Econometric tests of rationality and market efficiency

Richard T. Baillie

To cite this Article: Richard T. Baillie , 'Econometric tests of rationality and market efficiency', Econometric Reviews, 8:2, 151 - 186
To link to this article: DOI: 10.1080/07474938908800165
URL: http://dx.doi.org/10.1080/07474938908800165
ECONOMETRIC REVIEWS, 8(2), 151-186 (1989)

ECONOMETRIC TESTS OF RATIONALITY
AND MARKET EFFICIENCY

Richard T. Baillie*
Department of Economics, Michigan State University, East Lansing, Michigan 48824

ABSTRACT

Many economic theories give rise to restrictions between the future rational expectations of a set of variables. This paper describes how such theories can be tested from vector time series models. Particular attention is given to problems of nonstationarity and the use of the concept of cointegration in the modeling and testing procedure.

CONTENTS

1. Introduction
2. Restrictions Between Future Rational Expectations
   2.1 Observable Expectations
   2.2 The Martingale Difference Model of Asset Prices
   2.3 Expectations of the Term Structure
   2.4 Present Value Models
   2.5 The Fisher Equation
3. Testing Rational Expectations Restrictions from Multivariate Time Series Models
   3.1 Considerations of Non Stationarity
   3.2 Estimation of VARs
   3.3 Estimation of MARs
   3.4 Restrictions on MAR and VAR Coefficient Matrices
   3.5 Restrictions Based on Stationary Random Variables
   3.6 Estimation and Testing of the Restrictions
4. Single Equation Tests of the Restrictions
   4.1 Tests with Non Overlapping data

Copyright © 1989 by Marcel Dekker, Inc.
4. Tests with Overlapping data
5. Applications and Economic Interpretation
5.1 Implementation of the VAR Approach
5.11 Spot and Forward Rates Revisited
5.111 Economic Interpretation and Extensions

1. INTRODUCTION

Many economic theories specify relationships between the future rational expectations of a set of variables. Examples, such as the observable expectations hypothesis, the term structure of interest rates under rational expectations and the present value model are discussed in section 2.

This paper develops a general approach for testing such restrictions between the future rational expectations of a set of variables. The approach I advocate for testing these hypotheses is one of using vector autoregressions (VARs) under the maintained hypothesis, having first substituted out any cointegrating relationships to arrive at a vector process which is covariance stationary. The rational expectations restrictions are then expressed in terms of either the vector moving average representation (MAR) or VAR coefficients. Estimation of the process subject to the restrictions can then be easily accomplished and the rationality hypothesis directly tested. I set out the formal results required for this approach in section 3 and compare it with alternatives based on nonstationary processes.

Section 4 then discusses the relationship between the VAR approach and single equation methods. In the final section I provide one numerical example and attempt to assess the usefulness of the VAR approach to testing rationality and market efficiency and its overall contribution to our knowledge of economics. The development of more general alternative hypotheses which embody time varying risk premium for example, can in some cases still be expressed in terms of VAR models, albeit with time dependent conditional covariance matrices.

2. RESTRICTIONS BETWEEN FUTURE RATIONAL EXPECTATIONS

The most general set of restrictions to be considered in this paper is given by the equations

\[ E_{t-h} \sum_{i=1}^{g} \sum_{j=0}^{h} c_{ij} y_{i,t+j} = 0 \]  

\[ i = 1, \ldots, g, \quad h = 0, 1, 2, \ldots \]

where \( E_{t-h}(\cdot) \) represents a Muthian type of rational expectation conditioned on the set of all relevant and available information at time \( t-h \), and the \( c_{ij} \) are a set of parameters. The restrictions are between the conditional
ECONOMETRIC TESTS OF RATIONALITY AND MARKET EFFICIENCY

expectations of the future values of \( g \) random variables \( y_{1t} \) over a possible infinite time horizon, although in many cases this time horizon will be finite. The hypothesis in (1) may be tested against some general unrestricted data generating process such as a vector autoregression; or against a more general model derived from some specific economic theory.

In this paper I will consider alternative approaches to testing the validity of (1) as opposed to finding a solution for \( y_{1t} \). It is well known that even the simplest forward looking rational expectations equation of the form

\[ y_{2t} = \mathbb{E} y_{1t+1} \]

has a multiplicity of solutions; eg, see Gourieroux, Laffont and Monfort (1982), Whiteman (1983), Broze, Gourieroux and Szafarz (1985), Evans and Honkapohja (1986), Evans (1986), etc. Another feature of (1) is that the relationship is specified between the expectations of variables in the future, rather than in terms of the expectations of current or past variables. Inference in the latter case has been dealt with by Revankar (1980), Wallis (1980), Hoffman and Schmidt (1981), Wickens (1982), and others.

Some specific economic examples that give rise to equation (1) are the following:

2.1 Observable Expectations

\[ y_{2t} = \mathbb{E} y_{1t+1} \]  \hspace{1cm} (2)

where the observed variable \( y_{2t} \) is hypothesized to be the expectation of \( y_{1t+1} \) conditional on the set of all relevant and available information at time \( t \). Thus, corresponding to (1); \( g = 2, h = 0 \) and

\[ C_{1j} = \begin{cases} -1 & j = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad C_{2j} = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise} \end{cases} \]

The terminology used to describe (2) can be confusing. The hypothesis states that \( y_{2t} \) is an unbiased expectation of \( y_{1t+1} \) which is a necessary condition for rationality in the sense that no unbiased prediction of \( y_{1t+1} \) at time \( t \) has a smaller variance than that of \( y_{2t} \). Sargent (1973) has considered the concept of partial rationality where all the information used in forming the expectation is used efficiently. Partial rationality is thus a necessary condition for full rationality, where all available information is assumed to be used in an optimal manner. This is familiar as the concept of rational expectations due to Muth (1960, 1961).
Particular examples of (2) are when \( y_{2t} \) is derived from survey data of agents expectations; or in a forward or futures financial market, where \( y_{1t} \) is the logarithm of a spot exchange rate and \( y_{2t} \) is the logarithm of the corresponding forward exchange rate of an \( t \) period maturity time. The hypothesis (2) then implies that the forward rate is formed on the basis of rational expectations with risk neutral agents in the market. Hence (2) is essentially a joint hypotheses with any rejection being due to either the inappropriateness of the assumption of rational expectations and/or the presence of a time varying risk premia. The hypothesis has been variously described as "the unbiasedness of the forward rate," Hansen and Hodrick (1980), "the speculative efficiency hypothesis," Bilson (1981), and "the efficient market hypothesis," Baillie, Lippens and McMahon (1983).

2.11 The Martingale Difference Model of Asset Prices

A particular specialization of (2) concerns the standard model of asset pricing given by

\[
P_t = \beta E_t (p_{t+1} + d_{t+1})
\]

where \( p_t \), the price in period \( t \) depends on expected future price and dividend payments \( d_{t+1} \) next period and \( \beta \) is a constant ex ante real discount rate. Assuming the existence of the transversality condition that

\[
\lim_{n \to \infty} \beta^n E_t p_{t+n} = 0
\]

so that bubble effects are excluded, a standard solution gives

\[
p_t = \sum_{j=1}^{\infty} \beta^j E_t d_{t+j}
\]

In the absence of dividend payments and with \( \beta = 1 \), one solution is of the form

\[
P_{t+1} = p_t + \epsilon_{t+1}
\]

where \( \epsilon_t \) is a serially uncorrelated disturbance term with zero mean and constant variance \( \sigma^2 \). It follows from (4) that \( p_t = E_t p_{t+1} \) and is a special case of (2) with \( \lambda = 1 \).

Subsequent testing of (4) hinges on exploiting the Martingale property of \( \epsilon_t \) which is given by

\[
E_t \epsilon_{t+1} = E(\epsilon_{t+1} | \Omega_t) = 0
\]

where \( \Omega_t \) is the set of all relevant and available information at time \( t \).
The Random Walk version of (4) requires testing the hypothesis that the marginal distribution of \( \epsilon_{t+1} \), i.e. \( f(\epsilon_{t+1}) \), is equivalent to the conditional density

\[
f(\epsilon_{t+1}) = f(\epsilon_{t+1} | t)
\]

Hence the Random Walk model has stronger implications than the Martingale model and excludes the possibility of there being serial dependence within the higher conditional moments of \( \epsilon_t \) such as the Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982).

