NBER TECHNICAL WORKING PAPER SERIES

TESTING FOR COMMON FEATURES

Robert F. Engle

Sharon Kozicki

Technical Working Paper No. 91

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1990

We would like to thank the NSF for financial support and gratefully acknowledge helpful comments from Anil Bera, Clive Granger, Niel Ericsson, David Hendry, Jim Stock, Mark Watson and Hal White as well as many others. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #91
October 1990

TESTING FOR COMMON FEATURES

ABSTRACT

This paper introduces a class of statistical tests for the hypothesis that some feature of a data set is common to several variables. A feature is detected in a single series by a hypothesis test where the null is that it is absent, and the alternative is that it is present. Examples are serial correlation, trends, seasonality, heteroskedasticity, ARCH, excess kurtosis and many others. A feature is common to a multivariate data set if a linear combination of the series no longer has the feature. A test for common features can be based on the minimized value of the feature test over all linear combinations of the data. A bound on the distribution for such a test is developed in the paper. For many important cases, an exact asymptotic critical value can be obtained which is simply a test of overidentifying restrictions in an instrumental variable regression.

Robert F. Engle
Department of Economics/D-008
University of California
San Diego, CA 92093

Sharon Kozicki
Department of Economics
University of California
San Diego, CA 92093

COMMON FEATURES

I. Introduction

Economic time series have many distinctive characteristics. Generally, they exhibit serial correlation, trends, seasonality, often heteroskedasticity, skewness, kurtosis and various other features. In order to detect each of these features in a data set, a variety of tests are available each of which takes the particular feature in question as the alternative to a simpler representation under the null.

In this paper tests will be proposed to investigate such characteristics in multivariate data sets. In particular, tests will be presented to examine whether such features are shared in common in sets of time series. More specifically, tests will be developed for the hypothesis that there exists a linear combination of economic time series which does not exhibit the feature.

For example, consider US and Canadian real GNP growth rates. The Box Pierce test for serial correlation using 12 autocorrelations for 126 quarterly data points from 1957 III to 1988 IV is 24.3 for the US and 15.0 for Canada. However the linear combination of Canadian growth minus US growth, is only 7.7 which is well below the 5% critical value for such a test. In fact, the minimum statistic over all linear combinations is 6.1 which is achieved with a coefficient of 1.4. The distribution of such a test statistic will be studied below and its properties obviously depend upon how the linear combination is chosen. But even this simple illustration shows the reduction in the value of a test statistic which is possible.

Both of these series have only moderate amounts of serial correlation so that a better test is a test for zero coefficients in a low order vector autoregression. When the real growth rate in US GNP is regressed on lagged US and Canadian growth rates with one lag, the F–statistic is 6.52 which is significant with a p value of .002. When Canada is regressed on the same variables, the F–statistic is 5.83 which has a p value of .0038. In both cases, there is strong evidence of serial correlation relative to this bivariate information set. However, the minimized value of this F statistic over all linear combinations of growth rates is merely .297 which is far from significant at reasonable levels. The coefficient at which this minimum occurs is 1.02 which suggests real cycles in the US and Canada are common and of the same amplitude.

In contrast, the US and UK real growth rates do not appear to have cycles in common. The F test for no serial correlation against the bivariate US,UK information set is 8.25 for the US and 5.74 for the UK. The minimized value of this statistic over all linear combinations is however only 5.71, slightly below the UK itself. This occurs with a coefficient of −.20 indicating that cycles may even be out of phase.

The statistical model which motivates these tests is the unobserved components model. Two economic time series, y_{1t} and y_{2t} might be generated by the model:

$$(1) \qquad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix} \omega_t + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

where ω is serially correlated but the ϵ's are not. In this case, $y_{1t} - \lambda\, y_{2t}$ will be serially uncorrelated. Notice that no assumption need be made about the contemporaneous covariances among the components. In (1) it is clear that the dynamics of the two series are common. In the US/Canada example, we might think of ω as the business cycle component which is then common to both countries.

The statistical model in (1) does not contain a statement about the process of ω. If it has any detectable feature, such as trends, seasonality, etc. then this feature will be common and can be eliminated by taking a particular linear combination. Thus the same general statistical model can be used to describe the null distribution on many of the common feature tests discussed in this paper.

II. A Simple Testing Procedure

Let s(y) be a test statistic which can be used to detect a particular feature. That is,

H_0: No Feature

H_1: Feature

can be tested with s(y) for a data series $\{y_t\}$. For s and a particular choice of size such as 5%, the decision rule will be defined by a critical region $\{s(y)>c\}$ where c is defined by

$$(2) \qquad P_{H_0}[s(y)>c] \leq 5\%$$

We reject the null hypothesis of no feature or "find a feature" if $s(y)>c$.

