DRAFT ONLY

MODELLING AGGREGATE IMPORT DEMAND IN BARBADOS

by

Roland Craigwell
CENTRAL BANK OF BARBADOS

Prepared for Presentation at the
Research Department, Central Bank of Barbados
Annual Review Seminar
July 14 - 15, 1995

July 1994

Modelling Aggregate Import Demand in Barbados

I: Introduction

One of the important features of the Barbados economy has been its heavy reliance on imports, not only for consumption and production purposes, but also as a major source of government revenue. With such heavy reliance on imports[1] an analysis of how imports flows react to changing economic conditions becomes of critical importance. Moreover, such an understanding is essential to the success of a structural adjustment programme like the one being currently undertaken by the Barbados Government under the austerity of the International Monetary Fund. According to Faine *et al* [1992, p.279] "Being able to predict import flows more accurately can help policymakers assess more confidently the overall sustainability of an adjustment programme, determine the appropriate speed of the trade liberalization process, and avoid the possibility of unexpected foreign exchange constraints jeopardizing the adjustment effort".

This paper considers the empirical modelling of aggregate imports for the small open economy of Barbados. Previous efforts on Barbados (Cox and Worrell [1978], Joefield-Napier [1982], Worrell [1988] and Boamah [1994]) specified either a traditional static import demand function with real income and relative prices as regressors or a dynamic model which assumes a partial adjustment mechanism where a lagged dependent variable appears as one of the regressors[2]. Here an import demand function that incorporates the traditional variables as well as non-traditional variables like credit and money is examined. The analysis is conducted by means of relatively recent developed econometric concepts. Among these is the focus on the so-called "general to specific" approach (Hendry [1987, 1993]) in the context of data series whose (non-)

stationary properties are investigated. The notion of cointegration (Engle and Granger [1987]) of a set of variable is analyzed in a single equation framework. The multivariable method proposed by Johansen [1988, 1989] and Johansen and Juselius [1990] is used to check the rank of the cointegration vector. The Johansen method and non-nested tests are undertaken to choose between three rival models of import demand, namely, a model with price homogeneity, a model with symmetric price effects and a model without the price restrictions.

This paper consists of the following sections. Section II presents the import demand function. Section III provides the basic tools that will be used in the empirical analysis. Section IV gives the empirical results. Some conclusions are reported in the final section.

II: The Import Demand Function

The traditional specification of the import demand function corresponds to that of the imperfect substitute model (Goldstein and Khan [1983]) which implies the existence of both imports and domestic production, as well as intra-industry trade. By assuming zero degree homogeneity, the general form of the import function is:

$$M_t = f(Y_t, RP_t) \tag{1}$$

where M_t is the real quantity of imports in volume terms, Y_t is real income or domestic economic activity variable and RP_t is the ratio of import prices to domestic prices.

Three modifications are made to this function here. One, the domestic price index is disaggregated into traded (PT) and non-traded (PN) price indices as for example in Goldstein *et al* [1980] and Kholi [1982]. This decomposition reflects that fact that both domestic tradeable and non-tradeable goods are potential competitors with imports in the consumer budget. Second, credit from the banking system is considered. If income rises, it is expected that given a high propensity to import the level of imports will increase. Ready sources of credit (CR) is necessary to accommodate the increase in spending. Third, a dummy for the multi-national firm, Intel, which left Barbados in 1987, is included. Thus the equation actually estimated is:

$$m_t = \alpha_o + \alpha_1 y_t + \alpha_2 pn_t + \alpha_3 pt_t + \alpha_4 pm_t + \alpha_5 ct + \alpha_6\ intel \tag{2}$$

where lower case letters denote the natural logarithms of the corresponding variables. Log-linear form is assumed following the standard assumption of trade theory which relates imports to the explanatory variables through a multiplicative form that can be derived within a cost minimisation framework (Urbain [1992]). Empirical tests using the Box-Cox transformation (see Khan and Rose [1977] and Boylan *et al* [1980]) also support this. It should be noted that infinite supply elasticity has been assumed, allowing this function to be estimated in a single equation framework. In fact, although theoretically finite supply elasticity would require a simultaneous equation model as in Boamah [1994], it should be noted that all that is required, from an econometrical point of view, is that the regressors be at least weakly exogenous (in the sense of Engle *et al* [1983]). This hypothesis of weak exogeneity is tested below and it was found that it could not be rejected.

