

Ilona Jankowska

Uniwersytet Łódzki

ANALIZA DOKŁADNOŚCI WYNIKÓW SYMULACJI STOCHASTYCZNEJ NA PRZYKŁADZIE RÓWNAŃ REGRESJI

1. Wstęp

Rozwiązanie skomplikowanych problemów w dziedzinie ekonomii nie zawsze można uzyskać przez zastosowanie metod analitycznych. Często jedynym sposobem otrzymania wyników jest technika numeryczna.

Artykuł zawiera analizę dokładności wyników symulacji stochastycznej na przykładzie rozkładów estymatorów parametrów równania regresji. Podczas różnych eksperymentów, związanych nie tylko z ekonomią, korzysta się z symulacji. Współcześnie wykonuje się je z pomocą komputerów (z wykorzystaniem dużych mocy obliczeniowych), co znacznie zwiększa zakres ich zastosowań. Są one niezmiernie przydatne w otrzymywaniu numerycznych (liczbowych) rozwiązań problemów. Referat poświęcono temu tematowi, gdyż warto wiedzieć, czy wyniki, które uzyskuje się w wyniku przeprowadzanej symulacji, są zbliżone do rzeczywistych rozwiązań.

Chcemy określić, kiedy symulacja stochastyczna daje wiarygodne wyniki, na których można się oprzeć i których nie podaje się w wątpliwość. Choć jest to metoda czasami niezbędna i dająca ogromne możliwości prowadzenia badań, posiada również swoje wady [Hendry 1984, s. 943-945]. Główną z nich jest kosztowność eksperymentów, związana ze znaczną liczbą replikacji wykonywanych podczas próbkowania modelu. Im więcej replikacji, tym badanie staje się bardziej czasochłonne. Oczekiwania decydenta co do relacji kosztów obliczeń do dokładności otrzymanych rezultatów symulacji są ważne i wpływają na dobór liczby powtórzeń podczas przeprowadzania eksperymentu.

Interesujące jest, jak liczba replikacji wpływa na rozbieżności pomiędzy rozkładami populacji generalnej a rozkładami empirycznymi otrzymanymi w wyniku stosowania symulacji stochastycznej. Czy otrzymywane histogramy się zmieniają? Czy nieregularności są bardziej wygładzone? Czy efekty pracy są lepsze? Czy warto w takim razie, kosztem nakładu pracy i czasu, wykonywać większą liczbę replikacji? Określenie stopnia dokładności wyników na przykładzie rozkładów esty-

matorów parametrów zależy od liczby replikacji wykonywanych podczas eksperymentu symulacyjnego. O coraz większej precyzji wyników powinny świadczyć bardziej wygładzone histogramy prezentujące estymatory współczynników równania. Takie rozważania powinny pomóc w ocenie przydatności wykorzystywania symulacji stochastycznej również w innych typach badań.

2. Opis eksperymentu

Dokładność wyników, opisywana jako dokładność rozkładu aproksymowanego rozkładem uzyskanym w następstwie zastosowania symulacji stochastycznej, jest wprost proporcjonalna do pierwiastka z liczby replikacji [Gajda 2001, s. 109]. Wynika z tego, że dokładność rośnie wolniej niż liczba uzyskanych rozwiązań symulacji. Chcąc zatem osiągnąć wysoką dokładność rezultatów, narażamy się na znaczny wzrost kosztu eksperymentu.

Ważną zaletą symulacji stochastycznej jest to, że szybko daje ona przybliżone wyniki. Należy zawsze rozważyć, czy warto ponosić tak wysokie koszty. W wielu sytuacjach może się okazać, że przeprowadzenie większej liczby replikacji jest bezcelowe. Równocześnie wielość nieskomplikowanych obliczeń może w konsekwencji okazać się bardzo kosztowna, głównie ze względu na jej czasochłonność. Ponadto znaczna liczba wyników wymaga również ich statystycznego opracowania.

Badanie przeprowadzono według modelu jednorównaniowego:

$$y_t = a_1 + a_2 y_{t-1} + a_3 x_{1t} + a_4 x_{2t} + u_t, \quad \text{gdzie } t = 1, \dots, T,$$

którego zakłócenie podlega schematowi autoregresyjnemu pierwszego rzędu:

$$u_t = \rho u_{t-1} + \xi_t,$$

gdzie: a_2 – współczynnik autoregresji równania,

ρ – współczynnik autokorelacji reszt.