Now consider applying this procedure to the variable $u = y_1 - \delta\, y_2$ for various values of δ. The coefficient of y_1 has been normalized to be one which is computationally and notationally convenient although presents difficulties if it should be zero. In the bivariate case this is impossible although in more complicated cases the normalization could be important. The use of general linear combinations could everywhere be substituted in the rest of the paper.

The distribution of the minimand of s(u) over δ when the null hypothesis in (1) is true satisfies a simple inequality:.

$$(3) \qquad s(\hat{u}) \equiv s(y_1 - \hat{\delta} y_2) \equiv \min_{\delta} s(y_1 - \delta y_2) \leq s(y_1 - \lambda y_2) = s(\epsilon_1 - \lambda \epsilon_2).$$

Hence,

$$(4) \qquad P_{H_0}[s(\hat{u})>c] \leq P[s(\epsilon_1 - \lambda \epsilon_2)>c] \leq 5\%$$

By minimizing the test statistic over δ we get a new statistic which again has a size less than or equal to the nominal size. Thus one can simply apply the same test to the minimized value of the statistic and be confident that the null will be rejected less than or equal to 5% of the time when it is true. Of course, this inequality could be very strict, in which case the test might have no power. Thus we must now turn to a discussion of the power of this procedure.

As we have not fully described the joint density of the y's, it is not possible to establish that this class of tests is optimal in any sense. In fact, relative to procedures which have access to this specification, the present test is bound to be inefficient. Nevertheless, we can expect it to have substantial power. Notice that for large samples or very strong features in ω, the test statistic s(u) will diverge to infinity for all values of δ except $\delta=\lambda$ under the null. The question now is how it will behave under the alternative.

Suppose the alternative is that there are two common features:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix} \omega_{1t} + \begin{bmatrix} \phi \\ 1 \end{bmatrix} \omega_{2t} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \quad (5)$$

with $\lambda \neq \phi$ and with ω_1 not perfectly correlated with ω_2. Then every linear combination will contain some ω and the test statistic will diverge with sample size and with non–centrality. This heuristic discussion suggests that the test statistic will be consistent against all alternatives of the form in (5). A more precise discussion will follow the presentation of the most useful version of the tests where asymptotic distributions can be determined.

III. Regression Based Tests for Common Features

Many of the interesting features in economic data can be defined in terms of regression hypotheses. In the model:

(6) $$y_t = x_t\beta + z_t\gamma + \epsilon_t$$

we can define a feature in terms of whether γ is equal to zero. That is:

H_0: $\gamma = 0$, No Feature

H_1: $\gamma \neq 0$, Feature

provides a hypothesis testing basis for detecting a feature in the series $\{y_t\}$. In order to form a test of this hypothesis, we typically will need to add assumptions on the joint distribution of {y,x,z} which insure that an F type of statistic has at least a limiting chi squared distribution.

What are some of the examples of this set–up? To test for serial correlation,{z} would be lags of y_t and possibly other variables, while {x} could be simply a constant and relevant trends or could be other weakly exogenous variables. To test for seasonality, {z} could include seasonal dummies and seasonal lagged dependent variables, while {x} could be other lags as well as trends and other important variables. To test for flexible deterministic trends, {z} would include polynomials in time or a set of piecewise linear trends, and {x} would include the relevant lagged dependent variables. We could even specify the feature as a particular correlation with some other variables. For example, we might be interested in whether a particular stock return has a non–zero coefficient on a set of observable state variables. If so, we might ask whether linear combinations (portfolios) of such stocks can be found which are uncorrelated with all of these state variables. Here {z} would simply be the observed state variables and {x} would be the riskless rate.

If we assume that {x,z} are weakly exogenous or lagged dependent variables with respect to $\{\beta,\gamma\}$ and have conditional expectation with respect to $\{x_t, z_t, \mathcal{F}_{t-1}\}$ given by (6) where $\mathcal{F}_{t-1}$ is the sigma field generated by past values of y, and that {y,x,z} are jointly stationary, then the estimate and covariance matrix of γ can be written, using the notation

