CHAPTER 5

RESULTS OF EMPIRICAL ANALYSIS

In time series regression, positive correlation of residuals which are adjacent in time is frequently a problem. Serial correlation of the residuals, as it is called, is endemic in time series work (Hall, 1984).

We however, proposed an error-components model for the present study in chapter 4 which assumes that the individual error-components are uncorrelated with each other and are not autocorrelated across both cross-section and time-series units and therefore this chapter presents the results of fitted regression model proposed in equation 4.2 in the preceding chapter which go through a series of steps to test the specifications of the regression model as well as identifying any possible problem of multicollinearity in order to arrive at the goal of this research study. Lastly, the results of error component and weighted ordinary least square methods are worked out and compared in order to justify and confirm the selection of the technique used for the present study.

5.1 Result of Initial Test Run of Multiple Regression

Scatter plots (not shown) of the dependent variable as against each of the individual variables using TSP computer program do not show any serious departure from the linearity assumption of linear regression function. The study of the Matrix of Simple correlation of the variable (Table 5.1) shows correlation between some of the independent variables below -0.7. Makridakis et al (1989) suggested the picking of independent variables whose simple correlations are not bigger than 0.7 or smaller than -0.7 in order to avoid encountering the problem of multicollinearity. They however, cautioned that this rule of the thumb concerning multicollinearity does not always hold; this means that a value smaller than 0.7 or larger than -0.7 may result in multicollinearity, while a value larger than 0.7 or smaller than -0.7

may not, since the correlations do not so much from this rule of the thumb, the problem of multicollinearity are identified using any abnormalities in the coefficients of the variables during the regression analysis as suggested in section 4.4 (Chapter 4). Among all the independent variables only incidence rate, R, has the weakest correlation (-0.149) with the output variable, Q.

Table 5.1 Simple Correlation Matrix for Demand for Malaria Services

Variable	Q	R	X	S	log I	log G
Q	1	-0.149	0.345	0.980	0.489	-0.707
R	-0.149	1	0.602	-0.080	-0.782	0.319
X	0.345	0.602	1	0.458	-0.316	-0.060
S	0.980	- 0.08	0.458	1	0.463	-0.704
log I	0.489	- 0.782	-0.316	0.463	1	-0.521
log G	-0.707	0.319	-0.060	0.704	-0.521	1

Having met the linearity assumption of regression functions, the procedure adopted to estimate the proposed regression model (4.2) is demonstrated below.

For the model (4.2), the estimated regression function is as follows.

$$Q = b_0 + b_1 S + b_2 X + b_3 \log G + b_4 \log I + b_5 R \qquad (5.1)$$

where b_0, b_1, b_2, b_3, b_4 and b_5 are the least squares estimators of the corresponding parameters. The equation (5.1) contains all five independent variables theoretically derived in chapter 3. The results of applying TSP computer programme by using the foregoing regression equation and the data in Table 4.2 are shown in Table 5.2

Table 5.2 Regression Equation for Demand Analysis for Malaria Service

Dependent Variable = Q				
Variable	Parameter value	Standard Error	T. Ratio	Significant
Constant	0.49	7.56	0.06	No
S	0.35	0.02	18.97	yes
X	-2.22	0.52	-4.28	yes
log I	-1.33	0.74	-1.80	yes
log G	1.38	11.68	0.82	yes

R^2 = 0.978 ; t value from table (α = 0.05) = 1.706
Adjusted R^2 = 0.975 ; F value from table (α = 0.05) = 2.59
Durbin-Watson statistic = 1.502 ; F statistic = 239.35

The estimated regression function is seen to be:

$$Q = 2.30 + 0.355 - 2.03X - 20.31\log I + 1.57\log G - 0.10R \quad (5.2)$$

As can be seen in Table 5.1, not all the coefficients in this regression equation are significant. Looking at the t ratios and comparing them to the corresponding values from the table, the significance of the coefficients is determined at 95% level. It can be seen that independent variables log G and R are not significantly different from 0 in terms of their impact on Q. This result could be due to a lack of a significant relationship between variables log G and R and Q or it could be due to multicollinearity between some of the variables, since both R^2 and F test are large.

