A Cointegrating VAR Analysis of the Expectations Formation Process of US Output

by

Dimitrios Papaikonomou and Jacinta Pires*

(University of Leicester)

Abstract

Actual and expected output are analysed for the United States within a cointegrating VAR framework over the period 1969q2-2001q2. We make use of real-time data on actual output and direct measures of expectations at different time horizons provided by the Survey of Professional Forecasters. The questions of stationarity of the expectational errors, long-run efficiency, unbiasedness and rationality are addressed taking into account the $I(1)$ properties of the data. We also investigate the persistence of expectational errors, as well as the dynamics involved in the expectations formation process. Asymptotic inference is complemented by finite-sample results obtained with the use of bootstrap methods. We generally find expectational errors to be stationary, efficient, unbiased and consistent with rationality in the long-run, with a persistence of approximately 14 quarters.

JEL Classifications: C15, C32, C42.

Keywords: Output Expectations, Survey Data, Rationality, Long Run Structural VAR, Bootstrap Methods.

*Version dated December 2004. We are most grateful to Professor K.C. Lee for his invaluable help and guidance.

1 Introduction

Economic theory and policy decision-making are both crucially affected by the way that economic agents form their expectations. Direct measures of expectations provide the means by which to study explicitly the expectations formation process and have, therefore, attracted a considerable amount of attention. Examples include, *inter alia*, Lee (1993), Demetriades (1989) and Carlson and Parkin (1975), who have analysed qualitative types of survey data, whereas Baghestani (1994), Ball and Croushore (1998), Fair and Shiller (1989) and Keanu and Runkle (1990) have studied quantitative survey data.

However, most of this literature is primarily concerned with rationality tests that do not account explicitly for the time-series properties of the data. In particular, most of these tests have been carried out under the implicit assumption that the expectational errors are stationary.[1] Furthermore, rationality tests on output expectations, for instance, are frequently carried out using the most recently revised data on actual output, which can be quite dissimilar to the series available when the expectations were formed.[2]

This paper addresses both these issues in an analysis of the expectations formation process of US output. We utilise the multivariate cointegration framework of Johansen (1988) and its subsequent generalization in Pesaran, Shin and Smith (2000) and Pesaran and Shin (2002), in order to simultaneously model US actual output and direct measures of expected output over the period 1969q2-2001q2. Actual output is measured in real time, thus avoiding the criticisms associated with data revisions. Our direct measures of output expectations consist of the current-period, the one, two and three-period-ahead forecasts provided by the Survey of Professional Forecasters. These variables are modelled in a five-dimensional cointegrating VAR with transparent long-run structure and unrestricted short-run dynamic behaviour.[3]

Our primary aim is to study the long-run properties of the expectations formation

[1] One exception is the joint work of Lahiri and Chun (1989), where long-run unbiasedness of expectations is tested using the residual based approach by Engle and Granger (1987).

[2] Patterson and Heravi (1991), for example, illustrate that different vintages of UK GDP components need not even be cointegrated.

[3] For more details on the long-run structural VAR approach see Garratt *et al* (2000, 2003).

mechanism of US output over different time horizons and in particular, to revisit the issues of *(i)* stationarity of the expectational errors, *(ii)* expectational efficiency, *(iii)* unbiasedness and *(iv)* rationality. Our modelling framework further allows us to *(v)* investigate the short-run dynamics of output expectations and *(vi)* gain an insight as to how long the long run actually is. These issues are investigated in a way that is consistent with the time series properties of the data, thus avoiding the pitfalls associated with earlier work in this area. Particular attention is paid to the well-documented finite-sample bias associated with asymptotic inference in cointegrating VAR models and recently developed bootstrap methods are employed in order to simulate finite-sample distributions.[4]

The rest of the paper is organized as follows: section 2 describes the modelling framework, motivates the use of a cointegrating VAR and illustrates how the hypotheses *(i)*-*(iv)* above may be formulated in terms of model parameters; section 3 discusses the empirical findings on the issues *(i)*-*(vi)* and section 4 concludes.

2 Modelling Framework

2.1 Definition of the Expectational Errors

We consider the expectational errors $\eta_{i,t}$, $i = 1, 2, 3, 4$, $t = 1, 2, ..., T$, given by

$$\begin{aligned} \eta_{1,t} &= {}_ty^e_t - {}_{t+1}y_t,\ \eta_{2,t} = {}_ty^e_{t+1} - {}_{t+2}y_{t+1},\ \eta_{3,t} = {}_ty^e_{t+2} - {}_{t+3}y_{t+2}, \\ \eta_{4,t} &= {}_ty^e_{t+3} - {}_{t+4}y_{t+3},\ t = 1, 2, ..., T, \end{aligned} \qquad (1)$$

where ${}_ty_{t-1}$ is the (real-time) actual output in time $t-1$, made available to the respondents of the survey at time t, while ${}_sy^e_p$ denotes the expectation formed at time s on the value of output in time p, $s \leq p$.[5] The expectational errors in (1) may alternatively be written as

$$\eta_{1,t} = {}_ty^e_t - {}_ty_{t-1} - \Delta({}_{t+1}y_t),$$

[4] See for example Cheung and Lai (1993), Mantalos and Shukur (1998) and Greenslade *et al* (2002) on the limitations of asymptotic inference on cointegration rank and Fachin (2000), Johansen (2000a, b) and Jacobson *et al* (2001) for the limitations of χ^2 tests.