$\mathbf{x} = [x_1', \ldots, x_T]'$, and the projection matrix $M_x = I - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$, as

$$(7) \qquad \hat{\gamma} = (\mathbf{z}'M_x\mathbf{z})^{-1}\,\mathbf{z}'M_x\mathbf{y}$$

$$(8) \qquad V(\hat{\gamma}) = \sigma^2\,(\mathbf{z}'M_x\,\mathbf{z})^{-1}$$

so that the test statistic is given by

$$(9) \qquad s(y) = \mathbf{y}'M_x\mathbf{z}\,(\mathbf{z}'M_x\mathbf{z})^{-1}\,\mathbf{z}'M_x\mathbf{y}\,/\,\hat{\sigma}^2$$

where $\hat{\sigma}^2$ is a consistent estimate of the variance of ϵ. Typically,

$$(10) \qquad \hat{\sigma}^2 = e'e/T,$$

where $e_t = y_t - x_t\hat{\beta} - z_t\hat{\gamma}$ for a Wald type of test and $e_t = y_t - x_t\tilde{\beta}$ for an LM version where $\tilde{\beta}$ is the least squares fit of y on x. Notice the simplification if there are no $\{x\}$ in the model. Then the test is simply

$$(11) \qquad s(y) = \mathbf{y}'\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{y}/\,\hat{\sigma}^2$$

which is TR^2 of the regression of y on z using the LM version of the residuals.

Now suppose that two series y_1 and y_2 are each tested for this feature using

$$(12) \qquad \begin{aligned} y_{1t} &= x_t\beta_1 + z_t\gamma_1 + \epsilon_{1t} \\ y_{2t} &= x_t\beta_2 + z_t\gamma_2 + \epsilon_{2t} \end{aligned}$$

where {x,z} are assumed the same for both series. To test whether such a feature is common to these data series, one tests whether there is a δ such that $u_t = y_{1t} - \delta y_{2t}$ which does not have the feature. As before, minimize s(u) with respect to δ to get:

$$(13) \qquad s(\hat{u}) \equiv \min_{\delta} s(y_1 - \delta y_2) = \hat{u}'M_x z(z'M_x z)^{-1} z'M_x \hat{u} / \hat{\sigma}^2$$

where $\hat{u}_t = y_{1t} - \hat{\delta} y_{2t}$, and from straight forward evaluation of the first order conditions, one obtains the closed form solution for $\hat{\delta}$ to be:

$$(14) \qquad \hat{\delta} = [\mathbf{y}_2'M_x\mathbf{z}(\mathbf{z}'M_x\mathbf{z})^{-1}\mathbf{z}'M_x\mathbf{y}_2]^{-1}\, \mathbf{y}_2'M_x\mathbf{z}(\mathbf{z}'M_x\mathbf{z})^{-1}\mathbf{z}'M_x\mathbf{y}_1.$$

Notice that the estimate in (14) is simply the two stage least squares estimate of δ in the model:

$$(15) \qquad y_{1t} = \delta\, y_{2t} + x_t\beta + \epsilon_t$$

where the instrument list is {x,z}. The test in (13) will be called the regression common feature test or *regression cofeature test.* This test is however simply Sargan's(1958)(1959),or Bassman's (1960) test for the validity of instruments, and also Hansen and Singleton's(1983) test for the validity of a rational expectations restriction in a GMM framework. A further interpretation is that (15) and the second equation of (12) describe a structural equation system which incorporates the common feature. The test is then a test for a reduced form restriction which has been called reduced form encompassing.

Here, $\hat{\sigma}^2$ is a consistent estimate of the residual variance which is typically given by $\hat{\sigma}^2 = e'e/T$ where e is the vector of residuals either under the null $(y_{1t} - \hat{\delta} y_{2t} - x_t\tilde{\beta})$ or under the alternative $(y_{1t}-y_{2t}\hat{\delta}-x_t\hat{\beta} - z_t\hat{\gamma})$. Using the residuals under the null, $M_x u = M_x e$ and the test statistic (13) can be computed as TR^2 of the regression of e on {x,z}. The degrees of freedom of the test will be the number of new regressors introduced which is the number of overidentifying restrictions given by the number of variables in {z} (call this K) minus the number of right hand side endogenous variables,(one). This is easily seen in (15) since in 2SLS, y_{2t} is replaced by a linear combination of x_t, z_t so a test of whether z has zero coefficients is only a test that variables 2 through K are significant in explaining y_1 or equivalently e.

The statistic $s(\hat{u})$ will be asymptotically distributed as Chi Square(K−1) when the feature is common, a non–central Chi square otherwise. The tests in section II were shown to have a distribution with critical value less than or equal to Chi Square(K). The regression tests in this section have an exact asymptotic critical value which of course satisfies this inequality.

From the distribution of the test statistic it is clear that the test will have no power if only one instrument is used as every such feature will always appear to be common. This is particularly apparent if the feature is very simple such as a linear time trend where it is well known that a linear combination of two series can always remove such a trend.