Na potrzeby numerycznej implementacji doświadczenia stworzono program oparty na pakiecie ekonometrycznym EViews 3.1. Z tego względu konieczna była transformacja wyjściowej postaci modelu do równoważnej formy [Malinvaud 1968, s. 467-472]. Przekształcona postać równania jest wykorzystana w programie do estymacji parametrów modelu:

$$y_t = a_1(1 - \rho) + (a_2 + \rho)y_{t-1} - a_2\rho y_{t-2} + a_3(x_{1t} - \rho x_{1,t-1}) + a_4(x_{2t} - \rho x_{2,t-1}) + \xi_t.$$

Wartości parametrów a_1 , a_3 i a_4 są odpowiednio równe 10, -0,5 i 0,5.

Podczas uruchamiania symulacji każdorazowo było konieczne ustalanie parametrów eksperymentu. Poza współczynnikami autoregresji równania i autokorelacji reszt należą do nich również:

T – wielkość próby w eksperymencie,

r – liczba replikacji wykonywanej symulacji.

W doświadczeniu przyjęto arbitralnie zbiory wartości dla każdego z czterech parametrów. Współczynnik autoregresji równania a_2 mógł być równy jednej z dwóch wartości $\{-0,5; 0,5\}$. Zdecydowano natomiast o większej różnorodności wartości współczynnika autokorelacji zakłóceń $\rho \in \{-0,9; -0,5; 0; 0,5; 0,9\}$, zatem w doświadczeniu jest możliwych dziesięć kombinacji parametrów a_2 i ρ , dla których zostały wykonane symulacje.

Na potrzeby eksperymentu dane do próby zostały wygenerowane losowo tak, aby nie charakteryzowały się trendem ani wahaniami cyklicznymi. Liczba obserwacji w próbie w każdej symulacji jest taka sama – równa 52¹. Taka liczba obserwacji gwarantuje, że nie jest ona za mała (np. 30) ani za duża (100 obserwacji). Przyjęta liczebność próby jest najbardziej prawdopodobna do uzyskania w przypadku modeli ekonometrycznych.

Parametrem najbardziej interesującym jest liczba wykonywanych replikacji. Symulacje wykonano dla czterech różnych replikacji, więc $r \in \{5\,000; 10\,000; 20\,000; 50\,000\}$.

W przeprowadzanym doświadczeniu wykorzystano symulację z reestymacją [Hall, Henry 1988, s. 232]. Zaburzano model wzorcowy (model z ustalonymi arbitralnie parametrami), a następnie szacowano parametry a_2 i a_3 zaburzonego modelu. Wykorzystano cztery metody estymacji:

- metodę najmniejszych kwadratów (LS)²,
- metodę najmniejszych kwadratów z autokorelacją reszt (LA),
- podwójną metodę najmniejszych kwadratów (TL),
- zmodyfikowaną podwójną metodę najmniejszych kwadratów (TY).

Wartości estymatorów otrzymanych zmodyfikowaną podwójną metodą najmniejszych kwadratów pokryły się z wartościami estymatorów metody najmniejszych kwadratów. W związku z tym przy omawianiu uzyskanych rezultatów symulacji stochastycznej zmodyfikowana podwójna metoda najmniejszych kwadratów zostanie pominięta, gdyż nie okazała się trafnym sposobem estymacji współczynników równania.

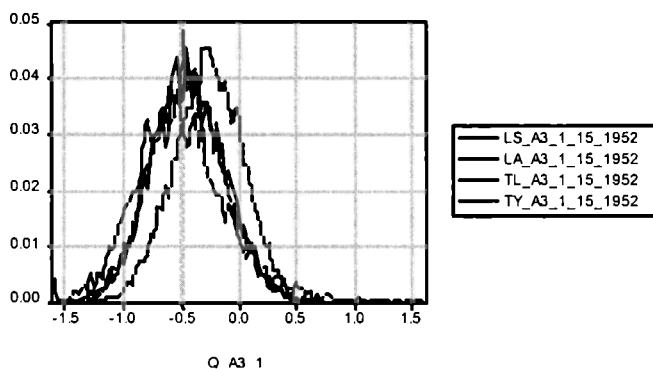
3. Prezentacja wyników i ich analiza

Na podstawie wstępnej analizy wyników eksperymentu zdecydowano się skupić się na otrzymanych rezultatach dla liczby replikacji $r \in \{5\,000; 10\,000; 50\,000\}$. Wykresy uzyskane dla 20 000 replikacji nie prezentowały tak wyraźnych zmian jak histogramy dla 50 000 replikacji.

