WORKING PAPER SERIES

Predictability in International Asset Returns:
A Reexamination

Christopher Neely
Paul Weller

Working Paper 1997-010D
http://research.stlouisfed.org/wp/1997/97-010.pdf

PUBLISHED: Journal of Financial and Quantitative Analysis. December 2000, pp. 602-20

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

Predictability in International Asset Returns: A Reexamination

February 1999

Abstract

This paper argues that inferring long-horizon asset-return predictability from the properties of vector autoregressive (VAR) models on relatively short spans of data is potentially unreliable. We illustrate the problems that can arise by re-examining the findings of Bekaert and Hodrick (1992), who detected evidence of in-sample predictability in international equity and foreign exchange markets using VAR methodology for a variety of countries over the period 1981-1989. The VAR predictions are significantly biased in most out-of-sample forecasts and are conclusively outperformed by a simple benchmark model at horizons of up to six months. This remains true even after corrections for small sample bias and the introduction of Bayesian parameter restrictions. A Monte Carlo analysis indicates that the data are unlikely to have been generated by a stable VAR. This conclusion is supported by an examination of structural break statistics. Implied long-horizon statistics calculated from the VAR parameter estimates are shown to be very unreliable.

Keywords: Vector Autoregression, Asset Price, Exchange Rate, Forecasting

JEL Classification: C32, F30

Christopher Neely, Economist
Research Department
Federal Reserve Bank of St. Louis
411 Locust Street
St. Louis, MO 63102
neely@stls.frb.org

Paul Weller, Professor
Department of Finance
College of Business Administration
University of Iowa
Iowa City, IA 52240
paul-weller@uiowa.edu

The authors wish to thank Robert Hodrick for supplying the data used in Bekaert and Hodrick (1992), Morgan Stanley for providing us with additional financial data and Kent Koch for excellent research assistance. We also thank Chuck Whiteman, Dick Anderson, Bob Rasche, Gene Savin and Dan Thornton for helpful comments and Robert Hodrick and Paul Sengmuller both for helpful comments and for pointing out an error in one of our programs. Paul Weller would like to thank the Research Department of the Federal Reserve Bank of St. Louis for its hospitality during his stay as a Visiting Scholar, when this work was initiated.

Predictability in International Asset Returns: A Reexamination

February 1999

Abstract

This paper argues that inferring long-horizon asset-return predictability from the properties of vector autoregressive (VAR) models on relatively short spans of data is potentially unreliable. We illustrate the problems that can arise by re-examining the findings of Bekaert and Hodrick (1992), who detected evidence of in-sample predictability in international equity and foreign exchange markets using VAR methodology for a variety of countries over the period 1981-1989. The VAR predictions are significantly biased in most out-of-sample forecasts and are conclusively outperformed by a simple benchmark model at horizons of up to six months. This remains true even after corrections for small sample bias and the introduction of Bayesian parameter restrictions. A Monte Carlo analysis indicates that the data are unlikely to have been generated by a stable VAR. This conclusion is supported by an examination of structural break statistics. Implied long-horizon statistics calculated from the VAR parameter estimates are shown to be very unreliable.

Keywords: Vector Autoregression, Asset Price, Exchange Rate, Forecasting

JEL Classification: C32, F30

Christopher Neely, Economist
Research Department
Federal Reserve Bank of St. Louis
411 Locust Street
St. Louis, MO 63102
neely@stls.frb.org

Paul Weller, Professor
Department of Finance
College of Business Administration
University of Iowa
Iowa City, IA 52240
paul-weller@uiowa.edu

The authors wish to thank Robert Hodrick for supplying the data used in Bekaert and Hodrick (1992), Morgan Stanley for providing us with additional financial data and Kent Koch for excellent research assistance. We also thank Chuck Whiteman, Dick Anderson and Bob Rasche for helpful comments and Robert Hodrick and Paul Sengmuller both for helpful comments and for pointing out an error in one of our programs. Paul Weller would like to thank the Research Department of the Federal Reserve Bank of St. Louis for its hospitality during his stay as a Visiting Scholar, when this work was initiated.

Introduction

Early empirical work on stock prices and exchange rates found it very difficult to reject the hypothesis that these prices followed a random walk (Fama, 1970; Meese and Rogoff, 1983). Subsequently, however, research has produced evidence for the existence of transitory, or mean-reverting components in both equity and foreign exchange markets (Summers, 1986; Campbell, 1987; Fama and French, 1988; Poterba and Summers, 1988; Mark, 1995). Although the earlier findings were the subject of some dispute on statistical grounds (Lo and MacKinlay, 1988; Nelson and Kim, 1993; Richardson, 1993) more recent research has tended to support and further refine the evidence for return predictability (Hodrick, 1992; Lamont, 1998). Most of these studies, however, have focused exclusively on *in-sample* evidence of predictability.[1]

Evidence of in-sample predictability implies out-of-sample predictability only if the estimated structural relationships are sufficiently stable over time. This is an argument for focusing on data collected over long time spans. If a relationship has persisted over fifty, or better yet, a hundred years, then one will clearly have more confidence that the relationship will continue to hold in the future. But recently a number of authors (Hodrick, 1992; Bekaert and Hodrick, 1992; Campbell, Lo and MacKinlay, 1997) have advocated an alternative approach, succinctly described as follows:

> "*An alternative approach [to looking directly at the long-horizon properties of the data] is to assume that the dynamics of the data are well described by a simple time-series model; long-horizon properties can then be imputed from the short-run model rather than estimated directly.*" (Campbell, Lo and MacKinlay, 1997, p.280):

These authors are careful to point out that such procedures are critically dependent on the

[1] Mark (1995) and Lamont (1998) are exceptions.

assumption of stability of the estimated parameters. However, it is rare in studies of asset price predictability to find any formal investigation of this issue.

The objective of this paper is to point out the potential pitfalls inherent in this approach, and to argue that more attention needs to be paid to the question of structural stability in studies of asset price predictability. Meese and Rogoff (1983) made this point forcefully in the context of structural models of the exchange rate. They showed that good in-sample performance of such models was accompanied by very poor out-of-sample performance.[2] We proceed by re-examining the findings of Bekaert and Hodrick (1992) on international asset price predictability. Using two-country VARs including the U.S. and either the U.K., Germany or Japan, Bekaert and Hodrick analyze the performance of dividend yields, forward premia and lagged excess stock returns as predictors of excess returns in stock and foreign exchange markets over the period 1981-89. They find evidence supportive of previous findings that domestic dividend yields forecast excess stock returns and that forward premia forecast excess foreign currency returns, and new evidence both that dividend yields have predictive power for currency excess returns, and that forward premia have predictive power for excess stock returns.

Bekaert and Hodrick (1992) also use the parameter estimates from the VARs to produce estimates of implied long horizon statistics such as slope coefficients from OLS regressions, variance ratios and R^2's. This is an application of a methodology advocated in Hodrick (1992). That paper presented Monte Carlo evidence showing that, if the assumed VAR structure is correct, the approach can produce accurate estimates of long horizon slope coefficients and R^2's without the need to use a very long series of data.

We are interested in investigating the stability of the relationships estimated by Bekaert and Hodrick (1992). If they are indeed stable, we should be able to detect evidence of predictability out-of-sample. To determine whether this is the case, we examine the forecasting performance of the two-country VARs over the out-of-sample period 1990-96. In all cases it is rather poor, and the predictions from the VARs are inferior to those from a simple benchmark model. We also find that modifying the forecasting procedure to use rolling and expanding data samples, a Bayesian approach and/or endogenous-regressor bias corrections fails to generate results that outperform the benchmark model. Monte Carlo analysis indicates that the data over the full sample period were unlikely to have been generated by a stable VAR. Examination of structural break statistics also provides strong evidence that the estimated relationships are not stable. Finally, further experiments show that long-horizon regression coefficients, R^2's and variance ratios implied by the VAR parameter estimates are subject to great uncertainty.