IV. Multivariate Regression Tests

The procedures of the previous section will here be generalized to apply to a vector time series, $\{Y_t\}$ where Y_t is Nx1. The multivariate model can be written as:

(16) $$Y_t = Bx_t' + \Gamma z_t' + \mathcal{E}_t$$

where Γ is now a NxK matrix which defines the features to be found in the individual series. If any row of Γ is zero, the corresponding variable is said to not show the feature. If there is a vector δ such that $\delta' Y_t$ does not have the feature, then δ will be called a *common feature vector* or a *cofeature vector*. Any vector with the property that $\delta'\Gamma = 0$ will therefore be a cofeature vector. Linear combinations of cofeature vectors will still be cofeature vectors. If the left null space of Γ has rank r, then there will be r linearly independent cofeature vectors and r will be called the *common feature*,or *cofeature rank*. Conversely if there are r linearly independent cofeature vectors, then the rank of Γ must be N–r. Since Γ is NxK, the reduced rank of Γ is only a restriction if $K \geq N$.

These naming conventions and analytics are obviously and intentionally very similar to those used in cointegration by Engle and Granger(1987). This analogy can be carried even further. If Γ has rank N–r, then it can be written as the product of two matrices of rank N–r with the following dimensions:

$$(17) \qquad \underset{(NxK)}{\Gamma} = \underset{(NxN-r)}{\Lambda} \; \underset{(N-rxK)}{\Phi}$$

and defining $\Phi z_t' = \Upsilon_t$, a N–rx1 vector, each element of which is a linear combination of z_t', equation (16) can be rewritten as:

$$(18) \qquad Y_t - B x_t' = \Lambda \Upsilon_t + \mathcal{E}_t$$

which is a components model with N–r common components which exhibit the feature and a factor or maybe "feature" loading matrix Λ which has rank r. This representation is exactly analogous to the common trends representation of Stock and Watson(1988). Estimation and testing however differ because Υ_t no longer has infinite variance, but is assumed to be a linear combination of observable instruments.

When testing for a common feature in a multivariate setting, the natural generalization of (13) and (14) can be used when the null hypothesis is N–1 common features or a cofeature rank of 1 against the alternative of N common features or a completely unrestricted Γ matrix. In particular, one simply regresses y_1 on the remaining y's(say $\tilde{Y}$), and x treating $\tilde{Y}_t$ as endogenous with a set of instruments {x,z}. If there are fewer than N–1 common features, then the 2SLS regressor matrix will be asymptotically singular since $\hat{\tilde{Y}}$ will have only rank N–r but dimension N–1. Although the regression program may rebel, the residuals will be well defined and the test statistic can be computed. It will as usual have the degrees of freedom of overidentification, however this will be K–(N–r) rather than the apparent K–(N–1). A more natural approach at this point is presumably to apply the methods and test statistics of reduced rank regression which tests the rank of the Γ matrix by testing for zero cannonical correlation coefficients between Y and z. See for example Reinsol(19) or Johansen(1988) and in particular, Tiao and Tsay(1988).

V. Testing for a Serial Correlation Common Feature

One of the most interesting features to test for is serial correlation. In the introduction, a test was formulated using the Box Pierce test statistic to define the serial correlation feature. This test essentially checks to see whether a series is an innovation relative to its own past. A more useful definition for the common feature test is to check whether a series is an innovation relative to a multivariate information set. A series y_1 is said to show serial correlation relative to $\{y_1, y_2\}$, if a test for the the hypothesis $\gamma=0$ is rejected in the model:

$$\begin{aligned} y_{1t} &= \beta_0 + \gamma_{11} y_{1t-1} + \gamma_{21} y_{2t-1} + \dots + \gamma_{1p} y_{1t-p} + \gamma_{2p} y_{2t-p} + \epsilon_{1t} \\ &= x_t\beta + z_t\gamma + \epsilon_{1t} \end{aligned}$$

The variables in x_t would typically be just an intercept or perhaps a trend and the variables in z_t would depend on the multivariate information set and the lag length. The test statistic is given by (9). The version in (11) can be used for demeaned or detrended series.

The estimator of the cofeature vector is given by 2SLS of y_{1t} on y_{2t} after normalizing $\delta_1=1$, and the test for a common feature is (13) which is simply computed as TR^2 of the regression of the 2SLS residuals on $\{z\}$ which is the past of both variables. This will again be asymptotically Chi squared(2p–1) under the null that the feature is common.

