NORTH-HOLLAND

Modeling and Forecasting Cointegrated Variables: Some Practical Experience

Timothy A. Duy* and Mark A. Thoma

Although the issue of identifying cointegrating relationships between time-series variables has become increasingly important in recent years, economists have yet to reach an agreement on the appropriate manner of modeling such relationships. In this paper, we attempt to distinguish between modeling techniques through a comparison of forecast statistics, while focusing on the issue of whether or not imposing cointegrating restrictions via an error-correction model improves long-run forecasts. We find that imposing cointegrating restrictions often improves forecasting power, and that these improvements are most likely to occur in models which exhibit strong evidence of cointegration between variables. © 1998 Elsevier Science Inc.

Keywords: Cointegration; Forecasting; Vector autoregressions

JEL classification: C32, C53, E17.

I. Introduction

How should economists model cointegrating relationships? Although macroeconomists have yet to reach a consensus on this question, Hamilton (1994) has described three approaches embraced by the literature. Assuming the presence of variables which are nonstationary and cointegrated, the first approach is to estimate the model in levels and allow the data to impose its own restrictions. This method results in consistent coefficient estimates, but some efficiency losses may occur because appropriate coefficient constraints (due to cointegration) are not imposed. A second approach is to construct an error-correction model where the cointegrating relationships are estimated from the data. The third approach also constructs an error-correction model, but in this case, the cointegrating relationships are derived from theoretical considerations rather than from estimation [e.g., Cochrane (1994)]. The possibility, however, of imposing false restric-

Both authors are affiliated with the Department of Economics, University of Oregon, Eugene, Oregon.
Address correspondence to: Dr. T. A. Duy, Department of Economics, University of Oregon, Eugene, OR 97403-1285.

0148-6195/98/$19.00
PII S0148-6195(98)00005-8

tions and, thus, biasing the coefficient estimates (and, consequently, the dynamics of the model) often leaves economists wary of the latter two approaches and favoring the unrestricted levels model[1] [see, for example, Bernanke and Blinder (1992)]. Engle and Granger (1987, p. 259), however, warned against rejecting the error-correction model: ". . . vector autoregressions estimated with co-integrated data will be misspecified if the data are differenced, and will have omitted important constraints if the data are used in levels. Of course, these constraints will be satisfied asymptotically but efficiency gains and improved multi-step forecasts may be achieved by imposing the constraints." Controversy over the power of available specification tests for unit roots, such as the Dickey-Fuller test, only adds to the problem of choosing an appropriate modeling technique. Which approach is best? In this paper, we attempt to distinguish between these three approaches to modeling cointegrating relationships by comparing the out-of-sample forecasting power at short-, medium-, and long-run horizons of various models.

The majority of research involving cointegration has focused on hypothesis testing, not forecasting, because the presence of long-run relationships among variables is often a prediction of a theoretical model. For example, the modern interpretation of the permanent income hypothesis [Hall (1978)] predicts that consumption follows a random walk, and that consumption and total income should be cointegrated. Cochrane (1994) showed that the consumption-income cointegrating vector enters significantly in a two-variable VAR model. Another example is the Fisher (1930) hypothesis, which predicts a cointegrating relationship between nominal interest rates and inflation. Shapiro and Watson (1988) imposed this relationship in a five-variable VAR model, and Mishkin (1992) extensively tested the cointegration hypothesis implied by the Fisher equation.

Previous work in forecasting variables in cointegrated systems includes that by Engle et al. (1989), Engle and Yoo (1987), Hall et al. (1992), and Fanchon and Wendel (1992). Most relevant to our research is the latter three papers. Engle and Yoo (1987) designed a Monte Carlo experiment to compare the forecasting power of an unrestricted vector autoregression model in levels versus that of an error-correction model where the cointegrating vector is estimated via ordinary least squares. They [Engle and Yoo (1987, p. 149)] claimed these two techniques are the "serious contenders" for an appropriate estimation technique and thus chose not to examine an error-correction model in which the cointegrating vector was derived from theory. Using mean square forecast errors to compare performance, they found the error-correction model to be superior in the long-run (7–20 periods), but not the short-run (<7 periods).[2] Hall et al. (1992) used theoretical error-correction, levels VAR, and naive (no change) models to forecast changes in the yields of U.S. Treasury Bills. Curiously, they examined only one-step ahead forecasts, thereby ignoring the potential of cointegrated variables to aid in long-run forecasting. Hall et al. (1992) found the error-correction model provided forecast improvements. Fanchon and

---

[1] Cointegrating restrictions and first differencing the data are both forms of linear subtractive restrictions, which when falsely imposed lead to problems of specification bias, including potential sign reversals. For a discussion of the dangers of subtractive restrictions, see Haynes and Stone (1981).

[2] Recent research has questioned the usefulness of ranking models with out-of-sample forecasting statistics. Specifically, Clements and Hendry (1993, 1995) have objected to the conclusions of Engle and Yoo (1987) by noting that the model rankings depend upon the variable forecasted and the transformation of that variable. Although we acknowledge this criticism, we feel that out-of-sample forecasting statistics provide useful results. As theory indicates that the levels of the variables in question are cointegrated, it is appropriate to compare models on the basis of their ability to forecast levels of those variables. Also, our choice of variables to forecast depended upon what we felt would be the most likely variable of interest in a model. Based on these criteria, we chose output, inflation, and nominal interest rates as our variables to forecast.

Wendel (1992) compared the power of levels VAR, Bayesian VAR, and error-correction models to forecast cattle prices. They found the VAR in levels model to yield the best forecasts.

The results of our research indicate that error-correction models in which theoretical cointegrating relationships are imposed tend to outperform those in which the relationships are estimated. These theoretical error-correction models may produce superior forecasts when compared to difference or levels models, but this result hinges critically on the supporting evidence suggesting cointegration. As would be expected, forecasts are most likely to be improved by applying error-correction techniques if the data strongly supports the cointegration hypothesis. Difference models tend to fare well across most models, while levels models in general produce very poor forecasts. We hypothesize that models not sufficiently restricted by theoretical assumptions or empirical evidence may have a tendency to overfit the in-sample data.

