Influential Commodities Using Hat Values in Stochastic Laspeyres Price Model with AR(1) Errors

Arfa Maqsood[1*], S. M. Aqil Burney[2], Tahseen Jilani[3] Suboohi Safdar[1]

[1] Department of Statistics, University of Karachi, Karachi.
[2]Institute of Business Management, Karachi.
[3] Department of Computer Science, University of Karachi, Karachi.

Abstract
This article considers the two structure of stochastic Laspeyres price model. One is the standard regression model of simple Laspeyres price index. While the other is based on an extended approach to the simple version that incorporates a systematic change in relative prices to the simple model. In both versions, the error structure is first order serial correlation. We use the general form of hat matrix to detect the influential commodities in estimating the Laspeyres index number. The results show that the corresponding weights of consumer items have a larger influence on parameter estimates. The extended version of the Laspeyres model investigates the influential commodities more accurately than the simple one as it depends on both the weights and the parameter of AR(1) process.

Keywords:
Laspeyres index Numbers, Serial Correlation, Autoregressive Process, Hat matrix, Influential Commodities

1. Introduction

Sometimes we are dealing with the observations in regression analyses that greatly disturb the results obtained from the analysis. These observations are called the influential points that may have an impact to the model estimates, fitted values and thus the residuals. For detailed description of outliers in statistical data, see Barnett and Lewis [1]. Influence diagnostics are developed and studied by many authors including Belsley et al. [3], Cook [7], Cook and Weisberg [8], Draper and John [9], and Draper and Smith [10]. They observed the effect of initial observations on the estimation of model parameters. Some of these techniques are based on the values of hat matrix where the diagonal elements h_{ii} are very worthwhile rational towards detecting influential observations. The properties of hat matrix are well discussed by Hoaglin and Welsch [11].
Several studies are available on the detection of influential observations in linear regression model when the errors are serially correlated. The important references in this context include the contribution of Puterman [15], Stemann and Trenkler [17], Barry et al. [2] and Özkale and Açar [14]. Puterman [15] investigated the impact of first transformed observation in linear regression model on the parameter estimates. Stemann and Trenkler [17] extended the approach of Puterman [15] to the regression model with more than one regressor and showed that the effect of the presence of a constant term on a leverage point when the magnitude of error correlation was large. Barry et al. [2] extended the study of influential observations to the regression model with AR(2) errors and developed the diagnostic technique using a hat matrix. Burney and Maqsood [5] used the analytical tools of hat matrix and DFBETA measures to identify the influential observations in estimating the Divisia price index number model with AR(1) errors. Maqsood and Burney [12] extended the approach to the simple Laspeyres price model with AR(2) errors.
The stochastic regression model of index numbers are developed to estimate the indices and many times this modeling helps to detect the influential commodities in consumer basket. Burney and Maqsood [4] estimated the extended version of Paasches price model which is the extension of the work done by Selvanathan [16] on extended version of Laspeyres price index numbers. Their work assumed the errors are uncorrelated. Another significant contribution in this series by Clements and Izan [6] who initially extended the Divisia index to augment the model with a systematic component. In this paper, we consider the two version of Laspeyres price model. First is the simple Laspeyres price model used by Maqsood and Burney [12]. The other is the extended version of simple model that is obtained by incorporating the systematic change in relative prices to the model, which helps in controlling the variation due to using various unequally important commodities. The first objective of this paper is to estimate the parameters of underlying Laspeyres price models. On the other hand, the influential commodities are examined using the hat values. The paper is organized as follows. Section 2.1 and 2.2 introduces respectively the simple Laspeyres price model and extended Laspeyres price model. The parameter estimators and the hat values are obtained for both the versions in their respective sections. An application is presented with reference to the price data of Pakistan in section 3. Section 4 presents the conclusion.