Applying this technique to the growth of GNP in the US and Canada with p=1, gives an estimate of:

$$\hat{\delta} = \underset{(.303)}{0.99}, \quad s(\hat{u}) = .610 \sim \text{Chi–square}(1)$$

The coefficient is similar to that obtained by minimizing the Box Pierce even though here only one lag of a bivariate information set is used while the Box Pierce used 12 univariate lags. The test is again unable to reject that the serial correlation is common for these two countries. Applying the same technique to US and UK, the coefficient of the US growth rate is .003 with a standard error of .4, and a test statistic of 10.8 which is highly significant as a Chi–square(1). There does not appear to be a common serial correlation feature between US and UK.

VI. Vector Autoregressions, Causality Tests, and Cointegration

It is useful to investigate the relation between this test procedure, vector autoregressions, and other time series tests such as Granger causality and cointegration. If a multivariate time series Y_t is generated by a finite order vector autoregression, then it can be represented by:

(19) $$Y_t = A_1Y_{t-1} + \ldots + A_pY_{t-p} + \epsilon_t$$

If δ is a cofeature vector, then $\delta'Y_t$ is orthogonal to all past Y. Projecting $\delta'Y_t$ on $\{z_t\}=\{Y_{t-1},\ldots,Y_{t-p}\}$, gives:

(20) $$E(\delta'Y_t|Z_t) = \delta'A_1Y_{t-1} + \ldots + \delta'A_pY_{t-p}$$

which will be zero if and only if $\delta'A_{t-i} = 0$ for all $i=1,\ldots,p$. That is, all the autoregressive matrices have a null space which includes δ. In terms of matrix polynomials in the backshift operator, B, (19) can be concisely expressed as $A(B)Y_t = \epsilon_t$. The condition for a common feature becomes:

(21) $$\delta'A(B) = \delta'A(0) = \delta'$$

so that δ is a left eigenvector for every value of B. If a different normalization were chosen for the VAR, then only the first equalty of (21) would hold; this is sufficient for the common feature restriction.

The vector autoregression in (19) can be rewritten in unobserved components form. Suppose the common feature rank is r and that δ is a Nxr matrix which is a basis for the null space of each of the A_j. Let $\delta^{\perp}$ be the NxN–r orthonormal matrix which completes the basis so that $(\delta,\delta^{\perp})$ is an orthonormal basis which spans $\mathbb{R}^N$. Each of the A's can then be written as the product of $A_j = \delta^{\perp} a_j'$ where a_j is an NxN–r matrix. Now defining

$$\Upsilon_t = \sum_{j=1}^{p} a_j'Y_{t-j}, \text{ gives}$$

$$Y_t = \delta^{\perp} \Upsilon_t + \epsilon_t$$

which is the unobserved components form in (18). If in addition, $a_j = a\phi_j$ where ϕ is an rxr matrix and a is an Nxr vector, then

$$\Upsilon_t = a' \left[\sum_{j=1}^{p} \phi_j Y_{t-j} \right]$$

which is an example of the observable index model of Sargent and Sims(198).

It is now easy to establish that in a vector autoregression with no Granger causality, there cannot be a serial correlation common feature. The assumption of no causality, means that all of the A_j matrices must be diagonal, possibly with zeroes on the diagonal. The only diagonal matrix which gives zero when premultiplied by a non–zero vector δ is the null matrix. However, if all of the A_j's are null, then the individual series will not exhibit the feature. Thus common serial correlation implies at least one way causality.

A rank 1 *cointegrated* system generated by a vector autoregression can be represented using the Engle–Granger(1987) notation as:

$$\begin{aligned} (22) \qquad A(B)Y_t &= A(1)Y_t + A^*(B)(1-B)Y_t \\ &= \gamma\alpha' Y_t + A^*(B)\,\Delta\, Y_t \end{aligned}$$

where α is the cointegrating vector and $-\gamma$ is the vector of coefficients of the error correction term in each equation. $A^*(B)$ is an autoregressive polynomial with $A^*(0) = I - \gamma\alpha'$. Premultiplying by δ' gives

$$\begin{aligned} (23) \qquad \delta' A(B) &= \delta'\gamma\alpha' + \delta' A^*(B)\Delta \\ &= \delta'\gamma\alpha' + \delta'[\overset{*}{A}_0 + \overset{*}{A}_1 B + \overset{*}{A}_2 B^2 + \ldots + \overset{*}{A}_{p-1} B^{p-1}](1-B) \end{aligned}$$

which will in general be a function of B. For δ to be cofeature vector, it is necessary that the highest power of B be zero which means that $\delta' A^*_{p-1}=0$ as this is the coefficient of B^p. By the same argument $\delta' A^*_j = 0 \ \forall \ j = 0,\ldots,p-1$. Because $A(0) - \gamma\alpha' = A^*_0$, this condition becomes $\delta' A(0) = \delta'\gamma\alpha'$. Under the normalization A(0)=I, only if $\alpha'\gamma=1$ is there a solution which is $\delta= \alpha$. It has been shown that only very special cointegrated systems will have a common feature; and if such a system is found, the common feature vector and the cointegrating vector will be the same(when A(0)=I).