2.iii Expectations of the Term Structure

A further well known exact linear rational expectation theory concerns the term structure of interest rates between \( r_t \), a one period bond and \( R_t \) an \( n \) period bond. Then the standard linearized rational expectations model of the term structure gives

\[
R_t = \frac{1}{n} \sum_{j=0}^{n-1} E_t [R_{t+j}]
\]

If \( y_{1t} = r_t \) and \( y_{2t} = R_t \) then (5) corresponds to (1) with \( g = 2 \) and

\[
\begin{align*}
C_{1j} &= \begin{cases} 
-1/n & j = 0,1,\ldots,n-1 \\
0 & \text{otherwise}
\end{cases} \\
C_{2j} &= \begin{cases} 
1 & j = 0 \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

It should be noted that similarly to (2), equation (5) specifies a joint hypothesis of the existence of rational expectations and a zero time varying liquidity premium. Rather than test these restrictions on a pair of interest rates it is also possible to consider the restrictions required to hold on a set of interest rates of different maturity times.

2.iv Present Value Models

Another familiar economic theory providing a set of restrictions of the form of (1) is the present value model

\[
p_t = a(1-\beta) \sum_{j=0}^{\infty} \beta^j E_t d_{t+j}
\]

where \( p \) is the price of an asset or stock, \( d \) is the dividend payment, \( \beta \) is the discount factor and \( a \) is a constant of proportionality. The model is also clearly similar to the permanent income theory of consumption under rational expectations. A transformation, which turns out to be useful, is to consider the discounted revisions in successive expectations,

\[
\epsilon_{t+1} = -\beta \sum_{j=0}^{\infty} \beta^j (E_t d_{t+1+j} - E_t d_{t+1+j})
\]
On substitution of (6) into (7), the discounted revisions in expectations variable $\xi_{t+1}$ can be shown to be

$$\xi_{t+1} = -\beta(\alpha(1-\beta))^{-1}(p_{t+1} - \xi_{t+1}), \tag{8}$$

which is thus proportional to the unanticipated change in the asset price.

Campbell and Shiller (1987) have suggested taking the spread $S_t$ which is defined in terms of the differenced price variable $\Delta p_t$ as

$$S_t = \left(\frac{\beta}{1-\beta}\right) E_t \Delta p_{t+1}. \tag{9}$$

The relationship can also be expressed as

$$S_t = a \sum_{j=0}^{\infty} \beta^j E_t \Delta d_{t+j}$$

$$= (p_t - \alpha d_t). \tag{10}$$

On defining

$$y_t = (\Delta d_t, p_t - \alpha d_t) = (y_{1t}, y_{2t})$$

it can be seen that (10) is of the same form as (1) with $g = 2$, $h = 0$ and

$$c_{1j} = \begin{cases} \alpha \beta^j & j = 0, 1, \ldots, \\ 0 & j > 0 \end{cases}$$

$$c_{2j} = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise} \end{cases}$$

One interesting aspect of the above present value relationship as formulated in (10) is that it depends upon the unknown parameter $\alpha$. More will be said of this later.

Before leaving the present value relationship it is worth noting that (6) occurs in many different areas of economics. For example the linearized expectations model of the term structure (5) is also valid for bond rates approaching consol rates in which case (5) becomes

$$R_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t \xi_{t+j}. \tag{11}$$

A further example is the monetary model of exchange rate determination. If $s_t$ is the logarithm of the nominal spot exchange rate, $m_t$ and $m^*_t$ are the logarithms of domestic and foreign money supplies and $y_t$ and $y^*_t$ are the logarithms of domestic and foreign real incomes, and on assuming instantaneous purchasing power parity, uncovered interest parity and stable and static demand for real balances in both countries with the same functional forms and parameter values; then
where $X$ is the and $x_t$ are the forcing variables given by

$$x_t = (m_t - m^*_t) + \alpha(y_t - y^*_t)$$

A solution of (12) is then

$$s_t = \sum_{j=0}^{\infty} \beta^j E_t X_{t+j}$$

where $\beta = \lambda (1 - \lambda)^{-1}$ and is clearly of the same form as (6). See, for example Hoffman and Schlagenhauf (1983) for further details of the derivation of this model and an interesting example of its estimation.

2.v The Fisher Equation

As a final example of (1) I will consider the well known Fisher equation, given by

$$i_t = \mu + E_t \Delta p_{t+1} + \epsilon_t$$

where $i_t$ is the nominal interest rate on a one period security, $\mu$ is the real rate of interest and $E_t \Delta p_{t+1}$ is the expectation at time $t$ of next period's change in the logarithm of the price level. The disturbance term $\epsilon_t$ may be due to variation in the real rate or error in measuring the nominal rate. The presence of the disturbance means that equation (14) does not postulate an exact relationship between future rational expectations and hence is different to equation (1). However, since $\epsilon_t$ is assumed to be a serially uncorrelated process it follows that an exact hypothesis between expectations variables can be found from merely conditioning on information at a previous time period. On conditioning on information available at time $t-1$ equation (14) becomes

$$E_{t-1} i_t - E_{t-1} \Delta p_{t+1} = 0 .$$

3. TESTING RATIONAL EXPECTATIONS RESTRICTIONS FROM MULTIVARIATE TIME SERIES MODELS

The approach I will advocate for testing the restrictions given by (1) is to develop a general, relatively unrestricted model under the maintained hypothesis. This model will include the current and past histories of all the variables entering the restrictions in equation (1) and indeed any other variables thought to make an important contribution to the information set. These variables will generally be dictated by economic theory and constitute a
g dimensional vector $y_t$. Suppose, initially that $y_t$ is a linear, non deterministic, jointly covariance stationary process with Wold decomposition; then following from standard Wiener Kolmogorov theory, $y_t$ will possess a unique infinite order moving average representation (MAR)

$$y_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j},$$

where $\varepsilon_t$ denotes an uncorrelated vector process with $E(\varepsilon_t) = 0$,

$$E(\varepsilon_t \varepsilon_s') = \begin{cases} \Omega & \text{if } s=t \\ 0 & \text{if } s\neq t \end{cases}$$

and $B_0 = I$.

The rational expectation of $y_{t+h}$ made at time $t$ will then be

$$E_{t-h}y_{t+h} = e_i \sum_{j=0}^{\infty} B_{i+h-j} \varepsilon_{t-h-j}$$

where $e_i$ is a $g$ dimensional row vector of zeros, except for unity in its $i$'th element. In this formulation all the information required to form the rational expectation is contained in the current and lagged vector of innovations at time $t-h$. Hence the rational expectation is equivalent to a minimum mean squared error prediction. Restrictions of the form of (1) will then generate sets of constraints on the MAR coefficient matrices.

Two major practical issues dominate the possible implementation of the above approach. One is to find a convenient finite parameterization, the second is how to deal with the likely non stationarity of components of the $y_t$ vector. In the rest of this paper I will argue strongly in favor of using a vector autoregression (VAR) approach. The VAR methodology has been popularized by many authors over the last fifteen years; particularly Sargent and Sims (1977) and Sims (1980, 1982 and 1986). The approach has also aroused substantial criticism, mainly from researchers who are understandably concerned about it replacing structural macro model building techniques. The use of VARs clearly makes assumptions concerning the linearity of the system, the lag lengths and the autocovariances of the data and with some series can face severe problems of overparameterization. However, in this paper I will advocate the VAR approach as being a reasonable strategy with much financial market data, and in some macroeconomic contexts when there is a genuine absence of theory in specifying alternative hypotheses. Furthermore, the VAR methodology is very flexible, relatively easy to handle and generates many useful statistics as by products of the overall procedure.

3.1 Considerations of Non Stationarity

In many practical cases the variables in (1) will be non stationary so that the MAR (16) will not be well defined. Following Engle and Granger
ECONOMETRIC TESTS OF RATIONALITY AND MARKET EFFICIENCY

(1987) a time series is said to be integrated of order \( d \), or \( I(d) \), if it requires differencing \( d \) times to make it stationary. This implies \( d \) repeated unit roots in the autoregressive polynomial of the random variable's univariate time series representation. More generally Engle and Granger (1987) have defined a vector random variable \( y_t \) to be cointegrated of order \((d,b)\); i.e.,

\[ y_t \sim CI(d,b) \text{ if} \]

(a) all the elements of \( y_t \) are \( I(d) \), that is \( y_t \sim I(d) \) and
(b) there exists a matrix \( a \), of cointegrating vectors such that

\[ z_t = a'y_t \sim I(d-b) \text{ and } b > 0. \]

If \( a \) is \( g \times r \) in dimension, and \( r < g - 1 \), there will be \( r \) linearly independent cointegrating vectors.

