5 UNIT-ROOT TESTS AND EXCESS RETURNS

Marie-Josée Godbout
Bank of Canada
Simon van Norden
École des Hautes Études Commercial

1. Introduction

When introducing students to the modern theory of financial markets, it is common to characterize the behavior of the log of asset prices as martingales, and their excess returns as being serially uncorrelated or even unpredictable. This is consistent with Fama's (1970) characterization of weak, semi-strong, and strong market efficiency. To be sure, there is an extensive literature documenting the deviations of asset prices from this characterization. Despite this, most would accept the proposition that asset prices contain a unit root in their time-series representation and that excess returns do not. Put another way, the stylized fact is that asset prices are integrated of order one (I(1)) and excess returns are integrated of order zero (I(0)). However, a small number of prominent recent papers have presented evidence that appears to reject this characterization in a surprising way. They show that, according to some tests, some excess returns appear to contain a unit root.

Of the four papers of which we are aware, three provide evidence from foreign exchange markets; Evans and Lewis (1993) (herinafter EL93), Evans and Lewis (1995) (herinafter EL95) and Crowder (1994).[1] All three begin with the premise that the log of the spot exchange rate s_t is I(1). Therefore, if it is not cointegrated with the log of the *k*-

period forward rate f_t, or, more specifically, if its cointegrating vector is not [1,-1], then excess returns to holding foreign exchange for k periods $x_t^k = s_{t+k} - f_t^k$ will be I(1).[2]

There is extensive published evidence that such returns appear to be I(0) and that spot and forward rates seem to be cointegrated with cointegrating vector [1, -1]. (For example, see Phillips, McFarland, and McMahon (1996) or the survey by Brenner and Kroner (1995)). However, Evans and Lewis carefully argue that the published evidence is not conclusive, noting the difficulties inherent in determining the asymptotic behavior of a series from a finite sample.[3] For example, if the permanent trend in x_t accounts for only a small fraction of its sample variance, even over the space of a decade or more, then unit-root tests on excess returns may incorrectly tend to reject the null hypothesis of a unit root, a fact documented in EL93 via Monte Carlo experiments.

More importantly, however, Evans and Lewis are careful not to interpret their results as literal evidence of unit roots in excess returns.[4] Instead, they argue that the presence of infrequent changes in regime that are rationally anticipated by the market (i.e., a "peso problem") will lead to very persistent "trends" in excess returns. Their simulations lead them to conclude that such changes in regime can cause tests to reject the null hypothesis of stationary excess returns in realistic sample sizes, even when they are truly stationary in large enough samples. Their preferred interpretation of the evidence they present is that there are highly persistent trends in excess returns that are consistent with regime-switching behavior in exchange rates.

On the other hand, Crowder (1994) interprets the results of unit root and cointegration tests more narrowly. He extends Baillie and Bollerslev's (1989) work on the apparent cointegration of nominal spot exchange rates, noting that such cointegration implies that changes in some foreign exchange rates must be predictable. Such predictability is commonly attributed to the existence of a risk premium. However, if the spot and forward exchange rates are not cointegrated, markets can only be economically efficient if the risk premium is I(1), whereas if spot exchange rates are cointegrated, then economic efficiency implies that the risk premium must be I(0). Since the risk premium cannot be both, Crowder interprets this as evidence against the economic efficiency of foreign exchange markets.

This paper reconsiders these new results about the persistence of excess returns. Specifically, we examine the literature on finite-sample problems in the application of cointegration tests and show how these may have affected the above conclusions. Section 1 begins by critically reviewing the above papers in more detail and explains why we believe that their conclusions may not be as sound as they appear. Section 2 attempts to replicate previously published results based on tests for the number of cointegrating vectors in a system of variables and applies a series of simulation experiments to assess their robustness. Section 3 does the same for tests of parameter restrictions on the cointegrating vector. Section 4 summarizes our conclusion and suggests possible extensions.

2. Literature Survey

We begin with a review of the empirical evidence presented by Evans and Lewis, which is based on tests of cointegrating relationships between spot and forward exchange rates.

These authors also present a model which predicts persistent trends in excess returns, based on the presence of regime-switching behavior in exchange rates. Such models are not entirely new and follow a well-established line of research on "peso problems" and related phenomena. However, in our minds the most striking contribution of EL93 and EL95 is the new empirical evidence they present on the apparent persistence of these trends. For reasons of brevity, we therefore limit our review to the evidence they present on these trends.

As noted in the introduction, Crowder (1994) presents similar evidence. However, his results are sensitive to the lag length selected, and he does no simulations to check the finite sample behavior of his tests. Moreover, his article is immediately followed by a rebuttal by Baillie and Bollerslev (1994). For these reasons, we think Evans and Lewis' results are the most convincing seen to date and we limit ourselves to their work. However, we should note that Crowder's methodology is essentially the same as that in EL93. Therefore, we suspect that the results we present on their work could apply equally to Crowder's.

Both of the studies by Evans and Lewis use the same data set, taken from Bekaert and Hodrick (1993).[5] This consists of log spot exchange rates against the U.S. dollar for the British pound, the German mark and the Japanese yen, and the three corresponding 1- or 3-month log forward exchange rates. The data are monthly and cover the period 1975 to 1989. Excess returns are taken to be the difference between the log forward rate and the log spot rate corresponding to the same settlement data as the forward rate.

