Instability of Return Prediction Models

Bradley S. Paye
Jones Graduate School of Management, Rice University

Allan Timmermann
University of California San Diego and CEPR

This Version: October 14, 2005

Abstract

This study examines evidence of instability in models of ex post predictable components in stock returns related to structural breaks in the coefficients of state variables such as the lagged dividend yield, short interest rate, term spread and default premium. We estimate linear models of excess returns for a set of international equity indices and test for stability of the estimated regression parameters. There is evidence of instability for the vast majority of countries. We then attempt to characterize the timing and nature of the breaks. Breaks do not generally appear to be uniform in time: different countries experience breaks at different times. We do identify a contemporaneous break for the US and UK indices in 1974. There is also some evidence of a break for a cluster of European nations during the 1978-1982 period. These breaks may relate to the oil price shock of 1974 and the formation of the European Monetary System in 1979. For the majority of intenational indices, the predictable component in stock returns appears to have diminished following the most recent break. We assess the adequecy of the break tests and model selection procedures in a set of Monte Carlo experiments.

1. **Introduction**

Predictability of stock returns has been well documented in the empirical finance literature and is now routinely used in studies of mutual fund performance (Christopherson, Ferson and Glassman (1998), Ferson and Schadt (1996)), tests of the conditional CAPM (Ferson and Harvey (1991), Ghysels (1998)) and optimal asset allocation (Ait-Sahalia and Brandt (2001), Barberis (2000), Brandt (1999), Campbell and Viceira (1998) and Kandel and Stambaugh (1996)). Variables commonly used to predict stock returns include the dividend yield, the short term interest rate, and term and default premia. Most studies assume a stable prediction model in which the coefficients on the state variables do not change over time.[1]

Recent empirical studies have, however, cast doubt upon the assumed stability of return forecasting models. In a forecasting model based on the dividend and earnings yield, Lettau and Ludvigson (2001) find some evidence of instability in the second half of the 1990s. Likewise, Goyal and Welch (2003) uncover instability in return models based on the dividend yield when data from the 1990s is added to the sample. Ang and Bekaert (2004) also find evidence of deterioration in predictability patterns in US returns in the second half of the 1990s.

Signs of instability in financial prediction models have also emerged from studies that specifically address the question of whether stock market investors could have exploited predictability to earn abnormal returns in real time. These studies have generally found that although stock returns were predictable *ex post* (or in-sample), the evidence of genuine *ex ante* (or out-of-sample) predictability appears to be much weaker. Bossaerts and Hillion (1999) find that stock returns on a range of US and international portfolios are largely unpredictable during an out-of-sample period (1990-95), while Cooper, Gutierrez and Marcum (2005) conclude that the relative returns on portfolios of stocks sorted on firm size, book-to-market value and past returns were not ex ante predictable during the period 1974-97.[2] Marquering and Verbeek (2005) study the economic significance of predictability in both the conditional mean and conditional variance of stock returns and conclude that the profitability of trading strategies they examine is concentrated in the first half of the sample period. Sullivan, Timmermann and White (1999) find that technical trading rules cease to identify profitable trading strategies for the period 1986-96, although there was some evidence that they managed to do so prior to this period.

While these studies find evidence of instability in return forecasting models, they do not determine the time where the return models may have changed, nor do they consider the possibility of

[1] An incomplete list of studies on predictability of stock returns includes Ait-Sahalia and Brandt (2001), Avramov and Chordia (2002), Bekaert and Hodrick (1992), Bossaerts and Hillion (1999), Brandt (1999), Campbell (1987), Campbell and Shiller (1988), Cochrane (1991), Fama and Schwert (1977), Fama and French (1988), Ferson and Harvey (1991), French, Schwert and Stambaugh (1987), Keim and Stambaugh (1986), Lamont (1998), Lettau and Ludvigson (2001), Lewellen (2001), Perez-Quiros and Timmermann (2000), Pesaran and Timmermann (1995), Whitelaw (1994). Bekaert (2001) discusses recent research on predictability.

