzez ich charakterystykę za pomocą metacechy.

Model wielorównaniowy

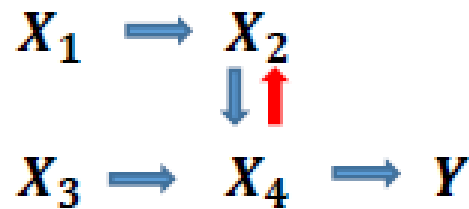


$$Y_1 = \beta_{12}Y_2 + \sum_{j=1}^4 \gamma_{1j}X_j + \varepsilon_1$$

$$Y_2 = \beta_{21}Y_1 + \sum_{j=1}^k \gamma_{2j}X_j + \varepsilon_2$$

Modele regresji wielokrotnej

$$y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$$



Zauważmy:

- ✓ nie mamy prostego podziału na zmienne zależne i niezależne: zmienna X_2 i X_4 są zarówno zmiennymi objaśnianymi, jak i objaśniającymi,
- ✓ w opisie relacji między wyróżnionymi zmiennymi może pojawić się konieczność uwzględnienia wzajemnego oddziaływania między zmiennymi X_2 i X_4 .

Klasyczna regresja nie daje możliwości analizowania tego typu związków między zmiennymi

MODELE RÓWNAŃ STRUKTURALNYCH, SEM

Zaliczane do grupy najważniejszych modeli statystycznych, (Cliff, 1983, Konarski, 2014)

- ✓ są zasadniczo podobne do modeli regresji wielorakiej,
- ✓ są bardziej elastyczne w przypadku opisu interakcji między zmiennymi, zwłaszcza jakościowymi, dla których nie są spełnione klasyczne założenia regresji,
- ✓ mogą być stosowane w opisie zależności nieliniowych, skorelowanych zmiennych niezależnych oraz zmiennych ukrytych.

MODELE RÓWNAŃ STRUKTURALNYCH

tworzą uogólniony model liniowy i zawierają poza regresją wieloraką także następujące procedury:

- **model wewnętrzny** przedstawiający **analizę ścieżkową**,
- **model zewnętrzny** reprezentujący **konfirmacyjną analizę czynnikową**.

MODELE SEM

- Umożliwiają konstrukcję modeli, w których dopuszcza się zarówno pośrednie jak i bezpośrednie powiązania strukturalne między zmiennymi obserwowalnymi i nieobserwowanymi.

Modelowanie równań strukturalnych SEM

Model ścieżkowy

sieć połączeń, związków przyczynowo-skutkowych między zmiennymi, również latentnymi, opisanymi przez zmienne mierzalne.

Zwykle wykorzystywane przez badania społeczne (psychologia, socjologia, marketing itp.) w celu zbadania tego rodzaju zmiennych nieobserwowalnych (na przykład: postawy, osobowość, wiara, satysfakcja itp.).

