

APPLICATION OF PRINCIPAL COMPONENT ESTIMATORS TO A SIMULTANEOUS EQUATION MODEL OF NIGERIAN ECONOMY.

G. E. I. NWORUH and J. C. NWABUEZE

(Received 20 January 2003; Revision Accepted 23 August 2004)

ABSTRACT

Six instrumental variable estimators are used to obtain consistent estimates of the structural parameters of a model developed for Nigeria economy called MAC III model. The performances of the estimators were ranked using Friedman test statistics based on four criteria namely: Root Mean Square Error (RMSE), Theils inequality coefficients, Bias and variance proportions. The result of this ranking show that the six (6) estimators rank differently or vary in performances. Based on this study, the preference ordering of the estimators is: $2SP_1C_4$, $2SP_2C_4$, $2SP_1C_5$, $2SP_2C_5$, $2SP_2C_6$ and $2SP_1C_6$.

KEY WORDS: Principal Component, Simultaneous, Model, Economy.

INTRODUCTION

The principal component method of estimation is applicable to linear or intrinsically linear equations, or also when the error process in one or more linear equations of a model are autoregressive. The method, inspite of its limitations which include being artificially orthogonal, still works well when compared with other methods, see Judge et al (1980). Principal component is a process in which a group of correlated variables are reduced in number to a more fundamental set of orthogonal variables. The reduced set of variables is used sometimes in a modified form for the first-stage regression: See Klock and Mennes (1960) and Klein (1969).

The estimators have gained wide applications particularly in simultaneous equation models. This is because, the simultaneous equation econometric model has increased in awareness because of the use of these models for policy analysis, the study of multiplier effects and for long and short term forecasting which are necessities for the promotion of rapid economic growth.

Again, at the macro level, simultaneous equation models in real life vary in complexity particularly in the number of structural parameters to be estimated and in the size of the samples. Therefore, to build an operational simultaneous equation or econometric model to describe a system is very important as well as to estimate the parameters of such a system. Unfortunately, the problem is not easily resolved because the estimators must not only be acceptable in terms of the basic statistical properties which include having small bias, root mean square error (RMSE) and being consistent, it must in addition, be able to replicate the original data series. Of particular importance, probably, more than statistical significance, are the signs and magnitudes of the estimates of the parameters. In a simultaneous equation model, each individual equation may have a very good statistical fit, but the model as a whole may perform poorly in reproducing the historical data. The study considers the simulation performance of six principal component estimators with instruments selected, such as $2SP_1C_4$, $2SP_2C_4$, $2SP_1C_5$, $2SP_2C_5$, $2SP_1C_6$ and $2SP_2C_6$. Where, $2SP_1C_4$: First four principal components and any predetermined variables that appear in the equations.

$2SP_2C_4$: First four principal components only.

$2SP_1C_5$: First five principal components and any other predetermined variables that appear in the equations.

$2SP_2C_5$: First five principal components only.

2SP₁C₆: First six principal components and any other predetermined variables that appear in the equations.

2SP₁C₆: First six principal components only.

PRINCIPAL COMPONENT (PC) METHOD OF ESTIMATION

Principal components are a process in which a group of related variables is reduced in number to a more fundamental set of orthogonal variables. The reduced set of variables is used sometimes in a modified form, for the first-stage regression, Klock and Mennes (1960) and Klein (1969). This reduction in dimensionality is achieved by imposing simple specific linear restrictions; the resulting transformed variables have certain alternative minimum variance properties, Greenberg (1975).

We define a vector

$$b = [b_1, b_2, \dots, b_h] \dots \dots \dots (2.1)$$

Where b is the $h \times h$ matrix, such that $b'b = 1$. The columns b_i are the orthogonal characteristics vectors of the normalized $x'x$. This is ordered to correspond to the relative magnitude of the characteristic roots of the

TABLE 1: The Use Of Theils Inequality Coefficient To Compare The Estimators In The Linear Model