1.1 Evans and Lewis (1993)

1.1.1 Summary

If each spot rate is cointegrated with its respective forward rate, then a system of three spot rates and three forward rates should have three cointegrating vectors in the system. If there are fewer than three cointegrating vectors, then at least one of the excess return series (i.e., the difference between the spot rate and its respective forward rate) must have a unit root. EL93 use the Johansen (1988) maximum-likelihood tests for the number of cointegrating vectors in such a three-exchange rate system, and they never find significant evidence of more than two cointegrating vectors.[6]

An alternative explanation for their failure to find evidence of three cointegrating vectors might be that the Johansen (1988) tests simply lack the power to detect them. EL93 address this problem with a series of Monte Carlo experiments (documented in Evans and Lewis (1992)). Simulating six-variable systems with three cointegrating vectors, they find that a lack of power in the Johansen test cannot explain their results. More precisely, they find that the value of the test statistic for the null hypothesis that there are no more than two cointegrating vectors lies well below the 99 per cent critical value of its simulated empirical distribution under the alternative hypothesis. Therefore, in their experiment where three cointegrating vectors exist, they find test statistics larger than those in the true data with a probability of over 99 per cent. They conclude that "...there is strong evidence of statistically significant trends in forward rates relative to spot rates."[7]

1.1.2 Critique

Since Evans and Lewis (EL) are dealing with a sample size of 180 observations, finite sample problems associated with cointegration tests might affect the results. However, the size-distortion problem commonly associated with this test should make the maximum-likelihood (ML) tests find evidence of too many cointegrating vectors, not too few.[8] Furthermore, EL attempted to compensate for these finite-sample effects with their simulation experiments. Despite this, we think that there are grounds for questioning the EL93 results.

First, contrary to the conclusion of EL93, EL95 (discussed in more detail below) state that these same spot rates are cointegrated with forward rates. Specifically, they say:

> Before running the regressions, we used the Johansen (1988) procedure to check that the pairs of individual spot and forward rates were cointegrated. The same procedure was also used to test for the number of trends in the vectors of three spot and three forward rates used in the system estimation. We could not eject the hypothesis that all these vectors contained three trends.[9]

EL95 do not reference their earlier work, so it is not obvious how these results are to be reconciled with those of EL93.

More importantly, however, we are concerned that the data-generating process (DGP) used in the simulations in EL93 and Evans and Lewis (1992) might give unrepresentative results. Their DGP is specified as $\bar{z}_t = [\bar{z}_{1t}{}', \bar{z}_{2t}{}']'$, where[10]

$$\begin{aligned} \bar{v}_t &= \bar{v}_{t-1} + \bar{\epsilon}_t & E_{t-1}(\bar{\epsilon}_t \cdot \bar{\epsilon}_t{}') &= Q_t \\ \bar{z}_{1t} &= \bar{v}_t + \Lambda \cdot \bar{u}_{1t} & E_{t-1}(\bar{u}_{1t} \cdot \bar{u}_{1t}{}') &= \Sigma_{1t} \\ \bar{z}_{2t} &= \Gamma \cdot \bar{z}_{1t} + \bar{u}_{2t} & E_{t-1}(\bar{u}_{2t} \cdot \bar{u}_{2t}\,) &= \Sigma_{2t} \end{aligned} \tag{1}$$

$\bar{z}_t$ is a d dimensional vector with m stochastic trends (d-m cointegrating vectors), Γ is the matrix of cointegrating vectors, $\bar{u}_{1t}, \bar{u}_{2t}, \bar{\epsilon}_t$ are mutually uncorrelated innovations of dimension m, d-m, and d.

Although EL vary the number of cointegrating relationships (m) in their simulations, we are concerned here with the case where we have three pairs of cointegrated series (so $d = 6$, $m = 3$.) They also considered both homoskedastic and heteroskedastic innovation processes, finding that they gave very similar results. We will restrict our attention to the homoskedastic case. They further assume that each of these three matrices is diagonal, and that both Γ and Λ are the identity matrix. Together, this implies that (1) is equivalent to three unrelated bivariate systems of the form

$$\begin{aligned} v_t^i &= v_{t-1}^i + \epsilon_t^i & E_{t-1}(\epsilon_t^i \cdot \epsilon_t^i) &= q^i \\ z_{1t}^i &= v_t^i + u_{1t}^i & E_{t-1}(u_{1t}^i \cdot u_{1t}^i) &= \sigma_1^i \\ z_{2t}^i &= z_{1t}^i + u_{2t}^i & E_{t-1}(u_{2t}^i \cdot u_{2t}^i) &= \sigma_2^i \end{aligned} \tag{2}$$

One can think of each of these systems as corresponding to the spot and forward rate for a particular currency *i*. Each system has two integrated moving-average IMA (1,1) variables that are jointly cointegrated with cointegrating vector [1,1]. To see that both series are IMA(1,1), note that

$$\Delta z_{1t}^{i} = \epsilon_t^i + \Delta u_t^{i} = e_t^{i} + \theta_1 \cdot e_{t-1}^{i}, \text{ where } \theta_1 = -\sigma_1^i/(\sigma_1^i + q^{\,i}) \tag{3}$$

for some innovation sequence e_t^{i}. The same logic applies to z_{2t}^{i}, except that its moving-average (MA) parameter will be $\theta_2 = -(\sigma_1^i + \sigma_2^i)/(\sigma_1^i + \sigma_2^i + q^{\,i})$. The process is then calibrated so that $\sigma_1^i = q^{\,i} = E(\Delta s^{\,i} \cdot \Delta s^{\,i})$ and $\sigma_2^i = E(\Delta f^{\,i} \cdot \Delta f^{\,i})$, where $s^{\,i}$ and $f^{\,i}$ are the log spot and forward exchange rates for currency *i*. If their first differences have roughly equal variances, then the MA parameters will be $\theta_1 = -0.5$ and $\theta_2 = -0.67$.

Evans and Lewis offer no justification for choosing this particular DGP or parameterization. It has at least one unrealistic feature: the presence of a strong negative MA component in both spot and forward exchange rates, when both are thought to closely approximate a martingale; for example, see Meese and Rogoff (1983). This difference could plausibly affect the outcome of their simulations. It is well understood that substantial MA components will require long autoregressive (AR) lags in the cointegration test to serve as a suitable approximation of the short-run dynamics. However, Evans and Lewis (1992) consider only the case of one and three AR lags. It is also understood that failure to adequately capture negative MA dynamics can make series appear to be more stationary than they are in reality. This in turn might cause cointegration tests on the simulated data to find fewer stochastic trends than are truly present, which would tend to exaggerate the power of the cointegration test. We are concerned that this, rather than the presence of very persistent trends in excess returns, might be the reason that EL93 find much less evidence of cointegration in the true data than in their simulation experiments with three cointegrating vectors.