In the following I will specifically consider the \( d = b = 1 \) case so that the time series vector \( y_t \) is \( I(1) \) and \( \Delta y_t \sim I(0) \) where \( \Delta = 1 - L \) and is the first differencing operator. Then \( \Delta y_t \) will have the MAR

\[ \Delta y_t = C(L)\varepsilon_t \]  \hspace{1cm} (19)

where \( C(L) \) is a possibly infinite order matrix polynomial in the lag operator, \( C(0) \) is lower triangular and \( E(\varepsilon_t \varepsilon_t' ) = I \). On using the identity

\[ C(L) = C(1) + \Delta C^*(L), \]

where \( C^*(L) \) is another matrix polynomial in the lag operator and is of one order lower than \( C(L) \), it then follows that

\[ \Delta y_t = C(1)\varepsilon_t + \Delta C^*(L)\varepsilon_t \]  \hspace{1cm} (20)

If there exists a matrix \( a \) of dimension \( g \times r \) with rank \( r \), such that \( a' C(1) = 0 \), then on premultiplying (20) by \( a' \) gives

\[ a'y_t = z_t = a'C^*(L)\varepsilon_t \]  \hspace{1cm} (21)

and \( z_t \) will generally be \( I(0) \). The \( r \) columns of \( a \) are known as the cointegrating vectors and \( z_t \) represents the vector of equilibrium errors.

Hence a necessary condition for cointegration is that \( C(1) \) has a reduced rank of \( g - r \). Equivalently the spectral density matrix of \( \Delta y_t \) is singular at the zero frequency.

Stock and Watson (1988) have considered an alternative common trends representation where \( y_t \) is the sum of a stationary component and a random walk
of dimension \( g - r \), which is defined as
\[
\tau_t = \tau_{t-1} + J' c_t
\]
(22)
where \( \tau_0 = 0 \), \( C(1) = GJ' \), and \( G \) and \( J \) are matrices of dimension \( gx (g - r) \).
From (20) it follows that
\[
y_t = y_0 + G \tau_t + C^*(L) c_t
\]
(23)
and \( y_t \) is therefore only driven by \( g - r \) unit roots or stochastic trends. Hence model (23) with \( g - r \) unit roots is consistent with \( r \) linearly independent cointegrating vectors.

Corresponding to (16) the autoregressive representation for \( y_t \) will be
\[
A(L) y_t = c_t
\]
(24)
where \( A(1) \) will have a reduced rank of \( r \). Engle and Granger (1987) use the transformation \( A(L) = A(1)L + A^*(L) \) and show that the autoregressive representation for \( A y_t \) will take the so called error correction form
\[
A^*(L) A y_t = - \gamma a' y_{t-1} + c_t
\]
(25)
A particularly important result is that (25) does not permit a standard VAR for \( A y_t \). The inclusion of the error correction term \( \epsilon_t = \gamma a' y_{t-1} \) is crucial in enforcing the long run relationship between the different elements of \( y_t \). Note that (25) only involves I(0) or stationary quantities.

3.1.1 Estimation of VARs

When all the elements of \( y_t \) are I(0), the MAR in (16) may be reasonably well approximated by the finite parameter vector ARMA model:
\[
y_t = \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t + \sum_{j=1}^{q} C_j \epsilon_{t-j}
\]
(26)
It is assumed that all the roots of \( 1 - A_j L^{j} \) and \( 1 + C_j L^{j} \) lie outside the unit circle. The relationship between the \( B_j \) MAR coefficients and the \( A_j \) and \( C_j \) coefficient matrices above is given by Baillie (1987). A parsimonious parameterization of vector ARMA models requires determining appropriate values for \( p \) and \( q \), the maximal lags; and also deciding on which zero restrictions to impose on the \( A_j \) and \( C_j \) coefficient matrices. Such model identification is typically nontrivial; see for example Tiao and Box (1981) and Tiao and Tsay (1983, 1984). For this reason, most empirical studies have employed the vector autoregression (VAR)
where $e_t$ is defined in (17). The VAR($p$) model has the advantage of being relatively simple to estimate. Many applications using (26) have tended to overparameterize on the grounds that this will only marginally reduce asymptotic efficiency of the parameter estimates. Studies by Lutkepohl (1982), Sims (1982) and Doan, Litterman and Sims (1984) have suggested procedures for reducing the parameter space.

For many purposes it is convenient to express (27) as the first order system

$$y_t = A y_{t-1} + \eta_t$$

(28)

where $y'_t = (y'_t, \ldots, y'_{t-p+1})$, $\eta'_t = (e'_t, \ldots, 0)$.

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I & \vdots & \vdots & \vdots & 0 \end{bmatrix}$$

and $I$ is the identity matrix of order $g(p-1)$. Hence the VAR in (27) is simply

$$y_t = N' y'_t = N' A y_{t-1} + e_t$$

(29)

where $N' = [I; 0]$ and is a $g \times gp$ dimensional matrix. Hence in unrestricted form the model has $g^p$ parameters and on using the column stacking operator 'vec', the parameter vector $\theta$ can be expressed as

$$\theta' = \text{vec}(N'A)$$

(30)

When all the elements of $y_t$ are I(0), i.e. stationary quantities, the maximum likelihood estimator $\hat{\theta}$ based on $n$ observations has the well known limiting distribution

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \Omega \otimes \Gamma^{-1})$$

(31)

where

$$\Gamma = \text{plim} \frac{1}{n} \sum_{t=1}^{n} y_{t-1}' y_{t-1}$$

(32)

In some cases it is of interest to consider the joint distribution of $\hat{\theta}$ and $\hat{\omega}$, where the latter are the unique parameters in the error covariance matrix $\Omega$. Then

$$\sqrt{n} \begin{pmatrix} (\hat{\theta} - \theta) \\ (\hat{\omega} - \omega) \end{pmatrix} \sim N(0, \Psi)$$
where

\[ \text{vec} \, \omega = J \omega \]

and \( J \) is a \( g^2 \times g(g+1)/2 \) dimensional selection matrix \( J \) of zeros and ones. For fuller details, see Baillie (1987).

Several recent papers have explored the distribution of parameter estimates of (27) in the presence of various forms of non stationarity. For example Sims, Stock and Watson (1987) consider the VAR model with unit roots and/or polynomial time trends in the companion form matrix (28) and where the elements of \( y_t \) may possibly be cointegrated. They examine the asymptotic distribution of OLS when the regressors are a linear transformation of \( y_t \), and find that OLS are consistent when an intercept is included in the regression. When the regressors are I(0) the parameter estimates converge to their limiting distribution at the usual rate of \( n^{1/2} \) as in (31). With non stationary regressors the parameter estimates tend to their limiting distribution at a rate which is faster than \( n^{1/2} \). Sims, Stock and Watson (1987) also show that while Wald tests tend to the usual asymptotic chi squared distribution when the VAR only includes I(0) variables, the presence of I(1) variables leads to the Wald statistic being a function of multivariate Wiener processes. Similar results for dynamic single equation models are provided by West (1987).

As mentioned at the beginning of section 3 I will advocate the use of VARs to test the relationship given by (1). However, in all five examples presented in section 2 most of the \( y_{t} \) variables will be non stationary and I(1). Early studies that used VARs for testing relationships such as (1) either ignored the problem of non stationarity or else rather arbitrarily differenced the series to reduce them to I(0) variables; eg, see Sargent (1979), Hakkio (1981a and b), Baillie, Lippens and McMahon (1983). With the development of the cointegration literature it is now known that if \( y_{t} \) is I(1) then a VAR in \( \Delta y_{t} \) must include the lagged error correction term as in equation (25). Studies by Melino (1983) and Shiller (1981) although not having the benefit of the theory of cointegration, nevertheless expressed concern about differencing \( y_{t} \).

There are several possible strategies for estimating a VAR on \( y_{t} \) which are I(1) and for subsequently testing restrictions of the form of (1). It is now certainly conceptually possible to estimate an unrestricted VAR in levels and to use the appropriate distribution theory developed by Sims, Stock and Watson (1987) and West (1987). Such distributions are complicated functions of multivariate Brownian motion and are currently computationally non trivial for a high order VAR system (ie large \( p \) and \( g \)).
An alternative approach is to estimate (25) having first identified the $r < g$ cointegrating factors of $z_t$. Engle and Granger (1987) describe how one cointegrating factor can be found and equation (25) then estimated by a two step procedure. Stock (1987) has shown that the OLS estimates of $\alpha$ will be super efficient, i.e. their limiting distribution converges at rate $n$ rather than $n^{1/2}$. Engle and Granger's (1987) procedure is then to first estimate $\alpha$ by OLS and then to use this superefficient estimate of $\alpha$ when estimating the rest of the system. Johansen (1988) has recently developed a technique for testing for $r > 1$ cointegrating factors; see Baillie and Bollerslev (1988b) for an application of the test to a seven dimensional vector of daily exchange rates. They find, on applying Johansen's test, the existence of only one cointegrating factor, which corresponds to a six dimensional random walk process in (22). When following this modeling strategy the estimation of the model in (25) will realize estimates of coefficient matrices that converge at rate $n^{1/2}$ except the coefficient matrix of $y_{t-1}$ which involves some elements that are estimated "superefficiently", Stock (1987), and hence converge at rate $n$.