To illustrate this case, consider the simple bivariate cointegrating system:

(24) $$y_{1t} = \pi\, y_{2t-1} + \epsilon_{1t}$$
$$y_{2t} = y_{2t-1} + \epsilon_{2t}$$

so the cointegrating vector is $(1,-\pi)$ and the error correction term $(y_{1t-1} - \pi y_{2t-1})$ enters into only the first equation and has a coefficient of 1. This is a special cointegrating system since the error correction term is not only stationary, it is white noise. This additional criterion imposes further restrictions on the system. An example where one might expect such a finding is in modelling forward and spot asset prices which might satisfy equation (24) with y_1 the spot rate and y_2 the forward rate.

An integrated vector autoregressive system which does not have cointegration can still be analyzed for common features, and in fact this is one of the most appealing applications. In such a case A(1)=0 so the vector autoregression is simply computed in the differences. Rewriting the unobserved components structure gives

(25) $$\Delta Y_t = \delta^{\perp}\, \Upsilon_t + \epsilon_t$$

where all series are stationary but Υ is serially correlated. Now integrating ΔY_t gives the common feature representation of Stock and Watson(1988) in terms of Υ but it also gives integrated ideosyncratic components from the cumulation of the ϵ_t. The variables in Y_t are not cointegrated because the ϵ_t are not zero. This interpretation shows that common features provide a measure of the comovement of sets of time series which is weaker than cointegration. Frequently time series which one expects to find are cointegrated, turn out not to be and the current procedure is an alternative measure which may reveal the importance of the common component.

Applying these ideas to the US and Canadian GDP series, the log of Canadian GDP is regressed on log US GDP and a constant getting a coefficient of .806. When a trend is included in this cointegrating regression, the coefficient drops to .664 which is very close to the numbers from the common feature test ($.664^{-1} = 1.50$). The Engle Granger(1987) test for cointegration is based on the residuals of these regressions. Regressing the change in the residuals on the lagged level and higher order lagged changes, no changes were ever significant. The t–statistic on the lagged levels was –2.2 when there was no trend, and –2.1 with trend included. Neither of these are significant at the 5% level according to the Engle and Yoo(1988) tables or more accurate tables by MacKinnon(1990), hence the null hypothesis of non–cointegration cannot be rejected. These tests find no cointegration even though there appears to be a common feature.

VII Testing for Factor ARCH

Thus far, all of the tests have been linear in the dependent variable; however the general approach extends beyond this case. For example, if heteroskedasticity is found in each of two series, one might ask whether it will be found in every linear combination of the series. A particular case is testing for ARCH in a multivariate setting. The FACTOR ARCH formulation specifies a simple covariance matrix which has the property that there will be linear combinations with no ARCH.

The FACTOR ARCH model was proposed in Engle(1987) and used in Engle Ng and

Rothschild(1990)(1989) in examining Treasury Bills and stock portfolios and by Lin(1989) in modelling exchange rate volatility. In its simplest form, a set of random variables Y, with a FACTOR ARCH structure with K factors has a conditional covariance matrix given by

$$E[(Y_t-\mu_t)(Y_t-\mu_t)'|\mathcal{F}_{t-1}] \equiv H_t = \Omega + \sum_{j=1}^{k} g_j g_j' \omega_{jt} \qquad (26)$$

where $E[Y_t|\mathcal{F}_{t-1}]=\mu_t$ and g_j are linearly independent vectors. In this case, the i^{th} element of Y will have time varying variances as long as not all the g_{ji} are zero. If the ω_{jt} are serially correlated, then a simple ARCH test on the i^{th} element of Y will reveal the heteroskedasticity. This would generally be accomplished by regressing y_{it}^2 on a constant and a series of its own lags and computing the TR^2 of the regression. This is the test developed by Engle(1982). An alternative is the Box Pierce test on the squares of y_i. A test which is attractive in this multivariate setting is to regress y_{it}^2 on an intercept and $\{y_{it-m}\, y_{jt-m};\ i\geq j=1,\ldots,N,\ m=1,\ldots,p\}$. Defining this potentially large set of regressors as $\{z_t\}$, and denoting $\mathbf{y_i^2}=[y_{i1}^2,\ldots y_{iT}^2]'$, and $z=[z_1',\ldots,z_T']'$, the test can be expressed as

$$s(y_i) = \mathbf{y_i^{2\prime}} M_x \mathbf{z}(\mathbf{z}' M_x \mathbf{z})^{-1} \mathbf{z}' M_x \mathbf{y_i^2} / \hat{\sigma}^2 \qquad (27)$$

where $M_x = I - \iota(\iota'\iota)^{-1}\iota$ with ι a column of ones, and $\hat{\sigma}^2$ a consistent estimate of the variance. This is just TR^2 of y_i^2 on Z and an intercept.