II. Methodology

The Models

Consider z_t, an $n \times 1$ vector of nonstationary ($I(1)$) time-series variables. Assume that two or more of the elements of z_t form a stationary ($I(0)$) linear combination. In other words, there exists cointegration between some of the elements of z_t. The error-correction model representing such a system of variables is[3]:

$$\Delta z_t = \alpha w_{t-1} + \sum_{i=1}^{j} \gamma_i \Delta z_{t-i} + \varepsilon_t^{EC},$$

where $w_t = \beta' z_t$ is the cointegration term; β is an $n \times 1$ vector of coefficients in the cointegrating vector[4]; α is an $n \times 1$ vector of coefficients, and ε_t^{EC} is a vector of Gaussian error terms. The elements of the cointegrating vector can be determined either through theory or estimation.

We employed two approaches for obtaining values for β. The first is to let theory determine the value. For example, the approach taken by Cochrane (1994) and Shapiro and Watson (1988) sets $\beta = [1\ -1]$. Although this theoretical vector may yield a series which appears to be $I(0)$, it does not necessarily capture the true relationship between the two series. If $[1\ -1]$ is not the true cointegrating vector, the empirical model is misspecified. Alternatively, a widely-used method for estimating the coefficients of a cointegrating vector is that of Engle and Granger (1987). The Engle-Granger (EG) method regresses the first element of z_t on the other elements, and the resulting OLS estimates serve as the coefficients of the cointegrating vector. We examined models containing both theoretical and estimated cointegrating restrictions. For clarity, we have denoted models with theoretical cointegrating restrictions as EC–T and those with empirical restrictions as EC–E. This convention is followed below with respect to other models.

---

[3] Constant terms are suppressed throughout this paper.

[4] This specification assumes only one cointegrating vector. If there exists more than one cointegrating vector, then β will be an $n \times r$ matrix with r being the rank of the cointegrating matrix.

Table 1. Summary of Models

Model	Description
EC–T	Error-correction model with theoretical cointegrating vector
EC–E	Error-correction model with estimated cointegrating vector
VARD	VAR model in pure differences
VARL	VAR model in unrestricted levels

Note: Four-variable VAR models were written in differences (M4) and levels (M5) only. Univariate models were in levels only. Two- and three-variable models were examined in all five versions.

If, however, cointegration is not present in the variables of z_t, then the above error-correction models will be misspecified. For this reason, we also examined the forecasting properties of two other models. First, consider a VAR model in pure differences:

$$\Delta z_t = \sum_{i=1}^{j} \gamma_i \Delta z_{t-i} + \varepsilon_t^D. \tag{VARD}$$

This type of model is often recommended when the elements of z_t are difference stationary but not cointegrated. Second, consider a VAR model in pure levels:

$$z_t = \sum_{i=1}^{j} \gamma_i z_{t-i} + \varepsilon_t^L. \tag{VARL}$$

As noted earlier, estimating the VAR in levels avoids the possibility of imposing false restrictions on the model. The tradeoff for ensuring consistent estimators, however, may be a decrease in out-of-sample forecasting power. Table 1 summarizes the various categories of models.

In selecting the elements of z_t, we focused on two pairs of variables commonly believed to be cointegrated: 1) consumption and output, and 2) nominal interest rates and inflation. Whereas consumption and output are generally accepted to be cointegrated, the relationship between nominal interest rates and inflation is still a matter of debate. Controversy surrounds the issue of whether or not the nominal interest rate and inflation data contain unit roots. If either the inflation rate or the nominal interest rate is stationary, then the two variables are not cointegrated. In the absence of cointegration, a model which imposes a cointegrating restriction between nominal interest rates and inflation should result in poor forecasts of either variable. The same holds true for consumption and output. Imposing a cointegrating restriction when it does not exist should harm forecasting power, particularly in the long-run.

It has been suggested that two-variable models of inflation and interest rates are misspecified due to government deficit spending [for example, see Correia-Nunes and Stemitsiotis (1995)]. Recall that real short-term rates were very low in the mid-1970s, but increased sharply after 1980 to a peak in 1985. It is very possible that the U.S. fiscal position, particularly large budget deficits in the 1980s, influenced movements of the real interest rate. If this is true, failure to include this information might result in inaccurate forecasts. To account for the U.S. fiscal position, we also considered a three-variable augmented Fisher model which attempts to capture the effect of government deficit

Table 2. Lag Length Tests: χ^2 Statistics

System	H_0:	Models			
		EC–T	EC–E	VARD	VARL
c_t, y_t	2 vs. 3 lags	2.61 (4)	2.82 (4)	3.40 (4)	4.25 (4)
i_t, π_t	2 vs. 3 lags	5.7 (4)	5.87 (4)	6.28 (4)	24.21* (4)
	3 vs. 4 lags	—	—	—	7.00 (4)
i_t, π_t, d_t	2 vs. 3 lags	10.69 (9)	10.54 (9)	9.57 (9)	38.53* (9)
	3 vs. 4 lags	—	—	—	9.70 (9)
y_t, m_t, p_t, r_t	2 vs. 3 lags	—	—	19.44 (16)	69.16* (16)
	3 vs. 4 lags	—	—	—	21.87 (16)

Notes: Degrees of freedom for each test are listed in parentheses. An * indicates significance at the 10% or smaller level.

spending on interest rates by including a deficit to GDP variable.[5] Including this specification yields three distinct theoretical models to form the elements of z_t. First, the Permanent Income hypothesis, in which the elements of z_t are consumption and output. Second, the Fisher equation, with nominal interest rates and inflation rates as the elements of z_t.[6] Third, an augmented Fisher equation, which adds the deficit to GDP ratio to the standard Fisher equation. The theoretical cointegrating vector for the first two sets of models is $\beta = [1\ -1]$, while for the augmented Fisher equation $\beta = [1\ -1\ \lambda]$, with λ to be estimated via OLS. For comparison with models implied by the Fisher equations, we estimated four-variable VAR models of output, money, prices, and nominal interest rates (the popular YMPR class of models) in differences (VARD) and levels (VARL). As a baseline, we also estimated univariate levels models of output, inflation and nominal interest rates.

