

 <p>ISSN NO. 2320-5407</p>	<p>Journal Homepage: -www.journalijar.com</p> <p>INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)</p> <p>Article DOI:10.21474/IJAR01/8887 DOI URL: http://dx.doi.org/10.21474/IJAR01/8887</p>	 <p>INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR) ISSN 2320-5407 Journal Homepage: http://www.journalijar.com Journal DOI:10.21474/IJAR01</p>
---	---	---

RESEARCH ARTICLE

AN ALGORITHM TO FIND SIGNIFICANT COMMODITIES IN STOCHASTIC LASPEYRES REGRESSION MODEL.

Arfa Maqsood¹, S. M. Aqil burney² and Suboohi Safdar¹.

1. Department of statistics, university of Karachi, Karachi.
2. Institute of business management, Karachi.

Manuscript Info

Manuscript History

Received: 11 February 2019

Final Accepted: 13 March 2019

Published: April 2019

Key words:-

Influential Commodities, Laspeyres Index Numbers, Serial Correlation, Autoregressive Process, Hat matrix, DFBETA.

Abstract

The main objective of this paper is to develop a step by step procedure to discover the influential commodities in stochastic Laspeyres regression model when the errors are assumed to be serially correlated with autoregressive process of order p. The two familiar methods of finding unusual observations the 'Hat matrix' and difference in parameter vector beta i.e. 'DFBETA' are considered in this algorithm. This algorithm facilitates the researcher to carry out their research in convenient way.

Copy Right, IJAR, 2019,. All rights reserved.

Introduction:-

The index number is one of the monetary measures that help us to understand the average change in price level over the two periods. Index number is of more reliable estimate if the relative importance to commodities is given instead of providing equal importance to each commodity in consumer basket. For example, basic food items like wheat, rice, and milk will always be more essential commodities in consumer's budget than the entertainment or education. Unequal importance of each consumer items as shown by its corresponding weights. For instance, food items are certainly given more weights as opposed to the entertainment and recreation. Similarly, wheat, rice, and other basic food items are undoubtedly offered greater weights as compare to non-basic food items such as soft drinks, coffee, milo, vinegar etc.

The prices corresponding to basic items should seem to be kept as low as possible and remain unchanged over the periods so that each individual in an economy could not be highly affected to fulfill their basic needs. For such reason, government have to pay attention not to change the price level with great margin via tight monetary policy and other fiscal measurements. They must carry some essential and critical steps to prevent the prices goes higher by the retailers on their own and intervene to overcome such a situation prevalent within a country. The price level could also be taken under control using the actions taken in the light of correct picture depicted by the price analysis and other fiscal measures.

The problem arises here is that among various consumer items, which one is more important to estimate price index numbers and thus price inflation. In other words, which commodities are more sensitive or influential to regression parameter estimates of index number model. For this purpose, we use the two methods Hat matrix, and DFBETA measures to find such items in the basket. Both detecting methods suggest to compare with calibration point in order to decide whether a particular commodity fall in influence zone.

Corresponding Author:-Arfa Maqsood.

Address:-Department of statistics, university of karachi, karachi.

Basically, the detecting methods have been proposed to find leverages and influential observations in regression analysis with one regressor and more than one regressors. The detailed description on unusual observations are given by Barnett and Lewis (1994). Belsley et al. (1980), and Chatterjee and Hadi (1986) classified the types of such abnormal observations with respect to regression analysis of two variables X and Y , via the help of graphical tools on two dimensional plots. The role of hat matrix (H) with reference to both regression analysis and analysis of variance (ANOVA), using a calibration point $2p/N$ is given by Hoaglin and Welsch (1978). They found the diagonal elements h_{ii} , a very worthwhile diagnostic towards detecting leverages in multiple regression. The properties of hat matrix are described and used by many authors like Hoaglin and Welsch (1978), Cook and Weisberg (1982), and Draper and Smith (1998). The chief role of hat matrix is in finding the leverages in simple regression as well as multiple regression analysis, Hoaglin and Welsch (1978) found hat values as a good indicator to detect the leverages. The larger the value of h_{ii} , the more significant value of regressor to estimate y value. We use the hat values in order to find the significant commodities in Laspeyres index numbers.