In the FACTOR ARCH expression in (26), there are linear combinations of Y which have no time varying components. Any linear combination δ with the property that $\delta' g_j=0$ for all j, will have constant conditional variance. Hence the test in (27) for the appropriate linear combination will have no ARCH relative to either a multivariate or univariate information set.

To test for FACTOR ARCH, one simply needs to minimize $s(u)$ in (27) where $\mathbf{u^2} = [(\delta'Y_1)^2, \ldots, (\delta'Y_T)^2]'$

$$(28) \qquad s(\hat{u}) = \min_{\delta} \mathbf{u^{2\prime}} M_x \mathbf{z}(\mathbf{z}' M_x \mathbf{z})^{-1} \mathbf{z}' M_x \mathbf{u^2} / \hat{\sigma}^2$$

While the solution to (28) requires non–linear optimization, the objective function is simple. In terms of the derivative of u_t^2 with respect to δ which can be written as:

$$(29) \qquad \partial\, \mathbf{u}^2 / \partial\, \delta' \equiv \mathbf{U},\ [\mathbf{U}]_{ti} = 2u_t Y_{it}$$

the first order conditions can be written as:

$$(30) \qquad U' M_x \mathbf{z}(\mathbf{z}' M_x \mathbf{z})^{-1} \mathbf{z}' M_x \mathbf{u}^2 = 0$$

with a simple iteration procedure being:

$$(31) \qquad \hat{\delta}^{n+1} = \hat{\delta}^{n} + [U' M_x \mathbf{z}(\mathbf{z}' M_x \mathbf{z})^{-1} \mathbf{z}' M_x U]^{-1} U' M_x \mathbf{z}(\mathbf{z}' M_x \mathbf{z})^{-1} \mathbf{z}' M_x \mathbf{u^2}$$

where U and u^2 are computed from iteration n and used to obtain the change in $\hat{\delta}$ for iteration n+1. As usual, the variance covariance matrix of the last step will give standard errors for $\hat{\delta}$.

Under the null that the data are generated by a Gaussian FACTOR ARCH model, (28) will have a limiting chi square distribution with degrees of freedom equal to the number of overidentifying restrictions. This follows generally from the standard theory of ARCH tests which recognize that under the null, the squared residuals are iid. The result however will be established in a more general setup in the next section.

VII Asymptotic Distribution

A general formulation of the problem includes each of the regression tests described above and many left undescribed. Let $\alpha = (\delta', \beta')$ and consider general non–linear instrumental variables estimators of α in a single equation with true value α_0:

$$F(Y_t, X_t, \alpha_0) = u_t.$$

Assume that there are instrumental variables Z_t including X_t, which satisfy:

(32) $$E[u_t | \mathcal{F}_{t-1}] = 0$$

for $\mathcal{F}_{t-1}$ is the sigma field generated by $\{Z_t, (Y_{t-1}, Z_{t-1}), (Y_{t-2}, Z_{t-2}), \ldots\}$. Further, assume that there exists a sequence, V_T, of uniformly positive definite O(1) matrices with the property that

(33) $$T^{-1/2} V_T^{-1/2} Z'u \overset{A}{\sim} N(0, I).$$

Suppose P_T is a sequence of uniformly positive definite O(1) weighting matrices. Then $\hat{\alpha}$ is the non–linear instrumental variables or GMM estimator if

(34) $$\hat{\alpha} = \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}}\ u'Z\, P_T\, Z'u$$

where $\mathcal{A}$ is a compact parameter set containing the true value α_0. Any interior solution will satisfy the first order conditions

$$\text{(35)} \qquad \mathbf{U'Z}\,P_T\,\mathbf{Z'u} = 0, \quad \mathbf{U} \equiv \partial\,\mathbf{u}/\partial\,a'$$

Under the assumptions in Hansen(1982) or Burguete, Gallant and Souza(1982), or Bates and White(1988) the limiting distribution of the estimator is given by

$$\text{(36)} \qquad \hat{D}_T^{-1/2}\sqrt{T}\,(\hat{a} - a_0) \overset{A}{-} N(0,I)$$

$$\text{(37)} \qquad \hat{D}_T = (U'Z\,P_T\,Z'U/T^2)^{-1}U'Z/T\;P_T\,V_T\,P_T\,Z'U/T\,(U'Z\,P_T Z'U/T^2)^{-1}$$