We investigate the latter problem below with our own simulation experiments. However, we think that some of the Monte Carlo work presented by Evans and Lewis (1992) is consistent with the reinterpretation of their results suggested above. Table B in Evans and Lewis (1992) shows that when there is no cointegration, asymptotic 95 per cent critical values would detect cointegration more than 50 per cent of the time in their three-variable system and more than 20 per cent of the time in their six-variable system. Such size distortion in the simulation experiment could exaggerate the estimated power of the test in the actual data. Furthermore, small increases in the number of lags used in their test have important effects on the simulation results. For example, their Table B reports that in a six-variable system with three cointegrating vectors and homoskedastic errors, increasing the number of lags from one to three lowers the 1 per cent critical value from 63.99 to 48.26. This is consistent with the hypothesis that the combination of an important MA component and few lags are driving their results.

1.2 Evans and Lewis (1995)

1.2.1 Summary

EL95 use the same data set as EL93 but, as discussed in endnote 9, work under the assumption that spot and forward rates are cointegrated. Instead of testing for the number of cointegrating vectors, EL simply test whether spot and forward rates move one-for-one in the long run (i.e., whether the cointegrating vector is [1, -1].)[11] EL95 use the dynamic ordinary least squares (DOLS) estimator proposed in Stock and Watson (1993) to test the cointegrating vectors. Tests of a single cointegrating vector require only ordinary least squares (OLS) estimation of

$$s_{t+k} = \alpha_0 + \alpha_1 \cdot f_t + \sum_{n=-3}^{3} \beta_n \cdot \Delta f_{t-n} + u_t \tag{4}$$

Hypotheses about α_1 can then be tested with the usual t-statistics, or if Newey-West standard errors are used, with a χ^2 (1) statistic. Testing the joint hypothesis that $\alpha_1 = 1$ for each of three exchange rates simply requires seemingly unrelated regression (SUR) estimation of three OLS equations (one for each currency), and the usual Wald statistic has a asymptotic $\chi^2(3)$ distribution under the null. EL95 show that t-tests for both 1- and 3-month forward rates allow forward rates allow them to reject the hypothesis that the cointegrating vector is [1,-1] for some currencies, and that the joint test always rejects with a marginal significance level of less than 0.1 per cent.

To address concerns about the finite-sample of their estimator, Table 3 in EL95 reports the results of a Monte Carlo experiment that examines the finite-sample properties of their test statistics under the null hypothesis that the cointegrating vectors are all [1, -1]. They conclude that linear DGPs cannot explain their rejection of the [1, -1] cointegrating vector, based on 1000 replications.

1.2.2 Critique

Stock and Watson (1993) find that the DOLS estimator tends to overstate the true precision of the estimated cointegrating vectors, and that this is worsened when non-parametric variance-covariance (VCV) estimators such as Newey-West are used. This may give grounds for believing that the marginal significance levels at which EL95 reject the cointegrating vector restriction are exaggerated. However, it does not explain why their Monte Carlo experiment would not detect such a problem.

EL begin their Monte Carlo experiment by estimating (4) and saving the residuals. They then bootstrap a new series of residuals and simulate an artificial series for s_{t+k} under the restriction that $\alpha_1 = 1$.[12] These artificial series are then used to determine the distribution of the DOLS test statistics under the null hypothesis that $\alpha_1 = 1$. This experiment therefore changes $\{s_t\}$ while holding $\{f_t\}$ constant. It also imposes the assumption that (4) is well specified.

The key to the validity of the Stock-Watson test is that the differences of the explanatory variable in (4) must produce errors that are orthogonal to the level of the explanatory variable; otherwise, estimates of the cointegrating vector will be biased. The design of the EL95 Monte Carlo guarantees that the generated errors will be independent of the forward rate, so that such bias cannot arise. Evans and Lewis justify this by noting that: "As discussed in Stock and Watson (1993), the residuals from this equation are independent of the entire sequence of the right-hand side variable and so can be treated as strictly exogenous."[13] This is true asymptotically if enough leads and lags of the differences are included. The extent to which this may hold in a finite sample for a given number of lags is not clear.

The conjecture that the endogeneity of the forward rate can account for EL95's rejections of the [1, -1] cointegrating vector between spot and forward rates is consistent with at least one other piece of evidence. EL95 present additional simulations of a regime-switching model, which manages to account for the large Stock-Watson test statistics they find in the true data. However, a potentially important difference between their simulations of (4) and their regime-switching model simulations is that, in the latter, spot and forward rates are modelled simultaneously instead of the forward rate being fixed. Therefore, the difference that they document in the behavior of their linear and non-linear models might not be evidence of regime-switching in exchange rates; it could instead be an artifact of finite-sample problems in the Stock-Watson test with endogenous regressors.

1.3 Conclusion

For the reasons given above, we think that the conclusions presented by Evans and Lewis may be premature, and that their results deserve closer scrutiny. In the following sections, we re-evaluate the results in EL93 and EL95 with the aid of further simulation experiments designed to test the hypotheses advanced above. Specifically, we examine whether the simulation results in EL93 on the power of the cointegration tests are sensitive to their assumptions on the exogeneity of the forward rate.

One additional aspect of this literature should be noted. Evans and Lewis use cointegration tests indirectly, to provide evidence of the importance of regime-switching in financial markets. There is no dispute over whether one can find evidence of regime-switching in financial markets and there is no shortage of direct empirical work on this. Apparently, their intended contribution is to show that this regime-switching gives rise to very persistent trends in excess returns as evidenced by the cointegration tests.

Viewed from this perspective, we think their work leaves several important questions unanswered. Is a cointegration test a more meaningful measure of persistence than autocorrelation coefficients? How is it superior to previous measures of return persistence? Is it particularly sensitive to the effects of regime-switching?