Variable	2SP ₁ C ₄	2SP ₂ C ₄	2SP ₁ C ₅	2SP ₂ C ₅	2SP ₁ C ₆	2SP ₂ C ₆
GDPA	(1) 0.00655	(4) 0.00696	(3) 0.00672	(2) 0.00656	(5) 0.00801	(6) 0.00830
GDPMCE	(1) 0.12454	(6) 0.12964	(2) 0.12468	(5) 0.12744	(3) 0.12479	(4) 0.12690
GDPMC	(2) 0.05773	(3) 0.06046	(5) 0.06465	(1) 0.05571	(6) 0.07728	(4) 0.06453
GDPTC	(1) 0.01813	(3) 0.02497	(2) 0.02395	(6) 0.03280	(5) 0.02791	(4) 0.02734
GDPC	(4) 0.02986	(6) 0.03212	(1) 0.02921	(5) 0.03072	(2) 0.02928	(3) 0.02420
GDPSV	(2) 0.09224	(1) 0.09144	(4) 0.09386	(3) 0.09297	(6) 0.10231	(5) 0.09976
GDPMI	(2) 0.09339	(6) 0.09668	(6) 0.0939	(5) 0.09494	(2) 0.09339	(4) 0.09355
CHMIP	(2) 0.05701	(4) 0.05766	(1) 0.05688	(3) 0.05718	(5) 0.05879	(6) 0.06979
CMMIPM	(2) 0.05057	(3) 0.05223	(5) 0.05287	(1) 0.04879	(6) 0.05655	(4) 0.05278
TMF	(3) 0.10443	(5) 0.10852	(4) 0.10667	(1) 0.09896	(6) 0.10859	(2) 0.10351
HITRP	(4) 0.06408	(3) 0.06355	(5) 0.06419	(2) 0.06286	(6) 0.06578	(1) 0.06154
GDR	(1) 0.10241	(3) 0.10327	(2) 0.10292	(6) 0.10730	(4) 0.10406	(5) 0.10440
GORP	(1) 0.04059	(4) 0.04466	(2) 0.04349	(3) 0.04369	(6) 0.04760	(5) 0.04559s
RPOP	(3) 0.00162	(1) 0.00133	(2) 0.00147	(6) 0.00208	(4) 0.00186	(5) 0.00189
UPOP	(3) 0.00565	(1) 0.00429	(2) 0.00516	(5) 0.00806	(6) 0.00814	(4) 0.00738
ECP	(1) 0.07792	(3) 0.08030	(2) 0.07878	(6) 0.08223	(4) 0.08076	(5) 0.08105
SFR	(1) 0.29794	(3) 0.30142	(2) 0.30011	(5) 0.30960	(4) 0.30463	(6) 0.31059
ICE	(1) 0.08199	(2) 0.08259	(3) 0.08326	(5) 0.08403	(6) 0.08498	(4) 0.08340
CFLA	(2) 0.01727	(1) 0.01726	(3) 0.01739	(6) 0.01856	(4) 0.01762	(5) 0.01836
P	(5) 0.11589	(4) 0.10859	(3) 0.06949	(2) 0.06688	(6) 0.35609	(1) 0.05039
PCOMP	(4) 0.00482	(6) 0.00732	(2) 0.00445	(5) 0.00489	(3) 0.00449	(1) 0.00440
GCOM	(1) 0.02024	(2) 0.02105	(4) 0.02339	(3) 0.02247	(6) 0.02993	(5) 0.02684
CFC	(2) 0.05096	(5) 0.05216	(3) 0.05098	(6) 0.05260	(4) 0.05184	(1) 0.05091
CFMET	(3) 0.14743	(5) 0.14825	(2) 0.14737	(4) 0.14773	(1) 0.14647	(6) 0.16435
TMS	(3) 0.02796	(2) 0.23541	(5) 0.02959	(1) 0.02057	(4) 0.02913	(6) 0.03164
Total Ranks	55	86	71	97	114	102
New Ranks	1	3	2	4	6	5