[5] All variables are measured in natural logarithms. For more details see the data appendix.

$$
\begin{aligned}
\eta_{2,t} &= {}_ty^e_{t+1} - {}_t y_{t-1} - \Delta({}_{t+1}y_t) - \Delta({}_{t+2}y_{t+1}), \qquad (2)\\
\eta_{3,t} &= {}_ty^e_{t+2} - {}_t y_{t-1} - \Delta({}_{t+1}y_t) - \Delta({}_{t+2}y_{t+1}) - \Delta({}_{t+3}y_{t+2}),\\
\eta_{4,t} &= {}_ty^e_{t+3} - {}_t y_{t-1} - \Delta({}_{t+1}y_t) - \Delta({}_{t+2}y_{t+1}) - \Delta({}_{t+3}y_{t+2}) - \Delta({}_{t+4}y_{t+3}),
\end{aligned}
$$

$t = 1, 2, ..., T$. For the m-vector $\mathbf{z}_t = [{}_ty_{t-1},\ {}_ty^e_t,\ {}_ty^e_{t+1},\ {}_ty^e_{t+2},\ {}_ty^e_{t+3}]'$ the vector of expectational errors $\boldsymbol{\eta}_t = [\eta_{1,t},\ \eta_{2,t},\ \eta_{3,t},\ \eta_{4,t}]'$ can be expressed as

$$\boldsymbol{\eta}_t = \boldsymbol{\beta}'\mathbf{z}_t - \mathbf{A}(L)\Delta\mathbf{z}_t \qquad , t = 1, 2, ..., T, \qquad (3)$$

where $\mathbf{A}(L) = A_1L^{-1} + A_2L^{-2} + A_3L^{-3} + A_4L^{-4}$ with

$$
A_1 = \begin{bmatrix} 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix},\ A_2 = \begin{bmatrix} 0&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix},\ A_3 = \begin{bmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix},
$$

$$
A_4 = \begin{bmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix},\ \text{and } \boldsymbol{\beta}' = \begin{bmatrix} -1&1&0&0&0\\ -1&0&1&0&0\\ -1&0&0&1&0\\ -1&0&0&0&1 \end{bmatrix}. \qquad (4)
$$

As is apparent from expression (3), the statistical properties of $\boldsymbol{\eta}_t$ depend on the properties of $\mathbf{z}_t$. In the event that $\mathbf{z}_t$ is a first difference stationary process, denoted $\mathbf{z}_t \sim I(1)$, then the expectational errors can only be stationary if the linear combination $\boldsymbol{\beta}'\mathbf{z}_t$ is an $I(0)$ process, i.e. $\mathbf{z}_t \sim CI(1,1)$ with cointegrating matrix $\boldsymbol{\beta}$. In the case that $\mathbf{z}_t \sim CI(1,1)$ with cointegrating matrix $\boldsymbol{\delta} \neq \boldsymbol{\beta}$, or in the absence of cointegration altogether, the expectational errors $\boldsymbol{\eta}_t$ would contain stochastic trends. Such a property would cause the first and second moments of $\boldsymbol{\eta}_t$ to be time-dependent and would thus imply a very disappointing performance on the part of economic agents in anticipating future values of output.

Thus, if the worst that economic agents can do is to be systematically wrong by some constant value, then the elements of $\mathbf{z}_t$ should *at least* have to be cointegrated with cointegrating matrix $\boldsymbol{\beta}$. For economic behaviour to be consistent with the Rational

Expectations Hypothesis (REH) forecast errors should also be equal to zero on average, which by (3) implies

$$E[\boldsymbol{\beta}'\mathbf{z}_t] = E[\mathbf{A}(L)\Delta\mathbf{z}_t], \tag{5}$$

where $E[.]$ denotes the expectations operator. In what follows we illustrate how the properties of the process $\boldsymbol{\beta}'\mathbf{z}_t$ and its implications on $\boldsymbol{\eta}_t$ may be investigated within a cointegrating VAR framework.