The third approach, and the one I will principally use in this paper is to substitute out the cointegrating factors and use the factors as separate elements of the $y_t$ vector so that $y_t \sim I(0)$. This procedure requires a priori knowledge of the cointegrating factors and may be available from economic theory and/or from previous statistical testing. The procedure uses cointegration theory and essentially bases tests of (1) on stationary quantities so that the standard asymptotic theory in (31) is appropriate. The approach is clean cut and straightforward and generally appears to give rise to low order VARs and empirically reasonable results. The implementation of this approach inevitably implies a different model under the alternative hypothesis. For example, instead of estimating a VAR on $(y_{1t} y_{2t})$, or $(\Delta y_{1t}, \Delta y_{2t})$ one may use $(\Delta y_{1t}, y_{1t} - y_{2t})$. The relative power of these alternate procedures is unclear and seems worthy of further investigation.

3.11 Estimation of MARs

Although I will test the restriction directly through a VAR model, it is also possible to perform a similar analysis with the MAR coefficients, which contain other information of economic interest. The MAR

$$y_t = \sum_{j=0}^{\infty} B_j^c y_{t-j}$$

for $y_t \sim I(0)$ is an important representation for a stationary time series process from Wiener Kolmogorov theory. If the data generating process is the VAR(p) in (27) with estimated parameter vector $\theta$ and with properties defined by (31) and (32); then on defining
\[ \beta_j = \text{vec} \left( \hat{\beta}_j \right) \]

where the \( j \)th estimated MAR coefficient matrix can be calculated from the estimated VAR coefficient matrices as

\[ \hat{\beta}_j = N A^j N, \]

where \( N = [I:0] \) and was previously introduced after equation (29).

The limiting distribution of the vectorized \( j \)th estimated MAR coefficient matrix is given by Baillie (1987) as

\[ \sqrt{n} \left( \hat{\beta}_j - \beta_j \right) \sim N(0, W) \]

where

\[ W = \sum_{k=0}^{1} B_k \sum B_k' N A^{-1} - k N A^{-1} - k N \]

Although the above provides a proper limiting distribution for the MAR coefficient estimates, so that standard errors are readily available, the covariance matrix will be singular for \( j > p \), so that for higher order MAR coefficients, it will not be possible to provide confidence intervals.

The result in (35) is further complicated when the model (16) is transformed to produce diagonality of its error covariance matrix \( \Omega \), or when the data generating process is a vector ARMA model. The \( B_j \) coefficients have been recognized by Sims (1980) as producing interesting economic information and are known as impulse response weights. Baillie (1987) has shown that transformations of the \( B_j \) coefficient matrices reflect unanticipated effects over different time horizons of different variables within the system.

\[ \text{3.iv Restrictions on MAR and VAR Coefficients} \]

Assuming \( y_t \sim I(0) \) with a MAR given by equation (1), then the rational expectation of \( y_{t+h} \) made at time \( t - h \), is from (18)

\[ E_{t-h} y_{t+h} = \sum_{j=0}^{l} B_{l+h-j} c_{t-h-j} \]

which is the minimum mean squared error prediction at time \( t - h \) of horizon length \( l + h \). In most practical applications \( h \) will be chosen as 0 or 1 so that the test is conditioned on either current information or information lagged one period.

If the true data generating process is assumed to be the VAR \( (p) \) model (27), then the equivalent prediction to (36) will be

\[ E_{t-h} y_{t+h} = e_1 A^{l+h} y_{t-h} \]
so that information used in determining the above minimum mean square error predictor \( L + h \) periods ahead is expressed purely in terms of the last \( p \) values of the vector process \( y_t \). To test the observable expectations hypothesis

\[
y_{2t} = E_t y_{1t+h}, \quad L > 1
\]

expectations can be conditioned on information available at time \( t-h \) to give

\[
E_{t-h}(y_{2t} - y_{1t+h}) = 0, \quad h = 0, 1, 2, \ldots
\]

Then from (37) and with \( h = 0 \)

\[
y_{1t+h} - E_t y_{1t+h} = e_1 \sum_{j=0}^{L-1} B^j e_{t+j}
\]

while the second component of the \( y_t \) vector, \( y_{2t} \) can be expressed as

\[
E_t y_{2t} = e_2 y_t = e_2 \sum_{j=0}^{\infty} B^j e_{t-j}
\]

Under the null hypothesis (2), the unanticipated component, or error associated with the observable expectations hypothesis is given by

\[
u_{t+h} = y_{1t+h} - y_{2t} = e_1 \sum_{j=0}^{L-1} B^j e_{t+j} + \sum_{j=0}^\infty (e_1 B^j e_{t+j} - e_2 B^j e_{t+j})
\] (38)

The first term on the right hand side of the above is the unanticipated component from using the optimal rational expectation, i.e., equation (37); while the second term in (38) reflects the impact of lagged innovations from using a sub optimal expectations scheme. Clearly, for the observable expectations hypothesis (2) to be valid it is necessary that the following restriction is satisfied on the MAR coefficients:

\[
(e_1 B^j e_{t+j} - e_2 B^j e_{t+j}) = 0, \quad j = 0, 1, 2, \ldots
\] (39)

Suppose the VAR(\( p \)) model (18) is used and predictions made from (37) then the restrictions made on the VAR coefficients will be given by

\[
r_h(\delta) = e_1 A^h - e_2 A^h = 0, \quad h = 0, 1, 2, \ldots
\] (40)

which are embodied in the \( gp \) dimensional restriction vector \( r_h(\delta) \).
It is straightforward to see that the above restriction on the VAR coefficient matrices is equivalent to the corresponding MAR restriction (39). From (34)

\[ B_k = N A^k N; \]

where \( N' = [I:0] \) and is of dimension \( g \times gp \), then the restrictions on the \( j \)th MAR coefficient matrices in (39) are given by

\[ e_1 ' N A^{j+1} N - e_2 ' N A^j N = 0 \]

\[ = (e_1 ' A^{j+1} - e_2 ' A^j) \]

and from the definition of \( N; \)

\[ = (e_1 ' A^{j+1} - e_2 ' A^j) \quad j = 0, 1, 2, \ldots \quad (41) \]

which is identical to (39). Hence imposing and testing the rational expectation restrictions on the VAR coefficient matrices will give rise to direct and equivalent restrictions on the MAR coefficient matrices. This is a general property for all linear expectations on VAR coefficients consistent with (1); that is, they imply equivalent conditions on the MAR coefficient matrices.

3.v Restrictions Based on Stationary Random Variables

In practical situations most of the \( y_{it} \) random variables are likely to be non-stationary and one of the modeling strategies discussed in section 3i will have to be employed. As mentioned previously I will concentrate on the third approach where stationary quantities from a priori known cointegrated relationships are substituted in the \( y_{it} \) vector in (16) or (18). Fortunately many of the examples considered do appear to exhibit cointegration.

Of all the relationships like (1) that have been examined, the most extensively analyzed is the observable expectations hypothesis (2); particularly where \( y_{2t} \) is the logarithm of the forward exchange rate \( f_t \) and \( y_{1t} \) is the logarithm of the spot exchange rate \( s_t \). Meese and Singleton (1983) apply the Augmented Dickey Fuller test, while Corbae and Ouliaris (1986) and Baillie and Bollerslev (1988) use the tests of Phillips and Perron to show that the hypothesis of a unit root in the autoregressive polynomial of the spot and forward rate univariate model cannot be rejected. Hence both series are \( I(1) \), and Baillie and Bollerslev (1988) also provide evidence that daily spot and forward rates are cointegrated with a cointegrating factor of (-1 1), so that the forward premium \( (s_t - f_t) \) is \( I(0) \).
The hypothesis (2) states that:

\[ f_t = E_t s_{t+k} \tag{2} \]

which can be expressed as

\[ (f_t - s_t) = E_t (s_{t+k} - s_t) \]

or

\[ (f_t - s_t) = E_t \sum_{j=0}^{\infty} \Delta s_{t+j} \tag{42} \]

Defining

\[ y_t = \{ \Delta s_t, (f_t - s_t) \} = \{ y_{1t}, y_{2t} \} \tag{43} \]

Then if \( s_t \) and \( f_t \) are I(1) and cointegrated with \( (f_t - s_t) \sim I(0) \), both the elements of \( y_t \) are I(0). The restriction is then

\[ E_{t-h} \sum_{j=1}^{\infty} y_{1,t+j} = 0 \quad h = 0,1,2,\ldots \]

which corresponds to (1) with \( g = 2 \) and

\[ c_{1j} = \begin{cases} -1 & j = 1,2,\ldots,2 \\ 0 & \text{otherwise} \end{cases}, \quad c_{2j} = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise} \end{cases} \]

which implies the following gp restrictions on the VAR parameters

\[ r_h(\theta)' = c_2 A^h - c_1 \sum_{j=1}^{\infty} A^{h+j} = 0 \quad h = 0,1,2,\ldots \tag{44} \]

Similarly the restriction corresponding to the MAR (16) is

\[ (e_1 \sum_{k=1}^{\infty} B_{k+j} - e_2 B_j) = 0 \quad \text{for } j = 0, 1, 2,\ldots \]

For practical purposes it appears simplest to deal with the restrictions (44) on the VAR parameters. Note that for the sake of generality I have allowed for \( g > 2 \) equations; this will extend the set of information beyond only the past history of spot and forward rates.