TABLE 3: THE USE OF VARIANCE PROPORTION CRITERION TO COMPARE THE PC'S ESTIMATORS OF THE LINEAR MODEL

Variable	2SP ₁ C ₄	2SP ₂ C ₄	2SP ₁ C ₅	2SP ₂ C ₅	2SP ₂ C ₆	2SP ₁ C ₆
GDPA	(1) 0.0435	(2) 0.0682	(3) 0.0722	(4) 0.0739	(5) 0.0793	(6) 0.0799
GDPMO	(1) 0.0435	(6) 0.0478	(4) 0.0478	(5) 0.0476	(2) 0.0473	(3) 0.0474
GDPMC	(2) 0.2831	(1) 0.2761	(3) 0.4116	(4) 0.441	(6) 0.5026	(5) 0.4893
GDPTC	(1) 0.1053	(4) 0.1178	(5) 0.1289	(6) 0.1354	(3) 0.1086	(2) 0.1073
GDPC	(1) 0.1027	(2) 0.1087	(4) 0.1225	(3) 0.1173	(5) 0.1285	(6) 0.1378
GDPSV	(1) 0.2204	(2) 0.2214	(3) 0.2499	(5) 0.2513	(4) 0.2504	(6) 0.2564
GDPMI	(1.5) 0.0530	(3) 0.0531	(1.5) 0.0530	(4.5) 0.0532	(4.5) 0.0532	(6) 0.0533
CHMIP	(1) 0.1859	(5) 0.1862	(6) 0.1863	(3.5) 0.1861	(2) 0.1860	(3.5) 0.1861
CMMIPM	(1) 0.229	(2) 0.2310	(4) 0.2576	(3) 0.2460	(6) 0.2593	(5) 0.2586
TMF	(1) 0.1128	(2) 0.1129	(4) 0.1324	(3) 0.1315	(5) 0.1334	(6) 0.1339
HITRP	(4) 0.1492	(1) 0.1386	(5) 0.1558	(6) 0.1559	(3) 0.1438	(2) 0.1431
GDR	(1) 0.1349	(2) 0.1352	(5) 0.1612	(6) 0.1614	(4) 0.1604	(3) 0.1601
GDRP	(1) 0.1294	(2) 0.1312	(4) 0.1589	(3) 0.1573	(6) 0.1613	(5) 0.1609
RPOP	(1) 0.1939	(2) 0.1942	(6) 0.2843	(3) 0.2764	(5) 0.2836	(4) 0.2814
UPOP	(1) 0.2120	(2) 0.2130	(3) 0.3579	(4) 0.3581	(6) 0.3691	(5) 0.3673
ECP	(1) 0.1682	(2) 0.1688	(3) 0.2004	(4) 0.2013	(5) 0.2046	(6) 0.2049
SFR	(5) 0.1674	(6) 0.1683	(4) 0.1562	(3) 0.1543	(2) 0.1536	(1) 0.1528
TCE	(2) 0.1203	(1) 0.1201	(4) 0.1445	(3) 0.1441	(5) 0.1456	(6) 0.1576
CFLA	(1) 0.0489	(2) 0.0561	(3) 0.0579	(4) 0.0589	(5) 0.0612	(6) 0.0614
P	(6) 0.0634	(5) 0.0631	(2) 0.0581	(1) 0.0561	(3) 0.0593	(4) 0.0594
PCOMP	(1) 0.0152	(2) 0.0163	(3) 0.0486	(4) 0.0489	(5) 0.0493	(6) 0.0499
GCOM	(1) 0.2953	(2) 0.3146	(4) 0.3476	(3) 0.342	(5) 0.3496	(6) 0.3499
CFC	(1.5) 0.1383	(1.5) 0.1383	(3) 0.1685	(4) 0.1688	(5) 0.1702	(6) 0.1705
CFMET	(4) 0.0731	(1) 0.0729	(2.5) 0.0750	(2.5) 0.0730	(5) 0.0742	(6) 0.0743
TMS	(2) 0.0235	(1) 0.0231	(3) 0.0346	(4) 0.0347	(5.5) 0.0349	(5.5) 0.0349
Total Ranks	44	61.5	69	95.5	111.5	121
New Ranks	1	2	3	4	5	6

obtained by deleting at least one of the h components listed in equation (2.3) above. This is because, using all the h components is the same as using all the related variables. Assume that Q has been partitioned into two parts

$X' [b_i^*: b_j^*] = [Q_1^*: Q_2^*]$. After deleting say Q_2^* and retaining Q_1^* , which accounts for a major proportion of the variance of the X'_s (the predetermined variables).

