Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.12, 2014

Estimating True Movement In Time Series As Linear Combination Of Two Indicators

Akeyede I., (Corresponding Author)
Department of Mathematics,
Federal University Lafia, PMB 146, Lafia
Nasarawa State, Nigeria.
Tel: +2348051389397; Email: imamakeyede@gmail.com

Babalola, O. A.
[3]National Population Commission Osogbo, Osun State
Tel: +2348038268363; Email: babalolaoa@yahoo.com

Chindo, A. A. And
[2]National Centre for Hydropower and Research Development
University of Ilorin, Ilorin, Nigeria
Tel: +2347030494872; Email: abdulrasaqchindo@yahoo.co.uk

Abstract

This paper describes an iterative procedure for estimating true change in a time series as a linear combination of two imperfect indicators with weight chosen to minimize errors particularly when the time series (dependent variables) is not or cannot be observed, this research work was compiled to estimate the coefficients of the β's responsiveness of indicators to change in true series, in order to determine the equation that connect variables. It also describes how an appropriate weight could be chosen for the two indicators as to error.

We found that; the best estimation for changing the number of caesarean section is;

$T_1 = 0.1720x + 0.967y_1 - 0.0287$

The length of stay in hospital series received a substantial weight compared to the weight of parity series. The final estimate of a_x to be positive and reduction of error expected from using the proposed procedure was estimated at about 6% and 27% per month compared to both parity and length of stay in hospital indicators respectively.

Key Words: Imperfect Indicators, True Series, Weighted Time series

1 Introduction

A time series could be defined as data collected at equal interval of time. Such interval could be weekly, monthly, quarterly or yearly. A true time series can be represented pictorially by constructing a graph of observation against time.

Experience with many examples of time series has revealed certain characteristics movements or variations, some or all of which are present to varying degrees. Analysis of such movement is of great value in many connections. One of which is the problem of forecasting future movements. It should thus come as no surprise that many industries, organizations and government agencies are vitally concerned with this important subject.

A problem that statisticians frequently face is to estimate the true time series movement in a time series on the basis of two or more imperfect indicators. The two or more indicators of the two series, each of which contains 'noise' that is the short run irregular movement as well as 'siginal', that is the responsiveness to the true series can generally be combined to give a better estimates of the true series than one indicator alone. The best way to combine indicators should depend on the relative strength of the signal content and the relative variance of their noise content.

The approach has some element in common with the work of Goldberger(1974) on "latent" or unobservable variables and work of stone (1977) on accounting system with measurement errors.

To circumvent this, Wald (1940) proposed to split up the observed sample into two sub-sample so as to apply the geometric principle of generalized two-point form for consistent estimation of the parameter (under the assumption of no correlation between measurement errors and between the values of the individual error series).

Wald stated that the key rule of this method was to make the sub- division independent of the errors. The idea amounted to solving the estimation problem by carefully creating new variable to help remove the dependence of the original regressors on the error but Wald did not seem to recognize the generality of his method. The formula derived here, however has no close parallel in any of this other work.

Indeed, the study describes a procedure for estimating truechange in a time series as a linear combination of two indicators with weight for indicators chosen so as to minimize error. An example used to analyse the procedure is, causes for a caesarian section as indicated (a) by number of conception sustained by woman (parity) and (b) by length of stay in hospitals

There are other examples not used in the work but susceptible to the same analytical treatments. For example, non agricultural employment as indicated by monthly survivor of households and by a monthly of employers; price paid for capital goods as indicated by official index based on whole sale prices of specific items of equipment and on construction cost.

1.1 Objective of the study

This research among other things examines the following objectives:

(a) To estimate a true movement in time series on the basis of two imperfect indicators.
(b) To choose weights for the two indicators so as to minimize errors. Note thatW_x and W_y are weights associated with indicators X and Y respectively.