With regard to the rational expectations model of the term structure (5) several authors such as Sargent (1979), Engle, Lilien and Robins (1987), Mankiw and Miron (1986) and Melino (1983) have observed that both short and long term rates, \( r_t \) and \( R_t \) respectively are I(1). However the change in the short term rate \( \Delta r_t \) and the spread \( (R_t - r_t) \) are I(0) if \( r_t \) and \( R_t \) are cointegrated with cointegrating factors \((-1, 1)\).
On subtracting \( r_t \) from both sides of (5) and rearranging

\[
\left( R_t - r_t \right) = \frac{1}{n} \sum_{t=1}^{n-1} \left( r_{t+1} - r_t \right) + \left( r_{t+2} - r_t \right) + \left( r_{t+3} - r_t \right) + \ldots + \left( r_{t+n-1} - r_t \right)
\]

and since

\[
\left( R_{t+k} - r_t \right) = \sum_{j=1}^{k} \Delta r_{t+j}
\]

it follows that (5) can be expressed as

\[
\left( R_t - r_t \right) = E_t \left( \frac{n-1}{n} \Delta r_{t+1} + \frac{n-2}{n} \Delta r_{t+2} + \ldots + \frac{1}{n} \Delta r_{t+n-1} \right)
\]

(45)

Although equivalent to (5), equation (45) is now expressed solely in terms of \( I(0) \) variables. On defining

\[
y_t' = \left( \Delta r_t \left( R_t - r_t \right) \right) = \left( y_{1t} \ y_{2t} \right)
\]

then (45) corresponds to (1) with \( g = 2 \), \( h = 0 \) and

\[
c_{1j} = \begin{cases} -1 & \text{if } j = 1, \ldots, n-1 \\ 0 & \text{otherwise} \end{cases}
\]

\[
c_{2j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases}
\]

Hence the following restrictions will be imposed on the VAR parameters:

\[
\tau_{h} = c_{2} \tau_{h} - c_{1} \sum_{j=1}^{n-1} \left( \frac{n-1}{n} \right) \Delta \tau_{h+j} = 0
\]

(46)

Equations (45) and (46) are similar to (42) and (44) in the sense they specify a joint hypothesis; in this case the existence of rational expectations and the absence of a time varying liquidity premium.

The theory can easily be extended to a set of interest rates where several equations such as (46) can be expected to hold simultaneously. For example, with three series \( r_t \), \( R_{3_t} \) and \( R_{6_t} \) which have one, three and six month maturity times respectively, a simple extension of (45) gives restrictions of the form

\[
\left( R_{3_t} - r_t \right) = E_t \left\{ 2/3 \left( \Delta r_{t+1} + 1/3 \Delta r_{t+2} \right) \right\}
\]

\[
\left( R_{6_t} - r_t \right) = E_t \left\{ 5/6 \Delta r_{t+1} + 2/3 \Delta r_{t+2} + 1/2 \Delta r_{t+3} + 1/3 \Delta r_{t+4} + 1/6 \Delta r_{t+5} \right\}
\]

and

\[
\left( R_{6_t} - R_{3_t} \right) = E_t \left\{ 1/2 \Delta R_{t+1} \right\}
\]
Finally, in the context of the present value model (6) Campbell and Shiller (1987) have shown that prices and dividends are both I(1) and cointegrated. Hence the coefficient $a$ in (10) can be estimated superefficiently and (10) corresponds to the cointegrating relationship. On defining

$$y_t = (\Delta d_t, \Delta t - \Delta d_t)$$

then

$$\gamma_h^{(\beta)} = \epsilon_2^\top A^h \epsilon_1 - \epsilon_1^\top \alpha \sum_{j=0}^{\infty} \beta^j A^j = 0$$

Since the discount parameter $B$ is such that $0 < B < 1$ and since $\lim_{j \to \infty} A^j = 0$ for any $y_t$ that is I(0) it follows that

$$\gamma_h^{(\beta)} = \epsilon_2^\top A^h \epsilon_1 - \epsilon_1^\top \alpha (I - BA)^{-1} = 0$$

The restrictions in terms of the VAR(p) parameters can then be represented by

$$\gamma_h^{(\beta)} = \epsilon_2^\top (I - BA) - \epsilon_1^\top \alpha = 0$$

(47)

3. vi Estimation and Testing of the Restrictions

I now assume that the VAR(p) model (18) has been estimated on a $y_t$ which is I(0) and it is desired to test the validity of the restrictions given by the following hypotheses

$$H_0: \gamma_h^{(\beta)} = 0 \text{ versus } H_1: \gamma_h^{(\beta)} \neq 0 \text{ for some } h = 0, 1, 2, ...$$

where the $\gamma_h^{(\beta)}$ may be of the form (44), (46), or (48) for example.

I assume that maximum likelihood estimates $\hat{\theta}$ are available for the unrestricted model such that

$$\sqrt{n} (\hat{\theta} - \theta) \to N(0, V)$$

In the case of the unrestricted VAR(p), the covariance matrix $V$ is $A \otimes \Gamma^{-1}$ as given by (31). These maximum likelihood estimates $\hat{\theta}$ can then be inserted in (28) to obtain $A$ and since $c_{ij}$ in (1) are known a priori, the corresponding MLE of $\tau(\theta)$ is easily obtained. The appropriate Wald statistic will then be of the form

$$W_h = \gamma_h^{(\hat{\theta})} \{D_h V D_h^\top\}^{-1} \gamma_h^{(\hat{\theta})}$$

where $D_h = \partial \gamma_h^{(\beta)}/\partial \theta$ and $D_h$ and $V$ are both evaluated at $\theta = \hat{\theta}$. Under the
null hypothesis that \( r_h(0) = 0 \), then \( W_h \) will have an asymptotic chi squared distribution with \( gp \) degrees of freedom. For example, to test (44) on spot and forward rates

\[
\begin{align*}
r_h'(0) &= e_2 A^h - e_1 \sum_{j=1}^{L} A^{j+h} = 0
\end{align*}
\]

where \( y'_t = [\delta t_t, e'_t - e_t, y_{3t}, \ldots, y_{gt}] \)

and \( y_{3t}, \ldots, y_{gt} \) include any other economically important variables.

On using a standard result on matrix differentiation, originally due to Schmidt (1973), the matrix of derivatives \( D_h \) is given by

\[
D_h = \begin{bmatrix}
\sum_{j=0}^{h-1} (e_1 \Lambda_j e_2 \Lambda_{h-1-j}) & \sum_{j=0}^{h-k-1} (e_1 \Lambda_j e_2 \Lambda_{h-k-1-j}) \\
\sum_{j=0}^{h-1} (e_2 \Lambda_j e_1 \Lambda_{h-1-j}) & \sum_{j=0}^{h-k-1} (e_2 \Lambda_j e_1 \Lambda_{h-k-1-j}) \\
\vdots & \vdots \\
\sum_{j=0}^{h-1} (e_3 \Lambda_j e_1 \Lambda_{h-1-j}) & \sum_{j=0}^{h-k-1} (e_3 \Lambda_j e_1 \Lambda_{h-k-1-j}) \\
\vdots & \vdots \\
\sum_{j=0}^{h-1} (e_g \Lambda_j e_1 \Lambda_{h-1-j}) & \sum_{j=0}^{h-k-1} (e_g \Lambda_j e_1 \Lambda_{h-k-1-j})
\end{bmatrix}
\]

which is a rather more general result than that given by Baillie, Lippens and McMahon (1983). Similar results for the parametric form of \( D_h \) can be straightforwardly derived for other restrictions; for example see Baillie and McMahon (1985, 1987) for the case of the term structure model where the initial variables are all \( I(0) \). Many of the parametric forms of \( D_h \) are discussed in the prediction literature, for example, see Baillie (1981) and Yamamoto (1982).

