Classical Pooling Model as Applied to Crimes Against Property in the Sudan

Abdalla Ahmed Alkhalifa Abdalla

University of Tabuk, Faculty of Science, Department of Statistics - Saudi Arabia, Tabuk
Email: *dr.a.alkhalifa[at]gmail.com*

Abstract: ***The research deals with data that is available for both time series - deferent periods of time for one individual - and cross section data – one period of time for deferent individuals in, with application to crimes against property in the Sudan, the study concluded that the estimates obtained from pooled model is more efficient than those obtained from separate analyses. Therefore, if data is available for both time series and cross section data it's better to pool them than to analyze them separately, and form the view of the case study it concluded that crimes against property in the Sudan is highly correlated with population size and most of the variations in crimes against property in the Sudan is due to population size, i. e. the hypotheses of the research are satisfied, the study recommended that efforts must be toward development in the states to encourage opposite migration and hence, distribution of population among states becomes normal and Increase the expenditure on security and control, particularly in the states with high density of population.***

Keywords: crimes against property, Sudan population, pooled model efficiency and time series data

1. Introduction

Last period Sudan suffered from local war and drought and desertification in many states, these lead to a great migration among states searching for security and essentials of living. As a consequence of this migration some states are crowded, so that there will be scarcity in jobs and many people live in the margin. Hence, one expects that the number of crimes against property increase, from which the aim of this study is to estimate the number of crimes against property in the Sudan due to population size change. Sudan is a very large country with twenty six states, and data found for each state over periods of time – time series data - and also for each period of time for all state – cross section data - , it is better to use pooled model to gain the advantages of both time series and cross section analysis and avoiding their disadvantages if they analyzed separately. Crimes against property include offenses, robbery, burglary (breaking & entering), larceny (theft), motor vehicle theft, arson, destruction of property, counterfeiting (forgery), fraud offenses, embezzlement, extortion (black mail), bribery and money laundry. And it has been under study because it affect widely the economy of the country, it is outside country crime and it is a composite of many types of crimes. The main source of data is "The Statistical year book for the year 2000" (Khartoum 2000) and "The Statistical year book for the year 2003" (Khartoum 2004). Both are issued by the Central Bureau of Statistics – Council of Ministers – The Republic of Sudan. And for reliability this data is supported by that of the General Administration of Criminal Investigation – Ministry of Interior – The Republic of Sudan. The research exclude all the southern states except the Upper Nile state, because of the local war the data was not recorded. So the research covers only seventeen states.

Pooling time series and cross section data:

Definition:
When data is available in both time series and cross sectional dimension is regarded as pooled data and takes the following shape [(1)]

Case 1, Time 1 $Y_{11}X_{111}X_{121} \ldots X_{1k1}$
Time 2 $Y_{12}X_{112}X_{122} \ldots X_{1k2}$
·
·
·
Time t $Y_{1t}X_{11t}X_{12t} \ldots X_{1kt}$
Case 2, Time 1 $Y_{21}X_{211}X_{121} \ldots X_{2k1}$
Time 2 $Y_{22}X_{212}X_{222} \ldots X_{2k2}$
·
·
·
Time t $Y_{2t}X_{21t}X_{22t} \ldots X_{2kt}$
·
·
·

Case n, Time 1 $Y_{n1}X_{n11}X_{n21} \ldots X_{nk1}$
Time 2 $Y_{n2}X_{n12}X_{122} \ldots X_{nk2}$
·
·
·
Time t $Y_{nt}X_{n1t}X_{n2t} \ldots X_{nkt}$

Where the first subscript indexes the case, the second indexes the variable, and the third indexes the time. Because Y is variable, there are only two subscripts (case, time).

Hence by pooling data, the dealing will be at one time with "N" firms, each with "T" observations, thus increasing the base of the data. These "NT" observations in a single more general model (i. e. pooled) provide more efficient parameter estimate as well as more meaningful interpretation of what is happening across all the "N" firms for the time period in question, also give the ability to make inference about the population from which the sample of the "N" firms is drawn.

However, when cross section units are more numerous than temporal units (N > T), the pool is often conceptualized as a "cross sectional dominant", conversely, when the temporal units are more numerous than spatial units (T > N), the pool is called "temporal dominant".