To determine appropriate lag lengths for the models, we relied on the likelihood ratio test outlined in Hamilton (1994, pp. 296–298). We began by testing the null hypothesis of two lags vs. the alternative of three lags. If the null hypothesis was rejected, we tested for the null of three lags vs. four lags. Table 2 lists the results. The likelihood ratio tests suggested a parsimonious specification for each model. In most cases, the tests suggested a two-lag specification, while three lags appears to be appropriate for the Fisher equation and augmented Fisher equation VARL models. These lag lengths follow Cochrane (1994), in which he estimated a cointegrated consumption/output model with two lags.

Forecasting

We used comparisons of out-of-sample forecasting ability to evaluate the alternative models. Whereas in-sample comparisons are prone to overfitting of the data, out-of-sample forecasting measures the predictive power of a model. The forecasting procedure was as follows: An initial estimation period was chosen for each system of equations. For

[5] We thank an anonymous referee for this suggestion.

[6] The Fisher hypothesis refers to *ex ante* real interest rates, the use of which requires a methodology for creating expected inflation rates. This complication can be overcome by noting that if the *ex ante* inflation rates are cointegrated with nominal interest rates, then so are *ex post* inflation rates. This follows from the rational expectations argument that the realized inflation rate from time t to $t + 1$ is the expected inflation rate plus an unforecastable stationary error term which is orthogonal to any information known at time t. For details, see Mishkin (1992). Also, any two cointegrated variables are also cointegrated at all leads and lags, although the cointegrating vector may differ [Engle (1991)].

consumption/output models, the initial estimation period was 1948:2–1965:2. This starting date closely corresponds with that of Cochrane (1994). Models containing interest rates were initially estimated over the period 1960:2–1972:2. The starting date for these models was limited by the lack of money data (for the four-variable model) prior to 1959:1. After estimation, values of output, nominal interest rates, and/or inflation were forecast from 1 to 40 quarters ahead. The model was re-estimated, with the end date moving forward one quarter, and variables were again forecast from 1 to 40 quarters ahead. This process continued until the end date of 1993:1, thereby generating a vector of 1 to 40 quarter ahead forecasts for each variable. A forecast horizon of 40 quarters was chosen to illustrate the forecasting power of the models in the short-, medium-, and long-runs. Because cointegration is a long-run property of the data, we expected the effects of correctly or incorrectly modeling cointegrated variables to be most striking at longer horizons. To evaluate the performance of the models, the root mean square error (RMSE) was calculated for each forecast period.

Some care must be taken when evaluating models based upon RMSE statistics. Problems arise when an outlier, in particular a very poor forecast, causes a model to exhibit large RMSE values at some forecasting horizons. To account for this issue, we employed the forecasting encompassing test of Chong and Hendry (1986). The forecast-encompassing technique addresses the issue of whether or not the forecast of one model contains useful forecasting information not contained in a competing model. For example[7], consider 2 time $t + j$ ahead forecasts of variable x (x_{t+j}), denoted by x^{f1}_{t+j} and x^{f2}_{t+j}. The forecasts were generated by models 1 and 2, respectively. The test was implemented through the regression equation:

$$x_{t+j} = \alpha x^{f1}_{t+j} + (1 - \alpha) x^{f2}_{t+j} + \varepsilon.$$

Model 1 encompasses model 2 if $\alpha \neq 0$, or, in other words, if the forecast generated by model 1 contains useful information not contained in the model 2 forecast. If $\alpha \neq 1$, then the opposite occurs and model 2 encompasses model 1. If $\alpha = 1$, then model 1 encompasses model 2 but model 2 does not encompass model 1. This indicates that model 1 contains superior forecasting information compared to model 2. Model 2 is superior to model 1 when $\alpha = 0$. In this case, model 2 encompasses model 1, but model 1 does not encompass model 2. Thus, the forecast-encompassing test is implemented by generating the competing forecasts and performing the above regression. The null hypotheses, $\alpha = 0$ and $\alpha = 1$, are then tested using a standard t test. If $\alpha = 1$ and $\alpha = 0$ or $\alpha \neq 1$ and $\alpha \neq 0$, the test is inconclusive and neither model displays clear superiority over the other.

III. Results

Data

The variables considered are consumption (c_t), output (y_t), the nominal interest rate (i_t), the money stock (m_t), inflation (π_t), and the deficit to GDP ratio (d_t). Following Cochrane (1994), consumption is defined as the log of consumer spending on nondurables and services. Also following Cochrane, output is the log of private GNP, defined as GNP minus government spending. Both consumption and output are real variables measured in

[7] This description follows Koenig (1996).

Table 3. Unit Root Tests

Series	Dickey-Fuller t Statistic
c_t	0.22
y_t	-1.51
π_t	-2.44
i_t	-2.11
d_t	-1.82
$i_t - \pi_t$	-2.17
$c_t - y_t$	-3.79**
$i_t - \pi_t - \lambda d_t$	-2.09
RES1	-2.10
RES2	-3.79**
RES3	-2.38

Notes: Following Shapiro and Watson (1988), the regressions for c_t, y_t, and $c_t - y_t$ included a constant term and a time trend, and had a 10% critical value of -3.12. All other regressions included only a constant term and had a 10% critical value of -2.57. An * indicates significance at the 10% level. A ** indicates significance at the 5% level.