III: The Econometric Methodology

Economic theory provides little guidance for the modelling of the short-run dynamics of aggregate import demand functions. Hence, a modelling strategy that incorporates information given by the time series properties of the data and economic theory is deemed necessary. Such a methodology has been advocated by Hendry and his associates and will be adapted here (see Hendry [1989, 1993] and Hendry and Richard [1982, 1983] for details).

The approach-called the "general to specific" approach - assumes - under hypotheses of conditional normality, linearity and time homogeneity - that the unknown data - generating process of the observable can be approximated by means of a finite - dimensional autoregressive distributed lag (ADL) model or a finite-dimensional error correction model (ECM)3. The error correction framework is preferred here as it is applicable where a suitable theory (long-run equilibrium or steady state theory) postulates proportionality between economic variables. Also, ECMs arise naturally from considering the temporal properties of the data using concepts such as integration and cointegration.

The second phase in the "general to specific" methodology considers the general ADL model or ECM is then considered as the maintained hypothesis and data - acceptable reductions and/or transformations performed in order to get a more parsimonious model. The final specification is called a tentatively adequate conditional data characterization if it is congruent with all the data evidence.

Following Hendry and Richard [1982], Hendry and Ericsson [1987], Ericsson *et al* [1990], a model is congruent with the evidence if it satisfies the following conditions: (a) data admissible, that is, it is logically possible for the data to have been generated from the model (b) exhibits parameter constancy; (c) data coherent, that is, the residuals should at least be white noise; (d) have at least weakly exogenous regressors in the sense of Engle *et al* [1983]; (e) encompasses rival models and (f) the empirical model must reproduce the theoretical model under the conditions assumed by the theory.

Trend Variables, Spurious Regressions, Integration and Cointegration

The fact that most variables used in traditional aggregate import demand functions display a powerful trend (see Appendix A for graphs, both in levels and first differences) has seldom been explicitly taken into account. Some authors (Khan and Ross [1975], Haynes and Stone [1983], Wilson and Tackacs [1979] and Goldstein et al [1980]) consider the decomposition of income into trend income ('potential income') and 'deviation from trend' in an attempt to separate the secular and the business-cycle components. This follows the practice of considering the strong pattern of secular growth over time displayed by many macroeconomic time series as arising from the presence of a deterministic trend. Random shocks have been only a temporal influence on the historical trajectory of the series. In an important paper, Nelson and Plosser [1982] posed the question as to whether the observed non-stationarity of macroeconomic time series was stochastic rather than deterministic. They found clear support for the stochastic nature of the non-stationarity. More precisely, the series could be reasonably well characterized as having a unit root and a drift. In the case of such a stochastic trend, the unit root implies an accumulation of past random shocks which have an enduring effect on the historical behaviour of the data series. This type of non-stationary variable is commonly called an integrated process.

Integration

Following Granger [1981], a time series Y_t which has a stationary, invertible, non-deterministic autoregressive moving average (ARMA) representation after differencing d times is integrated of order d and is denoted $Y_t \sim I(d)$. A simple example of an I(1) process is a random walk with drift, that is

$$Y_t = \alpha Y_{t-1} + \mu + \epsilon_t \tag{3}$$

where for simplicity $Y_o = 0$ and ϵ_t are identically independently distributed normal variates with zero mean and variance σ^2. If $\alpha = 1$ (unit root), then equation (3) can be rewritten as

$$Y_t = \mu_t + S_t;\ S_t = \sum_{t=1}^{j} \epsilon_t \tag{4}$$

These processes are linear functions of time (of slope μ) and the divisions from this function are non-stationary as they are the accumulation of past random shocks (through S_t). In the case of integrated processes, the sample moments converge to random variables and traditional central limit theorems are replaced by functional limit theorems (Dolado and Jenkinson [1987] and Galbraith *et al* [1993]). In regressions with integrated variables, t and F statistics may diverge as the sample size increase, the DW statistic tends to zero in the case of spurious correlation and so on (see Phillips [1987]). As a result, regressions involving integrated variables in levels often display serial correlation.