Let us assume that Q_1^* contains L principal components, we define

$$S = [Q_1^*, X_1] \dots \dots \dots (2.4)$$

Where, $Q_1^* = [Q_1, Q_2, \dots, Q_L]$; $L < h$ be the matrix of order $T \times L$ and Q_i^* is the i^{th} principal component. X_1 in equation (2.4) is the matrix of order $T \times h$ of observations on the predetermined variables included in the i^{th} equation. Thus, S is the $T \times (L+h)$ matrix of regressors for Y_1 . The inclusion of X_1 among the regressors in the first stage is a necessary and sufficient conditions, which ensures that two stage principal component (2SPC) estimators are consistent: see Braindy and Jorgenson (1971).

METHODOLOGY

From the correlations of the predetermined variables of the model with the component loadings and the foregoing conditions, we chose the first four principal components which explain 94 percent and the first six components which explain 98.3 percent of the total variations in the predetermined variables in model of

our study. The proportion of the variations accounted for by these two components are high and can be considered to be adequate for the regression.

Based on this, we considered each of the three principal components, $2SPC_4$, $2SPC_5$ and $2SPC_6$ in two forms; one, which includes other predetermined variables that appear in the equations and the other based only on the principal components as the instruments in the first stage regression, Klein (1974)

Consequently, we have the following six principal component estimators with instruments selected as indicated:

1. $2SP_1C_4$: First four principal components and any predetermined variables that appear in the equations.
2. $2SP_2C_4$: First four principal components only.
3. $2SP_1C_5$: First five principal components and any other predetermined variables that appear in the equations.
4. $2SP_2C_5$: First five principal components only.
5. $2SP_1C_6$: First six principal components and any other predetermined variables that appear in the equations.
6. $2SP_1C_6$: First six principal components only.

TABLE 4: THE USE OF ROOT MEAN SQUARE ERROR TO COMPARE THE PC'S ESTIMATORS OF THE LINEAR MODEL

Variable	$2SP_1C_4$	$2SP_2C_4$	$2SP_1C_5$	$2SP_2C_5$	$2SP_1C_6$	$2SP_2C_6$
GDPA	(1) 591.96114	(2) 593.9482	(4) 599.3513	(3) 596.361	(6) 601.1452	(5) 600.1986
GDPMO	(3) 2229.4416	(2) 2228.9642	(4) 2230.664	(1) 2228.6143	(6) 2232.1693	(5) 2231.864
GDPMC	(2) 633.16644	(1) 631.1764	(3) 670.355	(5) 672.1456	(4) 671.861	(6) 673.976
GDPTC	(1) 144.84474	(2) 153.9672	(3) 166.4725	(4) 169.861	(5) 170.1543	(6) 177.9634
GDPC	(5) 338.88705	(6) 340.1245	(4) 335.1945	(3) 335.0152	(2) 332.967	(1) 331.994
GDPSV	(2) 2494.0527	(1) 2491.0789	(3) 2515.8704	(4) 2637.1123	(5) 2894.1769	(6) 2976.1863
GDPMI	(2) 469.8223	(1) 468.7114	(3) 469.8234	(4) 470.1436	(5) 474.1689	(6) 476.6894
CHMIP	(5) 569.6088	(6) 570.9826	(4) 568.9900	(3) 567.8643	(2) 557.1934	(1) 554.1896
CMMIPM	(1) 172.34171	(2) 173.1489	(3) 176.2255	(4) 177.1345	(5) 178.1365	(6) 179.1563
TMF	(2) 201.6357	(2) 263.5576	(3) 264.9044	(4) 265.6394	(5) 273.8463	(6) 277.9434
HITRP	(1) 1150.4460	(1) 200.9783	(3) 201.8082	(4) 202.1576	(6) 206.8695	(5) 205.1345
GDR	(2) 488.63908	(2) 1150.3217	(3) 1153.2645	(4) 1185.1932	(5) 1188.3217	(6) 1189.4322
GORP	(3) 2.19397	(1) 486.1345	(6) 505.79901	(5) 503.8970	(4) 502.9345	(3) 501.8645
RPOP	(2) 1.36611	(4) 2.2167	(1) 2.08956	(2) 2.09342	(5) 2.2376	(6) 2.2489
UPOP	(3) 1150.4460	(3) 1.37694	(1) 1.30529	(4) 1.3983	(5) 1.4066	(6) 1.4472
ECP	(3) 1597.05332	(5) 1683.0674	(4) 1605.8711	(6) 1763.0035	(2) 1554.1345	(1) 1.5314.0067
SFR	(1) 1106.17706	(3) 1109.1783	(3) 1110.2073	(4) 1125.1763	(5) 463.8614	(6) 1469.7693
TCE	(1) 1662.3481	(2) 1663.9673	(4) 675.1499	(3) 1665.1234	(5) 1673.1345	(6) 1763.0023
CFLA	(1) 12.16203	(2) 12.1786	(4) 12.20331	(3) 12.1963	(5) 12.3164	(6) 12.7215
P	(5) 55.47327	(6) 58.663	(4) 42.95435	(1) 41.6934	(2) 42.6545	(3) 42.7663
PCOMP	(5) 1376.67425	(1) 432.6314	(1) 1322.38056	(2) 1324.8673	(4) 1372.1008	(6) 1377.1145
GCOM	(2) 438.64097	(5) 121.6343	(3) 171.58184	(4) 473.1578	(5) 484.1673	(6) 488.1963
CFC	(2) 920.21687	(1) 864.1973	(3) 920.3172	(1) 919.7363	(6) 922.6488	(4) 921.1764
CFMET	(3) 866.07238	(1) 875.9316	(2) 865.8741	(4) 867.1654	(5) 879.6634	(6) 881.9614
TMS	(2) 877.96624	(1) 875.9316	(4) 903.14312	(3) 100.1354	(6) 912.8670	(5) 909.1634
Total Ranks	55	66	80	85	115	123
New Ranks	1	2	3	4	5	6