2. Methods of study

The central idea is that the appropriate weights depend on being able to distinguish signal from noise in each indicator. The problem in applying this idea is finding a way to measure empirically these components of each indicator. To do so we begin by postulating a relation between one indicator, X and true series, T of the form;

$$X_i = a_x + b_x T_i + e_{xi} \text{...}(1)$$

Where i isa time subscript, a_x is a constant, b_x is assumed to be positive which measures the strength of the signal component of X_i(the actual signal component of X_i is $b_x T_i$) and e_{xi} assumed to have zero mean and to be uncorrelated with T. If we could estimate observe T_i directly, then we could estimate $a_x$and b_x by ordinary regression procedure and derive estimate of e_{xi} by subtraction. Of course if we could observe T_i there would be no need to use indicators. Not being able to observe T_i, we need some indirect way to measure the signal and noise components of X_i. The same problem applies to the relation between second indicator, Y and T.

$$Y_i = a_y + b_y T_i + e_{yi} \text{ ....} (2)$$

We can differentiate between noise and signal components by measuring the standard deviation of X and Y, and coefficient of b_x and b_y, so as to derive an appropriate weights for the two indicators. Since the coefficients of $b_x$and b_y are not directly observable, we can make inference about ratio of b_y to b_x

From (i)

$$T_i = \frac{X_i - a_x - e_{xi}}{b_x}$$

Substitute for T_i in (2)

$$Y_i = \left[a_y + b_y\left(\frac{X_i - a_x - e_{xi}}{b_x}\right) + e_{yi}\right] \text{ .... } (3)$$

Where the ratio of the b's is the coefficients rating of the observable variables X and Y. This cannot be estimated consistently by ordinary least squares since one component of the composite error term is not independent of the explanatory variable X.

In order to avoid the ordinary least squares bias, the procedure we propose for estimating equation 3 involves grouping data into high and low values so that the error terms have high probability of cancelling out, and then estimating $\frac{b_y}{b_x}$, from the group data. In practice, this technique will give different results depending on whether the grouping is based on sorting by values of X or Y, or by some weighted average of the two.

Once we know the ratio of $\frac{b_y}{b_x}$, then any information or assumption about the value of one of the two coefficients enable us to derive the other one. We do not recommend an econometrics techniques at this point; instead we suggest using the detailed knowledge about a series, its benchmark values, its known bias or other information to make judgment about judgement b_x and b_y, the same is true of the constant terms a_x and a_y, a judgement about one constant, in combination with other information needed to apply our procedure enables us to derive the constants.

Three techniques that have been proposed to avoid the ordinary least square bias in equation with errors-in-variable are:

(a) Weighted regression
(b) Instrumental variable
(c) Grouping method (Malinvaud 1974.Ch. 10).
(d) Weigthed regression has the serious drawback of requiring knowledge of the convariance matrix of errors. Instrumental variables the difficulty of finding appropriate to use as instrument. We propose to use a grouping method, which has the advantage of ease of calculation and consistency under a wide set of assumptions.

The grouping technique was first suggested by Wald(1940). Wald's proposal was to separate the observation of X_i and Y_i into group those for which X_i is above its median.

Bartlett (1949) proposed forming three groups instead of two, on the grounds that the greater differences between mean would improve the efficiency of estimation.

The technique we propose is to separate the observation into three subgroups –high, middle and low, and the estimate of$\frac{b_y}{b_x}$ on the means for the high, and low subgroups. Specifically,

$$\frac{b_y}{b_x} = \frac{Y_H - Y_l}{X_H - X_L} \qquad (4)$$

Where

Y_H is means of Y_i for the observation classified as high

Y_L is the means of Y_I for the observation classified as low

X_H is the means of X_i for the observation classified as high

X_L is the mean of X_i for the observation classified as low

The idea behind this technique is that mean values of e_{xi} and e_{yi} for groups of observation will tend to show less variation than individual values of e_{xi} ande_{yi}hence e_x and e_y will have less influence on parameter when the estimates are based on group means than when they are based on individual observation. In case when the number of observation is small, it may be desirable to divide the observation into two groups instead of three.

Separating observation into high, middle and low subgroups however raises its own statistical difficulties. If values of X_i are the basis for grouping, the resulting estimate of $\frac{b_y}{b_x}$ will in generalnot be the same as if value of y_i are the basis for grouping. Using some weighted combination of x_i will generally give still a different estimate. Selecting a basis for grouping can be as important as selecting an estimation technique.

The goal of selecting a basis for grouping is to separate the observation into high, middle and low groups based on the values of the unobserved true series,T_t. this goal requires (Malinvaud, 1976) assumptions that (a) the observations are classified into groups independent of e_{xi} and e_{yi} (or else the sample average ofe_{xi} or e_{yi} fails to approach zero as the number of observation increases).(b) the subgroup mean do not go to the same limit (or else a line could not be fitted).