Testing for Integration

The recent voluminous literature on unit roots has provided a variety of possibilities for detecting these in observed time series (Dolado and Jenkinson [1987] and Diebold and Nerlove [1990]). In this paper, the t-ratio $t_{\alpha=1}$ (from equation (3)) initially studied by Fuller [1976] and extended by Dickey and Fuller [1979, 1981] is considered. As the limiting distribution is non-standard, simulated critical values are available in Fuller [1976], Engle and Yoo [1987]) and more recently MacKinnon [1991]. However, the derivation of the limiting distribution of the t-statistic will crucially depend on the error term ϵ_t (Phillips [1987]). If they are correlated, then another model is required. This is an enlarged model of equation (4); a lag polynomial of ΔY_{t-1}, such that these terms capture the serial correlation and renders ϵ_ts empirically white noise is added to equation (4). This is called the ADF test[3]. Phillips [1987], Phillips and Perron [1988] and Perron [1988] provide an alternative non-parametric correction.

57 59

Cointegration

Over the recent years, theoretical and practical econometricians have shown a growing interest in the concept of cointegration which allows estimation and testing on "long-run" theoretical relationships between economic variables which are integrated time series. Intuitively, cointegration among a set of variables implies that there exist fundamental economic forces which make the variables move stochastically together over time. Formally, the components of the vector X_t are said to be cointegrated of order d, b, denoted CI(d,b), if (i) all the components of X_t are I(d) and; (ii) there exists a vector $\alpha(\neq 0)$ such that $Z_t = \alpha' X_t$ is I(d-b), b>o. The vector is called the cointegrating vector (See Engle and Granger [1987, p. 253]).

Granger [1981], Granger and Weiss [1983] and Engle and Granger [1987] have established a link and even equivalence between error correction and the concept of cointegration through the Granger Representation Theorem which states that if a series have an ECM representation, then they should be cointegrated, and conversely. Thus, cointegration provides formal statistical support for the use of ECM.

Consider the following regression model called the cointegration regression model

$$\alpha' X_t = Z_t \quad (5)$$

where all components of X_t are I(1). Stock [1987] and Engle and Granger [1987] show that the OLS estimates of α yields an excellent approximation to the cointegration vector (if one exist). The OLS estimates of the cointegrating vector converges to the true value extremely quickly (Stock [1987]), however its distribution is not asymptotically normal and the computed standard errors are meaningless. When X_t is a vector with more than two components, then if a cointegrating vector exists, it need not be unique since for k component there may be at most r ($r<k-1$) linearly independent cointegrating components (see Stock [1987]). Johansen [1988, 1989] and Johansen and Juselius [1990] proposed a unified maximum likelihood approach for the estimation of the number of cointegrating relations. Details of this approach is found in Appendix C.

Testing and Estimating Cointegration in a Single-Equation Framework

Testing for cointegration among a set of time series is similar to testing for unit roots. The only difference is that the null hypothesis is no-cointegration (that is, spurious regression) rather than unit root in an observed series. The tests are conducted on the constructed residual Z_t and the critical values used for unit-root tests have to be adjusted upwards if the null is not to be rejected very often. The DF tests are the most popular tests used by applied researchers. The critical values for the DF and ADF tests have been taken from McKinnon [1990] who tabulated these in the case of cointegration for different significant levels.

The cointegration regression model parametrizes the long-run relationship between the variables. In order to introduce short-run dynamics into the model, Engle and Granger [1987] propose to include a lagged value of X_t (equal the disequilibrium at each period) in a general dynamic model. If Y_t and X_t are cointegrated, then X_t will be I(1) and the problem consist of transforming the individual I(1) series into I(0) ones. This is their two-step approach. Alternatively, one can apply Hendry's approach in order to get an EC formulation in a dynamic model. Unless the variables are cointegrated the ECM term will be I(1) and have an estimated coefficient tending to zero. Monte Carlo results in Barnejee *et al* [1986] and Boswisk [1990] tend to show that the bias in the cointegrating vector may be much smaller in the dynamic ECM modelling strategy.

IV: **Data and Empirical Results**

All data are annual figures taken from the Central Bank of Barbados New Forecasting Model file. They cover the period 1960-1993. The following notation is used: m for real quantity of imports in volume terms, y_t is real gross domestic product at factor cost, pn_t and pt_t are traded and non-traded price indices, pm_t is the import price index and cr is real private sector credit. m and cr are deflated by the consumer price index while y is scaled by the GDP deflator, all at 1980 prices. All variables are expressed in logarithmic terms.