There is some evidence on the last question, but it does not tend to support the conclusions of Evans and Lewis. Evans (1991)[14] found that regime-switching tended to make explosive series appear stationary and to make non-cointegrated series appear to be cointegrated. Evans and Lewis are claiming the opposite, based not on the theoretical properties of cointegration tests, but instead solely on the Monte Carlo evidence they present. We think that this is another reason for re-examining the robustness of this evidence. We do this in the next two sections.

2. Evidence from Tests for the Number of Cointegrating Vectors

2.1 Reproduction

The first step in analyzing Evans and Lewis' (EL) results is to try to reproduce them using the same data set from Bekaert and Hodrick (1993). Table 1 shows that while we are not able to replicate EL93's results precisely, our statistics are quite close and produce the same conclusions when using asymptotic critical values with the maximum eigenvalue

Table 1. Replicating Evans and Lewis (1993): 1-Month Forward Rate Johansen (1988) Test Results (One Lag)

Number of cointegrating vectors	Maximum Eigenvalue			Trace Statistic				
	Godbout & van Norden	Evans & Lewis 1993	5% Critical Value[a]	Godbout & van Norden	Evans & Lewis 1993	5% Critical Value[a]	Reinsel & Ahn	Linear Combination Test
0	58.65	53.13	36.36	139.18	*118.00*	82.49	134.49	131.95
1	38.54	36.83	30.04	80.54	*64.91*	59.46	77.82	77.64
2	20.10	15.20	23.80	40.99	*28.09*	39.89	39.61	40.14
3	12.88	---[b]	17.89	20.93	---[b]	24.31	20.22	20.62
4	7.56	---[b]	11.44	8.05	---[b]	12.53	7.78	7.97
5	0.49	---[b]	3.84	0.49	---[b]	3.84	0.47	0.49

Notes: [a] Taken from Osterwald-Lenum (1992, Table 0). [b] Results not reported.

statistic.[15] However, using the Trace statistic, Reinsel and Ahn's scaled trace statistics or Pitarakis and Gonzalo's linear combination test (LCT) we obtain test statistics roughly equal to their 95 per cent asymptotic critical values for the test of the third cointegrating vector, whereas EL93 found evidence of only two cointegrating vectors. Given that their Monte Carlo evidence suggests that the asymptotic critical values are very conservative in this case, this weak evidence of a third cointegrating vector may not be convincing by itself. Furthermore, we find that this evidence vanishes if we add more lags to the Johansen test. Although not reported above, we also repeated these tests using the 3-month forward rates considered by EL93. In that case, regardless of the test statistic or the number of lags used, we always found exactly two significant cointegrating vectors.

2.2 Experimental Design

We would now like to repeat EL93's simulations to determine whether these results are consistent with the presence of three cointegrating vectors. However, for the reasons explained in the previous section, we will use a different DGP to perform these simulations. For each pair of spot and forward exchange rates, we begin by estimating the following bivariate system via ordinary least squares:[16]

$$\begin{bmatrix} s_t \\ f_t \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} \cdot \begin{bmatrix} s_{t-1} \\ f_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{st} \\ \epsilon_{ft} \end{bmatrix} \tag{5}$$

The restriction that the sum of the coefficients in each equation equals one ensures that the simulated spot and forward rates will be cointegrated with cointegrating vector [1, -1]. We do not impose any weak exogeneity restrictions, which in this case amount to restricting $\alpha = 1$ or $\beta = 1$. (Uncovered interest rate parity would imply $\alpha = 0$ for 1-period forward rates.) Since Evans and Lewis use few lags in their cointegration tests, we deliberately keep the dynamics of the system as simple as possible.

Having estimated (5) for each of the three currencies, we estimate the variance-covariance matrix of the six residual series. This estimate is then used to calibrate a mean-zero multivariate normal distribution, from which we make independent and identically distributed draws to simulate new residuals. These are then used in (5) to generate new data series of the same length as the original data. Johansen (1988) test statistics (using a single lag) are then calculated and the number of cointegrating vectors significant at the 95 per cent level tabulated. This was repeated for 1000 simulated data sets.

We feel that (5) is at least as plausible as DGP as that of EL93, since (5) relies on estimated rather than exogenously fixed parameters, and is consistent with the near-martingale properties of spot and forward exchange rates (whereas EL93's is not.) However, we also investigated a richer DGP, which added lagged differences of spot and forward rates to each equation in (5). We found that it produced similar conclusions, as explained below.

2.3 New Simulation Results

The results are very different from those reported by EL93. Although three cointegrating vectors are present by construction, in 1000 simulations we literally never find significant evidence of three such vectors, regardless of the test statistic used. Significant evidence of even two cointegrating vectors, such as we find in the true data, is relatively rare, occurring in less than 1 per cent of the trials. If we focus on the tests of the simple null hypothesis of less than two cointegrating vectors, we found that the test statistics were below their 95 per cent critical values in over 99 per cent of the simulations.[17] This means that the Johansen test would have effectively no power to detect the presence of three cointegrating vectors for the DGP we consider. Furthermore, in over 97 per cent of the trials we found that the simulated Lambda-max statistic was less than that produced by the true data. This suggests that, if anything, the true data show greater evidence of three cointegrating vectors than our artificial DGP.

Although not reported here, we similarly tested the data for the 3-month forward exchange rates. Again, the simulations literally never found evidence of three or more cointegrating vectors, despite their being present by construction. Again, we can reject the null of two or fewer cointegrating vectors in less than 1 per cent of the trials, and the true data give a Lambda-max statistic greater than the vast majority of the trials (in this case, 96 per cent).[18]

Based on the assumption that the DGP we used is at least as plausible as that considered by EL93, it seems that failure to find evidence of three cointegrating vectors may simply be due to the low power of the cointegration test. This in turn suggests that it is premature for EL93 to conclude that this is evidence of permanent of very persistent trends in excess returns.