1987 dollars. Nominal interest rates are the average of monthly rates of 90-day Treasury Bills. *Ex-ante* inflation is proxied by the difference in the log of the price level between period t and period $t - 1$, $\pi_t \equiv \ln p_t - \ln p_{t-1}$. The real interest rate is $r_t \equiv i_t - \pi_t$. The price level is the average of monthly data of the consumer price index with all items included. The money stock is the average of monthly data of M2. Models were estimated with quarterly U.S. data obtained from Citibank's *Citibase* data set.

Data Analysis

A necessary requirement for cointegration is that the variables in question are each integrated of the same order. Whether π_t and i_t contain unit roots is, however, a controversial issue.[8] That c_t and y_t contain unit roots is less controversial, but the low power of unit root tests still leaves room for doubt. As the unit root controversy is beyond the scope of this paper, we followed convention by listing Dickey-Fuller test statistics for c_t, y_t, π_t, i_t, and d_t in Table 3. The tests were based on data corresponding to model estimation dates using six autoregressive corrections.[9] In all five cases, the null of a unit root was not rejected at the 10% significance level. Thus, the Dickey-Fuller tests suggest that first differencing these series is appropriate.

Recalling that two series (integrated of the same order) are cointegrated if a linear combination of the series is a stationary ($I(0)$) series, Table 3 also lists Dickey-Fuller test statistics for $i_t - \pi_t$, $c_t - y_t$, and $i_t - \pi_t - \lambda d_t$. This is the method by which Shapiro and Watson (1988) tested for cointegration between i_t and π_t. As noted earlier, however, this formulation imposes the theoretical cointegrating vectors onto the variables. As an alternative, we also utilized the EG method for determining cointegrating vectors and regressed i_t on π_t, c_t on y_t, and i_t on π_t and d_t, and performed Dickey-Fuller tests on the residual series, RES1, RES2, and RES3, respectively. The results are also listed in Table 3. We failed to reject the null of a unit root in both the theoretical and estimated cointegrating relationships between i_t and π_t and i_t, π_t, and d_t. This indicates that there is

[8] Mishkin (1992) presented an extensive battery of stationarity tests on inflation and nominal interest rates.
[9] Alternative lag lengths failed to appreciably alter our results.

Table 4. Johansen Procedure Test Results

A. Multivariate Unit Root Tests

System	H_0:	λ_{max}	Trace
c_t, y_t	2 vs. 1 unit roots	23.86*	31.71*
i_t, π_t	2 vs. 1 unit roots	10.60	17.10*
i_t, π_t, d_t	3 vs. 2 unit roots	30.09*	42.02*
	2 vs. 1 unit roots	7.47	11.93

B. Tests of Linear Restrictions on the Cointegrating Vectors

System	H_0:	χ^2 (DOF)
c_t, y_t	$\beta = [1\ -1]$	0.38 (1)
i_t, π_t	$\beta = [1\ -1]$	0.52 (1)
i_t, π_t, d_t	$\beta = [1\ -1\ \lambda]$	9.53 (1)*

Notes: The 10% critical values for the λ_{max} statistic are 13.39 and 10.60 for the null of three unit roots and the null of two unit roots, respectively. The corresponding 10% critical values for the Trace statistic are 26.70 and 13.31. Full details concerning the test statistics are found in Johansen and Juselius (1990). Testing was performed using the CATS for RATS econometric package. The number of lags corresponds to those determined in Table 2. An * indicates significance at the 10% level.

no cointegration present among the elements of both the Fisher and augmented Fisher equations. We rejected the null of a unit root in both the theoretical and estimated c_t and y_t relationships. This indicates that these series are cointegrated.

Additionally, we used the Johansen maximum likelihood procedure to test for cointegration within the three systems of equations we analyzed. Table 4A lists the λ_{max} and trace statistics from multivariate unit root tests on each of the systems. The null hypothesis of no cointegration (two vs. one unit root) was rejected by both tests for the consumption/output system, providing evidence consistent with the Dickey-Fuller results in Table 3. The results for the interest rate/inflation rate system are contradictory. The λ_{max} statistic indicates no cointegration, while the trace statistic indicates the opposite. Stronger evidence of cointegration is found in the interest rate/inflation rate/deficit ratio system. Both test statistics rejected the null hypothesis of no cointegration (three vs. two unit roots), while the null of one cointegrating vector against two cointegrating vectors (two vs. one unit root) was not rejected, indicating there exists one cointegrating vector in the system. Table 4B lists likelihood ratio test statistics for tests of linear restrictions on the cointegrating vectors. In the two-variable systems, we were unable to reject the null that the cointegrating vectors are $[1\ -1]$. In the three variable system, we rejected the null that the cointegrating vector is $[1\ -1\ \lambda]$, where the λ corresponds with the parameter on d_t. Given the above results, it appears that there is evidence to support the imposition of the linear restrictions implied by the theoretical cointegrating vectors for the consumption/output systems, while less evidence exists for imposing the theoretical restriction in the remaining systems.

Based upon the above test results and following prior research, we expect to find the following results: First, if the cointegrating restrictions were correctly identified, error-correction models should yield the best forecasts. If theory correctly defines relevant restrictions, then models with the theoretical cointegrating vector (EC–T) should outperform those with the estimated cointegrating vector (EC–E). If the theory does not hold, but the series are still cointegrated, then the opposite may occur. Note that it is possible for the theoretical parameters to be incorrect, but still be closer than the estimated parameters

Table 5. Output Forecasts: RMSE

	Univariate	Two-Variable Models			
Period	Levels	EC–T	EC–E	VARD	VARL
1	0.011	0.098	0.010	0.010	0.010
2	0.018	0.015	0.016	0.017	0.017
3	0.025	0.020	0.023	0.023	0.023
4	0.032	0.024	0.028	0.028	0.028
8	0.054	0.036	0.044	0.047	0.043
16	0.080	0.047	0.057	0.062	0.057
24	0.10	0.053	0.066	0.069	0.066
32	0.13	0.062	0.077	0.080	0.082
40	0.17	0.077	0.092	0.097	0.11

to the true values. Second, if the series are not cointegrated and do not contain units roots, then the unrestricted levels models (VARL) will yield the best forecasts. If unit roots are present, then the best forecasts will come from pure difference models (VARD). Finally, the imposition of theoretical or empirical cointegrating restrictions, if true, should aid in long-run forecasts.