Manuscript received April 5, 2019
Manuscript revised April 20, 2019

2. Stochasic Laspeyres Price Models

2.1. Simple Version

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

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

Where $P_{it}^{o} = \frac{p_{it}}{p_{io}}$, ratio of current period price to the base period price for ith commodity, α_t common trend in the prices of all commodities at time t, and ε_{it} is the random component. The errors are assumed to be generated from the first order autoregressive scheme, that is $\varepsilon_{it} = \phi\varepsilon_{i,t-1} + u_{it}$, where $|\phi| < 1$, and $E(u_{it}) = 0,\ E(u_{it}u_{jt}) = \frac{\sigma^2}{w_i}\delta_{ij}$. This yields the error structure of model (1) as

$E(\varepsilon_{it}) = 0$ and variance-covariance

$$E(\varepsilon_{it}\varepsilon_{i,t-s}) = \begin{cases} \dfrac{\sigma^2}{w_i(1-\phi^2)} & s = 0 \\ \dfrac{\sigma^2\phi^k}{w_i(1-\phi^2)} & s > 0 \end{cases} \quad (2)$$

Defining more compactly in matrix notation as $E(\varepsilon) = 0$, $\mathrm{E}(\varepsilon\varepsilon^{'}) = \sigma^2 V$. Assuming V is known with symmetric and positive definite in nature, then the inverse of V can be decomposed using choleski decomposition to get $V^{-1} = Q^{'}Q$, where Q is a lower triangular matrix and obtained by Burney and Maqsood [5]. This can be expressed more compactly in matrix form as follows

$$P^{o}{}_{(nT\times 1)} = X_{(nT\times T)}\gamma_{(T\times 1)} + \varepsilon_{(nT\times 1)} \quad (3)$$

Where X is an (nT$\times$T) design matrix and γ is the vector of parameters. P^{o} and ε are, respectively (nT $\times$ 1) vectors of the observed Laspeyres index number and the errors It is well known that under the above assumption, the best linear unbiased estimator (BLUE) of γ in model (3) could be obtained by the generalized least square (GLS) approach as given below

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

The transformed model is obtained by multiplying both sides of equation (3) 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

$$\hat{\gamma} = \left(X^{*'}X^{*}\right)^{-1}\left(X^{*'}P^{o^*}\right) \quad (4)$$

The transformed matrix X^* is given in Burney and Maqsood [5], and the price relative vector is given as

$$P^{o^*} = \begin{bmatrix} \sqrt{1-\phi^2}\sqrt{w_i}\,p_{i1}^{o}\iota \\ \sqrt{w_i}\,p_{i2}^{o}\iota - \phi\sqrt{w_i}\,p_{i1}^{o}\iota \\ . \\ . \\ \sqrt{w_i}\,p_{iT}^{o}\iota - \phi\sqrt{w_i}\,p_{i,T-1}^{o}\iota \end{bmatrix} \quad (5)$$

Where
$\sqrt{w_i}p_{it}^{o}\iota =$
$[\sqrt{w_1}p_{1t}^{o} \quad \sqrt{w_2}p_{2t}^{o} \quad . \quad . \quad \sqrt{w_n}p_{nt}^{o}]$
, $\iota = [1 \quad 1 \quad . \quad . \quad 1]'$ and o is the vector of zero i.e. $o = [0 \quad 0 \quad . \quad . \quad 0]'$. Substituting the results in equation (4) provides the estimator of γ, the familiar Laspeyres index number, written as

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

The next step is to have an idea about the presence of influential observations and its impact on Laspeyres regression model. 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

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

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. $h_{it,it} = w_i\ ,\ i = 1,...,n$ 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.