The starting point of the analysis is to investigate the integration order of the series. Table 1 presents the results of the DF and ADF test procedures. It appears that all the data series are characterized as I(1) series. According to these results the decomposition of real GDP into a deterministic trend and a stationary deviation from this trend appears a questionable practice.

Table 2 shows the cointegration results of the imperfect substitute model (equation [2]) and several variants thereof. First of all, imposing domestic price homogeneity and symmetric price effects though significantly alters the coefficients of the model do not reject the hypothesis of no cointegration. Thus, any of the three equations in Table 2 contains a set of variables that are cointegrated. The likelihood ratio statistic in the Johansen [1988] procedure was used to test these price restrictions. The results indicate that these restrictions could not be accepted. As these are only long run results, short run dynamic models are estimated for all three equations. Note that the traditional models which contains only prices and real income as explanatory variables are not cointegrated, suggesting that previous studies are in error for specifying import demand in such a way.

The uniqueness of these vectors were examined using the Johansen (1988) maximum statistic which tests the null of at most r cointegrating vectors against r + 1 cointegration vectors. The length of the VAR used is 2. This was chosen by testing down, using a likelihood test, from a general VAR of length 3 until reducing the order by one lag could be rejected and the model was white noise. Table 3 shows that for equations (iv) and (v), the cointegrating vectors are unique but for equation (vi), the rank is 2. For the latter, the normalisation on m_t gave the most sensible economic meaning in terms of sign and size of coefficients. Hence, it is assume that this is the appropriate normalisation.

The next step in the empirical analysis is to apply the ECM proposed by Engle and Granger [1987][4]. Starting with the following general model

$$\Delta m_t = \delta_o + \alpha_1 \Delta m_{t-1} + \sum_{i=0}^{1} \beta \Delta x_{t-i} + INTEL + \sum_{i=1}^{3} \gamma_i \epsilon_{t-i}^{i} + U_t \quad (6)$$

where the maximum lag length is restricted to one given the small sample size, $x_t = (y_t, pn_t, pmt, cr)$ and the polynomial in ϵ_t represents the possibility of non-linear long run adjustment (see (Escribano [1987]), the model was reparametrized by dropping insignificant variables. The final specified error correction models for Barbados imports are

(iv) Price Homogeneity Equation

$$\Delta m_t = \underset{(0.93)}{-0.03} + \underset{(3.15)}{1.15 \Delta dp_t} + \underset{(3.36)}{1.32 \Delta y_t} - \underset{(4.53)}{1.00 \Delta pm_t} + \underset{(3.91)}{0.28 \Delta cr_t} \underset{(2.85)}{-0.59 ECM1_{t-1}}$$

$R^2 = 0.55$; s.e = 0.08; D.W = 1.71; SC [F(1,12)] = 0.55; Norm [$X^2(2)$] = 4.64; ARCH [F(1,12)] = 0.34; WHITE [F(1,29)] = 0.15; RESET [F(3,22)] = 0.29

CHOW..[F(6,19)] =0.65 CHOW 80 [F,6,19] = 0.51

HEND74 [(17, 8)] =0.96 HEND 80 [F(V1,14)] = 1.71

(v) Symmetric Price Equation

$$\Delta m_t = \underset{(1.64)}{0.03} + \underset{(3.33)}{1.418 \Delta y_t} \underset{(-1.56)}{-0.6870 \Delta y_{t-1}} + \underset{(2.66)}{0.243 \Delta cr_t} - \underset{(3.09)}{0.30 \Delta cr_{t-1}} \underset{(-4.05)}{- 0.90 ECM2_{t-1}}$$

$R_2 = 0.49$; s.e = 0.09; DW = 1.42; SC [F(1,24)] = 2.77 NORM [$X^2(20]$ = 1.11;

ARCH [f(1,24)] WHITE [F(1,29)] = 0.63; RESET f($3_0$21)]= 0.64

CHOW 74 [F(6,19)] = 0.31 HEND 74[F(17,8)] = 0.60

CHOW 80 [F(6,19)] = 0.78 HEND 80 [f(11,14)] = 1.10

(vi) Equation with no Price Restriction

$$\Delta m_t = \underset{(-1.12)}{-0.04} + \underset{(3.32)}{0.82 \Delta pn_t} + \underset{(3.22)}{1.23 \Delta y_t} + \underset{(2.69)}{0.25 \Delta cr_t}$$

$$\underset{(-4.78)}{-1.05 \Delta pm_t} + \underset{(2.39)}{0.59 \Delta pm_{t-1}} \underset{(2.09)}{-0.51 ECM3_{t-1}}$$