Wold suggestions of grouping by X does not satisfy (a), unless the values of e_{xi} are very small; the high subgroups will tend to over-represent low values of e_{xi} and the low subgroup will tend to under represent the values of X . This work suggested an alternative method of grouping based on weighted combination of X_i and Y_i

2.1 Procedure

The method suggested here is an iterative procedure in which successive iteration approach the goal of grouping observation according to the best estimate of $T_{i,}$ the true value of series. Base on this, we propose the following steps:

(1) Assume initially that $b_x = b_y$, so that the ratio $\frac{b_y}{b_x}$ =1.0 it implies that weight associated with indicator $Y(W_y)$ is negative ($W_y<0$).

If by assumption $W_y<0$ implies $\frac{b_y}{b_x} < {S_y}/{S_x}$ then, $\frac{b_y}{b_x} = \ell_{xy}\,{S_y}/{S_x}$ rather than 1.0

(2) Utilizing the result of step (1) and judgement that b_x be set at a constant value, then calculate b_y .
(3) Utilizing the result of step (2) and value of S_x, S_y and ℓ_{xy} directly observable from data, to derive estimate of W_x and W_y
(4) Calculate $\widehat{T_i} = W_xX_i + W_yY_i$ which is the required estimated value, T_i less a constant term.
(5) Use the result of step(4) to group the observation into high, middle and low subgroup (this is equivalent to grouping base on $T_{i,}$ since the missing constant term will not affect the rank of observation).
(6) Make a new estimate of $\frac{b_y}{b_x}$ from grouped mean, using equation 4
(7) Go back to step (2) using the new estimate of $\frac{b_y}{b_x}$ in place of the previous estimates
(8) The iteration continues until the estimate W_x and W_y in step (3) converge to stable values. However, it has not been proved that the procedure always converges. The only proposition about convergence that have been able to prove is that, if the procedure does converge, then the final estimate of $\frac{b_y}{b_x}$, will be the same regardless of initial choice of setting b_x values this is because the value of $\frac{b_y}{b_x}$ serves merely as a scaling factor in the procedure.

2.2 General method for forecast

After the transformation of the two indicators, the estimate of the true series T,based on indicators X_i and Y_i

$T_i = W_x X_i + W_y Y_i + C$ ……………….. ………………………………………………………………… (5)

Where

T_i is the estimate of T_i (true time series)

W_x and W_y are the fixed weights

C is a constant term

2.3 Derivation of weighting formula

The derivation of the weighting formula is divided into the following parts:

(i) Statement of the underlying assumption about the relation of the indicator series to the true series.
(ii) Derivation of a weighting formula that minimize the expected values of the squared difference between the estimated and the true series
(iii) The process of converting the formula into one expression in term of observable variables.
(iv) Determine how much reduction of errors can be expected from using the weighting formula, and proposes a method of estimating upper and lower limits to the reduction.

From equations (1), (2), (3) and (4), W_x and W_y were derived and given as follows:

$$W_x = \frac{\left[\frac{b_x}{b_y}\right]\left[\frac{S_y}{S_x}\right] - \ell_{xy}}{b_x\left\{\left[\frac{b_x}{b_y}\right]\left[\frac{S_y}{S_x}\right] + \left[\frac{b_y}{b_x}\right]\left[\frac{S_x}{S_y}\right] - 2\ell_{xy}\right\}} \dots\dots (6)$$

$$W_y = \frac{\left[\frac{b_y}{b_x}\right]\left[\frac{S_x}{S_y}\right] - \ell_{xy}}{b_y\left\{\left[\frac{b_y}{b_x}\right]\left[\frac{S_x}{S_y}\right] + \left[\frac{b_x}{b_y}\right]\left[\frac{S_y}{S_x}\right] - 2\ell_{xy}\right\}} \dots\dots (7)$$

Where

$$\ell_{xy} = \frac{S_{xy}}{S_x S_y}$$

Note that $W_x + W_y \neq 1$. The idea is to understand how the elements of the formula for W_x influence the weight for indicator X

Case 1: Role of coefficients S_x and S_y if $b_x = b_y = 1$ and $\ell_{xy} = 0$, then the formula reduces to

$$W_x = \frac{\left[\frac{S_y}{S_x}\right]}{\left\{\left[\frac{S_y}{S_x}\right] + \left[\frac{b_x}{b_y}\right]\right\}} \dots\dots (8)$$

Thus,

(a) The common sense of these relationship is that the more variable an indicator is the less reliable a guide is to movement in the true series
(b) If the standard deviation of X is high, then weight of X becomes low and vice-versa
(c) The short run variability in an indicator clearly implies more noise than a stronger signal.