One useful measure of change is difference in regression coefficients β (DFBETA), when calculated on the basis of full sample and reduced sample. Thus, defined as $\hat{\beta} - \hat{\beta}_{(i)}$ (see Miller (1974) for the derivation of DFBETA formula). A scaled version of DFBETA is proposed by Welsch and Kuh (1977), which is obtained by dividing $\hat{\beta} - \hat{\beta}_{(i)}$ to the standard error of $\hat{\beta}$. The values of DFBETAs are compared to the size of sample N and if these are greater than $2/\sqrt{N}$, the corresponding observations are considered as unusual. Özkale and Açar (2015) used these influence measures to find unusual observations in linear regression model with more than one regressors. Maqsood and Burney (2014) and Burney and Maqsood (2014) used the technique of hat matrix and DFBETA measure to find the influential commodities in Laspeyres index number model with autocorrelated errors.

The organization of the paper is as follows. Section 2 provides the stochastic Laspeyres regression model with brief description of how the Laspeyres index number estimators are obtained. The formulae of hat values and DFBETA values are also recalled in this section. An algorithm to find significant commodities is presented in section 3 and well displayed by flow chart. Lastly, section 4 gives the conclusion.

Stochastic Laspeyres Regression Model

The stochastic simple model of Laspeyres price index number is defined as follows;

$$P_{it}^o = \alpha_t + \varepsilon_{it} \quad i=1, \dots, n, \text{ and } t=1, \dots, T \quad (1)$$

Where $P_{it}^o = \frac{P_{it}}{P_{io}}$, ratio of current period price to the base period price for i th commodity, α_t common trend in the prices of all commodities at time t , and ε_{it} is the random component. The errors ε_{it} are assumed to be serially correlated and thus generated from an autoregressive process of order p .

$$\varepsilon_{it} = \phi_1 \varepsilon_{i,t-1} + \phi_2 \varepsilon_{i,t-2} + \dots + \phi_p \varepsilon_{i,t-p} + u_{it} \quad (2)$$

Assuming

$$E(u_{it}) = 0 \text{ and } E(u_{it}u_{jt}) = \frac{\sigma^2}{w_i} \delta_{ij} \quad (3)$$

Where $\phi_1, \phi_2, \dots, \phi_p$ are the parameters of autoregressive process, w_i are the weights of i th commodity and δ_{ij} is kronecker delta which takes the value one for $i=j$, and zero otherwise. For an stationary autoregressive process the roots of autoregressive polynomial must lie outside the unit circle.

The matrix formulation of model (1) is given as

$$P^o = X\gamma + \varepsilon \quad (4)$$

Where both P^o and ε are respectively $(nT \times 1)$ vectors of the observed Laspeyres index number and the error terms. The parameter vector γ consist of Laspeyres index number α_t so with the order $(T \times 1)$, and the design matrix X with order $(nT \times T)$. Assuming $E(\varepsilon) = 0$, and $E(\varepsilon\varepsilon') = \sigma^2 V$, where V is symmetric and positive definite in nature. The inverse of square matrix V can be decomposed using choleski decomposition to get $V^{-1} = Q'Q$,

where Q is a lower triangular matrix. Under these assumptions, the best linear unbiased estimator (BLUE) of γ in model (4) can be obtained using the method of generalized least square (GLS) as given below

$$\hat{\gamma} = (X'V^{-1}X)^{-1}(X'V^{-1}P^o)$$

The transformed model is obtained by multiplying both sides of equation (4) by Q , and then we apply the simple ordinary least square (OLS) estimator to the transformed data to obtain estimated generalized least square (EGLS). We have