2.2. Extended Version

We extend the simple Laspeyres price model in equation (2) as each price relative P_{it}^o is now written, as the sum of the common trend in all prices, α_t , a systematic component β_i , and a random error term, i.e.

$$P_{it}^o = \alpha_t + \beta_i + \varepsilon_{it} \quad \text{i=1, ..., n, and t=1, . . ., T} \quad (9)$$

Now rearranging equation and taking the mathematical expectation, we obtain the change in ith relative price as $E(\mathrm{P}_{\mathrm{it}}^{\mathrm{o}} - \alpha_t) = \beta_i$. This model is not identified. It can be seen by noting that an increase in α_t for each t by any number k and a lowering of β_i for each i by the same k does not affect the right side of equation (9). To identify the model, we impose the constraint

$$\sum_{i=1}^{n} \overline{w}_i \beta_i = 0 \quad (10)$$

This has the simple interpretation that a budget share weighted average of the systematic component of the relative price changes is 0. We write model (9) with the constraint (10) as follows

$$P_{it}^o = \begin{cases} \alpha_t + \beta_i + \varepsilon_{it} & \mathrm{i = 1,2,\ldots, n-1} \\ \alpha_t + \sum_{j=1}^{n-1} \left(-\dfrac{w_j}{w_n} \right) \beta_j + \varepsilon_{nt} & \mathrm{i = n} \end{cases} \quad (11)$$

The matrix formulation of this model is same as represented by (3). The parameter vector γ now consist of both Laspeyres index number α_t and commodity-specific component β_i , so now with the order $(T+n-1)\times 1$, and the design matrix X with order $nT \times (T+n-1)$.

We assume the error structure of extended model is same as we had for the simple Laspeyres model, described in equation (2). To apply the estimated generalized least squares (EGLS), we first transform the vector P^o and the design matrix X to the new vector $P^{o^*} = \hat{Q}P^o$ (given in equation (5)) and the matrix $X^{*} = \hat{Q}X$ is obtained as

$$\mathrm{X}^* = \left[\begin{pmatrix} \sqrt{1-\phi^2}\sqrt{w_i}\,\iota & o & . & . & o \\ -\phi\sqrt{w_i}\,\iota & \sqrt{w_i}\,\iota & . & . & o \\ . & . & . & . & . \\ . & . & . & . & . \\ o & o & . & . & \sqrt{w_i}\,\iota \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \sqrt{1-\phi^2}\sqrt{w_i}\, I_{(n-1)\times(n-1)} \\ -\sqrt{1-\phi^2}\,\dfrac{w_i}{\sqrt{w_n}}\iota' \end{pmatrix} \\ \begin{pmatrix} (1-\phi)\sqrt{w_i}\, I_{(n-1)\times(n-1)} \\ -(1-\phi)\dfrac{w_i}{\sqrt{w_n}}\iota' \end{pmatrix} \\ . \\ . \\ \begin{pmatrix} (1-\phi)\sqrt{w_i}\, I_{(n-1)\times(n-1)} \\ -(1-\phi)\dfrac{w_i}{\sqrt{w_n}}\iota' \end{pmatrix} \end{pmatrix} \right].$$

Using equation (4), we obtain EGLS estimator of α_t same as in equation (6) and β_i

$$\hat{\beta}_i = \frac{1}{k}\left[\left(1-\phi^2\right)a_{i1} + \left(1-\phi\right)\sum_{t=2}^{T}\left(a_{it} - \phi a_{i,t-1}\right) \right], \ \text{for } i = 1,2,\ldots\ldots, n \quad (12)$$

Where $k = (1-\phi^2) + (T-1)(1-\phi)^2$ and $a_{it} = p_{it}^o - \hat{\alpha}_t$. The estimator of the systematic component of ith relative price change, $\hat{\beta}_i$, is the change in this relative price a_{it} averaged using AR(1) process with parameter ϕ.