[2] In contrast, Avramov and Chordia (2002) report evidence of ex-ante predictability in individual stock returns over the period 1965-1999 by using standard predictor variables but also including firm-specific characteristics such as book-to-market ratio, turnover, previous-year returns and idiosyncratic volatility.

earlier structural breaks or the time of their occurrence. These are important issues to address since a plausible explanation for the discrepancy between the apparent strong in-sample predictability and the weak out-of-sample predictability is that the predictive relations are structurally unstable and change over time. Furthermore, if financial prediction models are unstable, the economic significance of return predictability can only be assessed provided it is determined how widespread such instability is both internationally and over time and the extent to which it affects the predictability of stock returns.

This study investigates these questions. Using data on a sample of excess returns on international equity indices we analyze both how widespread the evidence of structural breaks is and to what extent breaks affected the predictability of stock returns. We focus on ex post or full-sample predictability, while many earlier studies have studied ex ante predictability. There are several advantages of this approach over an ex-ante approach that splits stock return data into estimation and forecasting sub-samples (as is traditionally done in the literature). First, our approach allows us to date the possible time of changes in the return prediction models. In real time it is very difficult to identify such breaks and their timing can only be determined with the benefits of hindsight, i.e., by using the full sample of stock returns. Second, our approach is likely to have more power to detect changes in predictable relations. In a recent paper, Inoue and Kilian (2004) show that tests based on in-sample predictability typically have much better power than out-of-sample tests which generally use much smaller sample sizes. Indeed, it is possible that the absence of strong out-of-sample predictability in stock returns is entirely due to the use of relatively short evaluation samples. By using the full sample for our analysis, we gain sufficient power to address whether this explanation is valid or whether predictability genuinely has declined over time.

More specifically, we provide a systematic analysis of the stability of forecasting models using a dataset of monthly stock returns for ten OECD countries, including all members of the G7. With the exception of the default premium, local country forecasting variables are employed. We test for the presence of structural breaks in stock returns and characterize the timing and nature of the breaks. We find evidence of breaks for the vast majority of countries in multivariate regression models for excess returns. Further, our results indicate that the relationship between particular state variables and stock returns may change substantially following a break. Empirical evidence of predictability is not uniform over time and is concentrated in certain periods. For a number of the countries examined in our study ex post predictability appears to be substantially weaker after the most recent break, although a few exceptions exist. Using a longer historical dataset for the UK and US we find evidence of a common break around 1974-1975, which we relate to the oil price shock. Additionally, there is some evidence of a common break affecting a number of European markets during the period 1978-1982. We suggest that this break may be related to the introduction of the European Monetary System in 1979 and the associated constraints imposed on monetary and fiscal policy in member nations.

Our focus on international indices affords several advantages. First, the literature on stock return predictability is weighted toward US data with relatively few studies examining the question of predictability in global returns. Ang and Bekaert (2004) examine predictability for the US, UK, Germany and France while Campbell (2003) examines predictability in 11 countries using monthly data beginning in 1970. Hjalmarsson (2004) provides a comprehensive empirical investigation of global stock return predictability, using panel data that include over 20,000 monthly observations from 40 international markets, including 22 of 24 OECD nations. Rapach, Wohar and Rangvid (2002) examine both in-sample and out-of-sample performance of return prediction models for 12 countries. Broadly, the evidence reported in these studies suggests that the return predictability phenomenon extends to the global setting. Ang and Bekaert (2004), Rapach, Wohar and Rangvid (2002) and Hjalmarsson (2004) conclude that the short interest rate is a robust predictor of stock returns internationally, particularly at short horizons. The studies arrive at different conclusions, however, regarding the dividend yield as a forecasting variable. Ang and Bekaert (2004) find that the dividend yield predicts returns at short horizons when used in conjunction with the short rate, whereas Hjalmarsson (2004) concludes that there is no consistent evidence that the dividend yield (or earnings ratio) predicts returns for OECD countries.

While these recent studies address the question of international stock return predictability, to our knowledge this paper is the first to systematically address the question of whether globally documented predictive relationships are stable over time. Hjalmarsson (2004) touches briefly upon this issue by presenting results from rolling regressions using a 60-month window, however, formal tests of stability are not presented. Further, Hjalmarsson (2004) considers each regressor separately in turn, while we consider multiple regression models. Finally, following recent developments in breakpoint testing, we focus on occasional, large shifts in coefficients rather than a gradual evolution and we attempt to characterize the timing and nature of breaks, as well as investigate whether the timing of breaks appears to be uniform across countries.