2.2 Expectational Errors in a Cointegrating VAR Framework

According to, *inter alia*, Pesaran, Shin and Smith (2000), provided that $\mathbf{z}_t \sim I(1)$ with linear deterministic trending behaviour it may be approximated by the following $VAR(p)$

$$\boldsymbol{\Phi}(L)(\mathbf{z}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}t) = \mathbf{e}_t, \qquad t = 1, 2, ..., T, \tag{6}$$

where L is the lag operator, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are m-vectors of unknown coefficients, $\mathbf{e}_t \sim IN(\mathbf{0},\Omega)$, Ω positive-definite and the matrix lag polynomial of order p, $\boldsymbol{\Phi}(L) \equiv \mathbf{I}_m - \sum_{i=1}^{p} \Phi_i L^i$ is allowed to have roots that fall on, as well as outside the unit circle, i.e. $|\boldsymbol{\Phi}(\rho)| = 0$ for $|\rho| \geq 1$. Using the re-parameterization

$$\boldsymbol{\Phi}(L) \equiv -\Pi L + \boldsymbol{\Gamma}(L)(1 - L), \tag{7}$$

where $\Pi \equiv -\boldsymbol{\Phi}(1)$, and $\boldsymbol{\Gamma}(L) \equiv \mathbf{I}_m - \sum_{i=1}^{p-1} \Gamma_i L^i$ with $\Gamma_i = -\sum_{j=i+1}^{p} \Phi_j$, $i = 1, \ldots, p-1$, the $VAR(p)$ given by (6) may be written in VECM form as

$$\Delta\mathbf{z}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \Pi\mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta\mathbf{z}_{t-i} + \mathbf{e}_t, \qquad t = 1, 2, ..., T, \tag{8}$$

where

$$\mathbf{a}_0 = -\Pi\boldsymbol{\mu} + [\Pi + \boldsymbol{\Gamma}(1)]\boldsymbol{\gamma}, \tag{9}$$

$\mathbf{a}_1 = -\Pi\boldsymbol{\gamma}$ and $0 \leq rank[\Pi] = r < m$. Our economic priors that the vector of expectational errors $\boldsymbol{\eta}_t$ defined in (3) follows a stationary process suggest that $\mathbf{z}_t \sim CI(1, 1)$ with cointegrating matrix $\boldsymbol{\beta}$. In that case $r = 4$ and the Π-matrix may be written as $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ is $m \times r$, full column rank and $\boldsymbol{\beta}$ is given by (4). However, this is merely a necessary condition for time-independence of the first and second moments of $\boldsymbol{\eta}_t$ due to the presence of the deterministic trends $(-\Pi\boldsymbol{\gamma})t$.

The implications of (8) on the statistical properties of $\boldsymbol{\eta}_t$ become apparent by looking at the MA representations for $\Delta\mathbf{z}_t$, $\mathbf{z}_t$ and $\boldsymbol{\beta}'\mathbf{z}_t$. These may be obtained from (6) as

$$\begin{aligned} \Delta\mathbf{z}_t &= \boldsymbol{\gamma} + \mathbf{C}(L)\mathbf{e}_t, & (10)\\ \mathbf{z}_t &= \boldsymbol{\mu} + \boldsymbol{\gamma}t + \mathbf{C}(1)\sum_{i=1}^{t}\mathbf{e}_i + \mathbf{C}^*(L)\mathbf{e}_t, & (11)\\ \boldsymbol{\beta}'\mathbf{z}_t &= \boldsymbol{\beta}'\boldsymbol{\mu} + (\boldsymbol{\beta}'\boldsymbol{\gamma})t + \boldsymbol{\beta}'\mathbf{C}^*(L)\mathbf{e}_t, \qquad t = 1, 2, ..., T, & (12) \end{aligned}$$

where $\mathbf{z}_0 \equiv \boldsymbol{\mu} + \mathbf{C}^*(L)\mathbf{e}_0$, $\mathbf{C}(L) \equiv \sum_{i=0}^{\infty} C_i L^i = \mathbf{C}(1) + (1-L)\mathbf{C}^*(L)$, $C_0 = I_m$, $C_1 = \Phi_1 - I_m$, $C_i = \sum_{j=1}^{i} \Phi_j C_{i-j}$, for $i > 1$, $\mathbf{C}^*(L) \equiv \sum_{i=0}^{\infty} C_i^* L^i$, $C_0^* = I_m - \mathbf{C}(1)$, $C_i^* = C_{i-1}^* + C_i$, for $i > 0$, and according to *Granger's representation theorem* the cumulative effect matrix $\mathbf{C}(1)$ may be expressed as[6] $\mathbf{C}(1) = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma}(1)\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}'_\perp$, with $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_\perp$ being $m \times (m-r)$, full column rank such that $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \mathbf{0}$ and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$. Combining (3), (10) and (12) the MA representation of the expectational errors $\boldsymbol{\eta}_t$ takes the form of

$$\boldsymbol{\eta}_t = \boldsymbol{\beta}'\boldsymbol{\mu} + (\boldsymbol{\beta}'\boldsymbol{\gamma})t + \boldsymbol{\beta}'\mathbf{C}^*(L)\mathbf{e}_t - \mathbf{A}(1)\boldsymbol{\gamma} - \mathbf{A}(L)\mathbf{C}(L)\mathbf{e}_t, \tag{13}$$