Case 2: Role of coefficients b_x and b_y; If $S_x = S_y = 0$ and $\ell_{xy} = 0$

Then, the formula reduces to

$$W_x = \frac{\left[\frac{S_y}{S_x}\right]}{b_x\left\{\left[\frac{S_y}{S_x}\right] + \left[\frac{b_x}{b_y}\right]\right\}} \dots\dots (9)$$

Thus, the relationship between W_x and b_x reflect that the higher value of b_x given the standard deviation (S_x), implies

(a) W_x is increased relative to W_y
(b) The average of W_x and W_y are reduced, since otherwise their weight sum would be over responsive to the movement in the true series

Case 3: Role of ℓ_{xy}, the response of W_x to increase in ℓ_{xy} is ambiguous; it is positive if the ratio

$\frac{W_x}{W_y}$ *exceed the ratio* $\frac{b_x}{b_y}$ and negative in the opposite case.

The final step in the procedure is estimating the constant term in (5). The formula for the constant term becomes

$$C = W_x a_x - W_y a_y \dots\dots (10)$$

Where a_x *and* a_y are the constants of equation (1) and (2) respectively.

To estimate C, we need an independent judgment about a_x or a_y. This will be based on the same sources; benchmark data and detail knowledge about the indicators, as the judgment element in estimating the weights. Once one the a's has been specified, the order can be estimated by calculating means (overall observation for each term in (3))

$$\bar{Y} = a_y - \left[\frac{b_y}{b_x}\right] a_x + \left[\frac{b_y}{b_x}\right] \bar{X}$$

Where $\bar{Y}$ *and* $\bar{X}$ are the mean of X and Y, b_x and b_y have been estimated earlier.

3. Data Analysis

The data used for this study were collected from the maternity wing of University of Ilorin Teaching Hospital which covers the period between 2007 – 2011. The data collected includes number of previous conceptions sustained by mothers (Parity = X)and length of Stay in the hospital after delivery (Y) through Caesarean section. The two variable X and Y serve as the indicators

3.1 Transformation of data

The starting point of this procedure after assuming linear relationship between each of the two indicators and the true time series as in equation (1) and (2) is to eliminate all forms of pattern of variations by smoothening the raw data with the help of moving average. Table 1 and 2 shows observed values divided by the trend values left after the use of moving average of order twelve.

Table 1: Observed Values Divided By its Trend Values for Parity

Year	1	2	3	4	5	6	7	8	9	10	11	12
2007	-	-	-	-	-	-	1.96	0.00	1.92	2.71	0.90	2.30
2008	0.45	0.00	0.00	1.38	0.00	1.17	1.95	2.92	0.55	1.04	0.49	0.00
2009	1.12	1.95	0.00	2.55	0.48	0.00	0.00	0.80	2.58	1.11	1.16	1.50
2010	0.36	1.74	1.11	0.84	0.45	0.96	0.48	1.47	1.11	0.00	0.92	0.89
2011	1.75	0.00	0.78	2.40	1.24	0.89	-	-	-	-	-	-

Table 2: Observed Values Divided By its Trend Values for Length of Stay

Year	1	2	3	4	5	6	7	8	9	10	11	12
2007	-	-	-	-	-	-	0.93	0.86	2.64	0.78	0.73	0.66
2008	0.70	0.64	1.37	2.76	0.69	0.74	0.08	0.82	0.84	1.02	2.51	1.00
2009	0.77	0.92	0.64	0.92	0.89	0.86	1.30	1.10	1.37	1.04	0.92	0.77
2010	0.95	0.77	0.24	1.40	1.13	2.26	1.04	0.45	0.32	0.31	1.80	1.00
2011	1.54	1.16	0.88	1.05	0.81	0.80	-	-	-	-	-	-

Using the assumption that b_x=0.75 and initial assumption that $\frac{b_y}{b_x}$ = 1, and the weighted values, we then compute $W_xX_i + W_yY_i$, which is equal to the estimated true T_i, less a constant term. Remember the X_i and Y_i are adjusted seasonal variation as shown in table 1 and 2