TABLE 5: SUMMARY TABLE OF THE FRIEDMAN TEST STATISTIC BASED ON THE PC ESTIMATORS OF THE LINEAR MODEL

CRITERIA	T VALUES	DECISION
Bias proportion	83.99	Significant
RMSE	24.47	Significant
Variance proportion	69.07	Significant
Theils inequality coefficient	26.67	Significant

TABLE 6: SUMMARY OF THE RANKS SHOWING THE RELATIVE PERFORMANCES OF THE PC ESTIMATORS BASED ON THE FOUR CRITERIA

CRITERIA	Estimators in order of preference					
	2SP ₁ C ₄	2SP ₂ C ₄	2SP ₁ C ₅	2SP ₂ C ₅	2SP ₂ C ₆	2SP ₁ C ₆
Bias Proportion	1	2	3	5	4	6
Root Sq. Error	1	3	2	4	5	6
Variance Proportion	3	1	4	2	5	6
Theils Inequality Coefficient	1	2	2	4	5	6
	(1)	(3)	(3)	(4)	(5)	(6)

These six principal component estimators were used to obtain consistent estimates of model called MAC III model. The specification of the model and the structural equations are in Olofin (1985). The model has 25 stochastic equations and 4 identities, so that there are 29 endogenous variables in the model. There are 15 predetermined (endogenous and lagged endogenous) variables. All the equations are over-identified therefore, the number of predetermined variables excluded from the equation must be at least great as the number of exogenous variables included less one: see Johnson and Dinardo (1997). The performances of the estimators were compared using simulation statistics criteria. The interest of this study is not to look at the individual properties of an estimator but to use the properties in the comparison of the relative performance of the estimators. Thus, a ranking procedure is introduced as in Nehlawi (1977) in relation to the endogenous variables in the model.

A computer package for econometric studies called Time Series Processor (TSP) was used on an IBM computer to estimate the parameters of the model.

RESULTS AND ANALYSIS

The estimates of the structural parameters of the equations of the MAC III model were obtained using the six estimators. The variables of the model are described in table 7 in the Appendix. We use four criteria namely Root Mean Square Error (RMSE), Theils Inequality coefficients, Bias and Variance Proportions to compare these estimators. The results of these comparisons are displayed in tables 1-4 using the Friedman test statistic, we rank the performances of the estimators with these criteria and we obtain table 5 which summarizes the results.