$R^2 = 0.60$ s.e = 0.08; DW = 1.92; SC F[1,23)] = 0.02 NORM [$X^2(2)$] = 0.27;

ARCH F(1,23) = 0.98; WHITE F(1,29) = 0.20; RESET F[(1,23)] = 1.37

CHOW 74 [F(7,17)] = 0.19 HEND 74 [F(17,7)] = 1.45

CHOW 80 [F(7,17)] = 1.15 HEND 80 [F(11,13)] = 1.81

Each model was estimated by OLS. In brackets underneath the estimated coefficients are the t statistics. Other notations are: R^2 for the squared multiple correlation coefficient; s.e for the standard deviation of the regressor error; DW for the traditional Durbin Watson statistic, SC for the lagrange multiplier test for first order autocorrelation, NORM for Jarque and Bera test for normality, ARCH for first order autoregressive conditional heteroscedasticity effect, WHITE for heteroscedastic errors, RESET for third order misspecification test and HEND and CHOW for Hendry and Chow predictive failure and parameter constancy tests, respectively. Details of these diagnostic tests can be found in Godfrey (1988).

The fits of these models are graphed in Appendix C. They appear reasonable given that the models are in first difference form. These final reparametrizations should satisfy different criteria if they are to be considered as adequate data characterizations. The various residual diagnostic checks do not show evidence of serial correlation, ARCH effects, non-normality, heteroscadasticty or non-linearity. From and HEND , the models show good out-of-sample forecasting performance. They can thus be considered as having constant parameters. A within sample - constancy analysis of the estimated model, show constancy judged by the standard statistics such as Chow tests and recursive residuals (see Appendix D).

Although the valid conditioning can be analyzed using a Hausman-type specification test or an LM test for weak exogeneity, a proposal suggested by Richard [1987] and Hendry [1987] is followed here. If a conditional model has within it sample constant parameters, although it is known that the process generating the regressors have been subject to major changes (which is clearly in our case), then one may intuitively accept the hypothesis of super exogeneity of the regressors, and thus weak exogeneity.

Now that these equations are shown to be statistically adequate it is time to check their economic significance. In all three equations income and credit are significant variables while import price play an important part only in equation (iv) and (v). Domestic prices enter in (iv) and in (vi) through the non-traded price index. The signs are a prior expected except for the overall credit impact in equation (v) - credit should be positively related to import flows. The magnitudes of respective variables are also quite similar for equations (iv) and (v). However, each model has a different speed of long run adjustment; symmetric price eqution adjusts the quickest, followed by the price homogeneity equation, then the model with no price restrictions.

In summary, all three models are statistically and economically adequate characterizations of the data generation process. But which one is to be preferred? To do this non-nested and encompassing tests were carried out. The non-tested tests employed are the F test, calculated by embedding the null and alternative models in a comprehensive model, and the adjusted Cox-type tests, NT and W, which incorporate small sample corrections designed to bring actual significance levels close to the normal values (see Godfrey and Pesaran [1983]). Simulation evidence by Godfrey and Pesaran revealed that the NT and W tests have finite sample significance levels which are quite close to the normal value over a wide range of DGP. Moreover, these tests are more likely to lead to the correct decision of accepting the true model and rejecting the false model than the familiar F tests based upon the comprehensive model or regression parameter encompassing approaches. Note, however, that they only test variance encompassing and have the drawback that both models can be rejected. Parameter encompassing, theoreticaly, rectifies this negative aspect of the research and can be tested easily by transforming the more difficult Wald statistic to the classical F test of zero additional parameters in the "nested" model which embeds the maintained and the alternative hypotheses in a linear regression equation.