3. Evidence from Tests of the Value of the Cointegrating Vector

3.1 Reproduction

Table 2 shows the results of our attempt to reproduce the results reported in EL95. The first three lines show the estimates of α_1 in the OLS regression.

$$s_{t+k} = \alpha_0 + \alpha_1 f_t + v_{t+k} \tag{6}$$

Regardless of whether 1-month or 3-month forward rates are used, we always manage to reproduce the results in EL95 to within an accuracy of 0.2 per cent, and to reproduce them to the limits of published accuracy in half of the cases. While these estimates of α_1 are all close to 1.0, they are known to be biased downwards in small samples. For comparison, we also report unbiased estimates (α_1 from (4)) produced by the same estimation procedure used to test the parameter estimate. Shown in the third line (under "Corrected"), they are considerably closer to 1.0. Rejecting the hypothesis that they are equal to 1.0 implies that we can discriminate between an estimate of 0.997 and a value of 1.0.

In reporting tests of the hypothesis that $\alpha_1 = 1$, we report both the calculated *p*-value of the test (to allow comparison with EL95) as well as the corresponding χ^2 test statistic. We are able to replicate their conclusions, although we cannot precisely replicate their results.[19] We always obtain very strong rejections of the joint hypothesis that $\alpha_1 = 1$ for all three currencies, regardless of the method used to calculate the standard errors and regardless of whether 1-month or 3-month forward rates are used.

We also used the Johansen and Juselius (1990) methodology to test restrictions on the cointegrating vectors. We tested both two-variable systems, consisting of the spot and forward rate for a given currency, as well as the full six-variable system. The former tests were done under the maintained hypothesis of one cointegrating vector, while the latter assumed three cointegrating vectors. In all cases, we were able to reject the restrictions on the cointegrating vectors at the 1 per cent significance level using asymptotic critical values. Given the problems associated with size distortion in the ML tests for cointegration, we cannot say how reliable such a conclusion is in finite samples.

3.2 Experimental Design

To investigate the sensitivity of the simulation results to the design of the experiment, we performed two separate experiments. First, we simulated the data using the method used by EL95 for their linear DGP. Then we simulated the data using the same DGP as in Section 2.2, with one minor change. Instead of generating the errors from a multivariate normal distribution, we followed EL95 and used random draws from the residuals of (5) to simulate new series. In all experiments, we performed 1000 replications.

Table 2. EL95 Estimation of Cointegrating Vectors: Jan. 1975 to Dec. 1989 (Heteroskedasticity and Serial-Correlation Robust Results)

Term	Statistic	Standard errors	Source	Pound	Mark	Yen	Joint
k=1	$\hat{\alpha}_1$		EL95	0.971	0.986	0.991	
			GvN 96	0.973	0.986	0.991	
			Corrected	0.999	0.998	0.997	
	Test $\hat{\alpha} = 1$	Heteroskedasticity-consistent	EL95	0.209	0.010	<0.001	<0.001
			GvN 96	0.463	0.013	<0.001	0.001
			χ^2	0.54	**6.12**	**16.87**	**16.20**
		HAC robust (Newey-West)	EL95	0.463	0.013	<0.001	<0.001
			GvN 96	0.628	0.082	0.011	0.001
			χ^2	0.24	3.03	**6.48**	**15.97**
k=3	$\hat{\alpha}_1$		EL95	0.909	0.946	0.967	
			GvN 96	0.908	0.948	0.969	
			Corrected	0.992	0.996	0.993	
	Test $\hat{\alpha} = 1$	Heteroskedasticity-consistent	EL95	0.021	0.052	<0.001	<0.001
			GvN 96	0.021	0.052	<0.001	<0.001
			χ^2	**5.33**	3.78	**15.91**	**32.82**
		HAC robust (Newey-West)	EL95	0.297	0.057	0.052	<0.001
			GvN 96	0.145	0.195	0.014	<0.001
			χ^2	2.12	1.68	**6.04**	**32.49**
		HAC robust (flat kernel)	GvN 96	0.243	0.284	0.051	<0.001
			χ^2	1.36	1.15	3.82	**32.14**

Notes: **Heteroskedasticity-consistent** indicates *p*-values calculated using heteroskedasticity-robust standard errors. **HAC robust** indicates *p*-values are calculated using standard errors that allow for heteroskedasticity and serial correlation using the correction suggested by Newey and West (1987). **HAC robust (flat kernel)** indicates *p*-values are calculated using standard errors that allow for heteroskedasticity and serial correlation using an unweighted kernel in place of the weighted kernel suggested by Newey and West (1987). This is arguably better suited to the case where MA serial correlation is induced by overlapping observations (as with 3-month forward rates). **Corrected** shows the estimated parameter corrected for the downward bias in finite samples using the same Stock and Watson (1993) estimator used for the subsequent hypothesis tests. χ^2 shows the test statistics upon which the GvN96 *p*-values are based. Under the null, these are distributed $\chi^2(1)$ for individual currencies and $\chi^2(3)$ for the joint test.

Table 3. Percentage of Trials Where $\alpha_1 = 1$ Not Rejected at the 5 Per Cent Level When True in (4) (Based on 1000 Trials)

Forward (months)	Standard errors	Critical Values	DGP	Yen	Mark	Pound	Joint
1	Heteroskedasticity-consistent	95%	(4)	92.9	94.4	94.8	94.4
			(5)	41.8	53.1	39.0	12.3
		Data based	(4)	52.1	98.6	100.	100.
			(5)	14.6	64.9	71.4	31.2
	HAC robust (Newey-West)	95%	(4)	91.4	94.0	93.8	94.5
			(5)	60.2	72.1	57.5	12.3
		Data based	(4)	34.9	90.0	98.0	100.
			(5)	15.3	66.1	69.8	30.9
3	Heteroskedasticity-consistent	95%	(4)	92.9	94.4	94.8	94.4
			(5)	32.9	47.1	32.0	8.6
		Data based	(4)	97.3	94.4	100.	100.
			(5)	39.4	46.6	61.4	43.5
	HAC robust (Newey-West)	95%	(4)	91.4	94.0	93.8	94.5
			(5)	50.7	65.6	50.2	8.6
		Data based	(4)	83.1	79.5	97.7	100.
			(5)	39.2	47.5	59.7	43.1
	HAC robust (flat kernel)	95%	(4)	89.8	92.5	91.6	94.7
			(5)	59.2	75.3	59.1	8.6
		Data based	(4)	71.0	69.4	91.6	100.
			(5)	39.5	48.0	58.9	42.9

Notes: **Heteroskedasticity-consistent** indicates *p*-values calculated using heteroskedasticity-robust standard errors. **HAC robust** indicates *p*-values are calculated using standard errors that allow for heteroskedasticity and serial correlation using the correction suggested by Newey and West (1987). **HAC robust (flat kernel)** indicates *p*-values are calculated using standard errors that allow for heteroskedasticity and serial correlation using an unweighted kernel in place of the weighted kernel suggested by Newey and West (1987). This is arguably better suited to the case where MA serial correlation is induced by overlapping observations (as with 3-month forward rates). **95%** indicates that critical values are from the 95th percentile of the asymptotic distribution. **Data based** indicates that the critical values are the actual values of the test statistics shown in Table 2.