Examining the simple correlation matrix shown in Table 5.1 we can see that variable R has a relatively small correlation with Q.

Variable log G, however, whose coefficient was not significant in Table 5.2, seems to have a fairly high correlation with Q. The problem here is that multicollinearity does exist between logI and R. (The coefficient of correlation is -0.782). This means that we need to drop either log I or R from our regression equation. If we examine the correlation between Q and logI and R and Q, we see that the correlation is higher with log I. Thus we choose to eliminate variable R from our regression equation.

The new equation to be tested can be written as

$$Q = b_0 + b_1 S + b_2 X + b_3 \log I + b_4 \log G \qquad (5.3)$$

The results for this regression analysis are presented in Table 5.3.

Table 5.3 Regression Equation for Demand analysis of Malaria Services.

Dependent Variable = Q				
Variable	Parameter value	Standard Error	T. Ratio	Significant
Constant	0.49	7.56	0.06	No
S	0.35	0.02	18.97	yes
X	-2.22	0.52	-4.28	yes
log I	-1.33	0.74	-1.80	yes
log G	1.38	11.68	0.82	yes

R^2 = 0.978 ;
Adjusted R^2 = 0.974 ; Sum of Squared residual = 127.15
Durbin-Watson statistic = 1.382 ; F statistics = 290.95

Looking at the table 5.3 again it can be seen that the constant and the variable log G is not significantly different from 0 in terms of their impact on Q. Since from correlation matrix Table 5.1 there is a high correlation between Q and log G the problem is again multicollinearity between some of the variables, since both R^2 and F test are still large. From table 5.1 log G and S have a correlation of -0.704 a possible cause of multicollinearity according to the rule of the thumb stated above. This means that either log G or S should be eliminated. However, a look at the lag effect of log G on malaria control is very plausible from experience. Again, the correlation between S and log G(-1), the lag of log G is -0.642 and therefore the use of log G(-1) and S in the regression equation instead of log G and S is applied instead and the new equation becomes:

$$Q = b_0 + b_1S + b_2X + b_3\log I + b_4\log G(-1) \quad (5.4)$$

The results are presented in Table 5.4.

Table 5.4 Regression Equation for Demand analysis of Malaria Services.

Dependent Variable = Q				
Variable	Parameter value	Standard Error	T. Ratio	Significant
Constant	0.49	7.56	0.06	No
S	0.35	0.02	18.97	yes
X	-2.22	0.52	-4.28	yes
log I	-1.33	0.74	-1.80	yes
log G	1.38	11.68	0.82	yes

$R^2 = 0.979$; t value from table ($\alpha = 0.05$) = 1.706

Adjusted R^2 = 0.975 ; F value from table (α = 0.05) = 2.76

Durbin-Watson statistic = 1.131 ; F statistic = 288.57

Examining Table 5.4 the results are better and show no sign of multicollinearity among the variables. Yet the coefficient of log G(-1) is still not significant at the 95% level. The problem this time is due to the low Durbin-Watson statistic of 1.131 a sign of serial correlation of the residuals. If there is no problem of association between adjacent residuals, the Durbin-Watson statistic will be around 2. With positive serial correlation, the Durbin-Watson will fall below 2; in the worst cases, it will be near zero. Occasionally, negative serial correlation will occur and the Durbin-Watson statistic will lie somewhere between 2 and 4. Positive serial correlation is the principal hazard - with 50 or more observations and not too many independent variables, a Durbin-Watson below about 1.5 is a danger sign (Hall, 1984).

In our present regression, the Durbin-Watson statistic of 1.131 is a sure sign of trouble with positive serial correlation. The regression is plagued with an unexplained factor that moves slowly over time. Positive serial correlation has three implications. First, the reliability of the reported regression results is probably overstated. Regression theory rests on the assumption of zero serial correlation and so does the error-components model proposed for this research study in equation (4.2). The computed standard errors of the coefficients are generally too small when the assumption fails. Second, the regression is not the best prediction of the dependent variable if there is an association between adjacent residuals. Third, serial correlation is a sign of specification error in the regression. It is a hint that we should look for additional predictors of the sort that might explain the slowly moving errors. The last option is impossible in this study due to the obvious constraints of time, effort, nonavailability of data and budget.