To optimize this procedure, one must choose both an optimal weighting matrix P_T and optimal instruments. The optimal choice of P_T is V_T^{-1} so that expression (37) simplifies to

$$\text{(38)} \qquad \hat{D}_T = (U'Z\,V_T^{-1}Z'U)^{-1}$$

which is the smallest covariance matrix of all matrices of the form (37). In some cases, optimal instruments can be found following Bates and White(1988). If there exists a possibly stochastic diagonal TxT matrix B_T such that

$$E[u^{*2}_t \mid \mathcal{F}_{t-1}] = 1 \text{ ,where } u^* = B_T u,$$

then the efficient instrumental variables are given by $Z^* = (Z^{*\prime}_1, \ldots Z^{*\prime}_T)'$ where

$$\text{(39)} \qquad Z^*_t = E[U^*_t \mid \mathcal{F}_{t-1}].$$

Often only an estimator of Z^* is available, as it may depend upon unknown parameters.

Using consistent but inefficient estimates of these parameters, it is typically possible to find $\hat{Z}^*$ with the property that $\text{plim}[\hat{Z}^{*\prime}u/T] = 0$.

Finally, the validity of the overidentifying restrictions (32) can be tested by observing that

$$\text{(40)} \qquad \mathbf{u}(\hat{a})'\mathbf{Z}\, V_T^{-1}\, \mathbf{Z}'\mathbf{u}(\hat{a})/T \overset{A}{\sim} \text{Chi Square (rank } \mathbf{Z} - \text{rank } \mathbf{U})$$

VIII. Conclusions

This paper has developed the theory and implementation of methods for testing for common features in a multivariate data set. Obviously, it has only scratched the surface as there are many extensions and applications to be developed. Perhaps the most important future development is to identify important economic problems which fall conveniently into this structure and examine the usefulness of the test results.

REFERENCES

Basmann, R.L.(1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, pp 650–659

Bates, Charles and Halbert White,(1988), "Efficient Instrumental Variables Estimation of Systems of Implicit Heterogeneous Nonlinear Dynamic Equations with Nonspherical Errors", in *Dynamic Econometric Modelling*, (ed. Barnett, Berndt, and White), Cambridge University Press, pp 1–25

Burguete,J.A., R. Gallant and G. Souza,(1982), "On the Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews*, 1: pp151–212

Engle, R.F. (1987), "Multivariate ARCH with FACTOR Structures– Cointegration in Variance", U.C.S.D. Discussion Paper 87–27

Engle, R.F. , V.Ng, M.R. Rothschild, (1990), "A FACTOR ARCH Model of Treasury Bills", forthcoming *Journal of Econometrics*

____________________(1989),"A FACTOR ARCH Model of Stock Returns", U.C.S.D. Discussion Paper

Engle,R.F., and C.W.J. Granger,(1987), "Co–integration and Error Correction:Representation, Estimation and Testing", *Econometrica*, 55,pp 251–276

Hansen,L.P.(1982),"Large Sample Properties of Generalized Method of Moments Estimators",*Econometrica*, 50, pp 1029–1054

Hansen, L.P., and K.J. Singleton,(1983), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*,

Johansen, Soren, (1988) Statistical Analysis of Cointegration Vectors:, Journal of Economic Dynamics and Control, 12, pp231–234

Lin, Wen Ling,(1989), "The Sources of Intra–Daily Volatility in the Foreign Exchange Market– A Multivariate FACTOR GARCH Approach", in *Heteroskedsatic Intra–Daily Volatility in the Foreign Exchange Markets– Estimation, Identification and Forecasting*, Unpublished dissertation, UCSD.

Sargan, J.D. (1958),"The Estimation of Economic Relationships Using Instrumental Variables,", *Econometrica*, 26, pp 393–415

_______________, (1959), "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society B,* 21, pp91–105

Sargent, T.J. and C.A. Sims,(1977) "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory," in *New Methods in Business Cycle Research: Proceedings from a Conference*, ed, C.A.Sims, Minneapolis, Minn. Federal Reserve Bank of Minneapolis

Stock, J.H. and M.W. Watson,(1988), "Testing for Common Trends", *Journal of the American Statistical Association*, 83,pp 1097–1107

Tiao George, and Ruey Tsay, (1987) "Model Specification in Multivariate Time Series

Models," technical report, Carnegie–Mellon University, Dept. of Statistics.

Tsay, Ruey S.,(1989),"Parsimonious Parameterization of Vector Autoregressive Moving Average Models", JOURNAL OF BUSINESS AND ECONOMIC STATISTICS, 7, pp327–315