$t = 1, 2, ..., T$, where $\boldsymbol{\gamma}$ is the expected growth rate of $\mathbf{z}_t$ according to (10). This expression immediately reveals that under the assumptions discussed above $\boldsymbol{\eta}_t$ has a time-varying mean given by

$$E[\boldsymbol{\eta}_t] = \boldsymbol{\beta}'\boldsymbol{\mu} + (\boldsymbol{\beta}'\boldsymbol{\gamma})t - \mathbf{A}(1)\boldsymbol{\gamma} \qquad , t = 1, 2, ..., T. \tag{14}$$

As mentioned earlier, the hypothesis of long-run R.E. requires that $E[\boldsymbol{\eta}_t] = \mathbf{0}$, which by expression (14) is true in the current framework when

$$\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}, \tag{15}$$

also known as *the co-trending hypothesis*, and

$$\mathbf{A}(1)\boldsymbol{\gamma} = \boldsymbol{\beta}'\boldsymbol{\mu}, \tag{16}$$

which can be regarded as a *long-run unbiasedness hypothesis*. Provided that the cointegrating matrix $\boldsymbol{\beta}$ is given by (4) and for $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5]'$, co-trending is equivalent

[6] See Johansen (1995; Theorem 4.2; pp.49-52)

to $\boldsymbol{\gamma} = [\gamma_1, \gamma_1, \gamma_1, \gamma_1, \gamma_1]'$. In other words, the restrictions in (15) suggest that the individual series in $\mathbf{z}_t$ are not only driven by a common stochastic trend, but they also share the same deterministic trend with the real-time output series, ${}_ty_{t-1}$. Given (4) and (15), condition (16) provides a test of rationality in the long run. Noting that $\mathbf{a}_0$ takes the form of (9), and given (4), (15) and $\hat{\gamma}_1$, the restrictions in (16) may be expressed in terms of model parameters as

$$\mathbf{A}(1)\boldsymbol{\gamma} = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'[\boldsymbol{\Gamma}(1)\boldsymbol{\gamma} - \mathbf{a}_0]. \tag{17}$$

3 Empirical Analysis

In this section we apply the econometric tools discussed above in order to investigate the question of stationarity of the errors in the expectations of US output, as well as the empirical validity of the closely related concepts of rationality, efficiency and unbiasedness of output expectations. We make use of real-time data on actual output and direct measures of expectations for current period's output, the one, two and three-period-ahead forecasts, available on a quarterly basis over the period 1968q4-2001q2. Our empirical analysis is based on the VECM given by (8), augmented by the dummy vector $\mathbf{D}_t = [d71q4_t, d74q4_t, d81q1_t, d96q1_t, d99q4_t]'$, where $d71q4_t$ takes the value of one in 1971q4 and zero otherwise and the remaining variables are similarly defined. The dummies $d71q4_t$ and $d74q4_t$ are intended to capture the slowdown in economic activity at the end of 1971 and 1974, while $d81q1_t$, $d96q1_t$ and $d99q4_t$ are controlling for the effects of the comprehensive revisions in GDP that took place in December 1980, December 1995 and August 1999.

3.1 Statistical Properties of the Data

As mentioned in the previous section the econometric framework of (8) requires that the vector $\mathbf{z}_t = [{}_ty_{t-1}, {}_ty^e_t, {}_ty^e_{t+1}, {}_ty^e_{t+2}, {}_ty^e_{t+3}]'$ is an $I(1)$ process with linear deterministic trending behaviour. The requirement $\mathbf{z}_t \sim I(1)$ has been investigated in a model-consistent way with the use of multivariate ADF tests, reported in Tables 1 and 2. Table 1 clearly indicates that stationarity is rejected for all elements of $\mathbf{z}_t$, irrespective of the number of cointegrating relations, while Table 2 shows that $\Delta \mathbf{z}_t$ is a stationary process.[7] As illustrated, for example, in Pesaran, Shin and Smith (2000), the hypothesis that $\mathbf{z}_t$ contains only linear deterministic trends takes the form of the restriction of the trend coefficients according to $\mathbf{a}_1 = -\Pi\boldsymbol{\gamma}$ in (8). The LR test of this hypothesis is asymptotically χ^2 with $m - r$ degrees of freedom. Being agnostic at this stage regarding the empirically relevant value of r we carried out this test for all possible choices, $r = 0, ..., 4$, within a $VAR(2)$. The results are reported in Table 3 and indicate that the restriction of the trend coefficients according to $\mathbf{a}_1 = -\Pi\boldsymbol{\gamma}$ cannot be rejected, irrespective of the choice of r. The clear asymptotic evidence in favour of the trend restrictions appear to be particularly strong, taking into account the fact that χ^2 tests within the current framework are known to be biased in favour of rejection in samples of the size considered here. The lag-length, p, was set equal to 2 after testing for significance of additional lags within an unrestricted $VAR(8)$ in the level of $\mathbf{z}_t$, as well as with the use of the AIC and SBC. Empirically, we have found that $p = 2$ is sufficiently long to remove all traces of residual serial correlation.

