Tropical Agricultural Research Vol. 11 1999, 324-334

Estimation of Demand Function for Coconut Oil: A Cointegration Analysis

S. Samarajeewa and H.M. Gunatilake[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Peradeniya, Sri Lanka

ABSTRACT. *Very often, time series data show deterministic or stochastic trends. Such trends may give rise to spurious regressions with un-interpretable t-ratios and higher goodness of fit measures. Therefore, it is necessary to use cointegration analysis to make sure of the existence of long term equilibrium relationships in the data. This study uses 20-year time series data to estimate the demand function for coconut oil incorporating own price, three substitute prices and income. Standard econometric analysis shows somewhat acceptable results despite the presence of serial correlation and some unexpected signs. Results of the Dicku-Fuller and augmented Dicku-Fuller tests show that quantity consumed and price of palm oil are integrated of order zero, while prices of coconut and soy oil and income are integrated of order one. Order of integration of palm kernel oil is unknown. Residuals of the demand model show presence of a unit root. Therefore, a long-term equilibrium relationship among the variables in the model does not exist. A practitioner of standard econometrics would have erroneously concluded that coconut oil is a normal good and that its demand has no relationship to palm oil price. Therefore, examination of the stability properties of a demand model is vital if time series data are used for the estimation.*

INTRODUCTION

Time series data have been frequently used in estimating demand functions in standard econometrics (Patel and Vyas, 1971; Sidhu *et al.*, 1972; Viswanathan and Satyasai, 1977; Sihag, 1981; Mruthunjaya and Srinivasan, 1982; Pandey *et al.*, 1984). Very often, the time series data show deterministic or stochastic trends (Gordon *et al.*, 1993; Charemeza and Deadman, 1997). Trends, either deterministic or stochastic, may give rise to spurious regressions, which result in un-interpretable t-ratios and other

1 Department of Agricultural Economics, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.

statistics (Granger, 1986; Charemeza and Deadman, 1997). Although the goodness of fit measures are generally very high, valid inferences cannot be made from such regressions. Integration and co-integration tests have been proposed in order to avoid spurious regressions (Granger and Newbold, 1974).

In standard methods, econometrics is used to illustrate a theoretical model, which is believed to be correct. When serial correlation, multicollinearity, heteroscedasticity and simultaneity are encountered, the natural tendency is to 'fix' the problem without changing the model specification. Fixing the model is carried out until higher R^2 and significant t-statistics are obtained. This approach has been designated as 'data mining' and subjected to severe criticism in recent times (Dhrymes, 1970; Hendry, 1980; McAleer and Veall, 1995; Charemeza and Deadman, 1997).

In time series models, specification errors may be indicated by the presence of autocorrelation. When the auto-correlation is present, the suggested remedy in standard econometric methods is the Cochrane-Orcutt correction. As shown by Hendry and Mizon (1978) and Mizon (1983) that the Cochrane-Orcutt correction in a static model is valid only when there are common factors. Therefore, the convenient auto-correlation correction without checking the existence of common factors falls into the 'data mining' procedures of conventional econometrics. Mizon (1983) provides an important advice in the title of one of his articles "*A simple message to Autocorrelation correctors: don't*", and suggests that autocorrelation should be viewed as a broader specification problem. As an alternative to the data mining procedures in standard econometrics, General to Specific Modelling (GSM) approach has been suggested (Hendry and Richard, 1983; Pagan *et al.*, 1984; Gilbert, 1986; Charemeza and Deadman, 1997). This approach emphasises the proper specification of time series models starting with a broad general model with lags and narrowing it down to a specific model with appropriate statistical testing procedures.

The purpose of this paper is to illustrate the possible errors one would have made in estimating a demand function following the traditional method of econometrics. This paper uses a 20-year time series data to estimate the demand function for coconut oil using the traditional method. The data is checked for the stationarity using unit root tests. Engle-Granger two step procedure is then used to check the existence of long run equilibrium relationship of linear combination of variables. Finally the results of the traditional approach are summarised in the light of the possible mistakes one would have made by ignoring the use of non-stationary data and specification issues.