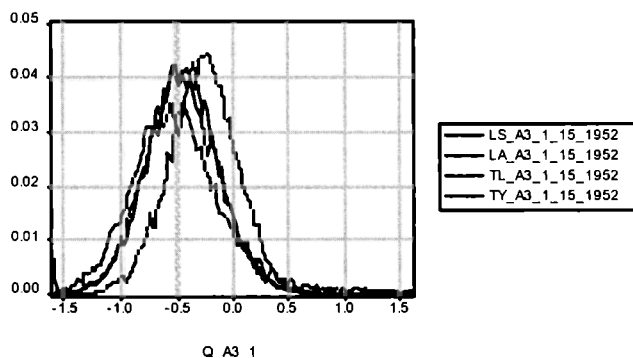
Dla każdego wariantu parametrów a_1 i ρ , oraz $r \in \{5\,000; 10\,000; 50\,000\}$, zaprezentowano kilka charakterystyk rozkładów estymatorów. Z pomocą arkusza kalkulacyjnego Microsoft Excel wyliczono miary związane z czterema pierwszymi

¹ Przy zastosowanych metodach estymacji w badaniu z próby 52 obserwacji otrzymujemy 50 wyników. Jest to skutek wykorzystanych technik obliczeniowych.

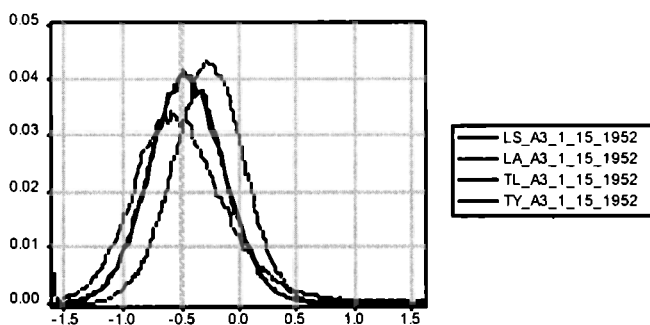
² Oznaczenie metod zaczerpnięte z kodu źródłowego programu.



Rys. 1. Rozkłady estymatorów parametru a_3 dla 5 000 replikacji; wariant I ($\rho = -0,9$; $a_2 = 0,5$)
 Źródło: opracowanie własne.



Rys. 2. Rozkłady estymatorów parametru a_3 dla 10 000 replikacji; wariant I ($\rho = -0,9$; $a_2 = 0,5$)
 Źródło: opracowanie własne.



Rys. 3. Rozkłady estymatorów parametru a_3 dla 50 000 replikacji; wariant I ($\rho = -0,9$; $a_2 = 0,5$)
 Źródło: opracowanie własne.

momentami rozkładu, które umożliwiają bardziej wnikliwą analizę rezultatów wykonanego doświadczenia. Są to: średnia, odchylenie standardowe, współczynnik asymetrii i miara koncentracji rozkładu (kurtoza).

Założeniem prezentacji tych czterech miar jest określenie zbieżności rozkładów empirycznych. Wyniki obliczeń mają pomóc w analizowaniu precyzyjności uzyskanych rezultatów eksperymentu symulacyjnego.

Aby móc ocenić stopień wygładzania nieregularności na wykresach częstości estymatorów współczynników, konieczne było wprowadzenie wzorcowego histogramu. Dla każdej z dziesięciu kombinacji parametrów a_2 i ρ wygenerowano rozkłady normalne restandaryzowane o takich samych parametrach jak rozkłady empiryczne wyników (odpowiednio dla parametrów a_2 i a_3) dla 50 000 replikacji. W ten sposób dla konkretnego wariantu parametrów symulacja jest wykonana dla trzech różnych replikacji 5000, 10 000 i 50 000. Natomiast parametry rozkładu normalnego (teoretycznego) uzyskano na podstawie danych otrzymanych w wyniku eksperymentu dla liczby 50 000 replikacji.