Here a technique - Cochrane-Orcutt technique or first-order autoregressive correlation - which deals with serial correlation as a statistic problem is used here. In microTSP, we estimate Cochrane-Orcutt by adding the specification AR(1) to the command for the regression (5.4). The results of the estimated regression function is seen to be:

$$Q = 21.19 + 0.34S - 1.20X - 2.03\log I - 3.46\log G(-1) \quad (5.5)$$
$$(3.404)\ (23.215)\ (-3.259)(-2.902)\quad (-2.543)$$
$$D\text{-}W = 1.878$$

R^2 = 0.985

Adjusted R^2 = 0.981

F-statistics = 295

Sum of squared residuals = 89.127

(t-values are shown in the parentheses).

As can be seen from this new regression equation, all the t ratios are statistically significant, which indicates that the term and each of the regression coefficients is significantly different from 0. The F test is greater than 2.76 so the entire regression equation is significant. Finally, the value of R^2 is equal to 0.985, which indicates that 98.5% of the fluctuations in demand for Malaria services are explained by the regression equation (5.5).

The equation does represent a linear relationship, the tests of significance are satisfied, the R^2 is good and it appears that the regression equation is a good representation of this situation. The value of the D-W test is 1.878 which lies within the allowable range of 1.72 to 2.28. However, this model as it is, does not take into account the existence of cross-section and time-series disturbances. It only combines the cross-section and time series data and perform ordinary least square regression on the entire data set. It therefore assumes an error term that consists of a single combined disturbance of which the intercept terms do not vary, either systematically, as in a covariance model or randomly, as in an error-components model. furthermore it does not take into account the differences of the regression coefficients and residual variances associated with the different districts. As postulated, we assumed that there is a unique residual variance associated with each of the three districts.

The next section looks at the results of the two techniques - error-components pooling procedure and weighted ordinary least square regression - which take into account these differences in the districts and then with a two-stage estimation method derive in each case a single relationship that serves as the best representation for all the districts and all the sector malaria clinics in Tak Province.

5.2 Results of Two-Stage Estimation of Weighted Ordinary Least Square Regression and the Error-Components Pooling Technique:

The methods and computations used in this section are explained in the preceding chapter, section 4.5. In the first stage, ordinary least squares is run on the entire pooled data in Table 5.5 using the regression equation (5.6).

$$Q = b_0 + b_1 S + b_2 X + b_3 \log I + b_4 \log G(-1) \qquad (5.6)$$

Table 5.5 Demand Analysis for Malaria Services at Selected Sector Malaria Clinics in Tak Province.

Obs.	M.C	Dist	Time	Q_{ijt} * 100	S_{ijt}	X_{ijt}	log I_{ijt}	log G_{ijt}
1	1	1	1991	10.06	57.41	6.6	1.386	4.259
2	1	1	1992	11.39	64.42	6.6	1.386	4.289
3	1	1	1993	7.74	60.39	6.6	1.386	4.452
4	2	1	1991	2.53	25.96	5.2	1.386	4.259
5	2	1	1992	2.27	23.68	5.2	1.386	4.289
6	2	1	1993	2.00	21.39	5.2	1.386	4.452
7	3	1	1991	7.26	40.22	5.3	1.386	4.259
8	3	1	1992	6.47	44.36	5.3	1.386	4.289
9	3	1	1993	4.86	44.31	5.3	1.386	4.452
10	4	2	1990	39.31	145.1	4.5	2.862	3.349
11	4	2	1991	23.86	87.07	4.5	2.882	3.261
12	4	2	1992	20.32	82.20	4.5	2.893	3.832
13	4	2	1993	11.39	65.93	4.5	3.068	3.872
14	5	2	1990	16.62	81.23	5.3	2.862	3.349
15	5	2	1991	50.78	172.5	5.3	2.882	3.261
16	5	2	1992	46.42	150.5	5.3	2.893	3.832
17	5	2	1993	35.60	143.9	5.3	5.370	3.872
18	6	2	1992	3.30	30.90	3.3	2.893	3.832
19	6	2	1993	8.8	49.51	3.3	3.068	3.068