3.2 Properties and Estimates of the Expectational Errors

Having found strong evidence in favour of the aforementioned assumptions on the statistical properties of $\mathbf{z}_t$, we may safely proceed with the estimation of the VECM given by (8). As discussed in the previous section, we expect to find four cointegrating relations among the elements of $\mathbf{z}_t$. Table 4 reports the $\lambda - trace$ and *maximal eigenvalue* cointegration rank statistics. Due to the well-documented tendency of these tests to over-reject the null in finite samples we follow a large number of authors and complement asymptotic results with simulated finite-sample critical values.[8] In general, we find that the simulated critical values exceed the asymptotic ones, thus verifying the presence of a finite-sample bias in favour of rejection reported, for example, by Cheung and Lai (1993), Harris and Judge (1998) and Greenslade *et al* (2002). Nevertheless, this appears to make no qualitative difference to the outcome of the tests, as all the evidence appear to be clearly in favour

[7]These results are also verified by the application of more standard univariate ADF and Phillips-Perron unit root tests, available on request.

[8]Details on the simulation of the finite-sample distributions of all tests can be found in Appendix B.

of our economic priors that $r = 4$.

Having obtained consistent evidence regarding the number of the cointegrating relations, we next turn our attention to the form of these four relations, which determines the properties of the expectational errors of US output. Noting that the deterministic trend in (8) enters the cointegrating space due to $\mathbf{a}_1 = -\Pi\boldsymbol{\gamma}$, it is possible to express the cointegrating relations as $\boldsymbol{\beta}'_*\mathbf{z}_{*t}$, where $\boldsymbol{\beta}'_* = [\boldsymbol{\beta}', -\boldsymbol{\beta}'\boldsymbol{\gamma}]$ and $\mathbf{z}_{*,t} = [\mathbf{z}'_t, t]'$. Table 5 reports the tests of three sets of over-identifying restrictions on $\boldsymbol{\beta}'_*$. The first set, denoted R_{OV1}, corresponds to the hypothesis commonly referred to as *long-run efficiency* and restricts all entries in the first column of $\boldsymbol{\beta}'$ to -1, while leaving $-\boldsymbol{\beta}'\boldsymbol{\gamma}$ unrestricted. The test of R_{OV1} leads to asymptotic non-rejection, indicating that $\boldsymbol{\beta}'$ is indeed consistent with (4). The set of restrictions denoted as R_{OV3} corresponds to the joint hypothesis of long-run efficiency and co-trending, the latter being defined by expression (15). Asymptotically this joint hypothesis is clearly rejected. With reference to the simulated finite-sample distribution, however, rejection is avoided at the 5% level, although not at the 10% level. In the light of the evidence reported in, *inter alia*, Fachin (2000), Johansen (2000a, b) and Greenslade *et al* (2002) regarding the finite-sample bias of such tests, it would appear risky not to put more weight on the simulation results. Stronger evidence is obtained for the hypothesis that the trend coefficients are absent from the cointegrating relations corresponding to $\eta_{1,t}$, $\eta_{2,t}$ and $\eta_{3,t}$ alone, denoted as R_{OV2}. These results indicate that the joint hypothesis of efficiency and co-trending appears to be consistent with the data as far as $\eta_{1,t}$, $\eta_{2,t}$ and $\eta_{3,t}$ are concerned. Regarding $\eta_{4,t}$ the evidence appear to be less conclusive and raise the suspicion that it could possibly follow a near trend-stationary process.