W celu porównania empirycznych wykresów częstości i wygenerowanych rozkładów normalnych wykorzystano miarę sumy kwadratów przyrostów wartości w kolejnych słupkach histogramów (tab. 2, 4). Przyrosty liczone są pomiędzy sąsiadującymi przedziałami szeregów rozdzielczych. Zatem suma kwadratów przyrostów opisywana jako:

$$\sum_i (\delta_i)^2,$$

pozwała na porównanie rozkładu wzorcowego (normalnego restandaryzowanego) z otrzymanymi rozkładami empirycznymi oszacowań współczynników równania.

Zastosowanie porównania rozkładu empirycznego, będącego produktem eksperymentu, do teoretycznego rozkładu normalnego (sztucznie wygenerowanego) może pomóc w analizie stopnia wygładzenia linii prezentujących rozkłady estymatorów.

Różnice między wykresami są lepiej widoczne, gdy porównuje się wielkości wyliczonej miary (sumy kwadratów przyrostów) stworzonej na podstawie wygenerowanego rozkładu normalnego dla każdego z wariantów parametrów.

Porównanie sum kwadratów dotyczy nie tylko zestawienia z sobą rozkładów empirycznych i teoretycznych. Przyrównano do siebie również odpowiednie sumy dla określonego wariantu parametrów w zależności od liczby wykonanych replikacji w eksperymentcie symulacyjnym.

Wartości w trzech ostatnich wierszach tabel (tab. 1, 3) pokazują, ile razy suma kwadratów przyrostów rozkładu empirycznego dla konkretnego przypadku jest większa od sumy kwadratów przyrostów rozkładu normalnego. Iloraz tych wartości świadczy o tym, czy rozkład empiryczny estymatora zbliża się do rozkładu teoretycznego. Zmniejszające się wartości wraz ze wzrostem liczby replikacji są pożądane.

Taki sam efekt uzyskano przyrównując sumy kwadratów przyrostów dla wszystkich replikacji do sumy kwadratów przyrostów dla 5000 replikacji (wartości w trzech przedostatnich wierszach tabeli). Również w tym przypadku malejące wartości potwierdziły wstępne przypuszczenia (tab. 1).

Tabela 1. Opracowane wyniki dla estymatorów parametru a_3 ; wariant 1 ($\rho = -0,9$; $a_2 = 0,5$)

	a_3	Liczba replikacji		
		5 000	10 000	50 000
Średnia	LS	-0,4587648	-0,4652464	-0,4658070
	LA	-0,2875520	-0,2860192	-0,2850835
	TL	-0,4975808	-0,4996624	-0,4990323
Odchylenie standardowe	LS	0,3195851	0,3159814	0,3154827
	LA	0,2967624	0,2969527	0,2990900
	TL	0,4173166	0,4164319	0,4196384
Współczynnik asymetrii rozkładu	LS	0,0700457	0,0434831	0,0298716
	LA	-0,0989263	-0,0798165	-0,0923290
	TL	0,7630635	0,7385750	0,7227782
Współczynnik kurtozy rozkładu	LS	2,9542370	2,9972944	2,9473819
	LA	2,9420236	2,9649170	2,9958279
	TL	4,6458375	4,7458867	4,6185058
Σ kwadratów przyrostów $\Sigma(\delta^2)$	LS	0,0008348	0,0003545	0,0001752
	LA	0,0004833	0,0004229	0,0001964
	TL	0,0004686	0,0002696	0,0001315
$\Sigma(\delta^2)/\Sigma(\delta^2)$ (5 tys. replikacji)	LS	1	0,4246287	0,2099138
	LA	1	0,8749793	0,4063235
	TL	1	0,5752177	0,2806017
$\Sigma(\delta^2)/\Sigma(\delta n^2)$	LS	5,6788040	2,4113829	1,1920591
	LA	2,8016311	2,4513693	1,1383684
	TL	7,5159161	4,3232876	2,1089791

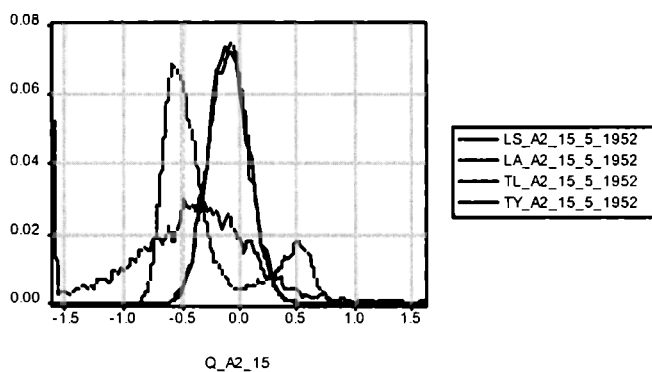
Źródło: opracowanie własne.