3.3 Simulation Results

Simulation results for tests of the hypothesis that the cointegrating vector between spot and 1-month forward rates is [1,-1] are shown in Table 3. The rows labeled (4) show the results of repeating EL95's simulation experiment, while those labeled (5) use the same DGP considered previously. If the tests are correctly sized, then using the 95 per cent asymptotic critical values, we should not reject the true null hypothesis in close to 95 per cent of the trials. This is what we find using EL95's simulation experiment, regardless of the currency, the horizon of the forward rates, and the method used to construct the standard errors; we cannot reject the null hypothesis in 89 to 95 per cent of the trials.

However, this conclusion is very fragile to the design of the experiment. When we use (5) as our DGP, we find much more evidence of size distortion. For individual currencies, we accept the correct null hypothesis in only 32 to 75 per cent of trials. The problem is more severe for the joint test, where the true null is accepted in only 12.3 per cent of trials with the 1-month forward rates, and 8.6 per cent with the 3-month forward rates. This is consistent with the hypothesis put forward in Section 1.2; that the endogeneity of the forward rate might cause significant finite-sample problems with Stock-Watson tests that would not be detected by EL95's simulation experiment.[20]

Of course, the presence of size distortion in the Stock-Watson test does not rule out the possibility that the cointegrating vector may be different from 1.0.; it simply means that asymptotic critical values cannot be used to test this restriction. However, since our simulation experiment using (5) gives us the approximate finite-sample distribution of the test statistic under the null, we can use this to determine the marginal significance level of the test statistics from the true data (i.e., those reported in Table 2). The results are shown in Table 3 in the rows labeled "data based" critical values. The values shown are the fraction of times the simulated statistics are smaller than the test statistics from the true data. A value greater than x per cent in this row would lead us to reject (conditional on the chosen DGP) the null hypothesis of $\hat{\alpha}_1 = 1$ at the $100 - x$ per cent significance level.

Based on tests for individual currencies, we find that the true test statistics have marginal significance levels that are always more than 25 per cent (i.e., the numbers in the table are always less than 75 per cent), and that the joint tests have marginal significance levels of just over 50 per cent (i.e., the numbers in the table are always less than 50 per cent.). Therefore, this DGP would imply that there is no evidence to reject the null hypothesis at any plausible level of significance. This strongly contrasts with the conclusion we would draw had we used the EL95 simulation method, which suggests that the joint tests were always significant at the 1 per cent level.

4. Conclusion and Directions for Research

We find that EL's conclusion of very persistent trends in returns in foreign exchange markets is premature, at least on the basis of the evidence from cointegration tests. We have suggested that the design of the Monte Carlo experiments in EL93 was flawed and that these flaws would tend to exaggerate the apparent power of the ML tests to detect three cointegrating vectors. Our simulations show that EL93's simulation evidence is very sensitive to specification of the DGP, and that less restrictive dynamics reverse their results. We find that the simulation results EL95 report on the finite-sample size of their tests are very sensitive to the experimental design. We show that a less restrictive design reverses their conclusions and suggests that their ability to reject the restrictions on the cointegrating vectors can be attributed to size distortion. This also casts doubt on the conclusions of Crowder (1994), whose test results are sensitive to the lag length used and are not checked for finite-sample distortions. Comparing our own test statistics with their estimated finite-sample distributions, we find no evidence that would lead us to reject the hypothesis that spot and forward exchange rates are cointegrated and that they move one-for-one in the long run. If anything, the results from our Johansen tests suggest that the evidence of cointegration found is to be expected given the apparently low power of the test.

To be convincing, future research in this area will have to pay more careful attention to the reliability of the tests used to investigate the long-run properties of the data. However, given the fundamental problems of distinguishing between stationary and non-stationary behavior in finite samples, we feel this line of research would also benefit from a more careful explanation of what is to be learned from such investigation. It is hard for us to understand why cointegration tests are a compelling way to test for the importance of regime-switching, or why they are a better measure of the persistence of excess returns than more conventional methods.

6. Appendix: Maximum-Likelihood Test for Cointegration

While there are many alternative tests for cointegration, Gonzalo (1994) suggests that the maximum-likelihood (ML) system estimation approach performs better than both single-equation and alternative multivariate methods in detecting cointegration.[21] This approach is also among the best-known and the most widely applied in empirical work. The starting point of these tests is a VAR specification for the $nx1$ vector of I(1) variables, namely,

$$X_t = \mu + A_1 X_{t-1} + \ldots + A_k X_{t-k} + u_t \tag{A1}$$

where u_t is assumed to be an independent and identically distributed Gaussian process. Note that we can rewrite (A1) as

$$\Delta X_t = \mu + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{k-1} \Delta X_{t-k+1} + \Pi X_{t-k} + u_t \tag{A2}$$

$$\begin{aligned} \Gamma_j &= -(I - A_1 - \ldots - A_t) \\ \Pi &= -(I - A_1 - \ldots - A_k) \end{aligned} \qquad j = 1, \ldots, k \tag{A3}$$