2. The Data

The data set consists of monthly data from four countries: the United States, Japan, the United Kingdom and Germany. The sample begins in 1981:1 and ends in 1996:10 for all countries except Germany, where the data run to 1995:6. Thus we extend the nine-year sample used in Bekaert and Hodrick (1992) by adding roughly seven years of additional data. The variables from each country will be identified by subscripts: 1 for the U.S., 2 for Japan, 3 for the United Kingdom and 4 for Germany. Let i_{jt} be the one-month nominal interest rate from time t to $t + 1$ in country j ($j = 1, 2, 3, 4$). Let r_{jt+1} be the continuously compounded one-month rate of return in excess of i_{jt} on the stock market in country j, denominated in the currency of that country. The log of the spot exchange rate for country j is s_{jt} (dollars per unit of currency in

[2] See Hansen (1992) and Stock and Watson (1996) for more recent evidence on instability in

country j) and the excess return in dollars to a long position in currency j held from t to $t + 1$ is rs_{jt+1}, where

$$rs_{jt+1} = s_{jt+1} - s_{jt} + i_{jt} - i_{1t} \quad (1)$$

Denoting the log of the one-month forward rate at time t as f_{jt}, then the one-month forward premium for currency j against the dollar is defined as $fp_{jt} = f_{jt} - s_{jt}$. Using the covered interest parity relation, $f_{jt} - s_{jt} = i_{1t} - i_{jt}$, the dollar excess return to holding currency j can be calculated from spot and forward rates as $rs_{jt+1} = s_{jt+1} - f_{jt}$. Finally, the dividend yield in country j is denoted dy_{jt}. More details on the data sources and construction are provided in the data appendix.

3. The Vector Autoregressions

We follow Bekaert and Hodrick (1992) in estimating a six-variable VAR for each of three two-country pairs, U.S.-Japan, U.S.-U.K. and U.S.-Germany.[3] The variables included in each VAR are the excess return on equity in each country $\{r_{1t}, r_{jt}\}$, the excess return to a long position in the foreign currency from the viewpoint of a U.S. investor $\{rs_{jt}\}$, the dividend yield in each country $\{dy_{1t}, dy_{jt}\}$ and the forward premium $\{fp_{jt}\}$ (equal to the one-month interest differential). We can write the first-order VAR regressions in vector notation as:

$$Y_t = \alpha_0 + AY_{t-1} + u_t \quad (2)$$

In the case of the U.S.-Japan VAR, $Y_t = \{ r_{1t}, r_{2t}, rs_{2t}, dy_{1t}, dy_{2t}, fp_{2t} \}$, α_0 is a vector of constants, A is the matrix of regression coefficients on the first lag of Y_t and u_t is an error vector.

We estimate the VARs by ordinary least squares over the whole sample (1981:1–1995:6/1996:10) and over each of the two subperiods (1981:1–1989:12 and 1990:1–

---

macroeconomic time series.

[3] Lag length was selected according to the Schwarz criterion. We found that optimal lag length was one in all cases, as did Bekaert and Hodrick for the shorter sample period.

1995:6/1996:10). The latter two periods correspond to our in-sample and out-of-sample periods for evaluating forecasting performance. For each variable in a given VAR we report the statistic for the joint test of whether all six coefficients on the lagged variables are zero. The results, presented in Table 1, conform closely to the findings of Bekaert and Hodrick (1992) over the period 1981–89. There is strong evidence of predictability for the U.S. excess stock return and for all three foreign currency returns. There is also evidence of predictability for the Japanese excess stock return. This pattern of predictability in stock and currency returns is broadly reproduced during the period from 1990 to the end of the sample. However, over the whole sample period, support for predictability emerges only in the currency markets, which suggests the presence of parameter instability. To throw further light on this issue we turn to considering the out-of-sample forecast performance of the VARs.

4. Out-of-sample forecast performance

If the VAR is well-specified and the covariance structure stable, then predictability implies that we should be able to use the parameter estimates from the period 1981-89 to forecast asset prices during the out-of-sample period 1990-95/96. To investigate the out-of-sample forecast performance of the VARs, we carry out the following exercise. We use the coefficient estimates from the sample period 1981:1-1989:12 to construct forecasts for each of the variables in the system over the out-of-sample period (1990:1-1996:10 for the US-UK and US-Japan, 1990:1-1995:6 for the US-Germany), at horizons of one, three and six months. That is, at each period we use the actual data at date t and the parameters as estimated over the fixed sample period and project the path of the system at dates $t + 1$, $t + 3$ and $t + 6$. We then update the data for the next period's set of forecasts. This gives us a set of 82 one-period-ahead forecasts, 80 overlapping three-period-ahead forecasts and 77 overlapping six-period-ahead forecasts for the US-UK and US-Japan.[4] There are 16 fewer forecasts for US-Germany at each horizon.

We first consider whether we can reject the hypothesis that the out-of-sample forecast errors have a mean of zero. The VAR residuals are constructed to provide an unbiased, in-sample forecast of the future value of the dependent variables. But if the estimated relationship is unstable, they may not have this property out-of-sample. Denoting the k-period-ahead forecast of variable Y_t, conditional on data at time t - k as $\hat{Y}_{t|t-k}$, we ask if the data are consistent with the following hypothesis:

$$E(\hat{Y}_{t|t-k} - Y_t \mid \Psi_{t-k}, \beta_0) = E(\hat{u}_{t|t-k} \mid \Psi_{t-k}, \beta_0) = 0 \qquad (3)$$

where $\hat{u}_{t|t-k}$ is the forecast error at time t, based on information Ψ_{t-k} through period t-k, and β_0 is the set of parameters estimated from the fixed sample period.

Panel A of Table 2 shows the mean errors and significance levels for tests of unbiasedness of the forecasts of the variables in each VAR. The hypothesis that the forecasts are unbiased is rejected at the five per cent level in 30 out of 54 cases, and at the one per cent level in 25 out of 54 cases. We also perform two further tests. The first considers whether the six forecasts from each VAR are jointly unbiased. The second considers whether the forecasts of each variable are jointly unbiased across the three VARs. Panels B and C of Table 2 report these results. The hypothesis of unbiasedness is rejected at any reasonable level of significance at all forecast horizons for the first test, and in all but two of eighteen cases for the second test.[5]

This evidence of bias means that the VAR forecasts are consistently under- or overpredicting the change in a variable during the out-of-sample period. For example, the

---

[4] The overlapping three and six-period forecast errors will have at least second and fifth order serial correlation in their errors which must be taken into account in the tests that follow.

forecasts fail to predict the very poor performance of the Japanese stock market during the out-of-sample period. The mean forecast error of 35.44 at the six-month horizon means that the VAR forecast overpredicts the excess return to holding Japanese stocks by 35.44 per cent per annum. The forecasts also miss the strong performance of the U.S. stock market. Two of the three six-month forecasts of excess returns in the U.S. stock market underpredict by ten percentage points or more.

It is important to emphasize that evidence of bias is not a sufficient reason on its own for concluding that a forecast is of no use, or that the variables in the system display no predictability. For example, the one-month forecast of the US dividend yield in the US-Germany VAR is quite accurate, with a mean forecast error of only 0.07 per cent per year. But the forecast is significantly biased because of the low variability of the error. Thus, even if the out-of-sample forecasts are biased, they may be valuable in the sense of being relatively more accurate — having a smaller average prediction error — than other forecasts available. Conversely, even an unbiased forecast may be of little use if it is very noisy. Notwithstanding these observations, it certainly appears that the magnitude of the forecast errors for all asset returns at all horizons is *economically* as well as statistically significant. In addition to the figures for US and Japanese stock returns mentioned above, we find that the mean forecast error for the excess return to German equity at the six-month horizon is 10.97%, and for the excess return to holding yen at the six-month horizon is 13.09%. In most cases the shorter horizon forecasts are even more inaccurate.

One way of providing further evidence on the economic significance of the forecast errors is to ask whether the VAR forecasts outperform a suitably chosen simple benchmark

---

[5] The statistical tests used in the paper are described in detail in the appendix.

model. We choose the natural one in which expected excess returns in equity and foreign exchange markets are assumed to be constant, and are therefore unpredictable. This assumption on expected excess returns is equivalent to imposing a constant risk premium, and is consistent with the standard representative agent asset pricing model. Expected excess returns are set equal to their respective sample means over the period 1981-89. In addition, in the light of the well-documented persistence of dividend yields and forward premia, we assume that these variables follow random walks.[6] We compare the VAR forecasts to the benchmark forecasts with the mean-squared prediction error (MSPE) criterion.[7] A simple way to compare the VAR and benchmark forecasts is to calculate the ratios of the MSPE from the VAR and benchmark models. A ratio less than one indicates that the forecasts from the VAR are more accurate, on average, than the benchmark forecasts.