For the hypotheses (44) and (46) the constraints vector has the form

\[
r_h'(0) = r_{h-1}'(0) A
\]

so that provided \( A \) is nonsingular it is possible by post multiplication through (51) by \( A \) or \( A^{-1} \) to successively obtain the restrictions conditioned on information at a previous or subsequent time period. In fact the informational content of the restrictions is summarized by the value of \( r(0) \) at any \( h \). While it is possible to recover all the restricted vectors at different time periods, there does not appear to be a simple relationship between the computed Wald test statistics given in (49). Hence it may be possible to reject the hypothesis conditional on information available at one time period, but not at another. When using the Wald statistic it is generally desirable to specify the test at \( h = 0 \) or \( 1 \) so that it is consistent with a particular set of economic information. To premultiply (51) by \( A \) or \( A^{-1} \) may lead to a numerically different value of the statistic; this problem
has been discussed in a related context by Gregory and Veall (1985a) and Phillips and Park (1988).

In some cases the assumption of independent disturbances \( \varepsilon_t \) in the VAR (16) will be inappropriate because time dependent heteroskedasticity will be present. This phenomenon is particularly likely for daily and weekly exchange rates, see Baillie and Bollerslev (1987a), Milhoj (1987) and for interest rates in the term structure equation (46), see Engle, Lilien and Robins (1987). In the presence of conditional heteroskedasticity, White (1984; pp. 133-135) has shown that the appropriate limiting distribution (31) should be replaced with

\[
\sqrt{n} (\hat{\theta} - \theta) \sim N(0, \Omega \otimes \Gamma^{-1} \mid \Gamma_{t-1} \gamma_{t-1})^{-1} \gamma_{t-1} \gamma_{t-1} (\Gamma'_{t-1} \gamma_{t-1})^{-1})
\]

where \( \Omega \) is a diagonal matrix with squared residuals on the diagonal. With no conditional heteroskedasticity the relationship (32) would be valid and the limiting distribution of the VAR parameters reduces to (31). On using the asymptotic covariance matrix in (52) a Wald statistic of the form (49) could be formed in the normal way.

The quantities used to form Wald statistics can also be used to construct asymptotically efficient two step estimates of \( \theta \) under the null hypothesis (1). The restricted estimates \( \hat{\theta}_h \) for the unrestricted VAR(p) model (26) are given by

\[
\hat{\theta}_h = \hat{\theta} - (\Omega \otimes \Gamma^{-1}) D_h [D_h' (\Omega \otimes \Gamma^{-1}) D_h]^{-1} D_h (\hat{\theta})
\]

and using the same arguments as given by Rothenberg (1973), \( \hat{\theta}_h \) can be shown to have the same asymptotic distribution as the restricted maximum likelihood estimator. In particular

\[
\sqrt{n} (\hat{\theta}_h - \theta) \sim N(0, \Omega_h)
\]

where

\[
P_h = (\Omega \otimes \Gamma^{-1}) - (\Omega \otimes \Gamma^{-1}) D_h [D_h' (\Omega \otimes \Gamma^{-1}) D_h]^{-1} D_h (\Omega \otimes \Gamma^{-1})
\]

Calculation of these restricted estimates then allows computation of likelihood ratio test statistics.

A third possibility would be to use the Lagrange multiplier or score test to test the restrictions and this has been implemented for single equation models with contemporaneous rather than future rational expectations by Gregory and Veall (1985b). The use of the score statistic to test restrictions of the form of (48) is not particularly attractive in this case since the model is easily estimated under both the null and alternative hypothesis.
4. SINGLE EQUATION TESTS OF THE RESTRICTIONS

4.1 Tests with Non Overlapping Data

In order to give this discussion a particular focus I will concentrate on the testing of equation (2) where the logarithms of spot and forward exchange rates are involved. Initially, studies chose the sampling interval of the data to exactly match the forecast horizon, i.e., the maturity time of the forward contract. In the following models

\[ s_{t+1} = \alpha + \beta f_t + u_{t+1} \]  

(56)

\[ (s_{t+1} - s_t) = \alpha + \beta (f_t - s_t) + u_{t+1} \]  

(57)

the notion of an unbiased forward rate, where rational expectations and risk neutrality are appropriate is consistent with the restrictions that \( \alpha = 0, \beta = 1, \) and \( u_t \) or \( u_t \) are uncorrelated. The above models were estimated by OLS or occasionally instrumental variables by Cornell (1977), Geweke and Feige (1979), Frenkel (1977, 1979 and 1981), Bilson (1981), Longworth (1981), Gregory and McCurdy (1984) and many others. Bilson (1981), Edwards (1983) and Bailey, Baillie and McMahon (1984) used a SURE framework and tested across several countries simultaneously. At one point equation (56) was one of the most widely used regressions in macroeconomics. Nearly all studies were unable to reject the \( (\alpha = 0, \beta = 1, u_t \) uncorrelated) hypothesis, and this was seen as strong evidence that forward foreign exchange markets were "efficient".

Given our current knowledge, these early studies were consistent with the notion of spot and forward rates being cointegrated, since the residuals were generally close to being uncorrelated and hence I(0), while \( s_t \) and \( f_t \) were invariably I(1) processes. Also the alternative regression (57) now seems more appealing since it can be seen from (25) that it is basically in error correction form with both \( \Delta s_{t+1} \) and \( (f_t - s_t) \) being I(0).

A third regression that was frequently run was of the form

\[ (s_{t+1} - f_t) = \alpha + \beta (s_t - f_{t-1}) + u_{t+1} \]  

(58)

so that the dependent variable is the forecast error and the lagged forecast error appears as an explanatory variable. The null hypothesis is now consistent with \( \alpha = \beta = 0, u_t \) uncorrelated in (58). Clearly (56), (57) and (58) are all equivalent under the null but not the alternative and considerations of power raise the same issues as I discussed at the end of section 3ii with the choice of stationary, nonstationary or cointegrated variables in the \( y_t \) process.

While many studies ground through the above regressions and were regularly able to report finding the forward rate to be an unbiased predictor
of the future spot rate, Gregory and McCurdy (1984) provided some more
critical comments and argued for the importance of checking the statistical
adequacy of the regression equations themselves and also argued in favor of
attempting to specify the \( (s_t f_t) \) process. For monthly data on the
Canadian/U.S. $ exchange rate, Gregory and McCurdy (1984) found substantial
evidence of heteroskedasticity and parameter instability.

Mishkin (1981, 1983) has performed tests in a wide variety of situations
which basically fall into the observable expectations category (2). If \( y_{1t} \)
is hypothesized to be the rational expectation of \( y_{1t+1} \) then Mishkin's approach
has been to estimate a regression of the form

\[
y_{1t+1} - y_{2t} = x_t^\prime \alpha + \xi_{t+1}
\]

and to test that \( \alpha = 0 \) and \( \xi_t \) is uncorrelated. An inability to reject the
null hypothesis leads to the conclusion that the markets expectation has
utilized all the information in \( x_t \). Essentially equation (58) is an example
of this approach. In fact only 'news' on \( x_t \) should be additional information
which is helpful in explaining the market's forecast error and Mishkin (1983)
and Abel and Mishkin (1983) have considered estimation of models of the form

\[
(y_{1t+1} - y_{2t}) = (x_{t+1} - E_{t} x_{t+1}) \beta + \zeta_{t+1}
\]

Hence the \( \beta \) coefficients indicate the relative importance of each item of
"news".

4.1.1 Tests with Overlapping Data

One of the main difficulties in computing regressions of the form of (56)
or (57) on spot and forward exchange rates is the loss of information involved
in making the frequency of observation equivalent to the maturity time of the
forward contract. Brown and Maitai (1981) and Hansen and Hodrick (1980) both
consider models of the form

\[
(y_{1t+k} - y_{2t}) = \psi x_t + u_{t+k}, \quad k \geq 1
\]

where the forecast error \( (y_{1t+k} - y_{2t}) \) is regressed on a set of \( k \) variables
\( x_t \) which represent a set of information known by market participants at time
\( t \). Under the null hypothesis of all information being incorporated into the
markets expectation, \( \psi = 0 \), and the autocovariance structure of \( u_t \) will be
such that

\[
\gamma_u(j) = 0 \quad \text{for } j > k,
\]

where

\[
\gamma_u(j) = \text{E}(u_{t+k} u_{t+k-j}).
\]
The simplest model in accord with (62) is the linear moving average process

\[ u_{t+1} = \varepsilon_{t+1} + \alpha_1 \varepsilon_{t+1} + \ldots + \alpha_{L-1} \varepsilon_{t+1} \]  

or, MA(\(L-1\)), where \( \varepsilon_t \) is a sequence of uncorrelated random variables with mean 0 and variance \( \sigma^2 \). Several authors, namely Hansen and Hodrick (1980) and Brown and Maital (1981), who partly attribute their awareness of the problem to Sims: have noted the inappropriateness of applying GLS to estimating the parameters in models such as (61). In particular many of the variables which an investigator would want to include in the information set \( x_t \) in (61) will be variables such as lagged forecast errors \( (y_{1t} - y_{2t-1}) \). These variables will provide useful information in forecasting future \( y_{1t+1} \) so that some of the elements of \( x_t \) will violate the strict exogeneity requirement of

\[ \mathbb{E}(u_{t+1} | x_{t-1}, x_t, x_{t+1}, \ldots) = 0 \]