Modelowanie równań strukturalnych SEM

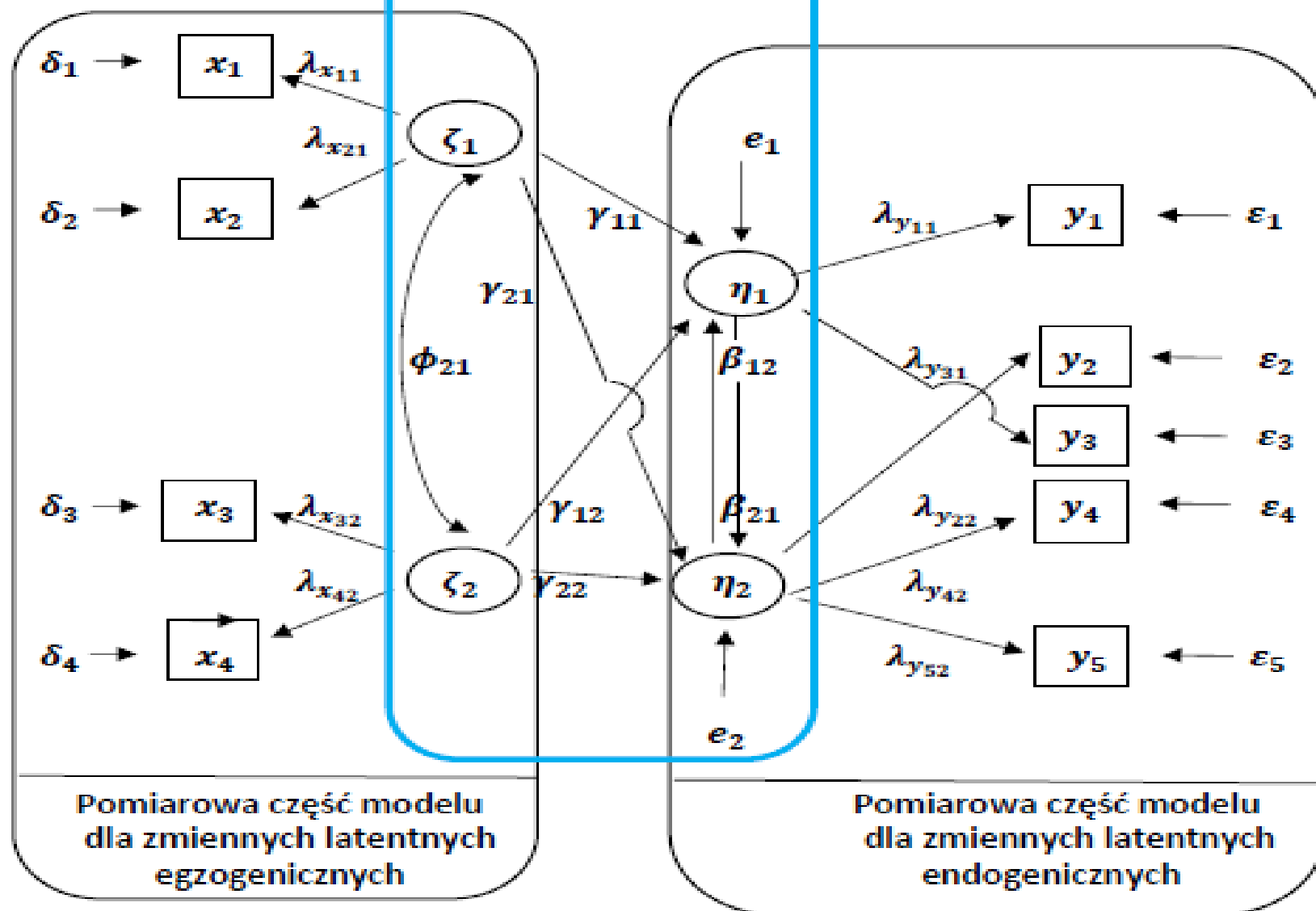
- **Do najważniejszych pozycji literaturowych w metodyce budowy SEM należą następujące prace:**

Bollena (1989), Kaplana (2000).

- **Wśród pozycji krajowych warto wyróżnić:**

Gatnar (2003), Brzeziński (2004), Osińska (2008, i inni 2011), Konarski (2014) – prace z zakresu psychologii, ekonomii i finansów behawioralnych.

Model SEM



Oznaczenia

$\eta_{m \times 1}$ – wektor endogenicznych zmiennych ukrytych,

$\zeta_{k \times 1}$ – wektor egzogenicznych zmiennych ukrytych,

$B_{m \times m}$ – macierz współczynników regresji przy zmiennych endogenicznych,

$\Gamma_{m \times k}$ – macierz współczynników przy zmiennych egzogenicznych,

$e_{m \times 1}$ – wektor składników losowych,

$y_{p \times 1}$ – wektor obserwowalnych zmiennych endogenicznych,

$x_{p \times 1}$ – wektor obserwowalnych zmiennych egzogenicznych,

Λ_x, Λ_y – macierze ładunków czynnikowych,

$\delta_{q \times 1}, \varepsilon_{p \times 1}$ – wektory błędów pomiaru.