The test statistics are shown below in Table 4. With W and NT equation (iv) and equation (v) rejects each other. The F test favours equation (v) against equation (iv). Likewise, equation (vi) and (iv) reject and equations (v) and (vi), respectively, each other with NT and W. Thus the negative aspects of the research implied by non-nested tests are manifested. The encompassing test presents no solution. According to this test, the null and alternative hypotheses respectively are valid [not valid] restricted forms of the embedding model which include both of them. Therefore we cannot say that one model (v and vi) [vi and iv] encompasses the other. The non-rejection of the models when they may be false may be due to the relatively low power of the F-test when applied to testing non-nested models with a large number of non-overlapping (common) variables (Godfrey and Pesaran (1983). However, the adjusted Cox test does not possess this feature (at least asymptotically).

Conclusion

This paper uses recently developed econometric concepts to model aggregate import demand in Barbados. Long run and short run equilibrium relationlships were derived, with the traditional hypotheses of price homogeneity and symmetry rejected by the data. This study also points to the inclusion of non-traditional variables like credit in modelling import demand in Barbados.

Table 1: Testing for Unit Roots

Variables	DFC(ADF) Statistics
m	-1.43 (-1.63)
Δm	-4.88* (-3.61)*
y	-0.68 (-0.57)
Δy	-4.57* (-2.87)**
pn	-1.26 (-1.51)
Δpn	-4.64* (-2.96)**
pt	-1.02 (-1.24)
Δpt	-4.80* (-3.77)*
cr	-2.12 (-2.17)
Δcr	-6.64* (-6-69)*
rp	-2.23 (-2.17)
Δrp	-5.99* (-4.58)*
dp	0.13(-0.56)
Δdp	-2.74**(-2.88)**

* Significant at 5% level
** Significant at 10% level
Δ change in the lag of the variable

Table 2: Cointegration Results

Equation	CONST	dp	rp	intel	pn	cr	pm	y	pt	R^2	DW	DF	ADF
(i)	-6.59							1.84		0.78	0.47	-2.13	-2.45
(ii)	2.88	0.56	-0.71				-0.23	0.22		0.81	0.36	-1.83	-2.18
(iii)	7.75				1.24		-0.53	-0.45	-0.49	0.93	1.04	-3.14	-2.43
(iv)	-1.86		-0.24	0.23		0.45		.0.75		0.95	1.75	-4.90	-4.59
(v)	0.01	0.31		0.22		0.48	-0.34	0.48		0.95	1.66	-4.69	-3.81
(vi)	3.46			0.20	0.74	0.27	-0.53	0.13	-0.22	0.97	1.80	-4.88	-3.50

Note:*indicates significant at the 5% level.

Table 3: Johansen Maximum Likelihood Results based on the Maximal Eigenvalue

Equation	Null	Alternative	Statistic	95% Entical Value
(i)	$\gamma=0$	$\gamma=1$	58.73	39.37
	$\gamma<=1$	$\gamma=2$	23.45	33.46
	$\gamma<=2$	$\gamma=3$	20.65	27.07
	$\gamma<=3$	$\gamma=4$	12.03	20.97
	$\gamma<=4$	$\gamma=5$	6.42	14.07
	$\gamma<=5$	$\gamma=6$	1.44	3.76
(ii)	$\gamma=0$	$\gamma=1$	122.71	94.16
	$\gamma<=1$	$\gamma=2$	63.99	68.52
	$\gamma<=2$	$\gamma=3$	40.54	47.21
	$\gamma<=3$	$\gamma=4$	19.89	29.68
	$\gamma<=4$	$\gamma=5$	7.86	15.41
	$\gamma<=54$	$\gamma=6$	1.44	3.76
(iii)	$\gamma=0$	$\gamma=1$	70.33	45.28
	$\gamma<=1$	$\gamma=2$	52.31	39.37
	$\gamma<=2$	$\gamma=3$	33.14	33.46
	$\gamma<=3$	$\gamma=4$	21.55	27.07
	$\gamma<=4$	$\gamma=5$	12.78	20.97
	$\gamma<=5$	$\gamma=6$	7.92	14.07
	$\gamma<=6$	$\gamma=7$	4.00	3.76

Table 4: Non-Nested and Encompassing Tests for Equations (i),(ii) and (iii)

Test Statistic	Distri-bution	(iv)against(v)	Distri-bution	(v) against (iv)
NT-test	N(0,1)	-1.59	N(0,1)	-3.36
W-test	N(0,1)	-1.43	N(0,1)	-2.78
Encompassing F-test	F(3,22)	2.15	F(3,22)	3.41