Advantages of pooling:

1) If data from a single cross section of firms at one point of time, or a single time series on one firm, were used, it is highly likely that estimation would be unsuccessful. Over time the level of inputs in a given firm are unlikely to vary sufficiently to permit accurate estimation of the constant term. On the other hand, if a single cross section of firms was used, it is unlikely that the coefficients could be estimated accurately. Much of the variation in the dependent variable can be attributed to whether variation and if the increased use of some inputs is to have a mitigated effect on output variation, observations over time are likely to be needed to capture this effect. Thus it would be desirable to use both time series and cross section data. ([2]
2) Identification of models and discriminating between competing hypotheses.
3) Eliminating or reducing estimation bias. [(4)]
4) Reducing problems of multicollinearity.
5) Gives the researcher a large number of observations, thereby increasing degrees of freedom.
6) Pooled analyses is supported because the possibility to capture not only the variations of what emerges through time or space, but the variation of these two dimension simultaneously. This is because instead of testing a cross section model for all individuals at one point in time or testing time series model for one individual using time series data, a pooled model tested for all individuals through time.

Complications of pooling:

Estimating pooled data using OLS procedure tend to generate the following complications:

1) Errors tend to be no independent from a period to a next – they might be serially correlated – such that errors in individual i at time t are correlated with errors in individual I at time t+i.
2) The errors tend to be correlated across individuals – they might contemporaneously correlate – such that errors in individual i at time t are correlated with errors in individual j at time t.
3) Errors tend to be hetroschesdastic.
4) Errors might be nonrandom across spatial and/or temporal units, because parameters are heterogeneous across subsets of units. In other words, since processes linking independent and independent variables tend to vary across subsets of individuals or/and period, errors tend to reflect some causal heterogeneity across space, time, or both.

Hence, the use of both of time series and cross section data is needed (i. e. pool them in one model), but the specification of the model will be more difficult, since the disturbance term is assumed to result in part from the effect of the omitted variables, from which the problem of the research arises.

The models:

The models that are used to analyze the pooled model of both time series and cross section data are: ([4])

1) Separate regression equation.
2) Seemingly unrelated regression.
3) Classical pooling.
4) The covariance or dummy variable model.
5) The error component model.
6) The random coefficient model.
7) The mixed random coefficient model.

The models from 1 to 3 are classified as extremes, because they either analyze each cross section unit individually or pool them classically.

Classical pooling model:

This model used if the parameters in the equation are equal for all N firms, then all N firms can be pooled performing a single O. L. S regression, ie estimating only K parameters.

The model is:

$$Y_{it} = \sum_{k=o}^{k-1} b_k X_{itk} + e_{it} \rightarrow (2.2.3.1)$$

The subscript I is omitted from the b_{ik}, since all coefficients are equal.

In matrix form the model is:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \cdots \\ Y_{Nt} \end{pmatrix} = b_k \begin{pmatrix} X_{111} & 0 & 0 & . & . & . & 0 \\ 0 & X_{222} & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & . & . & 0 & X_{NTk} \end{pmatrix} . + \begin{pmatrix} e_{1t} \\ e_{2t} \\ \cdots \\ e_{NT} \end{pmatrix}$$

But it criticized by that b_k may differ between cross - section, so in this case the model will be improperly specified and the coefficient estimate will be biased.

(1) The same: Look at the intercepts are they the “same” or “differ”, if they are differ then use the Error component model and if they are same use the classical pooling.

Statistical tests that are needed to determine the appropriate model:

To test whether slopes and intercepts for each of the N individuals are the same, so that the N firms can be pooled into a single regression (i. e. move from A to B or E, then from B to C or D in the flowchart). The statistic used is:

$$F = \frac{[SSE(R) - SSE(F)]/[df(R) - df(F)]}{SSE(F)/df(F)} \rightarrow (3.5.1.1)$$

Where,

SSE (R) is the residual sum of squares from the classical pooled regression.

SSE (F) is the full residuals sum of squares = $\sum SSE_i$ where SSE_i is the residual sum of squares from each separate regression.

Df (R) is the degrees of freedom associated with the classical pooled sum of squares.

Df (F) is the degrees of freedom associated with the full sum of squares.

Since there are NT observations used to estimate 2N parameters in the full residuals sum of squares then the degrees of freedom are df (R) = NT - 2N.

Paper ID: SR241209145249 DOI: https://dx.doi.org/10.21275/SR241209145249

And in the pooled regression there are only 2 parameters to be estimated then the degrees of freedom are df (R) = NT – 2.

Therefore, F statistic has an F distribution with (**NT – 2N, NT – 2)** degrees of freedom, to test the hypothesis:
H_o: All intercepts are the same and all slopes are the same.