Model SEM

Model wewnętrzny

$$\eta_1 = \beta_{12}\eta_2 + \gamma_{11}\zeta_1 + \gamma_{12}\zeta_2 + e_1$$

$$\eta_2 = \beta_{21}\eta_1 + \gamma_{21}\zeta_1 + \gamma_{22}\zeta_2 + e_2$$

Model zewnętrzny (model pomiaru)

Model pomiarowy ukrytej zmiennej egzogenicznej

$$x_1 = \lambda_{x_{11}}\zeta_1 + \delta_1, \quad x_2 = \lambda_{x_{21}}\zeta_1 + \delta_2$$

$$x_3 = \lambda_{x_{32}}\zeta_2 + \delta_3, \quad x_4 = \lambda_{x_{42}}\zeta_2 + \delta_4$$

Model pomiarowy ukrytej zmiennej endogenicznej

$$y_1 = \lambda_{y_{11}}\eta_1 + \varepsilon_1, \quad y_2 = \lambda_{y_{22}}\eta_2 + \varepsilon_2$$

$$y_3 = \lambda_{y_{31}}\eta_1 + \varepsilon_3, \quad y_4 = \lambda_{y_{42}}\eta_2 + \varepsilon_4$$

$$y_5 = \lambda_{y_{52}}\eta_2 + \varepsilon_5$$

Ogólny model SEM

Składa się z układu równań strukturalnych między konstruktami teoretycznymi (zmiennymi latentnymi) η i ζ

$$\eta = B\eta + \Gamma\zeta + e$$

oraz układu równań pomiarowych określających relację między zmiennymi ukrytymi a ich obserwowanymi wskaźnikami:

- równanie ukrytej zmiennej egzogenicznej,

$$x = \Lambda_x \zeta + \delta$$

- model pomiarowy ukrytej zmiennej endogenicznej

$$y = \Lambda_y \eta + \varepsilon$$

Uwaga

Wprowadzone do modelu SEM równanie zmiennych obserwowalnych nie służy do opisu zależności pomiędzy wyróżnionymi, mierzonymi podczas badania zmiennymi lecz do wydobywania z tych wielkości istotnych, w kontekście badania, informacji oraz eliminacja zakłóceń.

Intuicyjnie takie podejście wynika z traktowania zmiennej obserwowalnej jako odbicia (refleksu) zmiennej ukrytej:

$$y_o = y_p + \textit{błąd pomiaru}$$

gdzie

y_o - wartość obserwowana zmiennej Y ,

y_p - wartość prawdziwa zmiennej Y .

Standardowe założenia modelu SEM

- $E[e] = 0,$
- $cov[\zeta, e^T] = 0,$
- $cov[\zeta, \delta^T] = 0,$
- $\det(I - B) \neq 0,$
- $cov[\eta, \varepsilon^T] = 0,$
- $cov[\varepsilon_i, \varepsilon_{j \neq i}] = 0,$
- $cov[\delta_i, \delta_{j \neq i}] = 0,$
- $cov[\zeta_i, \zeta_{j \neq i}] = \varphi_{ji}.$

Modelowanie szarej strefy za pomocą metod pośrednich

– badania literaturowe (model MIMIC)