Table 3 shows that the out-of-sample prediction error ratios are, in 52 out of 54 cases, greater than one for every variable, for each country over every forecast horizon. That is, the VAR consistently predicts more poorly than the benchmark forecast. In the case of the forward premium against the DM, the out-of-sample, one-month MSPE for the VAR forecast is 41 times that of the benchmark forecast. In 29 of 54 cases, the differences have p-values of 0.01 or less. These figures are in stark contrast to those for the in-sample MSPE ratios, which range from 0.49 to 0.99 at all horizons.

It is possible that the poor performance of the VAR forecasts is at least in part a consequence of ignoring the presence of heteroscedasticity in the error term. To investigate this, we perform a Lagrange multiplier test for autocorrelation in the squared errors. First, we

[6] In our data first-order autocorrelations for the dividend yields and forward premia are all greater than 0.92 while those for the excess return variables are all less than 0.13.

estimate the VARs over the 1981-1990 sample period and select the optimal lag length for an autoregression of squared errors using the Schwarz criterion. Next, we rerun an autoregression of the squared residuals using this optimal lag length and calculate the TR^2 statistic, which is distributed as a chi-square random variable with number of degrees of freedom equal to the number of autoregressors under the null of no autocorrelation.[8] The test rejects the null of no autocorrelation in the squared errors at the 10% level in three of the 18 equations, the Japanese dividend yield and forward premium equations in the U.S.-Japan VAR and the forward premium equation in the U.S.-U.K VAR.

For each of these equations, a visual inspection of the error process revealed that a structural break in the variance process seemed at least as likely as a GARCH process to have generated the high LM statistics.[9] To distinguish these hypotheses, we compared the Schwarz criterion and Akaike information criterion obtained by modeling the variance process in three ways: 1) as a GARCH(1,1) process; 2) as having a structural break; or 3) as a homoscedastic process.[10] The date for the structural break was chosen in each case to maximize the likelihood function. In all three VARs, the Schwarz criterion and Akaike information criterion lead one to prefer the structural break model to either the homoscedastic or the GARCH model. Reestimating the three equations permitting a structural break in the variance term results in very little change in the forecasting performance of the VARs. We conclude that the poor forecasting

---

[7] All of the forecasting results in this paper are robust to using mean-absolute prediction error (MAPE) as the measure of forecast accuracy. These results are available from the authors.

[8] Engle (1983) proposed this test to detect ARCH errors. The results of this test are available upon request.

[9] It is well known that the presence of structural breaks in the variance process can bias the autoregressive coefficients of the squared error process just as breaks in the level of a series can bias unit root tests to the nonstationary alternative (see Perron (1990)).

[10] The GARCH(1,1) model is a parsimonious representation that has been shown to fit a number of data series well.

performance cannot be attributed to GARCH effects or to structural breaks in the error variance process.

The figures in Table 3 on the relative performance of currency excess return forecasts are worth noting. Bekaert and Hodrick found particularly strong evidence of predictability over the period 1981-89 for the dollar against the pound and the mark. They reported confidence levels of 0.999 or above. But the benchmark model outperforms the VAR at all horizons except six months for the mark. At the one-month horizon the MSPE ratio for the mark is 3.59. This suggests that findings on the predictability of foreign exchange excess returns are heavily dependent on the experience of the 1980s.

The results presented in this section show that although the VAR model is clearly superior to the benchmark model in-sample, the benchmark model consistently outperforms the VAR model during the out-of-sample period, and that the gain in performance is frequently statistically significant. We next turn to examining possible explanations for these results.

5. Modifications to the forecasting procedure

We will refer to the forecasting procedure used in the previous section, using a fixed data sample and OLS estimates of the parameters, as the *baseline case*. We examine several modifications designed to improve forecasting performance: the imposition of Bayesian restrictions on coefficient estimates, correction for the small sample bias caused by the presence of lagged endogenous regressors[11] and the use of additional data as the forecast date changes.

There is considerable evidence that combining Bayesian techniques with VARs is helpful

[11] Mankiw and Shapiro (1986) and Stambaugh (1986) discuss the small sample bias imparted by lagged endogenous regressors. Bekaert, Hodrick and Marshall (1997) use Monte Carlo procedures to correct for such a bias in term structure tests.

in forecasting (Litterman, 1986). We therefore consider a modified forecasting procedure in which we choose prior means for the coefficients of the A matrix to conform with our benchmark forecasting model. Thus, the own-lag coefficients on excess return variables have a prior mean of zero and the corresponding coefficients on dividend yields and forward premia have a prior mean of one. This is an adaptation of the well-known "Minnesota Prior" to the differenced variables in the model. In addition we assume a diffuse prior distribution for the constant term α_0. The standard deviation of the prior distribution for all own-lag coefficients is set to 0.2. So in the case of equity and foreign exchange excess returns the forecaster is assumed to attach a prior probability of 0.95 to the hypothesis that the own-lag coefficient lies between – 0.4 and 0.4. The standard deviation of the prior distribution for all off-diagonal elements of the A matrix is set equal to 0.1 multiplied by an appropriate scale correction (Litterman, 1986, Doan, 1996).

It is also important to recognize that estimating the coefficients from a fixed sample does not provide the VAR with all the information that would be available to the econometrician. So it may be possible to improve forecasting performance by updating the sample on which the forecasts are based, as new information becomes available. To investigate this question we consider two cases, an expanding data sample and a rolling data sample. If we are estimating a stable VAR relation, prediction with expanding sample sizes will provide the model with more useful information with which to make predictions. On the other hand, if instability in the underlying parameters is the cause of the inferior forecasting performance, using rolling samples may alleviate the problem.

In the case of an expanding sample, we re-estimate the VAR on all data available up to time t to generate forecasts at time t. This contrasts with the fixed sample approach used in the baseline case, where the VAR was estimated only once on data from 1981-89. In the case of a rolling sample, for a forecast at time t the VAR is re-estimated on a rolling window of data (up to time t) equal to the original sample size of 108 observations.

We summarize the results of comparing Bayesian and classical forecasts for fixed, expanding and rolling samples in Table 4. Since we have already demonstrated the substantial superiority of the benchmark model over the baseline VAR model, it is not surprising that we should see some improvement in forecast performance in the Bayesian case, whose priors push the predictions in the direction of the benchmark model. The use of a rolling sample leads to the greatest improvement in quality of forecast as measured by the number of cases in which the prediction error ratio is significantly greater than one. But the results are still in all cases inferior to those from the benchmark model.

We also test forecasts formed with coefficients adjusted for the bias arising from the presence of lagged endogenous regressors. The adjustments are calculated in a procedure similar to that used by Bekaert, Hodrick and Marshall (1997). Our adjustment procedure is as follows:

1. Using an OLS estimate A_{OLS} of the VAR parameter matrix A and covariance matrix from the period 1981-89, we generate 100,000 data sets of 108 observations—with initial conditions drawn from the unconditional distribution of the data.
2. We estimate the parameter matrix A of the VAR for each simulated data set using OLS, and calculate the average of those matrices, A_{MC}.
3. The bias-adjusted coefficient matrix is computed as:

$$A_{BA} = A_{OLS} + (A_{OLS} - A_{MC})$$

We find in all cases that the bias-adjusted matrices possess eigenvalues greater than unity, indicating that the VAR may be misspecified, but for completeness report the forecasts with the adjusted parameter matrices. The results are presented in the bottom line of the two panels of Table 4. In comparison to the baseline case, we find that the adjustment does not increase the number of MSPE ratios less than one at any horizon. The number of ratios significantly greater than one is reduced from 40 to 38. It is clear that this modification to the forecasting procedure produces virtually no improvement.

Given that the benchmark model outperforms the VAR it is reasonable to consider whether the VAR should be re-estimated with all the variables of an integrated order in an error-correction framework. Complicating such an approach, however, is the uncertainty about the nature of the processes generating dividend yields and forward premia. For example, all four dividend-yield series show evidence of a time-varying (declining) mean in the sample until 1988 (see Figure 1). It is not clear whether these series are simply very persistent or whether it would be appropriate to model them with a time trend, a structural break or by differencing. Indeed, research in the unit root and structural break literature has shown that no finite amount of data will enable the observer to distinguish between an integrated series and a "close" highly persistent stationary alternative. The lack of an obvious way to model these series prevents us from reformulating the model with any confidence that it is correctly specified.