If the statistic F has a value greater than that of F $(_{\alpha, NT - 2N, NT - 2})$ H_o will be rejected and vice versa.

This test can also be used to test the hypothesis that only slopes (intercepts) are the same.

Data and data description:
Data taken for each state into tow variables, the first one is the number of crimes against property as a dependent variable; the other variable is the population size as an independent variable. The data shown in the following tables:

Table 1: Khartoum State Data

Year	Population size (in thousands)	Number of crimes against property
1998	4372.00	82140.00
1999	4568.00	78049.00
2000	4740.00	73190.00
2001	4936.00	70843.00
2002	5139.00	53497.00
2003	5352.00	69821.00

Table 2: Northern State Data

Year	Population size (in thousands)	Number of crimes against property
1998	562.00	4780.00
1999	573.00	3892.00
2000	582.00	4303.00
2001	593.00	3503.00
2002	603.00	3451.00
2003	614.00	3950.00

Table 3: The River Nile State Data

Year	Population size (in thousands)	Number of crimes against property
1998	866.00	6555.00
1999	883.00	5722.00
2000	900.00	5203.00
2001	918.00	5069.00
2002	936.00	5469.00
2003	954.00	5810.00

Table 4: North Kordofan State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1483.00	10656.00
1999	1461.00	9747.00
2000	1483.00	9337.00
2001	1506.00	9397.00
2002	1530.00	8228.00
2003	1554.00	8256.00

Table 5: South Kordofan State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1081.00	5306.00
1999	1069.00	3436.00
2000	1111.00	3324.00
2001	1127.00	3174.00
2002	1143.00	2810.00
2003	1158.00	3263.00

Table 6: West Kordofan State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1087.00	5533.00
1999	1106.00	4544.00
2000	1124.00	5980.00
2001	1144.00	6050.00
2002	1164.00	3658.00
2003	1183.00	4874.00

Table 7: North Darfor State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1364.00	10708.00
1999	1412.00	8674.00
2000	1455.00	4213.00
2001	1503.00	4501.00
2002	1552.00	3758.00
2003	1603.00	5233.00

Table 8: South Darfor State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1364.00	10708.00
1999	1412.00	8674.00
2000	1455.00	4213.00
2001	2859.00	16197.00
2002	2960.00	15899.00
2003	3064.00	12014.00

Table 9: West Darfor State Data

Year	Population size (in thousands)	Number of crimes against property
1998	2575.00	16222.00
1999	2673.00	13302.00
2000	2760.00	15801.00
2001	1614.00	2231.00
2002	1653.00	1380.00
2003	1693.00	3072.00

Table 10: Gazeera State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1503.00	2389.00
1999	1541.00	2153.00
2000	1577.00	1961.00
2001	3477.00	17893.00
2002	3583.00	15860.00
2003	3692.00	15246.00

Table 11: Blue Nile State Data

Year	Population size (in thousands)	Number of crimes against property
1998	3177.00	21349.00
1999	3280.00	20664.00
2000	3374.00	18255.00
2001	655.00	4246.00
2002	675.00	4259.00
2003	696.00	5081.00

Table 12: White Nile State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1402.00	12218.00
1999	1441.00	11561.00
2000	1476.00	9958.00
2001	1515.00	10273.00
2002	1555.00	9665.00
2003	1595.00	8659.00

Table 13: Sinnar State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1114.00	13455.00
1999	1145.00	11216.00
2000	1173.00	10121.00
2001	1204.00	9302.00
2002	1236.00	8622.00
2003	1268.00	7482.00

Table 14: The Red See State Data

Year	Population size (in thousands)	Number of crimes against property
1998	713.00	7545.00
1999	717.00	6938.00
2000	712.00	6199.00
2001	724.00	6622.00
2002	728.00	7560.00
2003	723.00	7439.00

Table 15: Kassala State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1370.00	9630.00
1999	1420.00	9768.00
2000	1465.00	9607.00
2001	1515.00	11479.00
2002	1567.00	8845.00
2003	1621.00	6849.00

Table 16: Gadarif State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1426.00	4549.00
1999	1440.00	2064.00
2000	1453.00	1562.00
2001	1466.00	986.00
2002	1480.00	743.00
2003	1494.00	1566.00

Table 17: Upper Nile State Data

Year	Population size (in thousands)	Number of crimes against property
1998	1398.00	7641.00
1999	1435.00	7393.00
2000	1469.00	7571.00
2001	1507.00	7155.00
2002	1545.00	5890.00
2003	1584.00	6065.00

Table 18: Summary of population size and crimes against property in Sudan (1998 – 2003)

Year	Population size	Number of crimes against property
1998	26092000	227119
1999	26782000	204378
2000	27499000	191334
2001	28263000	188885
2002	29049000	159594
2003	29503000	174680

It is obvious that population size increases but the number of crimes against property decreasing from year to another. But this is a misleading result because it indicates that the relation between the two variables is an inverse relation, while it is a strong positive relation (0.85), from which the importance of pooling arising in capturing the variations among times and among states.