Results

For ease of exposition, we begin by summarizing our conclusions in the following list:

- Imposing a theoretical cointegrating restriction (EC–T models) often improves forecasting power.
- Theoretical cointegrating restrictions are preferred to empirical cointegrating restrictions (EC–E models).
- As we became more confident in the presence of cointegration between the variables examined, the benefits from imposing cointegrating restrictions increased.
- VARs in pure differences performed well across models.
- VARs in pure levels tended to have poor forecasting ability.

Table 5 contains output forecast results generated by the output/consumption models. The RMSE statistics reveal that short-run forecasts were very similar, but the models began to diverge after eight quarters. Most notably, the theoretical error-correction model (EC–T) yielded the best long-run forecasts. The remaining models, EC–E, VARD and VARL, all produced similar forecasts, while all models were superior to the univariate levels model.

Tables 6 and 7 report forecasting results for inflation/interest rate models. In comparison to the output/consumption models, the results are less clear. First consider the two-variable Fisher equation models. Again, we see a pattern of similar short-run forecasts with divergence after eight quarters. Unlike the output forecasts, however, the difference (VARD) forecasts appear to be superior to the EC–T models. Also, the estimated error-correction model (EC–E) performed poorly at long-run horizons. This indicates that the EC–T model is preferred to the EC–E model, or, in other words, imposing theoretical

Table 7. Inflation-Rate Forecasts: RMSE

	Univariate	Two-Variable Models				Three-Variable Models				Four-Variable Models	
Period	Levels	EC–T	EC–E	VARD	VARL	EC–T	EC–E	VARD	VARL	VARD	VARL
1	1.88	1.95	1.98	1.88	1.95	1.99	2.03	1.93	2.08	2.52	2.69
2	2.36	2.68	2.73	2.53	2.71	2.71	2.75	2.57	2.86	2.55	2.76
3	2.52	2.95	2.96	2.72	2.94	2.95	2.92	2.73	3.07	2.63	2.90
4	2.93	3.46	3.54	3.18	3.45	3.40	3.44	3.14	3.53	2.94	3.18
8	4.17	4.99	5.63	4.63	5.36	4.74	5.56	4.66	5.33	4.20	4.57
16	4.21	6.15	8.01	5.51	6.72	5.45	7.76	5.61	6.80	4.39	5.42
24	4.83	6.65	11.50	6.18	8.96	6.22	11.21	6.35	10.04	4.68	7.32
32	6.60	7.54	17.14	7.83	15.11	6.63	17.93	7.95	18.66	5.71	10.27
40	8.06	9.19	26.18	9.01	25.71	7.83	31.71	9.19	42.96	5.61	13.33

cointegrating restrictions appears to be preferable to estimated restrictions. Overall, however, a difference model is preferred to an error-correction model.[10]

The three-variable augmented Fisher equation models illustrate an attempt to increase forecasting power by imposing additional theory. Within this class of models, the EC–T model generally dominated the VARD model, particularly in the long-run. This result is opposite that of the two-variable models, indicating that the Fisher effect might be more significant when augmented to include the deficit to GDP ratio. The VARD model, however, performed quite well, and was superior to both the EC–E and VARL models.

Across all inflation/interest rate models, the best forecasts were generated by the four-variable VARD models, while the univariate levels model generated inflation forecasts comparable to those of the best two- and three-variable models. The VARL models

Table 6. Nominal Interest-Rate Forecasts: RMSE

	Univariate	Two-Variable Models				Three-Variable Models				Four-Variable Models	
Period	Levels	EC–T	EC–E	VARD	VARL	EC–T	EC–E	VARD	VARL	VARD	VARL
1	1.00	1.05	1.07	1.05	1.10	1.05	1.08	1.06	1.12	1.07	1.17
2	1.74	1.76	1.81	1.76	1.90	1.76	1.85	1.78	1.94	1.72	1.98
3	1.94	1.94	2.00	1.94	2.14	1.98	2.05	1.99	2.23	1.94	2.24
4	2.23	2.25	2.36	2.24	2.54	2.29	2.41	2.28	2.66	2.22	2.71
8	3.49	3.40	3.67	3.36	4.13	3.35	3.83	3.41	4.39	3.42	4.35
16	4.65	5.14	5.87	4.64	7.73	4.67	6.61	4.72	9.00	5.00	7.08
24	6.76	6.38	9.75	5.34	16.07	5.86	12.13	5.57	21.50	6.26	12.02
32	9.20	6.43	16.23	6.03	35.53	5.89	22.78	6.18	54.39	7.15	19.45
40	13.99	7.97	30.22	6.73	85.27	6.82	46.47	6.90	145.32	7.97	33.96

[10] Some readers might note that the cointegrating vector implied by the Fisher equation fails to take into account the effects of taxes on interest rates, a result known as the Darby (1975) effect. The effect of taxes implies that the cointegrating vector between inflation and nominal interest rates will be $[1 - 1/(1 - T)]$, where T is the lender's marginal tax rate on interest earned. Estimates of the appropriate coefficient on inflation ranged from 1.47 to 2.0 [see Ayanian (1983)]. To determine if changing the cointegrating vector $[1\ -1.47]$. The forecasts did not change sufficiently to have any qualitative effects on our results. This result is robust for values of the coefficient on inflation of 1.47, 1.57, and 1.67.