The paper is organized as follows. The second section presents the details of the data and model specification. The third section presents the results of the standard econometric analysis. The fourth section presents the results of integration and cointegration analyses and the changes in results when the stationary data are used for analysis. The final section presents the conclusions and policy implications.

MODEL SPECIFICATION AND DATA

Demand is a multivariate relationship. It shows the quantities of a product that consumers are willing and able to buy, over some given period of time, under different prices, *Ceteris paribus*. The most important determinants of the market demand for a commodity are its own price, consumers' income, prices of other commodities, consumers tastes, consumers' wealth, past levels of demand (Koutsoyiannis, 1979). In estimating a demand function, the choice of the model to be adopted depends upon the type of data and problem in hand. Fox (1953), enumerated situations where single equation models are justified in estimating demand functions. Many authors (Patel and Vyas, 1971; Sidhu *et al.*, 1972; Viswanathan and Satyasai, 1977; Mruthunjaya *et al.*, 1982; Pandey *et al.*, 1984) have estimated single equation demand functions having its own price, price of substitutes and per capita real income as the main explanatory variables. Balbir (1981) included time trend as an explanatory variable in estimating the demand function for rice and wheat considering the price variables as endogenous.

An appropriate functional form is a difficult choice in econometrics because theory does not provide direct guidelines for such a choice (Judge *et al.*, 1988; Gujarati, 1992, Green, 1993). Balbir (1981) suggested estimation of linear and log-linear forms and choosing whichever one provides better results in terms of signs, significance of co-efficient and better fit. Sidhu *et al.* (1972) estimated the demand functions for eggs using single equation linear and log-linear functions. Therefore, the demand for coconut oil was specified using single equation approach as described below. Note that, following the GSM approach all three substitute prices were included in the model.

$$D\ oil = f(Pc, Pp, Ppk, Ps, Income, T_t)$$

Where,

$D\ oil$ = demand for coconut oil (kg/head/year)
Pc = real price of coconut oil (Rs/kg)

Pp	=	real price of palm oil (Rs/kg)
Ppk	=	real price of palm kernel oil (Rs/kg)
Ps	=	real price of soy oil (Rs/kg)
Income	=	per capita real income (Rs)
T_t	=	time trend

Data required for the analysis were obtained from the annual reports of the Coconut Development Authority (price of coconut oil and per capita consumption), Department of Census and Statistics (prices of substitutes) and the Central Bank of Sri Lanka (per capita income). A 20-year data ranging from 1978 to 1997 for the relevant variables were used in the empirical estimation. Both linear and log-linear models of the above function were estimated using same explanatory variables.

RESULTS

Linear model seems better in terms of the significance levels and the goodness of the fit. Therefore, the linear specification was used for the rest of the analysis. Results of the linear model appear satisfactory as the own price and income variables show statistically significant negative and positive relations, respectively, to the quantity of coconut oil consumed (see Table 1). According to economic theory, price elasticity of demand for substitute goods should be positive. Palm kernel oil and soy oil prices, however, show a negative relationship to the coconut oil consumption. These relationships are not statistically significant. Moreover, the Durbin-Watson statistic indicates the presence of negative autocorrelation. Following the traditional approach the autocorrelation was corrected using the Cochrant-Orchut transformation. The results are presented in Table 2.

Autocorrelation corrected model also shows statistically significant negative own price effect and a positive income effect. Two out of the three substitute prices show unexpected signs. Of these, palm kernel oil variable shows a statistically significant negative effect on coconut oil consumption. Also note that both models have very high R^2 values and F-statistics indicating very good goodness of fit. Inspection of the correlation matrix of the variables indicates that income and prices of substitutes are highly correlated with time trend. This result indicates the possibility of the presence of trends in the time series data. Also the soy oil price and palm kernel oil price are highly correlated indicating the possibility for multicollinearity.