If one is prepared to maintain the joint hypothesis of long-run efficiency and co-trending, then the VECM can be estimated subject to R_{OV3}, which yields the estimates reported in Table 6. All equations appear to have reasonable explanatory power and the diagnostic statistics do not reveal any signs of model misspecification. The estimated expectational errors $\hat{\eta}_{i,t}$, $i = 1, 2, 3, 4$, are reported below Table 6.[9] As illustrated in previous sections, given R_{OV3}, the expectational errors will be consistent with long-run rationality

[9]The constants $\hat{\boldsymbol{\beta}}'\boldsymbol{\mu}$ have been retrieved from the intercepts $\hat{\mathbf{a}}_0$ using relation (9) as $\hat{\boldsymbol{\beta}}'\boldsymbol{\mu} = (\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}})^{-1}\hat{\boldsymbol{\alpha}}'[\hat{\boldsymbol{\Gamma}}(1)\hat{\boldsymbol{\gamma}} - \hat{\mathbf{a}}_0]$ for $\hat{\gamma}_1 = 0.0075176$.

if the hypothesis of long-run unbiasedness in (16) holds. In terms of our estimates this means that the constant appearing in $\hat{\eta}_{i,t}$ will have to equal $i\hat{\gamma}_1$, $i = 1, 2, 3, 4$, in absolute value, where $\hat{\gamma}_1 = 0.0075176$, is the average growth rate of ${}_t y_{t-1}$ in our sample.[10] An inspection of the estimates indicates that this hypothesis approximately holds. This is more formally verified by a test of (17), as the relevant Wald statistic can be found to be 2.691 with an asymptotic 5% critical value of 9.49. However, it should be stressed that the empirical support for long-run rationality provided by this test rests on the joint assumption of long-run efficiency and co-trending, for which the evidence were less convincing as far as $\eta_{4,t}$ is concerned.

3.3 Dynamic Behaviour of Output Expectations and Expectational Errors

In this section we illustrate the dynamic properties of the estimated VECM by evaluating *(i)* the speed with which expectational errors are eliminated in response to system-wide shocks and *(ii)* the dynamic response of real-time output and output expectations to news in real time output.

The first issue is addressed with the use of the Persistence Profiles (PP) introduced by Lee *et al* (1992) and Lee and Pesaran (1993). Unlike impulse responses, PPs are unique and do not depend on whether the shock under consideration is "structural" or reduced-form.[11] These measures have been computed for the estimated expectational errors $\hat{\eta}_{i,t}$, $i = 1, 2, 3, 4$, and are plotted in Figure 1. All profiles converge to zero, thus confirming the stationary nature of the expectational errors. They reveal that a unitary forecast error is eliminated slightly faster for the current period forecast and progressively slower as the forecast horizon increases. In particular, $\hat{\eta}_{1,t}$ is eliminated in a monotonic fashion with 95% of the adjustment process being completed within 10 quarters and 99% within 13 quarters. For $\hat{\eta}_{2,t}$, $\hat{\eta}_{3,t}$, and $\hat{\eta}_{4,t}$ the initial error increases during the first quarter by 24 to 32 percent. Thereafter, it is eliminated monotonically with 95% of the process being

[10] The reported value for $\hat{\gamma}_1$ has been obtained after excluding the observations corresponding to the dummies in $\mathbf{D}_t$. The full-sample average growth rate of ${}_t y_{t-1}$ can be found to be 0.0080834. All our results remain virtually unchanged using either estimate.

[11] For more details see also Pesaran and Shin (1996) and Pesaran and Smith (1998).

completed within 11 quarters and 99% within 14 quarters.

In order to address the dynamic response of the elements of $\mathbf{z}_t$ to an innovation in real-time output we employ the Generalized Impulse Responses (GIR) introduced by Koop *et al* (1996) and Pesaran and Shin (1998), which in this case are exactly equivalent to Sims' (1980) standard Orthogonalized Impulse Responses (OIR).[12] This means that the innovation considered here can be viewed both as: *(a)* a *typical* shock to ${}_ty_{t-1}$ by historical standards as measured by the system's covariance matrix, and *(b)* an *orthogonal* innovation in ${}_ty_{t-1}$ identified by imposition of the causal chain implied by the ordering of the variables in $\mathbf{z}_t$.[13] Figure 2 plots the responses to a one standard error shock in ${}_ty_{t-1}$, which are all found to converge to a common, non-zero value. The fact that this value is different from zero illustrates the $I(1)$ properties of the variables in $\mathbf{z}_t$ that cause a given shock to have a permanent effect. The fact that this value is the same for all variables illustrates the stationary nature of expectational errors, as the gap between actual and expected output is eventually eliminated. The impact effect of the shock is to raise real-time output by 0.62%, whereas the increase in output expectations is larger and ranges between 0.668% and 0.697%. As a result, the shock leads to over-prediction on impact. All variables continue to rise over the next 3-4 quarters, having overshot their long-run value in the first quarter. During this climb the positive expectational errors experienced on impact become negative around quarter 2. Consequently, the medium run is dominated by under-prediction that lasts until approximately quarter 14, when all expectational errors are eliminated.

[12]This is because real-time output, ${}_ty_{t-1}$, is the first entry in $\mathbf{z}_t$. For a proof of the fact that GIR = OIR when considering a shock in the first variable of the system see Pesaran and Shin (1998).