An alternative approach to improving forecasting performance is to proceed with step-wise elimination of insignificant parameter estimates to restrict the VAR coefficients. Li and Schadt (1995) analyzed the effect of this procedure on parameter estimates for the two-country VARs considered by Bekaert and Hodrick (1992). They examined the asset allocation strategy implied by out-of-sample forecasts at the six-month horizon over the period 1987-93. Based on an examination of Sharpe ratios they concluded that there was no evidence that the forecasts were informative. These results are consistent with our findings.

6. A Monte Carlo analysis of the forecasting performance of a stable VAR

The dividend yields and the forward premia series are very persistent (see Figure 1), with first-order autocorrelations over 0.92. The dividend yields also may have a time-varying (declining) mean in the sample. The extreme persistence found in dividend yields and forward premia is likely to produce poor small sample properties for the parameter estimates. This suggests that even a VAR estimated on stationary data with no structural breaks would not forecast particularly well. To investigate whether this persistence affects the forecasting performance of the VARs, we carry out the following Monte Carlo analysis. Using the VAR parameter estimates derived from the full sample (1981-1995/6) we generate 1000 new data sets of the same size as the full sample, using the data from 1980:12 as initial conditions.[12] We then produce the set of forecasting statistics shown in Table 4 using the simulated data sets. The results from the simulated VAR data are summarized in Table 5.

The forecasting performance of the stable VAR is not very impressive. At all horizons the Bayesian expanding sample case performs best, but beats the benchmark model on average in only 54 per cent of cases (see Panel A). However, despite this fact, we still find strong evidence against the hypothesis that a stable VAR process, even one with very persistent variables, generated the actual data. For one-month forecasts with an expanding sample (Panel C) we find that in both classical and Bayesian cases only five per cent of MSPE ratios from the simulated data are greater than those in the actual data. As one lengthens the forecast horizon these numbers increase. But this is simply a reflection of the reduced power of such comparisons at longer horizons to detect departures from a stable VAR process. Panel D compares the number of significance levels from the VAR-generated data that are less than those observed in the actual data. We again find that at short horizons there is strong evidence against the hypothesis that a stable VAR process generated the actual data.

The very poor forecasting performance relative to that of a stable VAR indicates that the estimated VAR does not well describe stable dynamic relationships between the variables. If the short-run dynamics are misspecified, this leads one to believe that the complex, nonlinear functions of the VAR parameters characterizing the long-run behavior of the system are of dubious usefulness.

7. Implications for Long-Horizon Statistics

Hodrick (1992) and Bekaert and Hodrick (1992) have argued that one of the advantages of the VAR-based approach is that, where evidence of predictability is present, one may use estimates of the parameters of the VAR to construct implied long-horizon statistics such as slope coefficients, R^2 values and variance ratios without having to rely on a long series of data. Hodrick (1992) presents evidence from a Monte Carlo study indicating that implied long-horizon statistics have good small sample properties. He considers a three-variable VAR in which lagged stock return, dividend yield and Treasury bill rate are used to predict stock returns. His data sample contains 431 monthly observations on US data, and so is substantially larger than the one we have available. We investigate the small sample properties of the implied long-horizon statistics for a data set of the size we have. Thus we pose the following question: under the assumption that the estimated VAR is the correct model, will a sample of 174/190 monthly observations provide us with satisfactory estimates of the long-horizon statistics of interest?

We estimate the VAR on the given data set (1981:1-1996:10 for US-Japan and US-UK

[12] Drawing initial conditions from the unconditional distribution of the data made no significant

and 1981:1-1995:6 for US-Germany) and use the parameter estimates to generate 10,000 simulated data sets drawing initial conditions from the unconditional distribution. For each simulated data set we obtain parameter estimates for the VAR and use them to construct long horizon statistics. In Table 6 we compare for the U.S.-Japan VAR (1981:1-1996:10) actual values of selected long-horizon statistics with the simulated distribution. In Panel A we see that implied coefficients in a regression of stock return on dividend yield for both the U.S. and Japan are very imprecisely estimated and subject to severe bias. In the U.S. case the 50th percentile of the empirical distribution is two to three times the actual value. The bias is similar in the Japanese case.[13] Only for the regression of foreign currency excess return on forward premium do we find that the empirical distributions are reasonably well centered on the actual value, although imprecisely estimated. A similar pattern emerges when we examine the implied R^2 values in Panel B. When we consider implied variance ratios we find that they are again very imprecisely estimated for U.S. and Japanese stock return, but fare better in the case of currency excess return. The results from the other two VARs are comparable and we do not report them.

So we find that even if the estimated VAR coefficients correspond to the true model, the implied long-horizon statistics will not be reliable for the sample size we consider. Our results stand in strong contrast to those of Hodrick (1992). The explanation for the difference in findings can be attributed to two factors. Hodrick's (1992) data set had over twice the number of observations and the VAR he analyzed had only half the number of variables of those we consider.

---

difference to the results.

[13] We found that it was necessary to increase the sample size to 5000 in order to approximately center the empirical distribution on the observed values in these two regressions.

8. Tests for Structural Breaks

The poor out-of-sample forecast performance of the VAR and the relative success of rolling forecasts indicate that the estimated parameters are unstable. This problem, a form of model misspecification, is common in time series regressions. To quantify the extent of the problem we test for a structural break at an unknown date by calculating the standard Wald test statistics for a structural break at each observation in the middle third of each sample. The supremum of these test statistics identifies a possible structural break in the series but will have a nonstandard distribution (Andrews, 1993). The critical value for the supremum is calculated from a Monte Carlo experiment.

Figure 2 shows a plot of these structural break statistics, along with the 1 per cent Monte Carlo critical values for the supremum of each series over the period from approximately the end of 1985 to the end of 1990.[14] In all three VARs the supremum lies comfortably above the 1 per cent critical value, indicating the presence of a structural break. The strongest evidence of a break for the US-Japan VAR and US-UK VAR occurs in the period from February to April 1987. To help identify the relationships in the VAR that are changing in this period, we calculate structural break statistics for each equation and for the individual coefficients in each equation. The evidence is not conclusive, but the equations for the US and Japanese equity returns and the Japanese dividend yield have large structural break statistics in this period.[15]

The US-Germany VAR structural break statistics are comfortably above the one per cent

---

[14] The Statistical Appendix describes the Monte Carlo procedures.

[15] Evidence from the individual equations and coefficients may be misleading in that the equation statistics lose covariance information compared to the statistics for the whole VAR. This may be important, as there is evidence of strong correlation in the coefficient estimators for the US equity return and US dividend yield equations in all three VARs. Similarly, the structural break statistics for individual coefficients may be misleading if there is significant correlation between the coefficient estimators.

critical value from early 1989 on, perhaps due to changes brought about by reunification. The forward premium equation statistic was quite high from 1989 through 1990 but the equations showing the strongest evidence of an increase in instability in August 1990 were those of the US equity return and dividend yield, which are highly correlated in each VAR.

9. Discussion and Conclusion

We have argued that the issue of structural stability of parameter estimates is generally ignored in studies that seek to demonstrate predictability of asset returns. This issue particularly complicates the inference of long-horizon properties of the data from relatively short time series. We have illustrated the problems that can arise by re-examining the findings of Bekaert and Hodrick (1992) on the predictability of international asset returns. We examine the out-of-sample forecasting performance of the VAR, and find it to be very poor. The VAR forecasts are conclusively outperformed by a simple benchmark model which assumes that excess returns in equity and foreign exchange markets are constant, and that dividend yields and forward premia follow random walks.

We consider several explanations for the very poor forecasting performance of the VAR.

1. Poor parameter estimates may be caused by pure sampling error.
2. The extreme persistence found in dividend yields and forward premia may result in poor small sample properties for the parameter estimates.
3. The VAR coefficients may exhibit small-sample bias due to the presence of lagged endogenous regressors.
4. The VAR parameters may have been subject to some underlying structural change.

The strong evidence of bias in the out-of-sample forecasts and their very poor performance relative to the benchmark model suggest that pure sampling error is not a plausible explanation. This is further supported by the results of the Monte Carlo analysis in which we generate simulated data sets with a stable VAR estimated over the sample period. Although Monte Carlo forecasts based on VAR parameter estimates are not very good, they significantly outperform the benchmark model and certainly do far better than forecasts based on actual data. The Monte Carlo analysis also indicates that while highly persistent variables may contribute to the poor forecasting performance of the VAR, they cannot adequately explain our results. The lack of an obvious way to model the highly persistent series prevents us from reformulating the model with any confidence that it is correctly specified.