By rewriting (A1) into (A2) we are able to summarize the long-run information in X_t by the long-run impact matrix, Π; it is the rank of this matrix that determines the number of cointegrating vectors. Note that under the null hypothesis of $r(0<r<n)$ cointegrating vectors, Π can be factored as $\Pi = \alpha\beta^T$, where α and β are nxr matrices. Therefore under the null we can write the process for X_t as

$$\Delta X_t = \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-1} + \alpha\beta^T X_{t-k} + u_t \tag{A4}$$

Johansen and Juselius (1990) demonstrate that β, the cointegrating vectors, can be estimated as the eigenvectors associated with the r largest, statistically significant eigenvalues found by solving the problem

$$|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0 \tag{A5}$$

where S_{00} represents the residual moment matrix from a regression of $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}; S_{kk}$ is the residual moment matrix from a regression of X_{t-k} on ΔX_{t-k+1}; and S_{0k} is the cross-moment matrix. These eigenvalues readily permit the formation of likelihood ratio tests to determine the value of r. Johansen and Juselius propose two tests with differing assumptions about the alternative hypothesis: (i) the Trace statistic tests the restriction $r \leq q \quad (q < n)$ against the completely unrestricted model $r \leq n$; and (ii) the λ^{max} statistic makes the alternative more precise by specifying that only one additional cointegrating vector exists $(r \leq q + 1)$. The log-likelihood ratio test statistics are formed as

$$\text{Trace} = -T \sum_{i=1+1}^{n} 1n(1 - \hat{\lambda}_i) \tag{A6}$$

$$\lambda^{max} = -T 1n(1 - \hat{\lambda}_{q+1}) \tag{A7}$$

The asymptotic critical values are non-standard and are tabulated in Osterwald-Lenum (1992). They depend on whether a constant is included in (A1), whether the restriction is imposed that any cointegrating vectors must also annihilate drift, and whether drift is present when it is allowed for. The test without a constant was originally proposed by Johansen (1988), the others by Johansen and Juselius (1990).

The critical values for these tests are based on their asymptotic distribution under the null hypothesis of interest. Recently, a number of authors have begun to examine how reliable this asymptotic approximation is in finite samples, and some have suggested modifications to the ML test.[22] Contributions to this research include Toda (1994, 1995), Gregory (1994), Gonzalo and Pitarakis (1994), Reinsel and Ahn (1988), and Cheung and Lai (1993). The importance of this problem will depend on the specific application. Edison, Gagnon, and Melick (1994) examine the size and power of ML cointegration tests in the context of testing purchasing power parity in the post-Bretton Woods period, and find important size distortions. Hendry (1995) examines the ML tests in the context of Canadian money demand and reaches similar conclusions. Ho and Sørenson (1994) review Durlauf's (1989) evidence on cointegration in U.S. sectoral value-added across one-digit standard industrial classification sectors and suggest that neither asymptotic critical values nor the finite-sample corrections suggested by Reinsel and Ahn (1992) or by Cheung and Lai (1993) are reliable in very short samples (40-50 observations). Godbout and van Norden (1997) show similar evidence from studies of cointegration in international financial markets. Edison and Melick (1995) find evidence of significant size distortions in their study of real interest rate parity. In arriving at these conclusions, these studies all use a simulation methodology similar to ours.

Notes

1. Evans and Lewis (1994) present similar evidence for bond markets. For compactness, we will limit our attention to the results from the foreign exchange markets, although our discussion may have implications for research on bond markets as well.
2. The timing distinction is irrelevant for the purposes of defining the cointegrating relationship, since (s_t, f_t^k) will be cointegrated with vector [1, -1] if and only if (s_{t+k}, f_t^k) is cointegrated with cointegrating vector [1,-1].
3. Evans and Lewis reference Campbell and Perron (1991), but more extensive discussions of this problem may be found in Blough (1994) and Cochrane (1991).
4. While this is true of Evans and Lewis (1993, 1995), it is less true of Evans and Lewis (1992, 1994).
5. These data were generously provided by Robert Hodrick and are more fully documented in Bekaert and Hodrick (1993). We used the average of the bid and ask prices throughout. Evans and Lewis noted that their conclusions were the same when the bid spot and ask forward prices were used instead.
6. See Evans and Lewis (1993, Table 1). Note that using the Johansen Trace test and 1-month forward rates, they find significant evidence of only one cointegrating vector. The Appendix provides more details on the construction of these tests and others.
7. Evans and Lewis (1993, p. 1011).
8. References to the literature on the size-distortion problem are given in the Appendix.
9. Evans and Lewis (1995, footnote 18). One possibility is that the failure to reject the hypothesis of three cointegrating vectors simply means that they could not find evidence of four or more cointegrating vectors.
10. The following based on Evans and Lewis (1992), Appendix, pages iii-iv.
11. As mentioned in endnote 9, EL do not explicitly discuss how to reconcile the assumption that spot and forward rates are cointegrated with the results in EL95. First, their assumption that spot and forward rates are cointegrated is a necessary condition for excess returns to be stationary. Therefore, this assumption should seem innocuous when (as they do) one tries to find evidence to reject the null hypothesis of stationarity. Second, the method that they use to test restrictions on the cointegrating vector is robust to the absence of cointegration, so that their inferences will be asymptotically valid whether or not cointegration is present. See Stock and Watson (1993).
12. In the case where $k > 1$, EL first fit an AR$(k-1)$ model to the residuals, then bootstrap the AR residuals to obtain a new series of residuals for (4). They also try generating errors from ARCH processes.
13. Evans and Lewis (1995, pp. 739-40).
14. Evans (1991) is by *George* Evans. *Martin* Evans coauthored Evans and Lewis (1992, 1993, 1995).
15. The accuracy of our statistical programs was checked by verifying the results of our own code against that produced by commercially available programs by Johansen, Juselius and Hansen, and by Phillips and Ouliaris.
16. Data series were demeaned prior to estimation of (5).
17. This is consistent with the results reported in the previous sentence. Using the sequential determination procedure suggested by Johansen, failure to reject the null hypothesis of one cointegrating vector leads to the conclusion of no more than one cointegrating vector, even if we can reject the null of two cointegrating vectors. Put another way, this implies that in those rare cases where we could reject the null of two cointegrating vectors, we were always unable to reject the null of either zero or one such vectors.
18. These conclusions were robust to variations in the number of lags in the DGP and in the cointegration test. We never found significant evidence of three cointegrating vectors in as much as 1 percent of the trials for any case we examined. Tests for the null hypothesis of two cointegrating vectors were never rejected in more than 2 per cent of the trials, and the true Lambda-max statistics were always greater than at least 78 per cent of their simulated counterparts.
19. With the 1-month forward rate data, we found that our results using the heteroskedasticity-robust standard errors were identical to the results EL reported for the serial-correlation-robust standard errors, suggesting a typographical error in the reporting of their results.
20. Additional simulations suggested that the conclusions were robust to minor changes in the number of lags used in (4) to construct the test statistics, the number of lags used to construct the