Table 5.5 continued

Obs.	M.C	Dist	Time	Q_{ijt} * 100	S_{ijt}	X_{ijt}	log I_{ijt}	log G_{ijt}
20	7	3	1989	15.36	65.50	4.9	2.282	3.846
21	7	3	1990	23.07	82.67	4.9	2.332	4.474
22	7	3	1991	11.28	58.34	4.9	2.397	4.267
23	7	3	1992	7.91	42.49	4.9	2.484	4.305
24	7	3	1993	6.09	34.67	4.9	2.525	4.366
25	8	3	1993	0.04	1.49	3.7	2.397	4.267
26	8	3	1992	0.34	11.61	3.7	2.484	4.305
27	8	3	1993	0.23	8.49	3.7	2.525	4.366
28	9	3	1991	0.86	8.64	2.4	2.397	4.267
29	9	3	1992	1.28	11.57	2.4	2.484	4.305
30	9	3	1993	1.030	9.33	2.4	2.525	4.366
31	10	3	1993	0.270	7.62	1.7	2.525	4.366

The ordinary least squares regression residuals derived by applying Micro TSP computer programme and the regression equation (5.6) are used in the methods described (chapter 4 section 4.5) to compute the weights. The results are shown in tables 5.6 and 5.7 respectively for ordinary least square and error-components pooling techniques.

Table 5.6 Weights for weighted least squares

Districts	n	S^2_j	C_j
1. Tasongyang	8	3.810	1.112
2. Phop Phra	10	2.555	0.910
3. Mae Ramard	12	2.884	0.967

n = the individual observations in the pooled data for each district

Table 5.7 Weights for error-components pooling technique

District	n	N_j	T_j	S^2_{wj}	S^2_{uj}	S^2_{vj}	S^2_j	$\tilde{C}_j$
1. Tasongyang	8	3	2.7	0.989	2.254	2.168	5.411	1.24
2. Phop Phra	10	3	3.3	0.848	1.0848	1.142	3.087	0.94
3. Mae Ramard	12	4	3	0.731	0.692	0.607	2.030	0.76

N_j = the number of selected sector Malaria clinics in district j

T_j = the average period time interval the pooled data was collected in district j.

As can be seen from the tables the C_j's in each of the techniques are quite different. The estimated weights in each case are used in the second stage in which the generalized least-squares parameter estimates are obtained. The transformed regression equations for both techniques are as follows:

For weighted ordinary least square (WLS) technique:

$$\frac{Q_{ijt}}{C_j} = b_0 \frac{1}{C_j} + b_1 \frac{S_{ijt}}{C_j} + b_2 \frac{X_{ijt}}{C_j} + b_3 \frac{\log I_{ijt}}{C_j} + b_4 \frac{\log G(-1)_{ijt}}{C_j} \qquad (5.7)$$

And for the error-components pooling procedure (ECP)

$$Q_{ijt} = b_0\ 1_{+}\ b_1\ S_{ijt+}\ b_2\ X_{ijt+}\ b_3\ \log I_{ijt+}\ b_4\ \log G(-1)_{ijt}$$

(5.8)

$$Cj \quad Cj \quad Cj \quad C_j \quad C_j$$

The results of applying Micro TSP programme by using equations (5.7 and 5.8) and the data in Tables 5.8 and 5.9 respectively (after correcting for any serial correlation) are shown in Table 5.10.