$$\hat{\gamma} = (X^{*'}X^*)^{-1}(X^{*'}P^{o*}) \quad (5)$$

Where $X^{*'} = \hat{Q}X$, and $P^{o*} = \hat{Q}P^o$. The transformation matrix Q is obtained by Burney and Maqsood (2014) when the errors are generated from autoregressive process of order p with $p=1$, and by Maqsood and Burney (2014) for $p=2$. With these assumptions Maqsood and Burney (2014) obtained the estimator of γ , the familiar Laspeyres index number, written as

$$\hat{\alpha}_t = \sum_{i=1}^n w_i P_{it}^o \quad \text{for } t = 1, 2, \dots, T \quad (6)$$

The order of autoregressive process does not have an impact on estimator of Laspeyres price index number. We have confirmed for $p=1$ and $p=2$ and get the same formula for Laspeyres as given in equation (6). The standard errors of estimated Laspeyres index numbers are derived by Maqsood and Burney (2017).

To find the influential observations and its impact on Laspeyres regression model we consider the two familiar influence measures hat matrix and difference in parameter vector beta (DFBETA). For this purpose, we find the hat matrix for transformed data using the equation given below

$$H = X(X'X)^{-1}X' \quad (7)$$

We get

$$h_{it,it} = w_i, i = 1, \dots, n \quad (8)$$

We use the subscript of hat values 'it,it' due to a matrix of order $nT \times nT$, where $nT=N$ are the total number of observations. The diagonal elements of matrix i.e. $h_{it,it} = w_i, i = 1, \dots, n$ clearly show that the weights of commodities determine how much the important of particular commodity is in order to find the Laspeyres index number. They are not affected by the parameter of autoregressive process. The greater the value of weight, the more influential the commodity is, irrespective of the time period. They are not affected by the parameter of autoregressive process.

Another measure to determine the influential observations is DFBETA, given by the following formula

$$DFBETA_{itj} = D_{itj} = \hat{\beta} - \hat{\beta}_{(i)} = \frac{(X^{*'}X^*)^{-1}x_i^*e_i^*}{1 - h_{it,it}} \quad (9)$$

We use D for DFBETA for the sake of convenience and subscript ' itj ' is used to describe ' i ' for commodities, ' t ' for time period, and ' j ' for parameters. DFBETA depends on the order of autoregressive process and a different estimator is obtained for $p=1$ and $p=2$. The DFBETA estimator is obtained by Burney and Maqsood (2014) for AR(1) error process and by Maqsood and Burney (2014) for AR(2) process. Burney and Maqsood (2014) derived the DFBETA estimator for Divisia index number which is same as we get for Laspeyres index number. The estimators and their explanations are given in these references.

Algorithm to Find Significant Commodities

To achieve our objective, we must pursue a step by step procedure to discover the influential commodities in Laspeyres index number. For the sake of convenience, we convince the reader to carry out an algorithm for employing the methods described in sections 2.

Figure 1 displays the algorithm using flow chart, in which the two phases are recommended to accomplish. Firstly, the first phase includes the parameter estimation of Laspeyres regression model, finding error series, verifying the stationary scenario of series, and fitting of an AR(p) model to error series. Then the second phase certainly comprises of utilizing methods to extract the influential commodities. The influential commodities in the sense, that it disturbs the regression estimates of index numbers and its related aspects. We use here the two widely used methods the hat matrix, and the DFBETA matrix, and the decision is made on calibration point that have been described in flow chart.

The hat values alone can not provide the accurate picture of the pattern that commodities in the consumer basket reflect. It may mislead the analyst to wrong and incorrect situation. Because the commodities, for which we are computing the higher hat values, do not necessarily have a big DFBETA value too. To clear that, we must start to establish the following hypothesis;

H₀:

ith commodity is not influential to estimate the underlying index number

H_A:

ith commodity is influential to estimate the underlying index number

To test this hypothesis, we work on hat values and DFBETA measures and take decision according to the figure 2. There are two possibilities based on the values of N, the total number of observations in analysis, and p, the number of parameters. We show these using two horizontal scales, which represents either $p < \sqrt{N}$, or $p < \sqrt{N}$. The first portion is acceptance zone, where both $h_{it,it}$ and D_{itj} are less than their respective threshold values, and we do not reject null hypothesis. The middle part is called the suspicious region where one of these two detecting methods exceeds its critical value and other does not. We, therefore recommend confirming the decision by computing other detecting value. On the contrary, the last portion is the rejection region, where we have strong evidence against null hypothesis in favor of alternative hypothesis. We also show these situations by the table 1.