The diagonal entries of hat matrix are obtained using equation (7) to the transformed data to get

$$h_{it,it} = \begin{cases} w_i + \dfrac{1-\phi^2}{k}(1-w_i) & t=1, i=1,\ldots,n \\ w_i + \dfrac{(1-\phi)^2}{k}(1-w_i) & t=2,\ldots,T, i=1,\ldots,n \end{cases} \tag{13}$$

We see here in the extended model where we cumulate the component incorporating the effect of particular consumer item, the value of $h_{it,it}$ depends on both the weights and autoregressive parameter. The hat values highly depend on the value of ϕ . We observe the following approximations when

$$\phi \to 0, \quad h_{it,it} = w_i + \frac{1-w_i}{k}, \text{ for all t}$$

$$\phi \to 1, \quad h_{it,it} = w_i, \qquad \text{for all t}$$

$$\phi \to \text{-}1, \begin{cases} h_{it,it} = w_i, & \text{for t} = 1 \\ h_{it,it} = w_i + \dfrac{1-w_i}{k}, & \text{for all t} > 1 \end{cases} \tag{14}$$

3. An Application

We take the monthly price data of Pakistan used by Maqsood and Burney (2014). The data is collected and published by the Pakistan Bureau of Statistics (PBS). The period of data is for ten fiscal years from July 2001 to June 2011, consisting of 374 consumer commodities that are further classified in ten commodity groups. The prices of first year from July 2001 to June 2002 are taken as base year prices and the prices of subsequent months are compared with prices of corresponding month of base year. The first phase of computation requires the estimation of parameter vector based on observed price data. Both versions of price models provide the same estimates of Laspeyres index numbers. These estimates of Laspeyres indices and commodity-specific components for ten groups along withtheir standard errors are presented in Maqsood and Burney [13].

After estimating the parameter vector based on observed price data, the residuals are obtained and plotted against observation numbers in figure 1(a) for simple Laspeyres case and in figure 1(b) for extended Laspeyres case. These graphs show the jumps along the constant central line. These have the longer ups and down as the time proceed far from the base period; however, these exhibit the stationary situation. The residuals parallel to the first thousand observations are comparatively less than the residuals recorded for the last thousands observations. In other words, the index number estimates corresponding to months that are close to base period are much more accurate and hence, with the smaller values of biasdness.

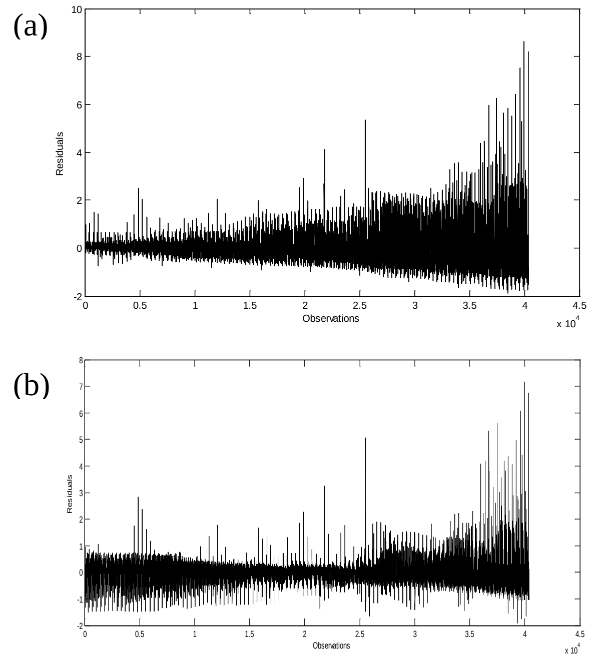

Fig. 1 Plot of residual series from (a) simple Laspeyres price model (b) extended Laspeyres price model

Table 1: Summary of results ADF test value, DW statistic, fitting of AR(1) process to residual series of simple and extended Laspeyres price models

	Simple Version	Extended Version
ADF Test Statistic	-108.718 (0.001)	-115.717 (0.001)
DW Statistic	0.906	0.996
AR parameter (ϕ)	0.5472	0.5021
AIC	-1.7429	-2.4279
RMSE	0.4183	0.2970

Table 2: Significant hat values corresponding to commodities for simple Laspeyres price model (cut-off value= 0.00534)