Table 8. Output Forecasts: Forecast-Encompassing Results: Significance Levels

	Two-Variable models					
	EC–T vs. EC–E		EC–T vs. VARD		EC–T vs. VARL	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.01*	0.52	0.00*	0.57	0.00*	0.59
2	0.00	0.09	0.00*	0.27	0.00*	0.54
3	0.00	0.01	0.00*	0.28	0.00*	0.14
4	0.00	0.00	0.00*	0.36	0.00*	0.3
8	0.00	0.00	0.00*	0.62	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00	0.00	0.00
32	0.00	0.00	0.00	0.09	0.00	0.00
40	0.00	0.00	0.00	0.04	0.00	0.00

Note: An * indicates instances when the EC–T model produced forecasts superior to the competing model, while a † indicates the opposite. The lack of an * or a † indicates the test was inconclusive.

and the EC–E models produced the poorest results. Univariate interest rate forecasts compared favorably at short- and medium-run horizons, but poorly in the long run. In contrast, the univariate inflation forecasts were superior to most other models, except the four-variable VARD model.

The RMSE statistics tend to suggest that imposing theoretical restrictions, when true, via an error-correction model, may yield superior forecasts. As noted earlier, however, RMSE statistics may be biased by the presence of an outlier, leading the researcher to erroneous conclusions. As an alternative to the RMSE statistics, we also performed forecast encompassing tests as outlined in Section II. Forecast-encompassing results for the output/consumption models are listed in Table 8. An * indicates the EC–T forecast was clearly superior to that of the competing models at the listed forecast horizon. A † indicates the opposite. Lack of either an * or a † indicates the test was inconclusive. Table 8 shows that none of the competing output/consumption models were superior to the EC–T model. The EC–T model, however, is preferred to the others only in the short-run. In the case of the EC–E model, the EC–T model is preferred at only the one-quarter horizon. Thus, while the forecast-encompassing results do not directly indicate that imposing theoretical cointegrating restrictions produces superior forecasts in the long-run, it is clear that the opposite is also not supported. Imposing the restrictions did not *harm* forecasts. The inability of the forecast encompassing tests to distinguish between models may also have been due to a lack of data.

Among the inflation/interest rate models, the most significant competition appeared between the EC–T and VARD models. Given this result, we focused the forecast-encompassing tests on these two models for the two- and three-variable models. Tables 9 and 10 list the results for the inflation and interest-rate models, respectively. An examination of these tables fails to yield a clear picture. At some horizons, the EC–T model was superior to the VARD model, while at others, the opposite was true. These results coincide with the murky RMSE results. It is unclear whether or not imposing cointegrating restrictions aids forecasting.

Tables 11, 12, and 13 report forecast-encompassing results for a comparison of univariate models vs. EC–T models. For output forecasts, the EC–T model was superior to the univariate model in the short-run, but the two models could not be differentiated in

Table 9. Inflation-Rate Forecasts: Forecast-Encompassing Results: Significance Levels

	Two-Variable Models		Three-Variable Models	
	EC–T vs. VARD		EC–T vs. VARD	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.14	0.00†	0.12	0.01†
2	0.03	0.00	0.06	0.00
3	0.02	0.00	0.02	0.00
4	0.05	0.00	0.09	0.00
8	0.36	0.00†	0.41	0.08†
16	0.97	0.00†	0.03*	0.36
24	0.04*	0.23	0.01*	0.92
32	0.08	0.05	0.00*	0.21
40	0.12	0.18	0.23	0.42

Note: An * indicates instances when the EC–T model produced forecasts superior to the competing model, while a † indicates the opposite. The lack of an * or a † indicates the test was inconclusive.

the long-run. Univariate inflation forecasts were almost always preferred to the EC–T inflation forecasts. In contrast, no clear difference emerged between univariate and EC–T interest-rate forecasts. Note that the univariate comparisons of the EC–T model follow the pattern of previous comparisons: Output models compared favorably, while the opposite was true for inflation/interest-rate models.

How well do these results match our predictions? Among the output/consumption models, RMSE statistics suggest that imposing the theoretical cointegrating restrictions improves long-run forecasts, while the forecast-encompassing tests provide no evidence to the contrary. In comparison, imposing the theoretical cointegrating restrictions did not consistently improve forecasts for inflation/interest-rate models. This result is evident from both the RMSE and forecast-encompassing statistics. The inability of cointegrating restrictions to significantly improve inflation/interest-rate forecasts initially seems puzzling. If interest rates and inflation are cointegrated, as is often assumed, then lack of an error-correction term should cause the differences model to be misspecified. If they are not cointegrated, one would expect models which impose the cointegrating vector to result in

Table 10. Nominal Interest-Rate Forecasts: Forecast-Encompassing Results: Significance Levels

	Two-Variable Models		Three-Variable Models	
	EC–T vs. VARD		EC–T vs. VARD	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.39	0.41	0.23	0.86
2	0.22	0.27	0.13	0.78
3	0.14	0.16	0.14	0.44
4	0.19	0.08†	0.23	0.19
8	0.17	0.01†	0.07*	0.18
16	0.43	0.00†	0.06*	0.15
24	0.01	0.00	0.34	0.00†
32	0.72	0.00†	0.67	0.00†
40	0.89	0.87	0.53	0.61

Note: See Table 9 note.

Table 11. Output Forecast-Encompassing Results: Significance Levels

	Two-Variable Model	
	UNI vs. EC–T	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.95	0.00†
2	0.94	0.00†
3	0.64	0.00†
4	0.76	0.00†
8	0.60	0.00†
16	0.00	0.00
24	0.00	0.00
32	0.05	0.00
40	0.00	0.00

Note: An * indicates the univariate model was superior to the EC–T model, while a † indicates the opposite.

poor forecasting ability. Instead, we see little difference between the EC–T and VARD models.