Hence GLS will produce inconsistent estimates of \( \psi \) in (61). As noted by Hansen and Hodrick (1980) and Brown and Maital (1981), OLS will realize consistent, albeit inefficient estimates of the regression model parameters. Both sets of authors independently show that the OLS estimates of \( \psi \) will have a limiting distribution of

\[ \sqrt{n}(\hat{\psi}_{OLS} - \psi) \sim N(0, \pi(X'X)^{-1} X'VX(X'X)^{-1}) \]  

where \( X \) is the matrix of observations on the explanatory variables \( x_t \) and \( V \) is an \( nxn \) covariance matrix of the residuals. The estimated asymptotic covariance matrix of the OLS estimator will be obtained by replacing \( V \) by an estimate in (64). In most applications the estimated \( V \) is formed by substitution of sample autocovariances from OLS residuals which are consistent with an MA(\(L-1\)) process for \( u_t \). A standard Wald test can then be used to test \( \psi = 0 \) in (61). This type of methodology has been used in several empirical studies, e.g., Evans and Gulumani (1984), Blake Beenstock, Brasse (1985), Pesaran (1985), etc.

Other methods for estimating (61) by consistent means have been provided by Hayashi and Sims (1983) and Cumby, Huizinga and Obstfeld (1983).

4.iii Extracting Information from VARs

While single equations such as (61) are readily interpretable, they are by no means more informative than a VAR. The important point is to express a VAR in an appropriate manner so that it reveals its relevant properties. For example when \( y_t \) contains \( g \) elements and the observable expectations hypothesis is valid on the first two elements, so that
$y_{2t} = \mathbf{E}Y_{1t+k}$

then from a simple transformation of (26):

$y_{1t+k} - y_{2t} = (e_1 A^t - e_2) Y_t + \sum_{j=0}^{k-1} \mathbf{e}_j B_j \mathbf{e}_{t+j-k}$  \hspace{1cm} (65)

There is a direct analogy between (61) and the transformed VAR in (65). The information set $\mathbf{x}_t$ is now $Y_t$, which includes current and lagged values of all the variables in the $Y_t$ process. Also the error term in (65) is equivalent to $u_{t+k}$ in (61) and is a MA($k-1$) process. The impact and size of different types of 'news' in (65) can be extracted from $u_{t+k}$; see Baillie (1987). The importance of different pieces of information can also be determined by the appropriate coefficient in the $Y_t$ vector. In particular the coefficient estimate $\mathbf{c}$ of $y_{k t-m}$ will be

$e_1 A^t e_{g(m-1)} + k$

and its limiting distribution is given by

$\sqrt{n} (c - e_1 A^t e_{g(m-1)} + k) \sim N(0, e_1 \sum_{j=0}^{k-1} A^j B_j A^{k-j} e_{g(m-1)} + k)$

Hence the coefficient and its asymptotic standard error can be readily determined. This is straightforward and quite an attractive feature of the VAR approach.

5. APPLICATIONS AND ECONOMIC INTERPRETATION

In this concluding section I will discuss some of the problems that arise in the implementation of the VAR testing approach and attempt to summarize some of the economic implications.

5.1 Implementation of the VAR Approach

While the theory discussed in section 3 is straightforward several practical problems such as choice of lag length, form of parameterization, size of information set, etc have only been briefly mentioned. Many possible strategies for the first two issues have been proposed. Doan, Litterman and Sims (1984) propose Bayesian prior schemes, Penm and Terrell (1982) advocate numerically burdensome search routines to find subset autoregressions, and many information criteria; eg Akaike, Schwartz, are also available. Studies by Sargent (1979) and Hakkio (1981a) were forced to choose $p = 4$ in (18) for computational reasons, although residual analysis suggested autocorrelation and hence model underfitting. Baillie, Lippens and McMahon (1983) used likelihood ratio tests on the VAR models order, also the Akaike information criteria and a multivariate portmanteau test for autocorrelation due to
Hosking (1980). Other forms of multivariate score tests by Poskitt and Tremayne (1981) would also be useful in this situation. By completing a range of tests Baillie, Lippens and McMahon (1983) were able to find maximal lag lengths for VARs fairly easily. However many studies using the VAR approach typically took differenced data without a lagged error correction term and typically fitted unrestricted VARs, eg Sargent (1979), Hakkio (1981a, b), Baillie, Lippens and McMahon (1983). The modeling of processes by means of unrestricted VARs on I(1) variables has also been popular in many other macroeconomic applications, eg. Dwyer (1981) when testing the Fisher effect in (7) and Evans (1986) when assessing the impact of budget deficits on the nominal exchange rate.

More recent experiences based on only using variables in \( y_t \) that are I(0) and in some cases represent residuals from cointegrating relationships suggest the order of the approximating VAR is frequently very low; particularly with financial market data.

One form of misspecification occurs when \( p \), the order of the VAR is underestimated. Lewis and Reinsel (1982) have shown that if \( p \) the order of the model goes to infinity at a much slower rate than the sample size, so that as \( p \to \infty \), then \( p^{3/n} = 0 \); and also that if

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=p+1}^{\infty} \text{trace} A_j^t A_j = 0
\]

then the \( g^p \) dimensional parameter vector \( \theta \) in (19), together with a \( g^p \) dimensional row vector \( k(p) \) with finite, bounded elements, will be such that

\[
\frac{\sqrt{n} - p k(p)' (\hat{\theta} - \theta)}{\sqrt{k(p)' (\Omega \otimes I^{-1}) k(p)}}
\]

has an asymptotic standard normal distribution.

Dufour (1985) has provided some interesting results which show that predictions from a misspecified VAR are still unbiased in small samples in the presence of a wide variety of error distributions.

As with virtually all econometric problems, substantial misspecification or omission of an important variable from \( y_t \) can lead to incorrect conclusions when testing hypotheses. When an important variable is omitted from the MAR (16) and tests are based on a MAR for the remaining variables, fairly complicated restriction on B(L) must hold, for the hypothesis tests to be unchanged. The practical importance of this problem is hard to assess.

5.11 Spot and Forward Rates Revisited

Tables 1 and 2 present the results of estimating a VAR(2) model on \( \{ \Delta s_t, f_t - s_t \} \) and \( \{ \Delta s_t, f_t, s_t, t_t, t_t^* \} \) respectively; where \( t_t - t_t^* \) represents the interest rate differential. The domestic country is the US, the foreign country is West Germany, and the data are weekly spot and 30 day
forward rates between April 1973 and June 1980 from the New York foreign exchange market. The interest rates series are Eurobond rates; also at 30 days maturity time. Tests of specification revealed that a VAR(2) model was appropriate for both the bivariate and trivariate case, and that $i - i^*$ was I(0). The data set matches that of Baillie, Lippens and McMahon (1983) who differenced the series. Similar analysis was also undertaken for the French, Swiss, British, Italian and Canadian currencies vis a vis the US dollar in two separate sample periods; 1973 to 1980 and 1980 to 1985. The first period contained 362 observations and the second had 267. The results of estimating bivariate and trivariate VARs revealed 18 models in which the most appropriate value of p was either 1, 2 or 3; although in two instances it was chosen as 4. At time t (i.e., $h = 0$) the restrictions could be strongly rejected in all cases and at time $t - 1$ the restrictions were generally rejected; although always at a much lower significance level. For reasons of space I only report the details for the West German deutschmark against the U.S. dollar. The estimated bivariate and trivariate models are given in Tables 1 and 2 respectively. Table 3 presents results on the importance of each item of information from the trivariate model for West Germany. Overall, the results I obtained from fitting these models gave broadly similar conclusions to previous studies; namely that the joint hypothesis of rational expectations and risk neutrality can be rejected in the context of spot and forward foreign exchange rates. The rejection is particularly marked once conditioning on contemporaneous information takes place.