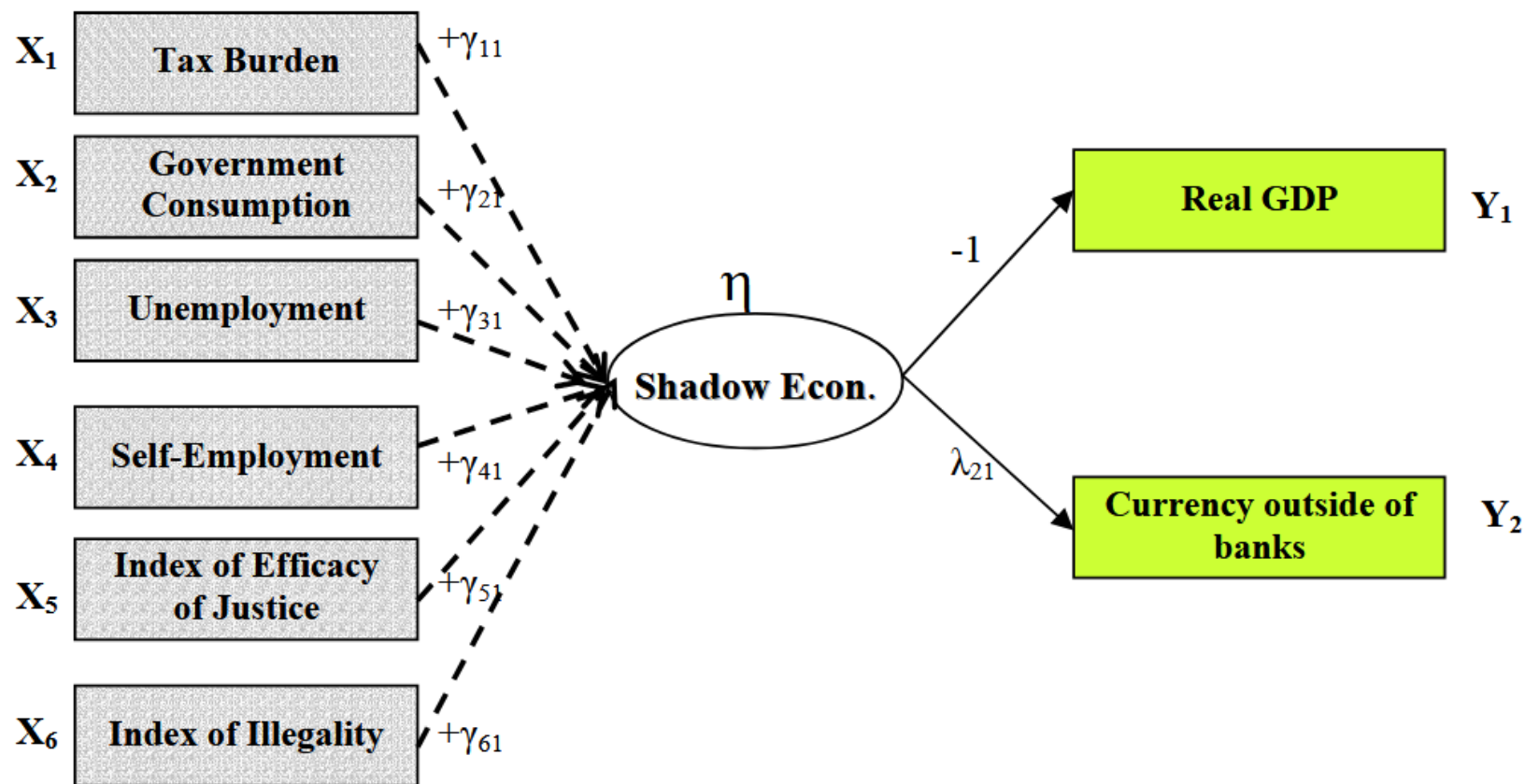
- w 1984 r. Frey i Weck-Hannemann zastosowali, po raz pierwszy model MIMIC do szacowania szarej strefy. Zaproponowali model dla 17 krajów OECD.
- w 1988 r. ich idea została rozszerzona przez Ainera, Scheidera i Ghosh'aw, którzy uwzględnili w modelu czas - dynamiczny model MIMIC tzw. DYMIMIC. Zaproponowany wykorzystali do oszacowania rozmiaru szarej strefy w USA.
- w 1999r. Giles wprowadził dalszą modyfikację, polegającą na uwzględnieniu możliwości wystąpienia kointegracji badanych szeregów czasowych. Rozważania teoretyczne wykorzystał do oszacowania szarej strefy dla Nowej Zelandii.
- w 2002 roku Gilles i Tedds oszacowali metodą MIMIC rozmiar szarej strefy w Kanadzie;
- rok później Dell Anno i Schneider przeprowadzili badania dla Włoch w zestawieniu z innymi krajami OECD.