Table 1. Preliminary OLS estimation results.

Variable	Coefficient	Standard errors	P-value
Price of coconut oil	-75.49	16.57	0.001**
Price of palm oil	20.21	22.08	0.377
Price of palm kernel oil	-22.75	13.67	0.120
Price of soy oil	-24.41	25.67	0.359
Income	0.44	0.19	0.043**
Trend	-0.43	0.15	0.012**
Constant	816.69	214.60	
R^2		0.83	
D-W statistic		2.88	

** - Significant at 5% level

Table 2. Results of the autocorrelation corrected model.

Variable	Coefficient	Standard errors	P-value
Price of coconut oil	-76.88	12.13	0.000**
Price of palm oil	20.74	15.19	0.195
Price of palm kernel oil	-20.08	10.29	0.073*
Price of soy oil	-22.55	17.55	0.221
Income	0.42	0.14	0.010**
Trend	-0.41	0.11	0.002**
Constant	816.69	214.60	
R^2		0.88	

** - Significant at 5% level

COINTEGRATION ANALYSIS

Although the results presented earlier show generally acceptable statistics, they may well be spurious regression results. Therefore, following the suggestion by Charemeza and Deadman (1997) the stationarity of the variables were checked applying Dicky-Fuller (DF) and Augmented Dicky-Fuller (ADF) tests. Results of these unit root tests are given in Table 3.

Table 3. Results of the unit root tests.

Variable	DF/ADF statistics	Asymptotic critical value	Remark
Quantity of coconut oil	-2 7982	-2.57	Reject null hypothesis
Price of coconut oil	-2.3516	-2.57	Do not reject null hypothesis
Price of palm oil	-2.9884	-2.57	Reject null hypothesis
Price of palm kernel oil	-2.4449	-2.57	Do not reject null hypothesis
Price of soy oil	-1.4391	-2.57	Do not reject null hypothesis
Income	0.4118	-2.57	Do not reject null hypothesis
Residual	-4.2489	-4.70	Do not reject null hypothesis

Note that in Table 3 when the variable is serially correlated the ADF test was performed with the appropriate lag length. Otherwise the DF test was performed without lags. The results given in Table 3 are without a trend term in the unit root test. Similar analysis was performed with a trend and the results are not reported to avoid unnecessary details. When a trend variable is included in DF and ADF tests, test statistics and critical values are different. However, the conclusions reached with the trend and without trend are the same. The results in Table 3 show that non-stationarity null hypotheses are rejected for two variables, namely quantity and price of palm oil. This indicates that these variables are integrated of order zero *I(0)*; they are stationary. Thus, prices of coconut oil, income, price of palm kernel oil and soy oil are non-stationary time series. Since the variables in the previous regression model are integrated of different orders, the test statistics derived from the model may not be valid if the linear combination of the variables is not cointegrated (Charemeza and Deadman, 1997).

Further unit root tests were carried out for prices of coconut oil, palm kernel oil and income variables, after first differencing them. The test results are shown in Table 4. These results indicate that all the variables are integrated of order 1, except for the palm kernel oil variable. Unit root test after further differencing of palm kernel oil variable provides very high t statistics indicating that this variable cannot be made stationary by differencing.

Table 4. Results of the Dicky-Fuller test for first-differenced variables.