The failure of bias adjustment to produce any detectable improvement in the VAR forecasts casts doubt on the third possible explanation, that lagged endogenous variables cause small sample bias. This leaves us with the fourth possibility, that one or more structural breaks occurred in the data. This hypothesis is strongly supported by the structural break statistics shown in Figure 2.

Examination of long horizon statistics inferred from the VAR parameter estimates has shown that *even if* the VAR were stable the estimates from a sample of 190 monthly observations would in general be badly biased and very imprecise. Thus the methodology advocated by Hodrick (1992) is shown to be sensitive to sample size and model size. The unreliability of the estimates is here exacerbated by the evident instability of the VAR.

Data Appendix

Excess equity returns were constructed by subtracting the continuously compounded 1-month Eurocurrency interest rate, collected by the *Bank for International Settlements* at the end of each month, from the total equity return provided by *Morgan Stanley Capital International* (MSCI). The MSCI total return series is subdivided into separate income return and capital appreciation series. The MSCI income return and capital appreciation series were used to calculate dividend yields as annualized dividends divided by current price for the U.S., Japan and the U.K. The German dividend yield series is taken from various issues of the *Monthly Report of the Deutsche Bundesbank*, from the column labeled "yields on shares including tax credit." Prior to 1993, this series could be found in Section VI, Table 6 of the statistical section. Starting in January 1993, this series was displayed in Section VII, Table 5 until June 1995. The spot and 1-month forward exchange rates were obtained from *DRI/McGraw-Hill*'s DRIFACS database. All bid and ask exchange rate data were sampled at the end of the month. Transaction costs were also incorporated by taking account of the fact that a currency is bought at the ask price and sold at the bid.

Statistical Appendix

A.1. Testing for bias in the VAR forecasts

To determine whether the mean errors that are reported in Table 2 are statistically significantly different from zero, we calculate the following statistic (Hansen and Singleton, 1982 or Diebold and Mariano, 1995):

$$T[\bar{e}' S_T^{-1} \bar{e}] \sim^A \chi^2(r) \tag{A.1}$$

which is asymptotically distributed as a χ^2 random variable with r degrees of freedom under the hypothesis that the errors have zero mean, i.e. the forecasts are unbiased. In this expression, $\bar{e}$ is the average forecast error during the out-of-sample period, $\bar{e} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t$, $\hat{e}_t$ is the r by 1 forecast error at time t, and S_T/T is the (r by r) covariance matrix of $\bar{e}$. The integer r is the number of forecasts being tested by the statistic. For example, for the tests of individual variables in panel A of Table 2, $r = 1$, but for the tests over a VAR in Panel B, $r = 6$ because we are testing six forecasts at a time. Similarly, for the tests across VARs, $r = 3$. Because the $\{\hat{e}_t\}$ series may be autocorrelated, we must take account of this in the construction of our estimate of the variance-covariance matrix of $\hat{e}_t$, S_T. We do this with the Newey-West estimator:

$$S_T = \Gamma_{0,T} + \sum_{v=1}^{q}[1-(v/(q+1))](\Gamma_{v,T} + \Gamma_{v,T}') \tag{A.2}$$

where

$$\Gamma_{v,T} = \frac{1}{T}\sum_{t=v+1}^{T}(\hat{e}_t - \bar{e})(\hat{e}_{t-v} - \bar{e})' \tag{A.3}$$

The parameter q is chosen to match the order of the serial correlation in the data with a sample-dependent formula suggested by Newey and West (1994). Experimentation with other lag length and weighting procedures produced similar results.

A.2. Testing for differences in the prediction errors of VAR and random walk models

To determine whether the prediction error ratios in Table 3 are statistically significant, we follow a similar procedure and calculate the following statistic (Hansen and Singleton, 1982):

$$T[\left(\bar{e}_{VAR}^2 - \bar{e}_B^2\right) S_T^{-1} \left(\bar{e}_{VAR}^2 - \bar{e}_B^2\right)] \sim^A \chi^2(r) \tag{A.4}$$

where $\bar{e}_{VAR}{}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{Y}_t^{VAR} - Y_t\right)^2$ is the MSPE from the VAR forecast ($\hat{Y}_t^{VAR}$), $\bar{e}_B{}^2$ is the comparable statistic from the benchmark forecast, S_T/T is the Newey-West estimate of the covariance matrix of the mean square difference ($\bar{e}_{VAR}{}^2$ - $\bar{e}_B{}^2$), and r denotes the dimension of $\bar{e}_{VAR}{}^2$ and $\bar{e}_B{}^2$, the number of forecasts we are testing. For the individual variables tested, $r = 1$. If the MSPEs from the two prediction methods are equal, the statistic is asymptotically distributed as a chi-square random variable with r degrees of freedom.

A.3. Testing for a structural break at an unknown break point

We test for an unknown break point in the middle third of the sample by first constructing a series of statistics for each observation in the subsample. The statistic at a given point, T_0, in a sample running from time zero to T, is calculated by estimating the VAR from time zero to T_0, and from T_0 to T, obtaining two sets of coefficient estimates, $\hat{\theta}_1$ and $\hat{\theta}_2$. If $\pi = T_0/T$ denotes the fraction of the total observations which are from the first part of the sample, then

$$\sqrt{T}\left(\hat{\theta}_1 - \theta_1\right) \sim N(0, V_1/\pi) \text{ and } \sqrt{T}\left(\hat{\theta}_2 - \theta_2\right) \sim N(0, V_2/(1-\pi))$$

where $V_1 = \Omega_1 \otimes Q_1^{-1}$, $V_2 = \Omega_2 \otimes Q_2^{-1}$, $Q_1 = \frac{1}{T_0}\sum_{t=1}^{T_0} y_{t-1}y_{t-1}'$, $Q_2 = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} y_{t-1}y_{t-1}'$, and

$\Omega_1 = \frac{1}{T_0}\sum_{t=1}^{T_0} u_t u_t'$ $\Omega_2 = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T} u_t u_{t-1}'$ (Hamilton, 1994). The test statistic for a break at T_0 is given by the quadratic function of the difference between the parameter estimates weighted by the inverse of their covariance matrix.

$$\lambda = T\left(\hat{\theta}_1 - \hat{\theta}_2\right)'\left[(\Omega_1 \otimes Q_1^{-1})/\pi + (\Omega_2 \otimes Q_2^{-1})/(1-\pi)\right]^{-1}\left(\hat{\theta}_1 - \hat{\theta}_2\right) \qquad \text{(A.5)}$$

The null of no structural break within a given subsample is rejected for sufficiently high values of the supremum of λ over the subsample. For tests of structural breaks in one equation or in an individual coefficient, statistics similar to that denoted by λ can be calculated using the appropriate coefficient estimate(s) and the variance-covariance matrix of those estimate(s).

We calculate the 1 per cent critical values from the following Monte Carlo experiment.

1. We estimate each VAR over the whole sample, saving coefficients and the covariance matrix.

2. Using the initial conditions (1980:12 data), estimated coefficients and covariance matrix we create 1000 new data sets of length T, where T is the length of the whole sample.

3. We compute 1000 time series of structural break statistics over the middle third of each of the simulated data sets, one series for each generated data set. Each time series of structural break statistics is approximately of length 63 ($0.33T$).

4. The ninety-ninth percentile of the distribution of suprema over these 1000 time series is the one per cent critical value used in Figure 2.

References

Andrews, Donald W. K., "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* (July 1993), 821-56.

Bekaert, Geert and Robert J. Hodrick, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," *Journal of Finance* (June 1992), 467-509.

Bekaert, Geert, Robert J. Hodrick and David A. Marshall, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, (June 1997), 309-48.

Campbell, John Y., "Stock Returns and the Term Structure," *Journal of Financial Economics* (June 1987), 373-99.

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.

Diebold, Francis X. and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* (July 1995), 253-63.

Doan, Thomas A., "RATS User's Manual Version 4," Estima, 1996.

Engle, Robert F., "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model," *Journal of Money, Credit, and Banking* (August 1983), 286-301.

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

Fama, Eugene F. and Kenneth R. French, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy* (April 1988), 246-73.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994.