Why does imposing the theoretical cointegrating vector assist long-run forecasts of output but not inflation or interest rates? As noted earlier, the existence of cointegration between interest rates and inflation is fairly controversial, particularly in light of visual analysis of the two series and the results of cointegration tests. Little evidence supports cointegration in the two-variable Fisher equation rate model, while Johansen tests appear to support cointegration in the three-variable augmented Fisher equation model. In contrast, evidence supports the presence of cointegration between output and consumption. Recall that the benefits from imposing the theoretical cointegrating restrictions were nonexistent in the two-variable Fisher equation model, but improved RMSE statistics in the three-variable augmented Fisher equation model. In neither case did forecast-encompassing results support the error-correction model. Cointegrating restrictions, however, clearly tended to improve output forecasts. Thus, it appears from our results that as the likelihood that a cointegrating relationship exists between two series increases, the

Table 12. Inflation-Rate Forecast-Encompassing Results: Significance Levels

	Two-Variable Model		Three-Variable Model	
	UNI vs. EC–T		UNI vs. EC–T	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.01	0.01	0.00	0.01
2	0.00*	0.12	0.00*	0.15
3	0.00*	0.65	0.00*	0.79
4	0.00*	0.55	0.00*	0.94
8	0.00*	0.10	0.00*	0.60
16	0.00	0.00	0.00*	0.32
24	0.00	0.07	0.00*	0.12
32	0.00*	0.85	0.53	0.00†
40	0.09*	0.27	0.16	0.52

See Table 11 note.

Table 13. Interest-Rate Forecast-Encompassing Results: Significance Levels

	Two-Variable Model		Three-Variable Model	
	UNI vs. EC–T		UNI vs. EC–T	
Period	H_0: $\alpha = 0$	H_0: $\alpha = 1$	H_0: $\alpha = 0$	H_0: $\alpha = 1$
1	0.06	0.00	0.04	0.00
2	0.07	0.01	0.04	0.01
3	0.30	0.01†	0.13	0.01†
4	0.11	0.06†	0.05	0.10
8	0.02	0.08	0.10	0.05
16	0.00*	0.94	0.00*	0.64
24	0.00*	0.81	0.01*	0.53
32	0.00	0.11	0.18	0.05†
40	0.62	0.24	0.47	0.14

See Table 11 note.

forecasting benefits from an error-correction model also increase. This result was expected and reveals precisely why the issue of cointegration is important to researchers modeling long-run economic relationships.

Another prominent result of our paper is the generally strong performance of difference models in contrast to the weak performance of levels models. The strength of VARD models is not surprising considering that Dickey-Fuller tests indicated the presence of a unit root in every series. Thus, first differencing of the data is appropriate.[11] In most cases, the predictive power of VARL models decreased quickly after eight quarters. Given the popularity of this class of models in the literature, this result is surprising. It is commonly accepted that the levels model will produce consistent coefficient estimates, although the estimates may be inefficient if the true model is in differences and/or requires an error-correction term. Our results indicate, however, that these models forecast poorly out-of-sample. An explanation for this result is that the models tended to overfit the data in-sample at the expense of out-of-sample forecasting power. Similarly, models containing the theoretical cointegrating vector (EC–T) tended to outperform those containing the estimated cointegrated vector (EC–E). Again, this may be an example of in-sample overfitting. Like the VARL models, the EC–E models have additional freedom to fit the data compared to more restricted alternatives. This additional freedom, however, may overemphasize fitting the data in-sample and, in the process, neglect determining the true dynamic nature of the model.

How do these results aid in the choice of appropriate modeling technique? It appears preferable to model consumption and income as cointegrated variables; doing so clearly improves long-run forecasts of output. Our results, however, call into question the usefulness of treating inflation and nominal interest rates as cointegrated.

A notable paper arguing in support of a long-run Fisher effect is Mishkin (1992). In his study, Mishkin found evidence to support the conclusion that there exists a long-run Fisher effect which is most evident when inflation and nominal interest rates both exhibit

[11] Another possibility for the strength of the VARD models is that difference models may be superior when the long-run underlying relationships are not stable. This may be of additional importance for interest-rate models considering issues such as changes in the U.S. fiscal position, capital mobility, and risk premiums. We thank an anonymous referee for identifying this point.

stochastic trends. The existence of a short-run Fisher effect, where changes in the nominal interest rate predict changes in inflation, was rejected. Furthermore, Mishkin (1992) found that the presence of cointegration is stronger in the theoretical Fisher equation than in an estimated equation. This finding implies that error-correction models in which the theoretical cointegrating vector is imposed will produce forecasts superior to models with estimated vectors. Indeed, this is exactly what we found.

Researchers should take note, however, that our results imply that caution should be used when describing inflation and nominal interest rates as cointegrated. Modeling inflation and interest rates as cointegrated variables does not appear to add any additional benefits for the purpose of modeling long-run forecasts of either variable. In fact, such a technique appears to be inferior to methods which, if cointegration were present, should be misspecified, most notably the two- and four-variable VARs in pure differences. It appears that differences models are most suited to describing the future path of inflation.

IV. Conclusion

In this paper, we estimated a variety of familiar models of cointegrated variables. The estimated models were used to create output, nominal interest rate, and inflation-rate forecasts from 1 to 40 quarters ahead, from which root mean square error and forecast-encompassing statistics were calculated. These statistics were then used to compare the predictive powers of the various models, with the goal of determining the effects of cointegrating restrictions on long-run forecasting.