Modelowanie szarej strefy za pomocą metod pośrednich

– badania literaturowe (model MIMIC)

- 2008 r. Buehn i Schneider wykorzystali model MIMIC do oszacowania szarej strefy we Francji. Przeprowadzili krótko- i długoterminową analizę, w podsumowaniu, której zasugerowali, że uzyskane wyniki mogłyby być wykorzystywane w procesie opracowywania strategii gospodarczej kraju.
- 2009 r. Zagoršek, Jaklič i Hribernik przedstawili analizę działań nieformalnych w Słowenii. Wskazali instytucje, które ich zdaniem są ściśle związane z rozwojem i funkcjonowaniem szarej strefy.
- 2018 r. Pasovic i Efendic opublikowali wyniki badania dynamiki rozmiaru gospodarki nieformalnej w Bośni i Hercegowinie w okresie 1998-2016.

Model MIMIC dla szarej strefy we Włoszech



Źródło: F. Schneider, R. Dell'Anno, *The Shadow Economy of Italy and other OECD Countries: What do we know?*, Journal of Public Finance and Public Choice, 2005.

Modelowanie szarej strefy za pomocą metod pośrednich

– badania literaturowe

Modele MIMIC (*Multiple Indicators and Multiple Causes*) są specjalnym przypadkiem ogólnego modelu SEM, zawierającym jedynie endogeniczne zmienne latentne η . O zmiennych egzogenicznych zakłada się, że są mierzone (dokładnie/ bez błędu pomiarowego).

W konsekwencji model MIMIC z jedną latentną zmienną endogeniczną:

$$y = \Lambda_y \eta + \varepsilon$$

$$\eta = \Gamma x + \xi$$

Modelowanie szarej strefy za pomocą metod pośrednich – badania literaturowe

Zauważmy, że w modelu zmienna latentna jest mierzona przez p endogenicznych zmiennych obserwowalnych $y^T = (y_1, y_2, \dots, y_p)$, i znajduje się pod wpływem q $x^T = (x_1, x_2, \dots, x_q)$.

Reguła MIMIC określa, że jeżeli $p \geq 2$ i $q \geq 1$, to model z jedną zmienną latentną jest zidentyfikowany (Bollen, 1989).

Propozycja

- Zastosować model SEM do zweryfikowania hipotez dotyczących identyfikacji zależności między wybranymi zmiennymi (tj. przyczynami i skutkami).
- Dokonać zbadania kierunku i siły ich wpływu na podjęcie decyzji o działalności mieszczącej się w kategorii szarej strefy.

SEM

- Jest elastycznym „narzędziem” modelowania relacji między zmiennymi, dzięki czemu stwarza możliwość uzyskania dowolnego modelu relacji przyczyn i skutków, dając wgląd w badane zjawisko w jego bardzo zaawansowanym wydaniu.
- Stosując modelowanie SEM warto rozważać różnorodne (koniecznie zgodne z teorią) struktury przewidywanych korelacji, alternatywne scenariusze relacji między zmiennymi, by móc wybrać model dopasowany lub porównywać modele alternatywne.

Metody estymacji modeli SEM

- Metoda największej wiarygodności (przy założeniu, że rozkład jest wielowymiarowym rozkładem normalnym),
- Uogólniona metoda najmniejszych kwadratów (wymaga dużej próby , tj. powyżej 2500 obserwacji),
- Metody asymptotyczne ADF (niewrażliwe na rozkład)

(Konarski 2014)

wzrost ilości pieniądza
gotówkowego w obiegu

obciążenie podatkowe

depozyty

konsumcja

wydatki

stopa bezrobocia

skuteczność systemu prawnego

społeczeństwo kolektywne

społeczeństwo kolektywne

liczba przestępstw karno-skarbowych

Ukrywanie wynagrodzeń

Zaniżanie dochodów właściciela

Zaniżanie kosztów ponoszonych
przez właściciela

samozatrudnienie

Moralność obywatelska

Zaufanie do państwa

Szara strefa

Tempo wzrostu PKB

Zużycie energii

rozwój wskaźników
pieniężnych

miernik swobody
działalności gospodarczej

Wzrost produkcji globalnej



„Czy jesteśmy w stanie podać „recepturę” pozwalającą przewidywać zachowania człowieka?