S.Nos.	Item Nos.	Items	h_i
1	2	Wheat flour fine/superior.	0.0134
2	3	Wheat flour bag	0.0377
3	16	Cooking oil (dalda)	0.0069
4	17	Vegetable ghee tin	0.0126
5	18	Vegetable ghee (loose)	0.0141
6	19	Sugar refined	0.0195
7	25	Tea loose kenya av.qlty 250g	0.0057
8	26	Milk fresh (unboiled)	0.0653
9	70	Toffee (hilal)	0.0101
10	71	Chowkelate candy (small size)	0.0099
11	86	Beef with bone av.qlty.	0.0161
12	87	Mutton av.qlty.	0.0109
13	88	Chicken farm broiler (live)	0.0092
14	91	Onion	0.0058
15	167	House rent index	0.2343
16	176	Elect.charges 301 - 1000 uni	0.0106
17	177	Elect.charges above 1000 uni	0.0274
18	180	Gas chrg 6.7438 - 10.1157mmb	0.0093
19	181	Gas chrg10.1157 - 13.4876mmb	0.0068
20	206	Household servant female p/t	0.0119
21	227	Petrol super	0.0173
22	266	Telephone charges local call	0.0083
23	268	Tel charges out side city	0.0083
24	286	School fee primary eng.med.	0.0081
25	287	School fee 2nd-ry eng.med.	0.0081
26	310	Washing soap nyl(135-160gms)	0.0054
27	333	Haircut charges for men	0.0067
28	374	Doctor (mbbs) clinic fee	0.0100

Table 3: Significant hat values corresponding to commodities for extended Laspeyres price model (cut-off value= 0.00238

S. No.	Item Nos.	Items	Weights	AR(1) Process
				h_i (t>1)
1	3	Wheat flour bag	0.037	0.046
2	19	Sugar refined	0.019	0.028
3	26	Milk fresh (unboiled)	0.065	0.073
4	86	Beef with bone av.qlty.	0.016	0.025
5	167	House rent index	0.234	0.241
6	177	Elect.charges above 1000 unit	0.027	0.036
7	227	Petrol super	0.017	0.026

To check the stationary scenario of residual series we apply Augmented Dickey Fuller (ADF) test. The test statistic values with p values are presented in table 1. The results show the small p values, indicating towards the rejection of null hypothesis of unit root and hence the stationary series. The presence of autocorrelation effect in these residual series is examined by the Durbin Watson (DW) statistics, reported in table 1. The DW statistics are significantly different from the standard value of 2, implying the existence of first order serial correlation in both residual series.

Confirming the serial correlation in the residual series, we estimate the AR(1) models using Yule Walker method. The estimated values of the autoregressive parameters are displayed in table 1. The accuracy measures Akaike information criterion (AIC) and Root mean square error (RMSE) are also computed to see the performance of fitting of an AR(1) process to the residual series. Both AIC and RMSE values for extended version are less than the values for simple version. It might reveal that the residuals from extended Laspeyres model can be well captured more accurately than the residuals from simple model.

The next task is to find the influential commodities to estimate Laspeyres index numbers. For simple Laspeyres price model, table 2 presents the hat values for various items exceeding the threshold value 0.005348. The largest hat value 0.2343 is corresponding to house rent index, implying its importance in estimating the Laspeyres index number. The other leading commodities includes milk fresh with $h_i = 0.0653$, wheat flour bag with $h_i = 0.0377$, and electric charges for the consumption of above thousand units with $h_i = 0.0274$. The same items remain significant for the extended Laspeyres model with AR(1) error process. Apart from these items, sugar refined, beef with bones, and petrol super are proved principal commodities with hat values greater than 0.023817 (see table 3). For extended Laspeyres price model, all 374 consumer commodities fall in significant zone for t=1 and only 7 items found significant for t>1. These seven significant items are listed in table 3.