Variable	DF/ADF statistics	Asymptotic critical value	Remark
Price of coconut oil	-2.8910	-2.57	Reject null hypothesis
Price of palm kernel oil	-2.4654	-2.57	Do not reject null hypothesis
Price of soy oil	-4.0806	-2.57	Reject null hypothesis
Income	-2.6390	-2.57	Reject null hypothesis

Although variables in a regression model are integrated in different orders, a linear combination of them can be cointegrated so that the residuals of the regression model is integrated of zero order. In that case the regression is not spurious and results are reliable. In other words, the variables in the model show long-term equilibrium relationships, if the cointegration condition is satisfied. In order for the cointegration to take place the order of integration of the dependent variable cannot be higher than the order of integration of any of the explanatory variables. Moreover, there must be either none or at least two explanatory variables integrated to an identical order higher than the order of integration of the dependent variable (Charemeza and Deadman, 1997). Since these requirements are met in the present case, further tests on stationarity of the residuals was carried out. Test results are given in the last row of Table 3. Since the non-stationarity of the residuals cannot be rejected, the regression results presented in Tables 3 and 4 are spurious. No valid inferences can be made from the analyses. In other words, there is no long-term equilibrium relationship among the variables in these models.

If a regression model shows a cointegration relationship, an error correction model can represent that model. The error correction form provides details of both long-run and short-run properties. Therefore, an error correction model can be used to study the dynamics of economic relationships (Charemeza and Deadman, 1997). In the absence of a cointegrating relationship, one cannot preserve both short-run and long-run properties. However, by differencing the data appropriately the residual of the model can be made stationary. Although differencing solves the problem of spurious regression, it leads to a significant loss of information contained in the original data. Especially, the differencing converts the original model to a growth rate model by imposing a restriction that initial dis-equilibrium does not enter into the decision making process of the economic agent (Charemeza and Deadman, 1997). Since this is clearly an unrealistic assumption, a model with differenced data may not be the ideal solution for spurious regression. However, in order to complete our exposition a model with appropriately differenced data was fitted. Given the higher collinearity of palm kernel oil with other variables and non-stationarity, this variable was dropped from the equation. The time trend was also dropped from the model since differencing has made necessary adjustments for the trends.

The results of the regression model with differenced data are given in Table 5. R^2 is much lower in this regression and the Durbin-Watson statistic shows no serial correlation. Own price and substitute prices show expected signs, while the own price and palm oil price are statistically significant. The income variable shows a negative relationship to the quantity consumed. However, this variable is not statistically significant. Since the

Table 5. Results of the model with stationary data.

Variable	Coefficient	Standard error	P-value
Price of coconut oil	-45.758	18.23	0.025**
Price of palm oil	65.921	22.34	0.011**
Price of soy oil	7.183	18.86	0.709
Income	-0.099	0.28	0.733
Constant	1.606	0.50	0.007**
R^2		0.497	
D-W statistics		1.886	

model with differenced data does not represent a demand model, further inferences are not warranted. Nevertheless, comparison of results in Tables 1 and 2 with those of Table 5 shows the difference between spurious and a true regression.

A practitioner of traditional econometrics would most probably have reported the results in Table 2. Her conclusions would have included a positive income effect on coconut oil consumption. As clearly shown in the previous analysis, regression results presented in the Table 1 and 3 are spurious. Therefore, at the aggregate level there is no statistical evidence to assert that coconut oil is a normal good. The de-trended data hints that coconut oil may be an inferior good. If the sample can be segregated for different income groups, the inferiority of coconut oil for higher income groups may emerge. Further, the conventional econometric analyses show statistically non-significant relation between coconut oil consumption and palm oil price. When the trends are removed the expected positive relationship emerges as shown in Table 5.

CONCLUSIONS

A 20-year time series data set was used to estimate the demand function for coconut oil in this study. Following the general to specific modelling approach, a demand model was specified using own price, three-substitute prices, income and time trend. Standard econometric analysis with correction for autocorrelation showed acceptable results, in terms of goodness of fit and significance levels. Higher correlation of the time trend with other variables, statistical non-significance and wrong signs of substitute prices and presence of serial correlation indicated possibility of spurious regression.