Newey-West standard errors, the use of heteroskedasticity-robust or conventional standard errors, a reversal of the roles of spot and forward rates in the Stock-Watson regression, and the use of bootstrapped or multivariate normal innovations.

21. See also Watson (1995), especially Section 3.d.

22. In addition, see the literature survey in Ho and Sørenson (1996). Gonzalo and Lee (1998) give examples of several "pitfalls" — cases where the size of the ML test for cointegration approaches 1 asymptotically.

References

Baillie, R. T. and T. Bollerslev (1989). "Common Stochastic Trends in a System of Exchange Rates," *Journal of Monetary Economics*, 44, 167-181.

Baillie, R. T. and T. Bollerslev (1994). "The Long Memory of the Forward Premium," *Journal of International Money and Finance*, 13, 565-571.

Bekaert, G. and R. J. Hodrick (1993). "On Biases in the Measurement of Foreign Exchange Risk Premiums," *Journal of International Money and Finance*, 12, 115-138.

Brenner, R. J. and K. F. Kroner (1995). "Arbitrage, Cointegration and Testing the Unbiasedness Hypothesis in Financial Markets," *Journal of Quantitative and Financial Analysis*, 30, 23-42.

Campbell, J. and P. Perron (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," in O. Blanchard and S. Fischer (eds.) *NBER Macroeconomics Annual*. Cambridge: MIT Press.

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

Cochrane, J. H. (1991). "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control*, 15, 275-284.

Crowder, W. J. (1994). "Foreign Exchange Market Efficiency and Common Stochastic Trends," *Journal of International Money and Finance*, 13, 551-564.

Durlauf, S. (1989). "Output Persistence, Economic Structure and the Choice of Stabilization Policy," *Brookings Papers on Economic Activity*, 69-137.

Edison, H. J., E. Gagnon, and W. R. Melick (1994). "Understanding the Empirical Literature on Purchasing Power Parity: the Post-Bretton Woods Era," Board of Governors of the Federal Reserve System International Finance Discussion Papers No. 465.

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

Evans, M. D. D. and K. Lewis (1992). "Trends in Excess Returns in Currency and Bond Markets," NBER working paper No. 4116.

Evans, M. D. D. and K. Lewis (1993). "Trends in Excess Returns in Currency and Bond Markets," *European Economic Review*, 37, 1005-1019.

Evans, M. D. D. and K. Lewis (1994). "Do Stationary Risk Premia Explain it All? Evidence from the Term Structure," *Journal of Monetary Economics*, 3, 285-318.

Evans, M. D. D. and K. Lewis (1995). "Do Long-Term Swings in the Dollar Affect Estimates of the Risk Premia?" *Review of Financial Studies*, 8, 709-742.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25, 383-417.

Godbout, M. and S. van Norden (1997). "Reconsidering Cointegration in International Finance: Three Case Studies of Size Distortion in Finite Samples," working paper No. 97-1, Bank of Canada.

Gonzalo, J. (1994). "A Comparison of Five Alternative Methods of Estimating Long Run Equilibrium Relationships," Journal of Econometrics, 60, 203-233.

Gonzalo, J. and T. Lee (1998). "Pitfalls in Testing for Long Run Relationships," *Journal of Econometrics*, 86, 129-154.

Gonzalo, J. and J. Pitarakis (1994). "Comovements in Large Systems," discussion paper No. 9465, CORE.

Gregory, A. (1994). "Testing for Cointegrating in Linear Quadratic Models," *Journal of Business and Economic Statistics*, 12, 347-360.

Hendry, S. (1995). "Long-Run Demand for M1," working paper No. 95-11, Bank of Canada.

Ho, M. S. and B. E. Sørenson (1996). "Finding Cointegration Rank in High Dimensional Systems Using the Johansen Test: An Illustration Using Data Based Monte Carlo Simulations," *Review of Economics and Statistics*, 78, 726-732.

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

Johansen, S. and K. Juselius (1990). "Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Meese, R. A. and K. Rogoff (1983). "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14, 3-24.

Newey, W. K. and K. D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Osterwald-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

Phillips, P. C. B., J. McFarland, and P. C. McMahon (1996). "Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence from the 1920s," *Journal of Applied Econometrics*, 11, 1-22.

Reinsel, G. C. and S. H. Ahn (1988). "Asymptotic Distribution of the Likelihood Ratio Test for Cointegration in the Nonstationary Vector AR Case," technical report, Department of Statistics, University of Wisconsin.

Reinsel, G. C. and S. H. Ahn (1992). "Vector Autoregressive Models with Unit Roots and Reduced Rank Structure; Estimation, Likelihood Ratio Tests, and Forecasting," *Journal of Time Series Analysis*, 13, 353-375.

Stock, J. H. and M. W. Watson (1993). "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.

Toda, H. Y. (1994). "Finite Sample Properties of Likelihood Ratio Tests for Integration When Linear Trends are Present," *Review of Economics and Statistics*, 76, 66-79.

Toda, H. Y. (1995). "Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Ranks in Vector Autoregressions," *Econometric Theory*, 11, 1015-1032.

Watson, M. W. (1995). "Vector Autoregressions and Cointegration," in Engle, R.F. and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4. New York: North Holland.