[13]In the first case the responses are unique and invariant to the ordering of the variables in $\mathbf{z}_t$, whereas in the second, one is committed to a lower-triangular contemporaneous matrix, and thus to a unique ordering of the variables. We believe the causal chain implied by the ordering of the variables in $\mathbf{z}_t$ to be realistic, however, we appreciate that this can be a matter of opinion. Should the reader object, he/she can always interpret the responses as if referring to a typical shock in real-time ouput.

4 Concluding Remarks

This paper illustrates how actual output and direct measures of output expectations over different time horizons may be studied simultaneously within a multivatiate cointegration framework. There is already a rich literature on the use of survey data for the investigation of the expectations formation mechanism and its consistency with the rational expectations hypothesis. However, very little work has been done on the simultaneous modelling of actual and expected series and most of the existing research ignores the cointegrating properties of the data.

In this study we have addressed these issues within a cointegrating VAR framework. This has allowed us to formulate and test the hypotheses of stationarity of expectational errors, expectational efficiency, unbiasedness and rationality in the long run, in a way that is consistent with the $I(1)$ properties of the data. Our findings have confirmed the quite frequent assumption of stationary expectational errors. Furthermore, all expectations may reasonably be argued to be efficient, unbiased and consistent with rationality in the long run, although the evidence is weaker in the case of the three-period ahead forecast. Our modelling framework has also enabled us to gain a useful insight on the dynamic properties of the expectations formation process. More specifically, we have found that expectational errors persist for approximately 13-14 quarters, with the adjustment process being slower the longer the forecast horizon. Also, a positive shock in actual output was shown to result on impact in an over-prediction that lasts for approximately two quarters. This is followed by a sustained period of under-prediction that stretches over approximately 3 years before expectational errors are eliminated.

A Data Appendix

Our direct measures of output expectations are provided by the Survey of Professional Forecasters. The SPF is a quarterly source of survey data on several economic variables for the US economy and is provided by the Federal Reserve Bank of Philadelphia. It has been carried out since 1968q4. The distinctive feature of this source is that it involves different forecast horizons. The respondents are provided with the last quarter's preliminary figure and are asked to put figures into their forecasts for the current quarter and the next five quarters. Here we use the current-period, the one, two, and three-period-ahead forecasts.

Our measure of actual output is GNP up to 1991q4 and GDP thereafter, in order to be consistent with the change in the survey's questionnaire. Both measures are available from the Federal Reserve Bank of Philadelphia on a quarterly basis. The data are subject to several revisions. Official figures are released fifteen days after the end of each quarter; revised figures are released 45 days after the end of each quarter and since 1974, further revised figures are released 75 days after the end of each quarter. Every July a more extensive revision is made reaching back between 6 and 15 quarters, while further comprehensive revisions take place occasionally. During the period 1968q4-2001q2 there were comprehensive revisions for GDP which took place in January 1976, December 1980, December 1991, December 1995 and August 1999. For our empirical work we chose the 15-day vintage, as this is the vintage provided to the respondents at the time the questionnaire is sent out and thus, reflects more accurately the information set available when the expectations are formed.

B Simulation of Finite-Sample Distributions

This Appendix illustrates the parametric and non-parametric bootstrap methods that have been employed for the simulation of the finite-sample distribution of the test statistics discussed in the empirical section. The approach is based on that of Fachin (2000) and can be summarised in four steps:

(i) The VECM in (8) (augmented by $\mathbf{a}_2\mathbf{D}_t$) is estimated under the null hypothesis of interest in order to obtain the estimates $\hat{\mathbf{a}}_0$, $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$, $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\Gamma}_i$, $\hat{\mathbf{e}}_t$ and $\hat{\Omega}$, $i = 1, ..., p-1$.

(ii) In the parametric case the bootstrap residuals $\mathbf{e}_t^j$, $j = 1, ..., B$, $t = 1, ..., T$, are obtained as random draws from $N_m(\mathbf{0}, \hat{\Omega})$. In the non-parametric version $\mathbf{e}_t^j$ is drawn with replacement from the (normalised) estimate $\hat{\mathbf{e}}_t$, except in the case of the tests of over-identification of $\boldsymbol{\beta}$, where the model is re-estimated under the alternative to produce the estimated residuals $\hat{\mathbf{e}}_{*t}$ and $\mathbf{e}_t^j$ is now drawn with replacement from $\hat{\mathbf{e}}_{*t}$.