The first inner cell of table 1 represents the situation shown by the first portion in figure 2, where both criteria take values less than calibration point and hence not influential case. The next two cells show the middle part of figure 2 i.e. suspicious zone. Similarly, the last cell confirms about the significant influence of respective commodity falling in rejection region. The analysis done by Maqsood and Burney (2014) is actually an example of determining influential commodities in Laspeyres index model using this algorithm. The first phase of computation requires the estimation of parameter vector based on observed price data. While in the second phase the hat values and DFBETA values are computed using the formulae given in equations (7) and (9). The results on Laspeyres index estimates and estimates of influential measures are presented in Maqsood and Burney (2014).

Table 1:-Decision Criteria for Hat Values and DFBETA Measures

		Hat Values	
		$h_{it,it} < 2p/N$	$h_{it,it} > 2p/N$
DFBETA values	$D_{itj} < 2/\sqrt{N}$	Not Influential	Suspecting Influential
	$D_{itj} > 2/\sqrt{N}$	Suspecting Influential	Influential

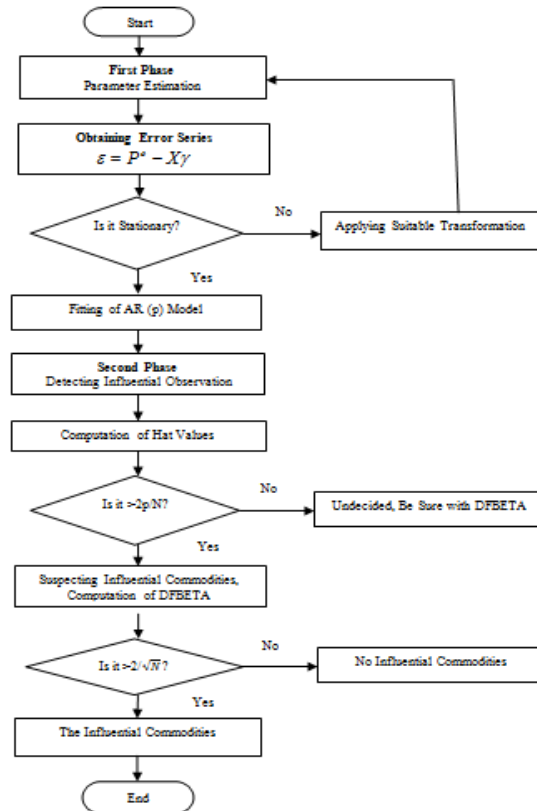


Figure 1:-Flow Chart to Find Significant Commodities in Stochastic Laspeyre Price Model