The extended version of Laspeyres price model filters out with more important commodities those are influential to estimate the underlying index number. It might be due to the reason that the hat values of simple version only depends on the weights of commodities, while of extended version depends not only on weights but also on AR(1) parameter ϕ and k (the function of ϕ). These seven items, named described earlier, are the core commodities alongwith the higher weights. Therefore, to monitor and control the prices of these core items should be of much more concern to the price regulatory authority in order to reduce the impact of price fluctuations particularly for the poor community of country.

4. Conclusion

In this paper, we considered the two versions of stochastic Laspeyres price model. First was the simple standard model and the second was based on the extended approach to the simple version. The extended version was obtained by augmenting the simple model with a systematic component that is responsible to explore the variation in underlying index numbers due to using various unequally important commodities. The estimators of the model parameters are obtained using GLS approach and we found the familiar Laspeyres index number from both versions of the models. Hence, the extended approach does not affect the estimates of Laspeyres index numbers as we got the same values for index number estimates.

The main task of the paper was to explore the influential commodities to estimate Laspeyres index numbers in serially correlated error models. A total of 28 commodities based on their weights were found significant in case of simple model (see table 2). In extended model, we observed all 374 consumer items significant for t=1 and only 7 items for t>1. This is noteworthy finding of this research that the extended version not only improves the method of parameter estimation but also investigates the core commodities more accurately. This is due to its hat values depend on both the weights of commodities and the parameter of AR(1) error process. By this conclusion, it might be possible to improve our results further when we would consider the higher order autoregressive process for the error structure.

References

[1] Barnett VD and Lewis T (1994). Outliers in Statistiocal Data, Third Edition, New York, Wiley.

[2] Barry AM, Burney SMA, and Bhatti MI (1997). Optimum Influence of Initial Observatins in regression Models with AR(2) Errors, Applied Mathematics and Computations, 82(1), 57-65.

[3] Belsley PA, Kuh E, and Welsch RE (1980). Regression Diagnostics, New York, John Wiley.

[4] Burney SMA and Maqsood A (2013). Extending the Stochastic Approach to Paasches Price IndexNumbers, Pakistan Journal of Engineering Technology & Science, 3(1), 1-17.

[5] Burney SMA and Maqsood A (2014). Influential Observations in Stochastic Model of Divisia Index Numbers with AR(1) Errors, Applied Mathematics, 5(6), 975-982.

[6] Clements KW and Izan HY (1987). The Measurement of Inflation: A Stochastic Approach, Journal of Business and Economic Statistics, 5, 339-350.

[7] Cook RD (1977). Detection of Influential Observations in Linear Regression, Technometrics, 19(1), 15-18.

[8] Cook RD and Weisberg S (1982). Residuals and Influence in Regression, New York: Chapman and Hall.

[9] Draper NR and John JA (1981). Influential Observations and Outliers in Regression, Technometrics, 23(1), 21-26.

[10] Draper NR and Smith H (1998). Applied Regression Analysis, Third Ed., New York: John Wiley.

[11] Hoaglin DC and Welsch RE (1978). The Hat Matrix in Regression and ANOVA, The American Statistician, 32, 17-22.

[12] Maqsood A and Burney SMA (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.

[13] Maqsood A and Burney SMA (2017). Standard Errors for the Laspeyre Index Number with Autocorrelated Error Models, Communication in Statistics: Theory and Methods, 46(21), 10607-10616.

[14] Özkale MR and Açar TS (2015). Leverages and Influential Observations in a Regression Model with Autocorrelated Errors, Communications in Statistics -Theory and Methods, 44(11), 2267-2290

[15] Puterman ML (1988). Leverage an2d Influence in Autocorrelated Regression Model, Journal of the Royal Statistical Society, 37(1), 76-86.

[16] Selvanathan EA (1993). More on the Laspeyers Price Index, Economics Letters, 43, 157-162.

[17] Stemann D and Trenkler G (1993). Leverages and Cochrane-Orcutt Estimation in Linear Regression, Communication in Statistics-Theory and Methods, 22(5), 1315-1333.