The tabulated χ^2 at 0.01 percentage level of significance and 5 degrees of freedom is 20.42. From the results in table 5, the null hypothesis that the estimator rank equally in their performance is rejected and we conclude that the estimators rank differently in their performances we now re-rank the estimators (as indicated in the last row of each of the tables (1-4) and obtain the summary of the preference ordering of the estimators as shown in table 6. From table 6, the ranking of the estimators in increasing order of preference is as follows:

2SP₁C₄, 2SP₂C₄, 2SP₁C₅, 2SP₂C₅, 2SP₂C₆ and 2SP₁C₆. From the analysis, it appears that the inclusion of the predetermined variables which appear in the equation as instruments has improved the performance of the estimators in comparison with when such variables are excluded.

REFERENCES

Johnston, J. and Dinardo, J., 1997. *Econometric Methods*, Fourth Edition, New York, Macgraw-Hill.

Judge, G.G. et al., 1980. *The Theory and Practice of Econometrics*, New York, John Wiley and Sons.

Klein, L. R., 1969. Estimation of Interdependent Systems in Macro Econometric. *Econometrica*, 37: 171-179

Klock, T. and Mennes, L.B., 1960. Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables. *Econometrics*, 28: 45-61.

Nehlawi, J. E., 1997. Consistent Estimations of Real Econometric Models with Undersized Samples: A study of the Trace (M K III) Econometric Model of the Canadian Economy. *Inter. Econ. Rev.*, 18: 163-179.

Olofin, S., 1985. *Modelling Nigerians Economic Development*. University of Ibadan Centre for Econometric and Allied Research. University of Ibadan, Nigeria.

MAC III MODEL* TABLE 7 LIST OF VARIABLES

<u>Endogenous Variables</u>		
GDPA	=	Gross Domestic Product in Agriculture (NM)
GDPMCE	=	Gross Domestic Product in Mining and Quarrying (NM)
GDPMC	=	Gross Manufactured Goods (NM)
GDPTC	=	Gross Domestic Product in Transport and Communications (NM)
GDPC	=	Gross Domestic Product in Construction (NM)
GDPSV	=	Gross Domestic Product in Services (NM)
GDPMI	=	Import of Manufactured Goods, Current, (NM)
CHMIP	=	Import of Machinery and transport Equipment, Current, (NM)
CMMIPM	=	Import of Construction Materials, Current, (NM)
TMF	=	Import of Machinery and Transport Equipment, Current, (NM)
HITRP	=	Import of Food Current, (NM)
GDR	=	Government Direct Revenue (NM)
GORP	=	Government Other Revenue (NM)
RPOP	=	Rural Population (Million Persons)
UPOP	=	Urban Population (Million Persons)
ECP	=	Exports of Crude Petroleum, Current (NM)
SFR	=	Stock of Foreign Reserve (NM)
TCE	=	Total Current Expenditure (NM)
CFLA	=	Capital Formation in Agriculture and Mining (NM)
P	=	Cost of Living Index 1975=100
PCOMP	=	Private Consumption (NM)
GCOM	=	Government Consumption (NM)
CFC	=	Capital Formation in Construction (NM)
CFMET	=	Capital Formation in Machinery and Transport Equipment (NM)
TMS	=	Total Money Supply (NM)
TGDF	=	Total Gross Domestic Product (NM)
TCR	=	Total Government Current Revenue (NM)
POP	=	Population (Million Persons)
TMI	=	Total Merchandise Imports (NM)
<u>Exogenous Variables</u>		
EXAG	=	Export Of Agricultural Products (NM)
TOCP	=	Total Output of crude petroleum (tonnes, million)
FEL	=	Level of Foreign Exchange (NM)
PM	=	Import Price Index 1975=100
CDEF	=	Government Deficit (NM)
DFR	=	Rate of Change of Foreign Reserve (NM)
GSS	=	Sale of Government Securities (NM)
T	=	Time in Years