Table 5.8 Transformed pooled data for Demand Analysis for Malaria Services - WLS Technique

Obs	$\frac{1}{C_j}$	$\frac{Q_{ijt}}{C_j}$	$\frac{S_{ijt}}{C_j}$	$\frac{X_{ij}}{C_j}$	$\frac{\log I_{ijt}}{C_j}$	$\frac{\log G(-1)_{ijt}}{C_j}$
1	0.899	9.047	51.628	5.935	1.247	3.831
2	0.899	10.243	57.932	5.935	1.247	3.856
3	0.899	6.960	54.308	4.676	1.247	4.004
4	0.899	2.275	23.345	4.676	1.247	3.831
5	0.899	2.041	21.295	4.676	1.247	3.858
6	0.899	1.799	19.236	4.766	1.247	4.004
7	0.899	6.529	36.169	4.766	1.247	3.831
8	0.899	5.818	39.883	4.766	1.247	3.858
9	0.899	4.371	39.847	4.766	1.247	4.004
10	1.099	43.198	159.396	4.945	3.145	3.681
11	1.099	26.220	95.681	4.945	3.167	3.585
12	1.099	22.330	90.330	4.945	1.179	4.212
13	1.099	12.516	72.451	4.945	3.371	4.256
14	1.099	18.263	89.264	5.824	3.145	3.681
15	1.099	55.802	189.517	5.824	3.167	3.584
16	1.099	51.011	165.384	5.824	3.176	4.212
17	1.099	39.121	158.132	5.824	5.901	4.256
18	1.099	3.626	32.956	3.626	3.179	4.212
19	1.099	9.670	54.407	3.626	3.371	4.256

Table 5.8 continued

Obs	$\frac{1}{C_j}$	$\frac{Q_{ijt}}{C_j}$	$\frac{S_{ijt}}{C_j}$	$\frac{X_{ij}}{C_j}$	$\frac{\log I_{ijt}}{C_j}$	$\frac{\log G(-1)_{ijt}}{C_j}$
20	1.034	9.670	67.735	5.067	2.360	3.978
21	1.034	23.857	85.491	5.067	2.412	4.627
22	1.034	11.665	60.331	5.067	2.480	4.410
23	1.034	8.180	43.940	5.067	2.570	4.453
24	1.034	6.298	35.853	5.067	2.612	4.561
25	1.034	0.041	1.541	3.825	2.480	4.413
26	1.034	0.352	0.352	12.01	3.826	4.452
27	1.034	0.238	0.238	8.780	3.826	4.516
28	1.034	0.889	0.889	8.935	3.826	4.413
29	1.034	1.324	1.324	11.96	2.481	4.453
30	1.034	1.065	1.065	9.648	2.481	4.516
31	1.034	0.279	0.279	7.880	1.758	4.516

C_j = the weights assigned to district j

Table 5.9 Transformed pooled data from Demand Analysis for Malaria services - Error-components pooling technique

Obs	$\frac{1}{\tilde{C}_j}$	$\frac{Q_{ijt}}{\tilde{C}_j}$	$\frac{S_{ijt}}{\tilde{C}_j}$	$\frac{X_{ij}}{\tilde{C}_j}$	$\frac{\log I_{ijt}}{\tilde{C}_j}$	$\frac{\log G(-1)_{ijt}}{\tilde{C}_j}$
1	0.805	8.101	46.223	5.314	3.430	3.430
2	0.805	9.171	57.869	5.314	3.454	3.454
3	0.805	6.232	48.623	5.314	3.585	3.585
4	0.805	2.037	20.902	4.187	3.430	3.430
5	0.805	1.610	19.066	4.187	3.454	3.454
6	0.805	1.610	17.222	4.187	3.585	3.585
7	0.805	5.845	32.383	4.267	3.430	3.430
8	0.805	5.209	35.709	4.267	3.454	3.454
9	0.805	3.913	35.676	4.267	3.585	3.585
10	1.066	41.908	154.638	4.797	3.571	3.571
11	1.066	25.437	92.825	4.797	3.477	4.477