Test Statistic	Distri-bution	(iv)against(vi)	Distri-bution	(vi) against (iv)
NT-test	N(0,1)	-1.76	N(0,1)	-0.87
W-test	N(0,1)	-1.56	N(0,1)	-0.82
Encompassing F-test	F(3,22)	1.58	F(2,22)	0.82

Test Statistic	Distri-bution	(v)against(vi)	Distri-bution	(vi) against (v)
NT-test	N(0,1)	-3.09	N(0,1)	-1.43
W-test	N(0,1)	-2.48	N(0,1)	-1.29
Encompassing F-test	F(4,21)	4.36	F(3,21)	3.02

APPENDIX B

Johansen Procedure

Consider the following model

$$Z_t=\sum_{i=1}^{k} A_i Z_{t-i}+\epsilon_t, \qquad \text{(AD.1)}$$

where Z_t contains all n variables of the model ϵ_t is a vector of random errors. The operator Δ denotes first differencing. Assume that all the variables in $\mathbf{Z}_t$ are integrated of the same order, and that this order of integration is either zero or one. This assumption is in fact not very restrictive. If some variables which are integrated of order higher than one are of interest, we may consider their appropriate differences, which themselves are integrated of order one, to be included in the VAR model. For simplicity, the deterministic part of ther VAR model (intercepts, deterministic trends, seasonals, etc.) are excluded.

The VAR model (AD.2) can also be represented in the form:

$$\Delta Z_t=\sum_{i=1}^{k-1} \Gamma_i \Delta Z_{t-i}+\Pi Z_{t-k}+\epsilon_t, \qquad \text{(Ad.2)}$$

where:

$\Gamma_i = - \mathbf{I} + \mathbf{A}_1 + \ldots + \mathbf{A}_i$ (I is a unit matrix),

$\mathbf{\Pi} = - (\mathbf{I} - \mathbf{A}_1 - \ldots - \mathbf{A}_k)$.

The equality of models (AD.1) and (AD.2) may be confirmed by adding $\mathbf{Z}_{t-1}$, $\mathbf{Z}_{t-2}$,..., $\mathbf{Z}_{t-k}$ and $\mathbf{A}_1\mathbf{Z}_{t-2}$, $\mathbf{A}_2\mathbf{Z}_{t-3}$,..., $\mathbf{A}_{k-1}\mathbf{Z}_{t-k}$ to both sides of (AD.1) and rearranging.

We will focus our attention on matrix $\mathbf{\Pi}$ of (AD.2) and particularly on its rank. Since there are n variables which constistute the vector $\mathbf{Z}_t$, the dimension of $\mathbf{\Pi}$ is $n \times n$ and its rank can be at most equal to n. It follows from the *Granger Representation Theorem* (see Engle and Granger (1987), or Johansen (1989)) that under some general conditions:

(i) If the rank of matrix $\mathbf{\Pi}$ is equal to n, that is, equal to the total number of variables explained in the VAR model, the vector process $\mathbf{Z}_t$ is stationary (that is all the variables in **Zt** are integrated of order zero);

(ii) If the rank of matrix $\mathbf{\Pi}$ is equal to $r < n$, there exists a representation of $\mathbf{\Pi}$ such that:

$$\mathbf{\Pi} = \alpha\beta', \qquad \text{(AD.3)}$$

where α and β are both $n \times r$ matrices

Matrix β is called the cointegrating matrix and has the property that $\beta' Z_t \sim I(0)$, while $Z_t \sim I(1)$. If the definition of cointegration is recalled, the straightforward conclusion is that the variables in Z_t are cointegrated, with the cointegrating vectors β_1 , β_2 ,..., β_r being particular columns of the cointegrating matrix β. Hence,in a VAR model explaining n variables there can be at most $r = n - 1$ cointegrating vectors.

We may note that model (AD.2) can be regarded as a multivariate generalization of a model in differences with an error correction mechanism. If the hypothesis concerning cointegration holds, that is, if in (AD.2) $\mathbf{\Pi} = \alpha\beta'$, the matrix $\beta' Z_{t-k}$ constistutes a set of r error correction mechanisms separating out the long and short run responses in the model.