5.111 Economic Interpretation and Extensions

Careful implementation of the above VAR approach has generally led to the sound rejection of simple theories assuming rationality and the absence of time varying risk premia. This is particularly marked in the spot and forward exchange rates arena where hypothesis (2) was generally supported by the data in original simple analyses, but was rejected by the data in more sophisticated work using overlapping data by Hansen and Hodrick (1980) and adherents of the VAR approach, Hakkio (1981b), Baillie, Lippens and McMahon (1983) and Nobay and Levy (1986). Similarly the hypothesis of rational expectations and constant liquidity premiums in the term structure has also been rejected by the VAR methodology, Attfield and Duck (1982), Baillie and McMahon (1985 and 1987). However, Dwyer (1981) has found evidence for the Fisher effect and in a detailed empirical study Campbell and Shiller (1987) could not reject the present value model given by equations (6) and (10), while a related study by Campbell (1987) examined the permanent income hypothesis under rational expectations. Overall the main achievement of using the VAR approach has been to provide relatively powerful tests and to obtain consistent econometric results over a wide set of data bases.
BAILLIE

Tests of Restrictions in equation (44):

The Unrestricted VAR

\[
\begin{bmatrix}
.346 & .313 \\
(.072) & (.086)
\end{bmatrix}
- \begin{bmatrix}
-.009 & -.201 \\
(.072) & (.085)
\end{bmatrix} L^2
= \begin{bmatrix}
\Delta s_t \\
\epsilon_{1t}
\end{bmatrix}
\]

\[
\begin{bmatrix}
-.054 & -.151 \\
(.061) & (.073)
\end{bmatrix}
- \begin{bmatrix}
.083 & .168 \\
(.061) & (.072)
\end{bmatrix} L^2
= \begin{bmatrix}
\Delta s_t \\
\epsilon_{2t}
\end{bmatrix}
\]

\[Q_2(20) = 13.46 (.76); Q_3(20) = 13.58 (.76), H(20) = 82.33 (.19) \log L = 3344.1\]

The Restricted VAR

\[
\begin{bmatrix}
-.213 & .208 \\
(.047) & (.058)
\end{bmatrix}
- \begin{bmatrix}
.013 & -.078 \\
(.029) & (.032)
\end{bmatrix} L^2
= \begin{bmatrix}
\Delta s_t \\
\epsilon_{1t}
\end{bmatrix}
\]

\[
\begin{bmatrix}
-.058 & -.054 \\
(.035) & (.064)
\end{bmatrix}
- \begin{bmatrix}
.008 & .004 \\
(.017) & (.022)
\end{bmatrix} L^2
= \begin{bmatrix}
\Delta s_t \\
\epsilon_{2t}
\end{bmatrix}
\]

Tests of Restrictions in equation (44):

\begin{align*}
\text{h} = 0 & \quad \text{h} = 1 \\
\text{LR} & \quad 141.46 (.00) \quad 10.10 (.04) \\
\text{W}_{h} & \quad 173.88 (.00) \quad 10.21 (.03) \\
\text{W}_{h}^* & \quad 179.87 (.00) \quad 11.21 (.03)
\end{align*}

\(W_{h}^*\) is adjusted for time varying heteroskedasticity by White's method.

p percentiles are in parentheses following test statistics.

The subsequent question of where this leads us as economists is more difficult to answer. The fact that the assumption of rationality and time invariant risk premium has been overturned for example poses a number of possibilities. The route currently pursued by the profession has been to model risk through using variations of Lucas' (1982) discrete time asset pricing model. There has currently been a development of ARCH related techniques, where conditional second moments can be simultaneously estimated along with structural parameters; see Engle (1982), Engle and Bollerslev


### TABLE 2

**The Unrestricted VAR**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.348</td>
<td>0.322</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.087)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>1 -</td>
<td>-0.013</td>
<td>-0.210</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.086)</td>
<td>(0.120)</td>
</tr>
<tr>
<td></td>
<td>0.036</td>
<td>0.078</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.040)</td>
<td>(0.055)</td>
</tr>
<tr>
<td></td>
<td>0.091</td>
<td>0.177</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.072)</td>
<td>(0.101)</td>
</tr>
<tr>
<td></td>
<td>-0.004</td>
<td>-0.033</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.040)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

L²:

\[ f_t = s_t = \epsilon_t \]

Qₙ(20) = 13.20, Qₛ(20) = 13.75, Qₚ(20) = 34.82, H(20) = 93.48

**The Restricted VAR**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.210</td>
<td>0.207</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.058)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>1 -</td>
<td>0.016</td>
<td>-0.082</td>
<td>-0.088</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.033)</td>
<td>(0.047)</td>
</tr>
<tr>
<td></td>
<td>0.055</td>
<td>-0.068</td>
<td>-0.179</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.040)</td>
<td>(0.063)</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>-0.009</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.030)</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
<td>0.070</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.040)</td>
<td>(0.054)</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.020</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.039)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

L²:

\[ f_t = s_t = \epsilon_t \]

**Tests of Restrictions**

\[ h = 0 \]

\[ LR \quad 142.07(0.00) \quad 11.84(0.06) \]

\[ \chi^2 \quad 174.78(0.00) \quad 11.69(0.06) \]

(1986), Bollerslev, Engle and Wooldridge (1988), etc. This has lead to the development of quite general simultaneous equation models of the form

\[
A(L) y_t = \text{Diag}(H_t) + B(L) \xi_t
\]

\[
\xi_t | \Omega_{t-1} \sim \mathcal{N}(0, H_t)
\]

which is vector ARMA with disturbances with a conditional covariance matrix \( H_t \) whose elements are also allowed to enter the mean. Univariate ARCH or GARCH
The $\psi_i$ weights applied to each piece of information $x_{it}$ in the model:

$$\pi_{bt} = f_t + \sum_{i=1}^{\infty} \psi_i x_{it} + u_{bt}$$

Note: the results are based on the VAR for West Germany in Table 2 with $g = 3$ and $p = 2$. The weights $\psi_i$ are elements of the vector $e_t^A - e_t^c$ from equation (65).

<table>
<thead>
<tr>
<th>Information:</th>
<th>$\Delta s_t$</th>
<th>$(f_{t-s_t})$</th>
<th>$(l_{t-1}^*)$</th>
<th>$\Delta s_{t-1}$</th>
<th>$(f_{t-l_{t-1}}^*)$</th>
<th>$(l_{t-1}^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight:</td>
<td>.002365</td>
<td>-.003353</td>
<td>-.003773</td>
<td>.000085</td>
<td>-.000604</td>
<td>.000914</td>
</tr>
</tbody>
</table>

Parameterizations are typically specified for the elements of $H_t$ and Baillie and Bollerslev (1987b) and Kaminsky and Peruga (1987) have both applied this to spot and forward exchange rates. However only minimal success in finding significant and economically meaningful coefficients consistent with the asset pricing model formulation has been observed. Engle, Lilien and Robins (1987) have applied a single equation version of (66) to the term structure and determined estimates of the liquidity premium. Hence these and other applications provide some interesting extensions of the VAR methodology and allow the possibility of separating out assumptions of rationality and models of risk. While VARs and tests of efficiency are useful first steps in many types of investigation, subsequent developments both in terms of economic theory and ARCH type models strongly suggest the need for future technical developments and refinements, such as models like (66). In short I believe the use of the VAR methodology for testing restrictions like (1) has played a useful role in our testing and understanding the limitations of various economic theories. They can also provide a useful benchmark on forecasting MSE under rational expectations. Subsequent extensions suggest greater rewards in testing theories of risk and informational efficiency independently.

ACKNOWLEDGMENTS

I am very grateful to Esfandiar Masoumi and Dale Poirier for their considerable editorial advice and encouragement during the preparation of this.
article. The paper has also substantially benefitted from the comments of five anonymous referees and particularly from discussions with Peter Schmidt, Rob Engle, Clive Granger, Tim Bollerslev and Patrick McMahon. Any remaining errors are of course my sole responsibility.

REFERENCES


Evans, P. (1986), Is the dollar high because of large budget deficits? Journal of Monetary Economics, 18, 227-249.

ECONOMETRIC TESTS OF RATIONALITY AND MARKET EFFICIENCY


Frenkel, J.A. (1979), Further evidence on expectations and the demand for money during the German hyperinflation. Journal of Monetary Economics, 5, 81-96.


Rothenberg, T.J., (1973), Efficient Estimation with A Priori Information, Yale University Press, New Haven, CT.

Sargent, T.J. (1973), Rational expectations, the real rate of interest and the natural rate of unemployment. Brookings Papers on Economic Activity, 2, 429-472.


Sargent, T.J. and C.A. Sims (1977), Business cycle modeling without pretending to have too much a priori economic theory. In New Methods of Business Cycle Research, edited by C.A. Sims, Federal Reserve Bank of Minneapolis.


West, K.D. (1987), Asymptotic normality when regressors have a unit root. Princeton University discussion paper.