Hansen, Bruce E., "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics* (July 1992), 321-35.

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* (September 1982), 1269-86.

Hodrick, Robert J., "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* (1992), 357-86.

Lamont, Owen, "Earnings and Expected Returns," *Journal of Finance* (October 1998), 1563-87.

Li, Hong and Rudi Schadt, "Multivariate Time Series Study of Excess Returns on Equity and Foreign Exchange Markets," *Journal of International Financial Markets, Institutions & Money* (1995), 3-35.

Litterman, Robert B., "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience," *Journal of Business and Economic Statistics* (January 1986), 25-38.

Lo, Andrew W. and A. Craig Mackinlay, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies* (Spring 1988), 41-66.

Mankiw, N. Gregory, and Matthew D. Shapiro, "Do we reject too often? Small Sample Properties of Tests of Rational Expectations Models," *Economics Letters* (1986), 139-45.

Mark, Nelson M., "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* (March 1995), 201-18.

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

Nelson, Charles R. and Myung J. Kim, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* (June 1993), 641-61.

Newey, Whitney K., and Kenneth D. West, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies* (October 1994), 631-53.

Perron, Pierre, "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics* (April 1990), 153-62.

Poterba, James M. and Lawrence H. Summers, "Mean Reversion in Stock Prices," *Journal of Financial Economics* (October 1988), 27-59.

Richardson, Matthew, "Temporary Components of Stock Prices: A Skeptic's View," *Journal of Business and Economic Statistics* (April 1993), 199-207.

Stambaugh, Robert F., "Bias in regressions with lagged stochastic regressors," *Center for Research in Security Prices* (January 1986), working paper number 156,

Stock, James H. and Mark W. Watson, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* (January 1996), 11-30.

Summers, Lawrence H. "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance* (July 1986), 591-601.

Table 1
Tests for predictability in the vector autoregressions

		1981-89		*1990-end of sample*		*1981-end of sample*	
		$\chi^2(6)$	p-value	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value
U.S.-Japan VAR	r_1	30.43	0.00	13.64	0.03	4.69	0.58
	r_2	14.75	0.02	16.81	0.01	8.87	0.18
	rs_2	15.27	0.02	15.73	0.02	19.07	0.00
	dy_1	2858.67	0.00	1973.64	0.00	5049.14	0.00
	dy_2	16076.11	0.00	521.27	0.00	13763.07	0.00
	fp_2	350.45	0.00	3213.09	0.00	1057.94	0.00
U.S.-U.K. VAR	r_1	17.57	0.01	16.06	0.01	6.07	0.42
	r_3	7.10	0.31	25.22	0.00	5.99	0.42
	rs_3	26.55	0.00	9.41	0.15	24.85	0.00
	dy_1	2511.45	0.00	1942.97	0.00	4410.49	0.00
	dy_3	2401.98	0.00	551.30	0.00	4002.55	0.00
	fp_3	518.91	0.00	2528.28	0.00	1310.00	0.00
U.S.-Germany VAR	r_1	10.08	0.12	17.70	0.01	2.79	0.83
	r_4	7.19	0.30	8.01	0.24	7.23	0.30
	rs_4	36.02	0.00	11.39	0.08	27.24	0.00
	dy_1	1901.93	0.00	755.96	0.00	4433.16	0.00
	dy_4	4557.45	0.00	308.15	0.00	4791.59	0.00
	fp_4	74.60	0.00	5144.52	0.00	2631.97	0.00

The variables from each country are identified by subscripts: 1 for the U.S., 2 for Japan, 3 for the United Kingdom and 4 for Germany. r_i is the excess equity return for country i; rs_i is the excess return in dollars to a long position in the currency of country i; dy_i is the dividend yield in country i; fp_i is the forward premium for the currency of country i. End-of-sample is 1996:10 for U.S.-Japan and U.S.-U.K, and 1995:6 for U.S.-Germany. The χ^2 test is a Wald test with heteroskedastic-consistent standard errors for the null hypothesis that the six slope coefficients in each equation are jointly equal to zero. Low p-values reject the null of no predictability in the VAR.

Table 2
Mean forecast errors and tests for unbiasedness

Panel A: Individual variables

		1 month		*3 month*		*6 month*	
		Mean	*p-value*	*Mean*	*p-value*	*Mean*	*p-value*
U.S.-Japan VAR	r_1	-11.25	0.23	-18.11	0.01	-10.98	0.03
	r_2	57.75	0.00	42.27	0.00	35.44	0.00
	rs_2	26.71	0.00	19.86	0.00	13.09	0.00
	dy_1	0.03	0.44	0.15	0.10	0.26	0.05
	dy_2	-0.06	0.00	-0.15	0.00	-0.24	0.00
	fp_2	0.15	0.16	0.40	0.15	0.67	0.15
U.S.-U.K. VAR	r_1	-12.89	0.11	-13.35	0.02	-10.04	0.04
	r_3	4.78	0.51	-1.85	0.76	-2.18	0.69
	rs_3	4.89	0.49	6.43	0.25	6.16	0.24
	dy_1	0.06	0.03	0.19	0.01	0.32	0.00
	dy_3	-0.04	0.12	-0.06	0.36	-0.08	0.42
	fp_3	-0.10	0.46	-0.17	0.62	-0.28	0.59
U.S.-Germany VAR	r_1	18.24	0.00	1.54	0.76	-4.84	0.34
	r_4	32.48	0.00	20.05	0.01	10.97	0.18
	rs_4	57.74	0.00	15.13	0.01	5.31	0.29
	dy_1	-0.07	0.00	-0.11	0.01	-0.06	0.35
	dy_4	-0.12	0.00	-0.27	0.00	-0.35	0.00
	fp_4	2.26	0.00	4.16	0.00	5.13	0.00

Forecast errors are measured in percentage points per annum. The p-values report the probability that the mean forecast errors were drawn from a distribution with zero mean. They were calculated with a Newey-West correction for serial correlation. See the appendix for a more detailed discussion of the tests. Low p-values reject the null of unbiased predictions by the VAR.

Table 2 (continued)
Mean forecast errors and tests for unbiasedness

Panel B: Variables pooled within each VAR

	1 month		*3 month*		*6 month*	
VAR	*Mean*	*p-value*	*Mean*	*p-value*	*Mean*	*p-value*
U.S.-Japan	73.33	0.00	44.41	0.00	38.24	0.00
U.S.-U.K.	-3.30	0.00	-8.81	0.00	-6.10	0.00
U.S.-Germany	110.53	0.00	40.50	0.00	16.17	0.00

The mean errors are averaged over all the variables in the VAR. The p-values are calculated for the null hypothesis that the mean errors for all variables within each VAR are zero. Low p-values reject the null of unbiased predictions by the VAR.

Panel C: Variables Pooled Across VARs

	1 month		*3 month*		*6 month*	
Variable	*ME*	*p-value*	*ME*	*p-value*	*ME*	*p-value*
r_I	16.21	0.00	-13.90	0.00	-15.68	0.77
r_j	106.95	0.00	74.40	0.00	57.41	0.00
rs_j	90.01	0.00	39.14	0.01	20.82	0.36
dy_I	-0.06	0.00	0.02	0.00	0.20	0.00
dy_j	-0.23	0.00	-0.53	0.00	-0.78	0.00
fp_j	2.47	0.00	4.90	0.00	6.43	0.00

The mean errors for each variable are averaged across the three VARs. The p-values are calculated for the null hypothesis that the mean errors for each variable in all VARs are zero. Low p-values reject the null of unbiased predictions across the VARs.

Table 3

A comparison of the mean square prediction error (MSPE) of VAR forecasts to those of benchmark forecasts.