Tests on integration shows the quantity consumed and price of palm oil are integrated of zero order, (*I(0)*) and income, prices of soy oil and coconut oil are integrated of order one (*I(1)*). Order of integration of the price of palm kernel oil is unknown. Since it is possible for a linear combination of these variables to be integrated, the stationarity of the residuals of the model was checked. The result shows the presence of a unit root in the residuals. This indicates that there is no long-run equilibrium relationship among the variables in the model. Thus, the regression results obtained from the standard econometrics are spurious and they may have led to the erroneous conclusion that coconut oil is a normal good whose demand is not affected by price of palm oil. A regression model with stationary variables further highlights misleading goodness of fit measures and significance levels associated with the spurious regression. Therefore, it is crucial to check the stability properties of a demand model when time series data are used for estimation.

REFERENCES

Balbir, S.S. (1981). Estimation of income and price elasticities of demand for wheat and rice on a state wise basis. Ind. J. Agric. Econ. XXXVI(3).

Charemeza, W.W. and Deadman, D.F. (1997). New Directions in Econometric Practice. Edward Elgar, Cheltenham, UK.

Dhrymes, P.J. (1970). The game of maximizing. Aust. Econ. Ppr. 9: 177-185.

Fox, K.A. (1953). The analysis of demand for farm products. US Dept. of Agriculture, Technical Bulletin. 1081: 11-14.

Gilbert, C.L. (1986). Professor Hendry's econometric methodology. Oxford Bulletin Econ. and Stat. 48: 283-307.

Gordon, D.V., Hobbs. J.E. and Kerr, W.A. (1993). A test for price integration in the EC lamb market. J. Agric. Econ. 44(1): 126-134.

Granger, C.W.J. (1986). Developments in the study of co-integrated economic variables. Oxford Bulletin Econ. and Stat. 448: 213-228.

Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics. J. Econometrics. 35: 143-159.

Green, W.H. (1993). Econometric Analysis, Macmillan Publishing Company. New York.

Gujarati, D. (1992). Essentials of Econometrics. McGraw-Hill, Inc. New York.

Hendry, D.F. (1980). Econometrics - alchemy or science? Economica. 47: 387-406.

Hendry, D.F. and Mizon, G.E. (1978). Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England. Econ. J. 88: 549-563.

Hendry, D.F. and Richard, J.F. (1983). The econometric analysis of economic time series. Int. Statistical Review. 51: 111-163.

Judge, G.G., Hill, R.C., Griffiths, W.E. and Lütkepohl, T. (1988). Introduction to the theory and practice of econometrics. John Willey and Sons, New York.

Koutsoyiannis, A. (1979). Modern Microeconomics. 2nd Edition, St. Martin Press, New York.

McAleer, M. and Veall, M.R. (1995). Data mining and the con in econometrics - the US demand for money revisited. Mathematics and Computer in Simulation. 39: 329-333.

Mizon, G.E. (1983). A Simple Message to Autocorrelation Correctors: Don't. European University Institute.

Mruthunjaya, A. and Srinivasan, V.R. (1982). Demand for potato in Karnataka - policy alternatives. Ind. J. Agric. Econ. 37(4).

Pagan, D.F., Hendry, A. and Sargan, J.D. (1984). Dynamic Specification. Vol. 2, North-Holland.

Pandey, U.K., Sardana, P.K. and Monocha, V. (1984). An estimate of demand for and supply of cereals, coarse grains, pulses and oil seeds in Haryana. Ind. J. Agric. Econ. 39(1).

Patel, A.S. and Vyas, V.S. (1971). An estimate of demand for cereals and pulses in the coming decade. Ind. J. Agric. Econ. XXVI(2).

Sidhu, D.S., Guptha, M.P. and Sirohi, A.S. (1972). Estimation of demand function for eggs in Punjab - An econometric study. Ind. J. Agric. Econ. XXVII(3).

Viswanathan, K.U. and Satyasai, K.J.S. (1997). Fruits and vegetables: Production trends and role of linkages. Ind. J. Agric. Econ. 52(3).