Table 5.9 continued

Obs	$\frac{1}{\tilde{C}_j}$	$\frac{Q_{ijt}}{\tilde{C}_j}$	$\frac{S_{ijt}}{\tilde{C}_j}$	$\frac{X_{ij}}{\tilde{C}_j}$	$\frac{\log I_{ijt}}{\tilde{C}_j}$	$\frac{\log G(-1)_{ijt}}{\tilde{C}_j}$
12	1.066	21.663	87.633	4.797	3.084	4.086
13	1.066	12.143	70.288	4.797	3.271	4.129
14	1.066	17.719	86.599	5.650	3.051	3.571
15	1.066	54.136	183.859	5.650	3.072	3.478
16	1.066	49.488	160.448	5.650	3.084	4.086
17	1.066	37.953	153.412	5.650	5.725	4.129
18	1.066	3.5181	32.942	3.518	3.084	4.086
19	1.314	9.381	52.783	3.518	3.271	4.129
20	1.314	20.184	86.071	6.439	2.991	5.054
21	1.314	30.315	108.633	6.439	3.065	5.879
22	1.314	14.723	76.662	6.439	3.151	5.608
23	1.314	10.394	55.834	6.439	3.265	5.658
24	1.314	8.003	45.5581	6.439	3.319	5.738
25	1.314	0.053	1.958	4.862	3.151	5.608
26	1.314	0.447	15.256	4.862	3.265	5.658
27	1.314	0.302	11.156	4.862	3.319	5.738
28	1.314	1.130	11.353	3.154	3.151	5.608
29	1.314	1.682	15.204	3.154	3.265	5.658
30	1.314	1.353	12.260	3.154	3.319	5.738
31	1.314	0.355	10.013	2.339	3.319	5.738

$\tilde{C}_j$ = the weights assigned to district j

Table 5.10 WLS and ECP coefficients for Demand for malaria services data (n = 29)

Variable	WLS			ECD		
	Coeff	S.E	t	Coeff	S.E	t
S	0.329	0.017	19.314	0.332	0.014	23.006
X	-1.373	0.834	-1.647	-1.427	0.553	-2.5804
log I	-1.744	0.768	-2.270	-2.243	0.709	-3.160
log G(-1)	-5.353	1.932	-2.770	-3.375	1.418	-2.3805
Constant	13.33	14.41	0.925	13.17	4.295	3.066

R_2 = 0.987 S^2 = 93.76 R^2 = 0.987 S^2 = 92.20
F test=273.93 D-W = 2.064 F test=270.13 D-W = 1.867

t value from table (α = 0.05) = 1.711

F value from table (α = 0.05) = 2.78

These results are obtained by running a regression of the WLS and

$\frac{Q_{ijt}}{C_j}$ versus $\frac{1}{C_j}$, $\frac{S_{ijt}}{C_j}$, $\frac{X_{ijt}}{C_j}$, $\frac{\log I_{ijt}}{C_j}$, $\frac{\log G(-1)_{ijt}}{C_j}$ for

$\frac{Q_{ijt}}{\tilde{C}_j}$ versus $\frac{1}{\tilde{C}_j}$, $\frac{S_{ijt}}{\tilde{C}_j}$, $\frac{X_{ijt}}{\tilde{C}_j}$, $\frac{\log I_{ijt}}{\tilde{C}_j}$, $\frac{\log G(-1)_{ijt}}{\tilde{C}_j}$ for the ECP respectively.

As can be seen from Table 5.9, both WLS and ECP have their values of R^2 equal to 0.987 which indicates that 98.7% of the fluctuations in demand for malaria services are explained by the regression equations shown in Table 5.10. Also, F test in both cases are higher than 2.78 (F value from table). Therefore the entire regression equations are significant. However, there are considerable differences between the two regression equations. First, the residual sum of squares associated with WLS is a little bit higher than EPC's. Second, standard errors of the coefficients of the independent variables as well as the constant are larger than those of EPC thus having considerable effects on the t-ratios. Whilst the t-ratios in EPC are all statistically significant, indicating that the term and each of the regression coefficients (b_1, b_2, b_3 b_4) is significantly different from 0, this is not true for all the t-ratios of the WLS. In the latter, it can be seen that the independent variable X and the term are not significantly different from 0 in terms of their impact on demand for malaria services. Third, though the D-W tests are different they are all within the allowable range of 1.72 to 2.28 indicating that there are no serial correlations in both cases. The residual plots show no serial correlations in both cases. The residual plots of both WLS and EPC show that their residuals are about constant. They are neither larger nor smaller in the beginning or at the end, as can be seen in appendices 1 and 2 respectively. Consequently, on the basis of statistical testing, it is adequately shown that EPC is more efficient than WLS and hence the choice for the present study. The rest of the analysis from now on is based on the regression equation of the EPC.