		1 month		*3 month*		*6 month*	
		MSPE	*p-value*	*MSPE*	*p-value*	*MSPE*	*p-value*
U.S.-Japan VAR	r_1	1.86	0.00	1.54	0.02	1.18	0.02
	r_2	1.28	0.00	1.15	0.00	1.09	0.01
	rs_2	1.45	0.01	1.23	0.05	1.07	0.42
	dy_1	2.56	0.00	4.20	0.02	5.04	0.04
	dy_2	2.09	0.00	3.38	0.00	3.85	0.00
	fp_2	1.55	0.01	2.62	0.00	2.70	0.01
U.S.-U.K. VAR	r_1	1.58	0.00	1.30	0.01	1.15	0.02
	r_3	1.25	0.00	1.07	0.10	1.02	0.45
	rs_3	1.31	0.16	1.09	0.54	1.02	0.89
	dy_1	2.33	0.00	3.64	0.01	4.61	0.01
	dy_3	1.39	0.01	1.29	0.23	1.40	0.34
	fp_3	2.06	0.00	2.89	0.00	2.60	0.01
U.S.-Germany VAR	r_1	1.35	0.01	1.03	0.48	1.04	0.10
	r_4	1.17	0.02	1.04	0.11	0.98	0.10
	rs_4	3.59	0.00	1.10	0.52	0.94	0.44
	dy_1	1.63	0.00	1.51	0.08	1.14	0.68
	dy_4	1.46	0.01	1.64	0.06	1.34	0.29
	fp_4	41.12	0.00	30.34	0.00	13.37	0.00

Columns headed MSPE give the MSPE ratio. Ratios are constructed as VAR MSPE/benchmark model MSPE. A ratio less than one indicates that the VAR forecast is more accurate on average than the benchmark forecast. The columns labeled "p-value" show the probability of obtaining at least as extreme a test statistic given that the MSPEs from each model are equal. Ratios greater than one indicate that the simple benchmark forecasts are better than those of the VAR, low p-values reject the null of equal MSPE. See the appendix for a more detailed discussion of the tests.

Table 4
Summary of the performance of alternative forecasting methods relative to the benchmark model

Panel A	*Sample*	*MSPE ratios < 1*			
		1 month	3 month	6 month	Total (of 54)
Classical (baseline case)	fixed	0	0	2	2
Classical	rolling	0	0	3	3
Classical	expanding	0	0	2	2
Bayesian	fixed	0	0	2	2
Bayesian	rolling	0	0	2	2
Bayesian	expanding	0	0	2	2
Bias corrected.	fixed	0	0	1	1

Panel B	*Sample*	*MSPE ratios: number significant*			
		1 month	3 month	6 month	Total (of 54)
Classical (baseline case)	fixed	17	9	9	35
Classical	rolling	7	3	2	12
Classical	expanding	10	7	8	25
Bayesian	fixed	14	9	8	31
Bayesian	rolling	4	1	2	7
Bayesian	expanding	9	5	5	19
Bias corrected.	fixed	17	12	9	38

Panel A: Ratios are constructed as forecast MSPE/benchmark model MSPE. A ratio less than one indicates that the particular forecast is more accurate on average than the benchmark forecast. The panel records the number of MSPE ratios less than one over all variables and country pairs, by horizon and forecasting method. Ratios less than one indicate that the VAR outperforms the simple benchmark.
Panel B: Number of MSPE ratios significantly greater than one at the 5% level over all variables and country pairs, by horizon and forecasting method. Significant MSPE ratios indicate that the benchmark model is superior to the particular VAR model in a statistically significant way.

Table 5
Monte Carlo forecasting performance with a stable VAR generating the data.

Panel A	*Sample*	*% MSPE ratios < 1*			
		1 month	3 month	6 month	Total
Classical	fixed	37.34	42.73	46.30	42.12
Classical	rolling	35.49	36.66	41.23	37.79
Classical	expanding	48.67	49.51	51.04	49.74
Bayesian	fixed	46.73	46.39	48.30	47.14
Bayesian	rolling	48.51	45.50	47.69	47.23
Bayesian	expanding	57.34	52.78	52.84	54.32

Panel B	*Sample*	*MSPE ratios: % significant*			
		1 month	3 month	6 month	Total
Classical	fixed	17.65	16.54	15.14	16.44
Classical	rolling	6.79	7.37	7.21	7.12
Classical	expanding	4.01	4.05	4.53	4.20
Bayesian	fixed	12.78	13.11	13.04	12.98
Bayesian	rolling	4.03	4.59	5.38	4.66
Bayesian	expanding	2.71	3.25	4.12	3.36

Panel C	*Sample*	*% MSPE ratios > actual*			
		1 month	3 month	6 month	Total
Classical	fixed	7.21	14.18	23.60	14.99
Classical	rolling	10.82	21.19	25.88	19.30
Classical	expanding	5.25	12.87	19.82	12.65
Bayesian	fixed	8.16	15.36	23.53	15.68
Bayesian	rolling	9.10	17.31	23.86	16.76
Bayesian	expanding	5.12	13.29	20.48	12.96

Panel D	*Sample*	*% MSPE significance levels < actual*			
		1 month	3 month	6 month	Total
Classical	fixed	7.41	16.11	19.73	14.41
Classical	rolling	11.19	25.75	26.44	21.13
Classical	expanding	3.92	4.01	4.52	4.15
Bayesian	fixed	8.57	17.15	19.87	15.20
Bayesian	rolling	13.75	22.91	25.08	20.58
Bayesian	expanding	6.04	13.48	16.01	11.85

1000 data sets of 190 observations were generated from VAR parameter estimates for the sample 1981-96. Out-of-sample forecasts were constructed as in Table 4 and MSPE ratios relative to the benchmark model were calculated. Panel A reports the percentage of MSPE ratios less than one in VAR-generated data. Panel B reports the percentage of MSPE ratios significantly greater than one at the 5% level. Panel C reports the percentage of MSPE ratios from VAR-generated data greater than the corresponding ratios from the actual data. A low percentage of MSPE ratios greater than the actual ratios indicates that the data were unlikely to have been generated under the null of a stable VAR. Panel D reports the percentage of significance levels from VAR-generated data less than corresponding significance levels from the actual data. A low percentage of MSPE significance levels less than the actual ratios indicates that the data were unlikely to have been generated under the null of a stable VAR.

Table 6
Long-horizon statistics: Comparison of actual values with empirical distribution assuming that the estimated VAR is the true model: US-Japan VAR 1981:1- 1996:10

Panel A: US-Japan slope coefficients

	Horizon (months)	*1*	*3*	*6*	*12*	*24*	*36*	*48*	*60*
U.S.: Excess Stock Return and Dividend Yield	Actual	1.08	4.29	9.87	22.99	50.46	73.11	88.82	98.45
	5%	-1.34	-2.96	-3.74	-1.29	8.03	13.25	16.38	17.05
	50%	4.51	14.96	32.02	67.47	126.39	161.93	178.16	182.82
	95%	17.16	51.71	97.22	165.66	237.74	259.78	263.12	260.85
Japan: Excess Stock Return and Dividend Yield	Actual	14.01	43.38	84.57	160.43	288.09	387.01	461.31	515.48
	5%	-2.61	-4.78	-4.60	4.32	38.08	73.70	101.87	122.50
	50%	29.04	89.31	174.19	326.50	558.05	700.09	774.63	810.13
	95%	92.84	266.73	475.29	761.04	1055.17	1183.15	1244.56	1273.63
Currency Excess Return and Forward Premium	Actual	-4.56	-12.83	-23.18	-37.65	-49.57	-49.49	-45.14	-40.27
	5%	-6.74	-18.79	-33.97	-56.76	-83.07	-94.49	-99.73	-101.63
	50%	-4.45	-12.47	-22.27	-35.43	-44.73	-43.90	-40.31	-37.32
	95%	-1.95	-5.43	-9.23	-13.07	-10.80	-3.99	2.15	6.57

Table 6 (continued)
Long-horizon statistics: Comparison of actual values with empirical distribution assuming that the estimated VAR is the true model: US-Japan VAR 1981:1- 1996:10

Panel B: US-Japan implied R^2's

	Horizon (months)	*1*	*3*	*6*	*12*	*24*	*36*	*48*	*60*
U.S.: Excess Stock Return	Actual	0.02	0.03	0.04	0.06	0.07	0.07	0.07	0.07
	5%	0.02	0.03	0.03	0.05	0.05	0.05	0.04	0.04
	50%	0.05	0.09	0.13	0.18	0.22	0.22	0.21	0.20
	95%	0.10	0.19	0.28	0.36	0.42	0.43	0.42	0.39
Japan: Excess Stock Return	Actual	0.04	0.03	0.04	0.05	0.08	0.09	0.09	0.10
	5%	0.03	0.03	0.04	0.05	0.05	0.05	0.04	0.04
	50%	0.07	0.09	0.13	0.18	0.21	0.21	0.20	0.18
	95%	0.14	0.18	0.27	0.35	0.41	0.41	0.40	0.39
Currency Excess Return	Actual	0.07	0.15	0.21	0.24	0.21	0.17	0.14	0.11
	5%	0.03	0.05	0.07	0.08	0.06	0.05	0.04	0.03
	50%	0.09	0.17	0.23	0.26	0.23	0.18	0.14	0.12
	95%	0.18	0.34	0.45	0.51	0.49	0.45	0.40	0.35