Pomimo stopniowego wygładzenia rozkładów estymatorów na histogramach w pojedynczych przypadkach (rys. 4-6.), z uzyskanych obliczeń w tab. 3 wynika, że poprawa nie następuje wraz ze wzrostem liczby replikacji.

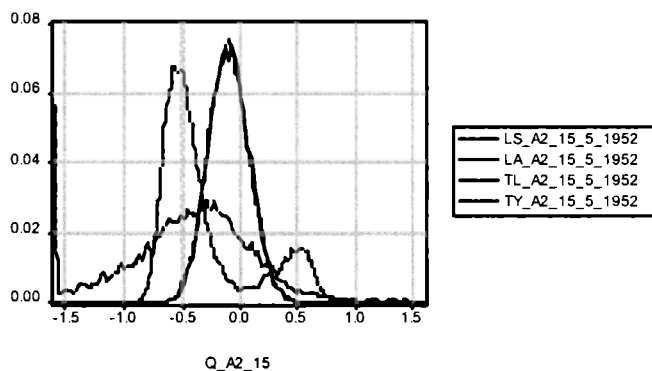
Tabela 2. Suma kwadratów przyrostów rozkładu normalnego $\Sigma(\delta n^2)$ dla estymatorów parametru a_3 ; wariant 1 ($\rho = -0,9$; $a_2 = 0,5$)

LS	0,0001470
LA	0,0001725
TL	0,0000624

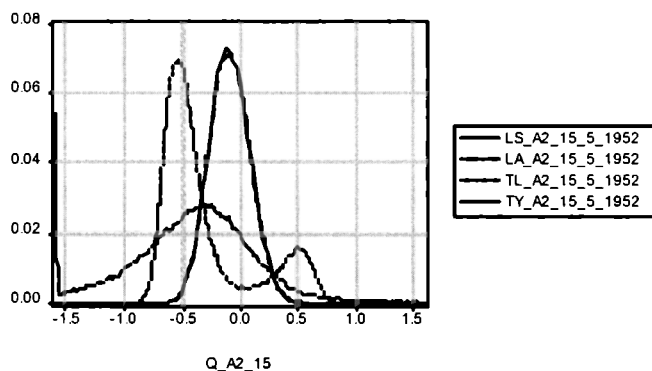
Źródło: opracowanie własne.



Rys. 4. Rozkłady estymatorów parametru a_2 dla 5000 replikacji; wariant 9 ($\rho = 0,5$; $a_2 = -0,5$)
 Źródło: opracowanie własne.



Rys. 5. Rozkłady estymatorów parametru a_2 dla 10 000 replikacji; wariant 9 ($\rho = 0,5$; $a_2 = -0,5$)
 Źródło: opracowanie własne.



Rys. 6. Rozkłady estymatorów parametru a_2 dla 50 000 replikacji; wariant 9 ($\rho = 0,5$; $a_2 = -0,5$)
 Źródło: opracowanie własne.

W większości przypadków otrzymano wartości satysfakcjonujące, potwierdzające oczekiwania dotyczące wyników symulacji. Wartości dwóch ostatnich miar w tabelach o niemalejącej tendencji mogą świadczyć o nieznaczącym przyroście wartości w kolejnych przedziałach na histogramach (tab. 1, 3).

Tabela 3. Opracowane wyniki dla estymatorów parametru a_2 ; wariant 9 ($\rho = 0,5$; $a_2 = -0,5$)

	a2	Liczba replikacji		
		5 000	10 000	50 000
Średnia	LS	-0,1084672	-0,1124224	-0,1106822
	LA	-0,3089664	-0,3204480	-0,3205619
	TL	-0,3730656	-0,3684448	-0,3676966
Odchylenie standardowe	LS	0,1689194	0,1708351	0,1708961
	LA	0,4053322	0,3969868	0,3981725
	TL	0,4678518	0,4636587	0,4678986
Współczynnik asymetrii rozkładu	LS	0,1057391	0,0897557	0,0956646
	LA	1,1452854	1,1983653	1,1983552
	TL	0,0156587	-0,0103146	0,0097531
Współczynnik kurtozy rozkładu	LS	2,7719677	2,8251786	2,8070542
	LA	2,9659860	3,1474267	3,1334812
	TL	3,6969045	3,7348179	3,7507522
Σ kwadratów przyrostów $\Sigma(\delta^2)$	LS	0,0010570	0,0010229	0,0009013
	LA	0,0011778	0,0011861	0,0011089
	TL	0,0003642	0,0002392	0,0000752
$\Sigma(\delta^2)/\Sigma(\delta n^2)$ (5 tys. replikacji)	LS	1	0,9677212	0,8526830
	LA	1	1,0069959	0,9415112
	TL	1	0,6566996	0,2063833
$\Sigma(\delta^2)/\Sigma(\delta n^2)$	LS	1,1465016	1,1094939	0,9776024
	LA	16,1019872	16,2146344	15,1602020
	TL	8,0940741	5,3153753	1,6704818