Wydaje się to mocno wątpliwe...

Jednak pomimo trudności ze wszech miar warto wysiłki takie podejmować.

Ludzie to nie tylko homo oeconomicus, to także, a może przede wszystkim istoty czujące, popychane do działania przez system wierzeń i wartości. A wartości to cele, ku którym dążą ludzie, to reguły wyznaczające sposób właściwego postępowania” (Zygan, 2013).

Literatura

Z. A. Błasiak, (2018), Przydatność metod ekonometrycznych w badaniach nad szarą strefą.

K. A. Bollen, (1989), Structural Equations with Latent Variables, Wiley.

Brzeziński J. (red.), Wielozmiennowe modele statystyczne w badaniach psychologicznych, PWN, Warszawa 1987

Frey S.B., Hanneman-Weck H., (1984), The Hidden economy As An 'Unobserved' Variable, European Economic Review, Vol. 26, No. 1-2.

E. Gatnar, (2003), Statystyczne modele struktury przyczynowej zjawisk ekonomicznych, Wydawnictwo Akademii Ekonomicznej, Im. Karola Adamieckiego w Katowicach.

K. Dymarski, (2016), *Estymacja rozmiaru szarej strefy w oparciu o dane ankietowe gospodarstw domowych*, Praca doktorska.

A. Domurat, (2009), Identyfikacja wartości osobistych w badaniach psychologicznych. Wartości jako cele działań i wyborów, Wyd. UW, Warszawa.

R. Konarski (2014), Modele równań strukturalnych. Teoria i praktyka, Wydawnictwo PWN SA, Warszawa.

Literatura

D. Kaplan, (2000), Structural Equation Modeling: Foundations and Extensions, Sage Publications.

J. Kotlińska, (2018), Szara strefa a dochody publiczne.

M. Leszczyński, (2018), Wzrost i rozwój gospodarczy a problem zatrudnienia w gospodarce czarnorynkowej.

P. P. Malaczewska, (2016), Determinanty i mechanizmy kształtowania się szarej strefy gospodarki.

M. Osińska, (2008), Ekonometryczna analiza zależności przyczynowych, Uniwersytet Mikołaja Kopernika, Toruń.

M. Osińska, M. B. Pietrzak, M. Żurek, (2011), Wykorzystanie modeli równań strukturalnych do opisu psychologicznego mechanizmów podejmowania decyzji na rynku kapitałowym, ACTA UNIVERSITATIS NICOLAI COPERNICI, Ekonomia XLII-Nauki Humanistyczno-Społeczne, Zeszyt 402, Toruń .

.

Literatura

T. Piecuch, E. Szczygieł, (2018), *Przedsiębiorczość szarej strefy* .

Ż. Popławska, V. Humenyuk, L. Seniv, (2018), *Wpływ państwowych regulacji biznesu na wielkość szarej strefy*

T. Tyszka, (2010), *Decyzje. Perspektywa psychologiczna i ekonomiczna*, Scholar, Warszawa 2010,

Szymańska A., (2017), *Problematyka hierarchiczności – wprowadzenie metacech w modelach SEM*, *Studia Psychologica UKSW* 17(1) 2017 s. 65–84

M. Zygan, (2013) *Ekonomia behawioralna – wprowadzenie do problematyki*, *Studia i prace Wydziału Nauk Ekonomicznych i Zarządzania* Nr 32, Uniwersytet Szczeciński

H. Żukowska, (2018), *Szara strefa w gospodarce a sektor finansów publicznych i system ubezpieczeń społecznych – sprzężenie zwrotne.*

Dziękujemy za uwagę