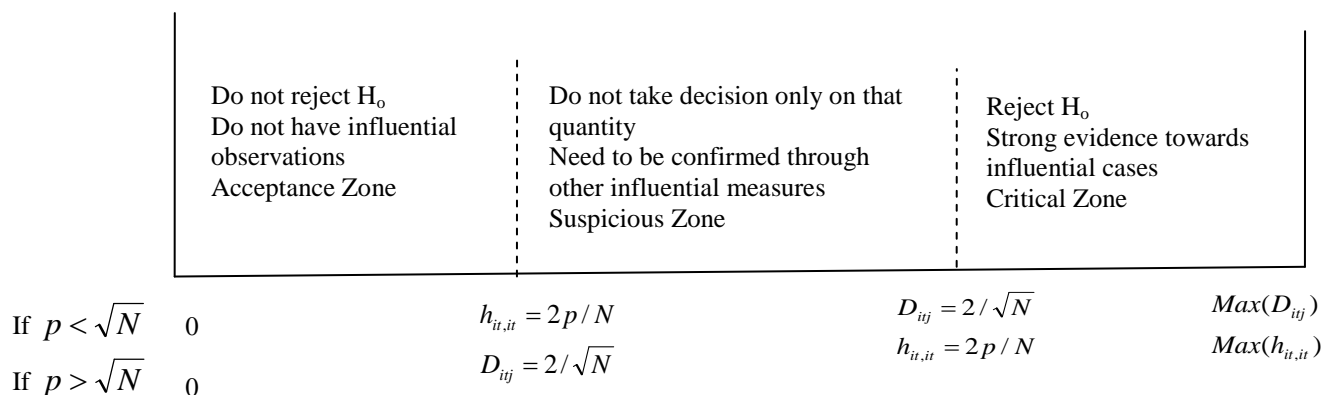


Figure 2:-Different Zones with Respect to Values of Hat Matrix and DFBETA Matrix

Conclusion:-

In this paper, we considered the general form of hat matrix and DFBETA measure to detect the influential commodities in estimating the stochastic Laspeyres index number when the errors are serially correlated with AR(p) process. For this purpose, we defined a step by step procedure consisting of two phases. The first phase computes the estimates of Laspeyres index numbers. The second phase helps to determine the influential commodities using hat values and DFBETA values. These two phases are well explained by flow chart. A hypothesis is generated for a specific commodity whether has significant impact on estimated index number in section 3. To check this hypothesis a figure representing the three zones of acceptance, suspicious, and critical is shown. The decision is made using the given calibration points for both hat values and DFBETA measures. One can reach a decision easily by looking the

different scenarios in table 1 for acceptance or rejection of the null hypothesis. This proposed algorithm is useful for the researchers working on influential commodities using these measures.

References:-

1. Barnett, V. D. and Lewis, T. (1994), Outliers in Statistical Data, 3rd Edition, New York: Wiley.
2. Belsley, P. A., Kuh, E. and Welsch, R. E. (1980), Regression Diagnostics, New York, John Wiley.
3. Chatterjee, S., and Hadi, A. S. (1986), Influential Observations, High-Leverage Points, and Outliers in Linear Regression, Statistical Science, 1(3), pp. 379-416.
4. Hoaglin, D.C. and Welsch, R. E. (1978), The Hat Matrix in Regression and ANOVA, The American Statistician, 32, pp. 17-22.
5. Cook, R. D. and Weisberg, S. (1982), Residuals and Influence in Regression, New York: Chapman and Hall.
6. Draper, N. R. and Smith, H. (1998), Applied Regression Analysis, Third Ed., New York: John Wiley.
7. Miller, R. G. (1974), An Unbalanced Jackknife, Annals Statistics, 2, pp. 880-891.
8. Burney, S. M. A. and Maqsood, A. (2014). Influential Observations in Stochastic Model of Divisia Index Numbers with AR(1) Errors, Applied Mathematics, 5(6), 975-982.
9. Maqsood, A. and Burney, S. M. A. (2014). Extracting the Influential Commodities in Stochastic Model of Simple Laspeyre Price Index Numbers with AR(2) Errors, Open Journal of Statistics, 4(3), 220-229.
10. Maqsood, A. and Burney, S. M. A. (2017). Standard Errors for the Laspeyre Index Number with Autocorrelated Error Models, Communication in Statistics: Theory and Methods, 46(21), 10607-10616.
11. Özkale, M. R. and Açar, T. S. (2015). Leverages and Influential Observations in a Regression Model with Autocorrelated Errors, Communications in Statistics -Theory and Methods, 44(11), 2267-2290.
12. Welsch, R. E. and Kuh, E. (1977), Linear Regression Diagnostics, Technical Report, Sloan School of Management, Massachusetts Institute of Technology, pp. 923-977.