(iii) For each of the $\mathbf{e}_t^j$, $j = 1, ..., B$, $t = 1, ..., T$, and using the starting values $\mathbf{z}_0, \ldots, \mathbf{z}_{p-1}$, the series $\Delta\mathbf{z}_t^j$ are dynamically simulated as

$$\Delta\mathbf{z}_t^j = \hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1 t + \hat{\mathbf{a}}_2\mathbf{D}_t + \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\mathbf{z}_{t-1}^j + \sum_{i=1}^{p-1}\hat{\Gamma}_i\Delta\mathbf{z}_{t-i}^j + \mathbf{e}_t^j. \quad (18)$$

(iv) For each of the simulated $\Delta\mathbf{z}_t^j$, $j = 1, ..., B$, $t = 1, ..., T$, the statistic of interest is computed in order to obtain its bootstrap distribution under the null.

Tables 1-3

Tables 4 & 5

Table 6

Figures 1 & 2

References

Cheung, Y. Wong and K.S. Lai (1993), "Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration", *Oxford Bulletin of Economics and Statistics*, 55, 3, 313-328.

Fachin, S. (2000), "Bootstrap and Asymptotic Tests of Long-Run Relationships in Cointegrated Systems", *Oxford Bulletin of Economics and Statistics*, 62, 4, 543-551.

Fair, Ray C., Shiller, R.J. (1989), "The informational content of ex-ante forecasts", *Review of Economics and Statistics*, 71, pp. 325-331.

Garratt, A., K. Lee, M.H. Pesaran and Y. Shin (2000), "The Structural Cointegrating VAR Approach to Macroeconometric Modelling", Chapter 5 in S. Holly and M. Weale (eds.) *Econometric Modelling; Techniques and Applications*, Cambridge University Press.

Garratt, A., K. Lee, M.H. Pesaran and Y. Shin (2003), "A Long Run Structural Macroeconometric Model of the UK", *Economic Journal*, Vol. 113, 412-455.

Greenslade, J.V., S.G. Hall and S.G.B. Henry (2002), "On the Identification of Cointegrated Systems in Small Samples: A Modelling Strategy with an Application to UK Wages and Prices", *Journal of Economic Dynamics & Control*, 26, 1517-37.

Harris, R.I.D. and G. Judge (1998), "Small Sample Testing for Cointegration using the Bootstrap Approach", *Economics Letters*, 58, 31-37.

Jacobson, T., P. Jansson, A. Vredin and A. Warne (2001), "Monetary Policy Analysis and Inflation Targeting in a Small Open Economy: A VAR Approach", *Journal of Applied Econometrics*, 16, 487-520.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.

Johansen, S. (2000a), "A Small-Sample Correction for Tests of Hypotheses on the Cointegrating Vectors", *Journal of Econometrics*, 111/2, 195-221.

Johansen, S. (2000b), "A Bartlett Correction Factor for Tests on the Cointegrating Relations", *Econometric Theory* 740-778.

Keanu, M.P., Runkle, D.E. (1990), "Testing the rationality of price forecasts: new evidence from a panel data", *American Economic Review*, 80, pp714-35.

Koop, G., M.H. Pesaran and S.M. Potter (1996), "Impulse Response Analysis in Nonlinear Multivariate Models", *Journal of Econometrics*, 74, 119-147.

Lahiri, K. and T.S. Chun (1989), "Some Tests for Unbiasedness in the Long-Run Using Survey Data", *International Economic Journal*, 3, 27-42.

Lee, K.C. and M.H. Pesaran (1993), "Persistence Profiles and Business Cycle Fluctuations in a Disaggregated Model of UK Output Growth", *Ricerche Economiche*, 47, 293-322.

Lee, K.C., M.H. Pesaran and R.G. Pierse (1992), "Persistence of Shocks and its Sources in a Multisectoral Model of UK Output Growth", *Economic Journal*, 102, 342-356.

Mantalos, P. and G. Shukur (1998), "Size and Power of the Error Correction Model Cointegration Test: A Bootstrap Approach", *Oxford Bulletin of Economics and Statistics*, 60(2), 249-255.

Patterson, K.D. and S.M. Heravi (1991), "Are Different Vintages of Data on the Components of GDP Co-integrated? Some Evidence for the United Kingdom", *Economics Letters*, 35, 409-413.

Pesaran, M.H. and Y. Shin (1996), "Cointegration and Speed of Convergence to Equilibrium", *Journal of Econometrics*, 71, 117-143.

Pesaran, M.H. and Y. Shin (1998), "Generalised Impulse Response Analysis in Linear Multivariate Models", *Economics Letters*, 58, 17-29.

Pesaran, M.H. and Y. Shin (2002), "Long-Run Structural Modelling", *Econometrics Reviews,* 21, 49-87.

Pesaran, M.H., Y. Shin and R.J. Smith (2000), "Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables", *Journal of Econometrics*, 97, 293-343.

Pesaran, M.H. and R.P. Smith (1998), "Structural Analysis of Cointegrating VARs", *Journal of Economic Surveys*, 12(5), 471-505.