For empirical analysis, the essential problems are in the determination of r, that is, in identifying the number of cointegrating vectors, and in estimating the cointegrating matrix β. The procedure outlined below is that of Johansen (1988, 1989). It is numerically quite complicated and we only describe the general principles. It consists of the following steps:

STEP 1: Regress ΔZ_t on ΔZ_{t-1} , ΔZ_{t-2} , ΔZ_{t-k+1}

Since there are n variables to explain in the VAR model, this is equivalent to performing n separate regressions. Construct the $n \times 1$ vector from the residual from each of these regressions at time t, and denote it by R_{0t}. Also, regress Z_{t-k} on ΔZ_{t-1} , ΔZ_{t-2} ,....., ΔZ_{t-k+1}. Construct the $n \times 1$ vector of the residual from each of these regressions at time t, and denote it by R_{kt}.

STEP 2: Compute the four n x n matrices S_{00}, S_{0k}, S_{k0}, and S_{kk} from the second moments and crossproducts of R_{0t} and R_{kt} as:

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it} R'_{jt}\ ,\ i\ ,j = 0, k. \quad (T = \text{sample size}).$$

STEP 3: Solve the equation:

$$|\mu S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0.$$

That is, find the roots or eigenvalues of the polynominal equation in μ obtained from the determinant above. This is a non-standard form of the eigenvalue problem. The solution yields the eigenvalues $\hat{\mu}_1 > \hat{\mu}_2 > \ldots > \hat{\mu}_n$ (ordered from the largest to the smallest) and associated eigenvectors $\hat{v}_i$ which may be arranged into the matrix $\hat{V} = [\hat{v}_1\ \hat{v}_2\ \ldots\ \hat{v}_n]$. The eigenvectors are normalized such that $\hat{V}' S_{kk} \hat{V} = I$. Hall (1989), pp. 216-217) provides some guidance on the computational aspects of these calculations.

If the cointegrating matrix β is of rank $r < n$, the first r eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_r$ are the cointegrating vectors, that is they are the columns of matrix β.

STEP 4: For each μ_i compute the LR statistic:

$$LR = -T. \sum_{i=r+1}^{n} ln\ (1 - \mu_i), \quad \text{(AD.4)}$$

which, under the null hypothesis that there are most r cointegrating vectors, has an asymptotic distribution whose qualities are tabulated by Johansen (1988) and Osterwald-Lenum (1990). Normally testing starts from r = 0, that is from the hypothesis that there are no cointegrating vectors in a VAR model. If this cannot be rejected, the procedure stops since no confirmationi of the existence of cointegrating vectors has been found. If it is rejected, it is possible to examine sequentially the hypotheses that $r \leq 1, r \leq 2$, etc. If the null hypothesis cannot be rejected for, say $r \leq r_0$, but it has been rejected for $r \leq r_0 - 1$, the straightforward conclusion is that the number of cointegrating vectors or, in other words, the rank of β is r_0. Johansen (1989) has shown that the first r estimated eigenvectors $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_r$ are the maximum likelihood estimates of the columns of β, the cointegrating vectors.

So far, we have concentrated on the elements of the matrix β (AD.3), since the columns of β have an economic interpretation as cointegrating vectors. That is, after normalization, they may be interpreted as long run parameters. The elements of α in (AD.3) are straightforward to obtain once the matrix β is known, as the first r columns in the matrix $S_{0k}\hat{V}$. These elements also have an economic interpretation. In general, they measure the speed of adjustment of particular variables with respect to a disturbance in the eqilibrium relation (see

Johansen (1989, p. 16)). Quite appropriately, the matrix α is called the *adjustment matrix* or, since we are dealing with a VAR model where the lagged values of the left-hand side variables enter the error correction mechanism, the *feedback matrix*.

ENDNOTES

. Iports is about 45% of GDP over the period 1960-93.

2. These models considered import demand from the consumer side. Boamah and Craigwell (1993) estimate imports as a factor of production.

4. Engle and Granger (1987) also tabulated critical values. However, these must be carefully interpreted since they were obtained by Monte Carlo simulation for a given DGP for the more than one variable case, Engle and Yoo (1987) provide some critical values for both DF and ADF tests. However, the test for cointegration depends on the number of regresors. The choice of the lag structure in the ADF case can also be of great importance. When the number of variables become important, procedures like those proposed by Johansen (1988) may be preferable.