Table 3: Computation of True Series with their Ranks

$\hat{T_i} = W_xX_i + W_yY_i$	Rank	X_i	Y_i
T_1 =1.2611	L	0.88	0.97
T_2 = 1.1397	L	0.88	0.85
T_3 = 0.9422	L	0.45	0.76
T_4 = 2.0479	H	1.70	1.49
T_5 = 1.0321	L	0.51	0.86
T_6 = 1.3821	L	0.72	1.14
T_7 = 1.1634	L	1.05	0.82
T_8 = 1.1935	L	1.24	0.79
T_9 = 1.7421	H	1.47	1.26
T_{10} = 1.1479	L	1.16	0.77
T_{11} = 1.7307	H	0.83	1.45
T_{12} = 1.2028	L	1.11	0.84

Thus the means for high and low ranked observations are as follows

X_H = 1.3333, X_L = 0.8889

$Y_H = 1.4000$ $Y_L = 0.8667$

Therefore,

$\frac{b_y}{b_x} = \frac{1.4000-0.8667}{1.3333-0.8889} = 1.2000$and initial setting of $b_x = 0.75$

$b_y = 1.2000 \times 0.75 = 0.9000$.

The above coefficient of b's and the usual values of S_x, S_y and ℓ_{xy} was then used to calculate corresponding weights for second iteration. The table 4 below shows Summary of Iteration for Causes of Caesarean Section.

Table 4: Summary of Iteration for Causes of Caesarean Section

Iteration	$\frac{b_y}{b_x}$	b_y	b_x	W_y	W_x	C
1	1.0	0.75	0.75	1.0115	0.3181	0.7856
2	1.2	0.9	0.75	0.9677	0.1720	-0.0287
3	1.2	0.9	0.75	0.9677	0.1720	-0.0287

It can be observed that the iteration converges at the third iterations. Indicating the fact that the final estimate of $\frac{b_y}{b_x}$are the same regardless of the initial choice of b_x.

4. OBSERVATION AND CONCLUSION

As discussed earlier that, time series analysis helps in forecasting future movement or change in pattern series. We obtained the best estimate for change in number of caesarean section as follows:

$$\hat{T} = 0.1720X + 0.9677Y - 0.0287$$

Subsequently, we observed that; the length of stay in hospital series received a weight of about 462% above the weight of parity series. The sum of the two weights exceeds 1.0. The final estimate of a is positive and b_y, the responsiveness of the length of stay in hospital after delivery to the true change in number of caesarean section is 0.9, which agreed with the value close to 1.0 that was suggested earlier in this work.

The application of method of estimating lower and upper limits to the reduction of error expected from using the proposed procedure leads to reduction about 6% error per-month, and a reduction of about 27% per month compared to both parity and length of stay in hospital indicator alone respectively.

The conclusion is based on the resulting evidence from the iterative procedure in which the successive iteration approach the goal of grouping observation according to the best estimate of T (true unobserved series)

(i) The statistical operation shows that the proposed procedure was very useful in estimating a true change in time series (unobserved) on the basis of two imperfect indicators (observed)
(ii) The use of proposed procedure in this manner, has also provide the estimate of a to be positive, suggesting the fact that some sources of true increase in mother with no previous conception has not been source of bias and indeed related to the average increase in number of caesarean section.
(iii) The estimate 0.9 for b suggests that estimate based on the length of stay in hospital after delivery tends to represent true increase in number of caesarean section.

REFERENCES

1. Barlett, M.S. (1947): "Fitting a straight-line when Both variables Are subject to Error". Econometrical 5, 207-212.
2. Goldberger, A.S. (1974): "Unobservable variables in Econometrics," in frontiers Econometrics, ed. Paul Zarembka, New York: Academic Press, 193-213.
3. Malinvaud E. (1974): Statistical Methods of Econometrics, Chicago: Rand McNally.
4. Murray, R.S (1972), Schaum's Outline of Theory and Problems of Statistics. Second Edition in S.I. units

5. Stone, R. (1977): Forward to G. Pyatt and A. Roe, Social Accounting for Development Planning with Special References to Sri Lanka, Cambridge: Cambridge University Press.
6. Theil, H (1971): Principle of Econometrics, New York.
7. Wald, A. (1940): "Fitting of straight lines if Both Variables are subject to Error". Annals of Mathematical Statistics, 11, 284-300.
8. Yode, O.(1990): College And University Text: Statistics. NPS Educational.