Table 19: Comparing variations in the three types of data

Data type	Mean of		Standard deviation of		Coefficient of variation C. V of	
	Population	crimes	Population	crimes	Population	crimes
Cross - section data	1618000	11255	1070000	16587	0.7	1.5
Time – series data	1639000	11235	15855000	1057	9.6	0.09
Pooled data	1664941	11306	1061831	15841	0.6	1.4

ő Which is a reasonable result since the C. V of the pooled data is laying between the other two i. e pooling captured the variations among times and among states.

Model selection:

Looking at the output of the separate analysis for each state, it can be observed that the coefficients are approximately equal and since the regression is done through the origin (no intercepts) all intercepts equal to zero and referring to the flowchart on page (42), the appropriate model is classical pooling.

2. Interpretation of the Results

The following table shows the summary of the results obtained from the analysis.

Table 20: summary of the results obtained from the analysis

	Type of data		
	Time series data	Pooled data	Cross section data
r	0.85	0.85	0.86
R^2	0.73	0.73	0.74
F	282	279	47
D. W	1.56	0.197	0.787
β	8.4	8.43E - 03	8.8
$S.e_{\beta}$	0.505	0.001	1.287
t	16.8	16.72	6.8

From the above table, the following results can be obtained:

1) The correlation coefficients (r) between the population size and the number of crimes against property in time series analysis, pooled analysis and cross section analysis are 0.85, 0.85 and 0.86 respectively, which implies a very strong and direct relationship.

2) The coefficients of determination (R^2) for time series analysis, pooled analysis and cross section analysis are 0.73, 0.73 and 0.74 respectively, which implies that the portion of the dependent variable explained by the independent variable is very high.
3) The overall (F) for time series analysis, pooled analysis and cross section analysis are 282, 279 and 47 respectively, which implies that the models are highly significant. The values of F for time series analysis, pooled analysis are higher than that of cross section analysis, because they have a higher degrees of freedom (101), while the cross section analysis has only (16) degrees of freedom.
4) Durbin Watson statistic (D. W) for time series analysis, pooled analysis and cross section analysis are 1.56, 0.197 and 0.787 respectively. Referring to table (2 - 1) and Durbin Watson tables, it implies that there is no serial correlation among residuals.
5) The standard errors of the coefficients for time series analysis, pooled analysis and cross section analysis are 0.505, 0.001 and 1.287 respectively, it implies that the pooled analysis has the lowest standard errors i. e. it is the efficient estimate.
6) The t - statistic values for time series analysis, pooled analysis and cross section analysis are 16.8, 16.7 and 6.8 respectively, indicating a very high significant coefficients.

3. Concluding Remarks

Investigating the results in chapter five it can be concluded that statistically, the estimate obtained from pooled model is more efficient than those obtained from separate analyses. Therefore, if data is available for both time series and cross section data it's better to pool them than to analyze them separately, and form the view of the case study it concluded that crimes against property in the Sudan is highly correlated with population size and most of the variations in crimes against property in the Sudan is due to population size, i. e. the hypotheses of the research are satisfied.

4. Recommendations

It is highly recommended that:
1) Efforts must be toward development in the states to encourage opposite migration and hence, distribution of population among states becomes normal.
2) Increase the expenditure on security and control, particularly in the states with high density of population.
3) Appling a hard punishment to the criminals.

References

[1] Federico Podesta, Recent development in quantitative comparative methodology: the case of pooled time series and cross section analysis, 2004.
[2] Griffths William E. and Anderson Jock R., using time series and cross section data to estimate production function with positive and negative marginal risk, Journal of the American statistical association (JASA), vol 77, September 1982, p. p 529 - 536.
[3] John L. Worral and Travis C. Pratt, estimation issues associated with time series – cross section analysis in criminology, western criminology review, 5 (1), 2004.
[4] Terry E. Dielman, pooled cross sectional and time series data analysis, 1989.