In contemporaneous research, Rapach and Wohar (2005) find complementary evidence of instability in return regressions using US data and a broad set of forecasting variables. They apply $SupF$-type tests to detect the presence of breaks and apply a method suggested by Bai and Perron to select models (as we do). We demonstrate via simulation experiments that the finite sample performance of $SupF$-type tests can be rather poor in the presence of persistent lagged endogenous regressors. This finding is clearly relevant in the context of stock return regressions since 'ratio' variables such as the dividend yield and price-earnings ratio satisfy this description. Fortunately, our simulation analysis illustrates that a recent test for instability suggested by Elliott and Müller (2003) possesses excellent finite sample size properties even in the presence of persistent lagged endogenous regressors. This test provides important corroboration regarding our evidence of breaks. As further corroboration, we present results for breaks in long-horizon return regressions using cumulated returns. The breaks identified at the single-month horizon carry over to multiple-horizon regression models in most cases.

The remainder of the paper is organized as follows. Section 2 introduces the breakpoint methodology applied in this study. Section 3 reports the outcome of Monte Carlo experiments for the small sample performance of break tests and model selection procedures. Section 4 describes the international returns data. Section 5 presents empirical results of tests for breakpoints and structural stability in international equity indices. Section 6 characterizes the nature of breaks, including the timing of the breaks, changes in the regressions coefficients and the predictable component of returns, and offers possible economic motivations for common breaks. Section 7 considers issues of robustness as well as several extensions of the basic results. Section 8 summarizes and further discusses our findings.

2. Motivation and methodology

In the context of linear regression models many empirical studies have documented the ability of a variety of economic variables to predict stock returns (see the references in footnote 1). To apply models of this type in practice, parameters must be estimated using historic data of returns and predictor variables. Besides determining which variables to include, a key decision when estimating return forecasting models is how much data to use.

Determining the sample size for the return prediction model can be very important if the coefficients are not constant over time and including pre-break data will lead to biased forecasts. For example, Brandt (1999, p. 1611) points out the importance of stability in the relation between state variables and stock returns: "Returns and forecasting variables must have a time-invariant Markov structure. If the relation between returns and forecasting variables is time-varying... conditional expectations cannot be estimated with conditional sample averages."

However, there are good reasons for suspecting instability. On theoretical grounds breaks or discrete changes in the parameters that relate security returns to state variables could arise from a number of factors, such as major changes in market sentiments or regime switches in monetary policies (e.g., from money supply targeting to inflation targeting). Institutional changes or large macroeconomic shocks that give rise to changes in economic growth or affect risk premia may also cause a break in the financial return models. Similarly, if predictability of returns partly reflects market inefficiencies and not just time-varying risk premia, then such predictive relationships should disappear once discovered provided that sufficient capital is allocated towards exploiting them. For example, Dimson and Marsh (1999) argue that the small-cap premium disappeared in the UK stock market after it became publicly known. Finally, in an international context, breaks may arise as a by-product of the ongoing globalization process, i.e., as markets become more integrated and, as in the case of the European Union, fiscal and financial policy constraints are introduced on member nations. These possibilities are important both because they introduce new sources of risk and because they fundamentally affect the extent to which returns are predictable.

There are also good empirical reasons to expect breaks to be important. In a thorough study of a large set of financial and macroeconomic time series, many of which are commonly used as state variables in financial models, Stock and Watson (1996) find breaks in the regression models for the majority of the variables they consider. Andreou and Ghysels (2003, 2002) also find evidence of breaks in the comovements of foreign exchange returns and the volatility dynamics of asset returns related to the Asian and Russian financial crises.

Some recent studies have considered breaks in the equity premium. Using a Bayesian framework, Pastor and Stambaugh (2001) examine a long history of annual returns on US stocks and find evidence of structural breaks in the equity premium in the form of high posterior probabilities that breaks occurred during certain months of the sample. As pointed out by Pastor and Stambaugh, detection of breaks in the mean of stock returns is made extremely difficult by the very noisy nature of stock market returns. Without conditioning (state) variables, tests for structural breaks are unlikely to have sufficient power to identify breaks in the equity premium of an economically interesting size even if they truly occurred. Pastor and Stambaugh deal with this problem by assuming that there is a concurrent relationship between the level of volatility and the equity premium. Since it is easier to identify shifts in the volatility of returns, this provides an instrument to identify the timing of the breaks. On the other hand, the extent of a conditional risk-return trade-off, and even the direction of such a trade-off, remains a contentious and active topic in empirical finance. While the combination of a Bayesian setup and this identifying assumption provides a way to identify breaks, the drawback is of course that the number and timing of breaks in the equity premium may be sensitive to the nature of prior beliefs.[3]