The precise equation of EPC taken from Table 5.9 is

$$Q_{ij} = 13.17 + 0.33S_{ij} - 1.43X_{ij} - 2.24\log I_j - 1.42\log G(-1)_j \quad (5.9)$$

Where

Q_{ij} = Annual demand for Malaria Services i.e. Malaria patients diagnosed positive and given treatment at sector malaria clinic i in district j (in units of hundreds)

S_{ij} = Annual blood slides examined at sector Malaria Clinic i in district j i.e. for both negative and positive blood slides examined (in units of thousands)

X_{ij} = mean distance from suspected patient's village/town to the sector Malaria clinic i in district j. (in km.)

I_j = the average annual household income for district j (in thousands of Bahts)

$G(-1)_j$ = lag of annual per capita government expenditure on malaria control activities in district j (in Bahts).

Since the annual blood slides examined constitute both patients microscopically confirmed as malaria patients and those who are not (but might have been suffering from other disease apart from malaria) and, a substantial time, effort and budget are spent on the blood slide preparations and examinations, (though those confirmed negative do not receive treatment) it may be of interest to know the annual blood slides to be examined ahead of time so that with it managers of the malaria clinics can plan for their budget allocation. Also since equation (5.9) already contains the variable S_{ij}, the total annual blood examined, the only independent variable whose estimate is not easy to make (as variables X, I and G(-1) can reasonably be estimated with relative ease), the most appropriate thing to do is to construct a forecast equation to extrapolate annual total blood examine every year ahead.

Consequently, as the only available data we have are those being used for this study, it is reasoned out that the first true forecasting regression uses the lagged value of annual blood slides examined, (S(-1)) the lagged value of log of per capita of government expenditure, (log G(-1) the mean distance travel, X and the lagged value of annual average household income, (log I(-1)) . It is a 77

pretty reliable rule of forecasting that the current value of a variable will help you forecast its future value (Hall, 1984).

The results of the forecasting equation using error-components pooling technique described in the preceding chapter and the data in Table 5.9 are shown below:

$$S = 340.58 + 0.53S(-1) - 58.10\log G(-1) - 21.42\log I(-1) + 6.13X \quad (5.11)$$
$$\quad (5.43) \quad (3.122) \qquad (-4.76) \qquad\qquad (-2.205) \qquad\qquad (0.87)$$

R^2 = 0.775   Residual square = 16707.94

Adjusted R_2 = 0.728   F Statistic = 16.51

D-W = 2.032

(t-values are shown in the parentheses)

As can be seen, though R^2, D-W and F test values are all statistically significant the coefficient of variable X is statistically insignificant i.e. not different from 0. So we remove the variable X from the equation and run the regression again using the same data set without X variables. The results are shown as follows:

$$S = 317.66 + 0.57S(-1) - 23.27\log I(-1) - 59.34\log G(-1) \quad (5.12)$$
$$\quad (5.94) \quad (4.69) \qquad (-3.54) \qquad\qquad (-4.96)$$

R2 = 0.763   F Statistic = 27

Adjusted R^2 = 0.736   Sum of squared residual = 17563

D-W = 1.980

(t-values are shown in the parentheses)

Now the result seems quite impressive. F test and R^2 have increased and D-W value is within acceptable range of 1.72 to 2.28. All the t 78 test values for the coefficients are now statistically significant. The coefficient on the lagged dependent variable S(-1), is under 1, which is typical of many forecasting equations - in effect we are forecasting the change of S from its current value. The regression is saying that we should expect the annual blood slides examined to rise in the year per capita government expenditure on malaria control activities and annual household income fall and to fall after high government expenditure and increase in annual household income reliability and the economic implications of the two error-components functions 5.9 and 5.11 the subject of discussion in the next chapter.