Panel C: US-Japan variance ratios

	Horizon (months)	*1*	*3*	*6*	*12*	*24*	*36*	*48*	*60*
U.S.: Excess Stock Return	Actual	1.00	1.05	1.06	1.07	1.04	0.98	0.91	0.85
	5%	1.00	0.88	0.82	0.71	0.53	0.42	0.35	0.30
	50%	1.00	1.04	1.04	1.01	0.90	0.78	0.68	0.60
	95%	1.00	1.22	1.33	1.48	1.57	1.53	1.44	1.35
Japan: Excess Stock Return	Actual	1.00	1.01	0.99	0.94	0.86	0.79	0.73	0.68
	5%	1.00	0.86	0.79	0.68	0.52	0.44	0.38	0.34
	50%	1.00	1.01	0.98	0.92	0.81	0.71	0.64	0.58
	95%	1.00	1.18	1.22	1.26	1.30	1.27	1.21	1.15
Currency Excess Return	Actual	1.00	1.19	1.40	1.73	2.17	2.41	2.54	2.60
	5%	1.00	0.99	1.05	1.15	1.25	1.29	1.32	1.32
	50%	1.00	1.18	1.38	1.70	2.10	2.32	2.42	2.47
	95%	1.00	1.41	1.86	2.63	3.76	4.46	4.88	5.14

10,000 data sets of 190 observations were generated from VAR parameter estimates for the sample 1981-96. The VAR was estimated on each data set and the parameter estimates were used to calculate the implied long-horizon regression slope coefficients, R^2s and variance ratios. The panels of the table report the actual value of the long-horizon statistic and the percentiles of the empirical distribution.

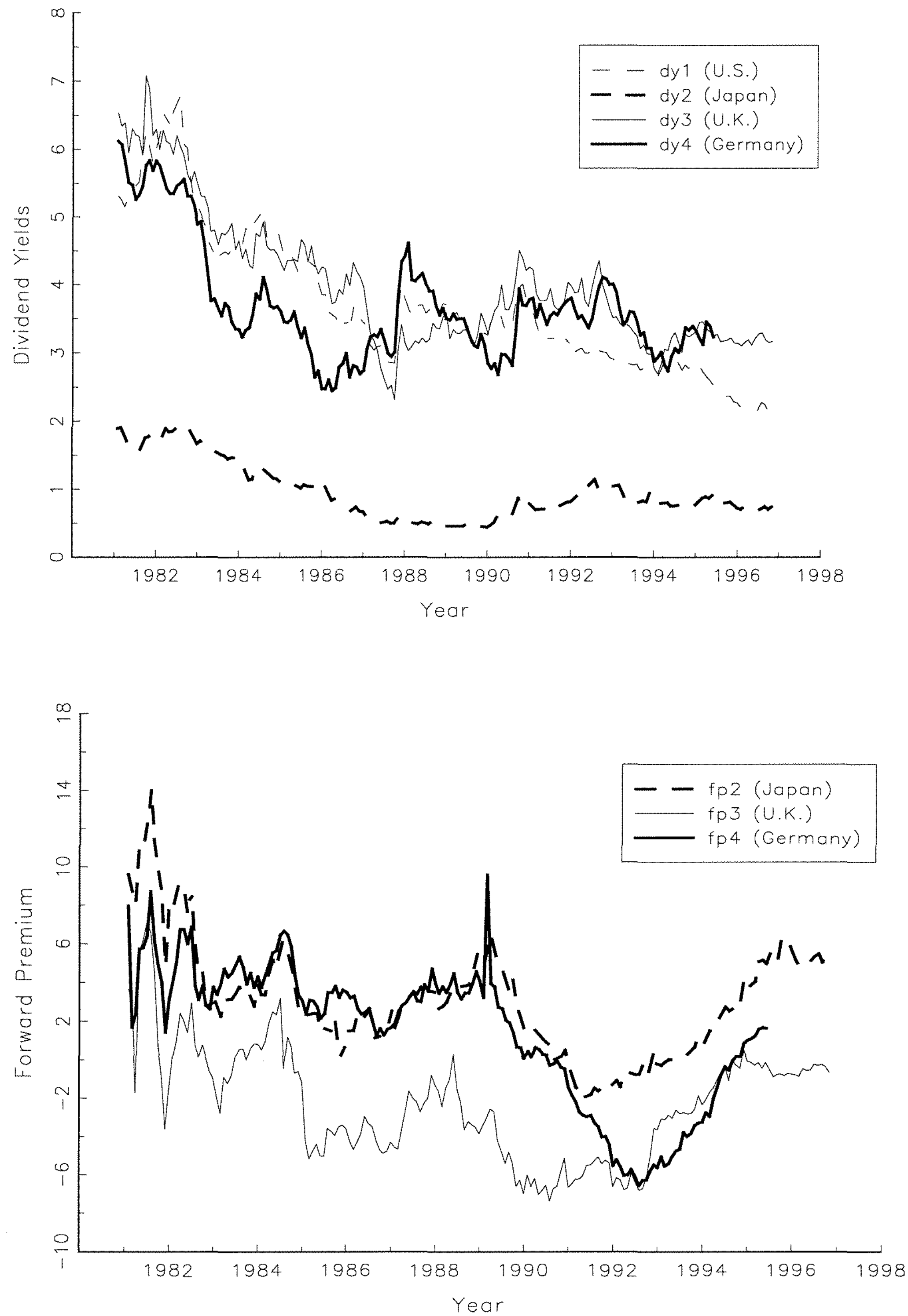

Figure 1: Time series of the dividend yields (top panel) and the forward premia (bottom panel), in annualized percentage terms.

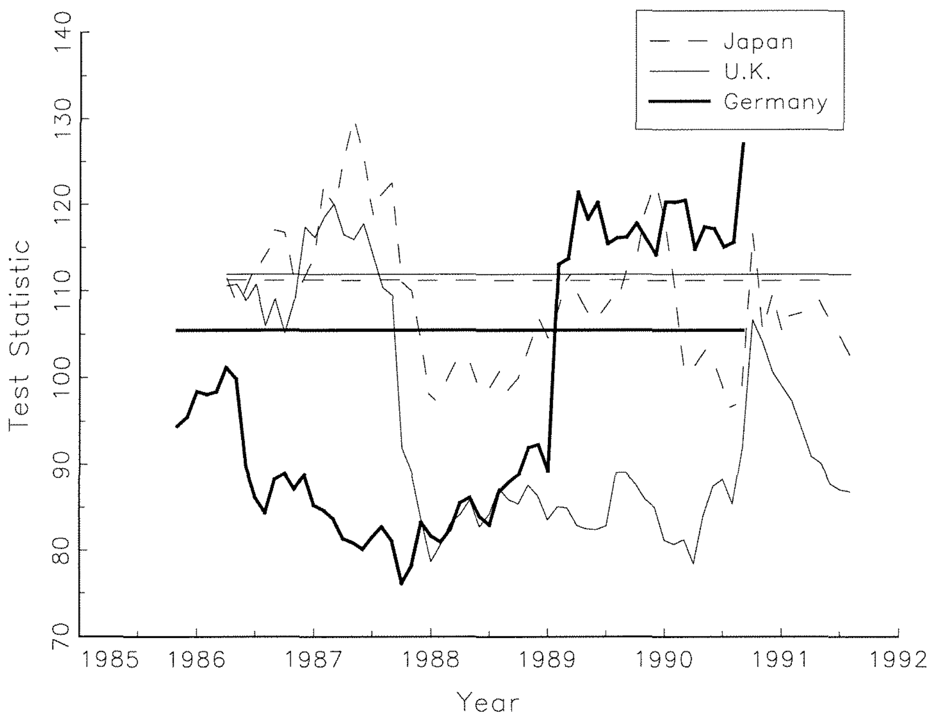

Figure 2: Structural break statistics for the VARs. The horizontal lines show one-percent Monte Carlo critical values, as described in the statistical appendix.