Our findings provide results researchers may find useful in modeling macroeconomic time series. A significant conclusion is that the failure to impose appropriate restrictions may be a misspecification as significant as imposing false restrictions. Models with fewer restrictions, such as levels models and those with estimated cointegrating restrictions tended to possess poorer forecasting performance than their more restricted counterparts. This is not meant to imply that theoretical restrictions unsupported by firm empirical evidence will aid in forecasting performance. For example, some theories have qualitative rather than quantitative predictions, or there may exist competing theories about the magnitudes of model parameters. What is implied is that restrictions based upon solid theoretical or empirical evidence should be closely examined before being discarded *a priori*. Theoretical cointegrating restrictions appeared to aid long-run forecasts of output in a two-variable error-correction model, but were detrimental to long-run forecasts of nominal interest rates and inflation in a similar two-variable model. RMSE statistics reveal that forecasts from the three-variable augmented Fisher equation system may be improved by implementing an error-correction technique, but this result could not be substantiated by the forecast-encompassing tests. We feel this discrepancy may be attributed to a weaker cointegrating relationship between interest rates and inflation than that between output and consumption, perhaps reflecting that asset-pricing relationships are more vulnerable to shifts than are consumption patterns. It is important to note that it is unlikely that the theoretical cointegrating restriction is a complete misspecification; if this had been the case, this model would not be ranked second best among the two-variable nominal interest-rates/inflation-rates models. It does appear, however, that treating inflation and nominal interest rates as cointegrated does not provide any additional benefits from a forecasting perspective. The opposite is true for consumption and output. Treating these two variables as cointegrated does aid in long-run forecasting performance.

What may be more clearly a poor choice of models is any of the levels models, particularly the four-variable YMPR model. Contrary to our predictions, these models performed fairly poorly. Difference models, in general, did very well, and in the case where cointegration was weak (inflation and interest rates), produced the best forecasts, dominating models with cointegrating restrictions. When viewed together, the two economic relationships examined here highlight conditions under which accounting for hypothesized long-run relationships may worsen (as with inflation and interest rates) or improve (as with consumption and income) our ability to forecast economic variables.

This paper has benefited from comments by Jo Anna Gray, George Evans, and two anonymous referees. The usual disclaimer applies.

References

Ayanian, R. Sept. 1983. Expectations, taxes and interest: The search for the Darby effect. *American Economic Review* 73(4): 762–765.

Bernanke, B. S., and Blinder, A. S. Sept. 1992. The federal funds rate and the channels of monetary transition. *American Economic Review* 82(4):901–921.

Chong, Y. Y., and Hendry, D. F. Aug. 1986. Econometric evaluation of linear macro-economic models. *Review of Economic Studies* 53(4):671–690.

Clements, M. P., and Hendry, D. F. Dec. 1993. On the limitations of comparing mean square forecast errors. *Journal of Forecasting* 12(8):617–637.

Clements, M. P., and Hendry, D. F. April–June 1995. Forecasting in cointegrated systems. *Journal of Applied Econometrics* 10(2):127–146.

Cochrane, J. H. Feb. 1994. Permanent and transitory components of GNP and stock prices. *Quarterly Journal of Economics* 109(1):241–265.

Correia-Nunes, J., and Stemitsios, L. Nov. 1995. Budget deficit and interest rates: Is there a link? International Evidence. *Oxford Bulletin of Economics and Statistics* 57(4):425–450.

Darby, M. R. June 1975. The financial and tax effects of monetary policies on interest rates. *Economic Inquiry* 13(2):266–276.

Granger, C. 1991. Developments in the study of cointegrated economic variables. In *Long-Run Economic Relationships: Readings in Cointegration* (R. F. Engle and C. W. J. Granger, eds.). New York: Oxford University Press, pp. 65–80.

Engle, R. F., and Granger, C. W. J. March 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55(2):251–276. Reprinted in *Long-Run Economic Relationships: Readings in Cointegration* (R. F. Engle and C. W. J. Granger, eds.). New York: Oxford University Press.

Engle, R. F., Granger, C. W. J., and Hallman, J. J. May 1989. Merging short- and long-run forecasts: An application of seasonal cointegrating to monthly electricity sales forecasting. *Journal of Econometrics* 40(1):45–62. Reprinted in *Long-Run Economic Relationships: Readings in Cointegration* (R. F. Engle and C. W. J. Granger, eds.). New York: Oxford University Press.

Engle, R. F., and Yoo, B.-S. May 1987. Forecasting and testing in co-integrated systems. *Journal of Econometrics* 35(1):143–159. Reprinted in *Long-Run Economic Relationships: Readings in Cointegration* (R. F. Engle and C. W. J. Granger, eds.). New York: Oxford University Press.

Fanchon, P., and Wendel, J. Feb. 1992. Estimating VAR models under non-stationarity and cointegration: Alternative approaches for forecasting cattle prices. *Applied Economics* 24(2): 107–217.

Fisher, I. 1930. *The Theory of Interest*. New York: Macmillan.

Hall, A. D., Anderson, H. M., and Granger, C. W. J. Feb. 1992. A cointegration analysis of Treasury Bill yields. *The Review of Economics and Statistics* 74(1):116–126.

Hall, R. Oct. 1978. Stochastic implications of the life cycle permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86(5):971–987.

Hamilton, J. D. 1994. *Time-Series Analysis*. Princeton: Princeton University Press.

Haynes, S. R., and Stone, J. April 1982. Spurious tests and sign reversals in international economics. *Southern Economic Journal* 48(4):868–876.

Johansen, S., and Juselius, K. May 1990. The full information maximum likelihood procedure for inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52(2):169–210.

Koenig, E. F. Third Quarter 1996. Capacity utilization as a real-time predictor of manufacturing output. Federal Reserve Bank of Dallas *Economic Review* 16–23.

Mishkin, F. S. Nov. 1992. Is the Fisher effect for real? A reexamination of the relationship between inflation and interest rates. *Journal of Monetary Economics* 30(2):195–215.

Shapiro, M. D., and Watson, M. W. 1988. Sources of business cycle fluctuations. In *NBER Macroeconomics Annual* (O. Blanchard and S. Fischer, eds.). Cambridge: MIT Press, pp. 111–141.

Sims, C. A. Jan. 1980. Macroeconomics and reality. *Econometrica* 48(1):1–48.