Źródło: opracowanie własne.

Częściowej analizy dokonano już po wstępnym przeglądzie kształtów histogramów. Jednak nie wszystkie szczegóły są dostrzegalne, dlatego konieczne było opracowanie danych i wykorzystanie miar do analizy wyników symulacji na potrzeby eksperymentu.

Stwierdzenie o coraz wolniejszym procesie wygładzania się wykresów częstości sugerowało skoncentrowanie się na badaniu stopnia ich gładkości. Po omówieniu wyników eksperymentu i analizie wykresów wysunięto wnioski dotyczące wartości otrzymanych niezgodnie z oczekiwaniami.

Wpływ na miary o niemalejącej tendencji mogą mieć:

- sposób wykonywanych obliczeń;
- przyjęcie rozkładu normalnego za wzorcowy, który nie najlepiej odwzorowuje rzeczywiste rozkłady wszystkich estymatorów parametrów przykładowego równania;
- kształt wykresu częstości, wynikający z charakterystyki rozkładu estymatora np. bimodalność.

Można również się spodziewać, że ludzkie oko nie dostrzeże niezgodności w wynikach prezentowanych graficznie, natomiast obliczenia mogą wskazywać na małe różnice, które nie odgrywają akurat w takim przypadku znaczącej roli.

Tabela 4. Suma kwadratów przyrostów rozkładu normalnego $\Sigma(\delta n^2)$ dla estymatorów parametru a_2 ; wariant 9 ($\rho = 0,5$; $a_2 = -0,5$)

LS	0,0009220
LA	0,0000731
TL	0,0000450

Źródło: opracowanie własne.

4. Podsumowanie

Symulacja stochastyczna szybko dostarcza wyników przybliżonych, co oznacza, że dokładniejsze rezultaty badań są niewspółmierne do nakładu kosztów poniesionych w celu ich uzyskania.

Stworzona miara sumy kwadratów przyrostów pozwala ocenić dokładność, z jaką rozkład otrzymany z symulacji stochastycznej aproksymuje rozkład empiryczny z populacji generalnej. Stosunek tej miary dla kolejnych replikacji stanowi zadowalającą miarę nieregularności. Zastosowanie rozkładu normalnego jako wzorcowego rozkładu nie jest niezbędne.

Szczególnie cennym osiągnięciem byłoby stworzenie pewnej uniwersalnej strategii oceny i porównywania nieregularności wykresów częstości otrzymywanych wyników.

Literatura

- Gajda J. B., *Prognozowanie i symulacja a decyzje gospodarcze*, Wydawnictwo C.H. Beck, Warszawa 2001.
- Hall S.G., Henry S.G. B., *Macroeconomic Modeling*, North Holland, Amsterdam, 1988.
- Hendry D.F., *Monte Carlo Experimentation in Econometrics*, [w:] *Handbook of Econometrics*. Vol. 2, red. Z. Griliches, M. D. Intriligator, North Holland, Amsterdam 1984.
- Malinvaud E., *Statistical Methods of Econometrics*, North Holland, Amsterdam 1968.
- Meier R.C., *Simulation in Business and Economics*, Prentice Hall, Englewood Cliffs, 1969.
- Naylor T.H., *Modelowanie cyfrowe systemów ekonomicznych*, PWN, Warszawa 1975.

ACCURACY ANALYSIS OF STOCHASTIC SIMULATION'S RESULTS APPLIED IN LINEAR REGRESSION

Summary

The paper describes an exemplary application of stochastic simulation to analysis of results' accuracy. Reliability and certainty of the solution is the subject of the discussion. The main purpose of the investigation is to determine the desirable number of replications using experimental simulations.