2.1. *Econometric approach*

The approach and focus in this paper are very different from those in earlier studies. First, as we are interested in breaks in the return forecasting models that are now widely used throughout finance, we test for breaks in the conditional equity premium as a function of a set of commonly used state variables. This is an important exercise given the widespread use of these models throughout finance (see the references in the introduction). Furthermore, we use the estimation and model selection framework for linear models with multiple structural breaks developed by Bai and Perron (1998).[4] This allows us to determine the number of breaks, confidence intervals for the time of their occurrence as well as the value of the coefficients around the time of the breaks. By considering instruments whose correlation with the equity premium is sufficiently strong to identify breaks we therefore do not need to impose any identifying restrictions on our model. Of course, this approach is also not without disadvantages and some of our results will be quite noisy given the low predictive power typical of return prediction models.

Suppose that (excess) returns at time $t+1$, Ret_{t+1}, depend linearly on a vector of state variables, $\mathbf{x}_t$, but that the model is subject to K breaks occurring at times $(T_1, T_2, ..., T_K)$. This gives the

[3] Kim, Morley and Nelson (2000) also apply a Bayesian framework and test for a structural break in a model of excess returns in which the equity premium responds to recurrent changes in volatility. They find evidence of a structural break in the Markov switching variance process in the early 1940s, but do not find evidence of breaks in the equity premium given the level of volatility.

[4] Computations in this paper related to the Bai and Perron (1998, 2003) methodology were carried out using GAUSS programs made available by Bai and Perron.

model

$$Ret_{t+1} = \begin{cases} \boldsymbol{\beta}_1'\mathbf{x}_t + \varepsilon_{t+1}, & t = 1, ..., T_1 \\ \boldsymbol{\beta}_2'\mathbf{x}_t + \varepsilon_{t+1}, & t = T_1 + 1, ..., T_2 \\ \vdots & \vdots \\ \boldsymbol{\beta}_K'\mathbf{x}_t + \varepsilon_{t+1}, & t = T_{K-1} + 1, ..., T_K \\ \boldsymbol{\beta}_{K+1}'\mathbf{x}_t + \varepsilon_{t+1}, & t = T_K + 1, ..., T \end{cases} \tag{1}$$

In many respects this is a simplified representation of the return generating model and shifts in the regression coefficients, $\boldsymbol{\beta}$, may well occur gradually over time rather than through the assumed step function. Nevertheless, it can be viewed as a useful approximation to more complicated representations of time-variation in the parameters linking the state variables to stock returns. In fact, some of the potential sources of breaks such as shifts in economic policy regimes, large macroeconomic shocks or publication of predictable patterns are likely to lead to rather sudden shifts in the parameters of the return forecasting model. Furthermore, as pointed out by Andrews (1993), Elliott and Müller (2003) and Sowell (1996), tests for a single break also have power against alternatives such as a sequence of smaller breaks, so our tests have the ability to detect instability of a more general form.

The key objectives are of course to test for the presence of breaks, determine the number of breaks, K, and estimate both the time of their occurrence, $(T_1, T_2, ..., T_K)$, as well as the parameters around the time of the breaks, $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', ..., \boldsymbol{\beta}_{K+1}')'$.[5] Bai and Perron (1998) provide a least-squares method for optimally determining the unknown breakpoints as well as the resulting size of shifts in the parameter values. The basic principle involves searching over the possible K-partitions $(T_1, T_2, ..., T_K)$ of the data to compute the minimizer of the sum of squared residuals. For a set of K breakpoints, $(T_1, T_2, ..., T_K) = \{T_j\}$, the coefficient estimates $\hat{\boldsymbol{\beta}}_{k,\{T_j\}}$ are chosen to minimize the sum of squared residuals

$$S_T(\{T_j\}) = \sum_{k=1}^{K+1} \sum_{t=T_{k-1}+1}^{T_k} \left(Ret_t - \hat{\boldsymbol{\beta}}_{k,\{T_j\}}' \mathbf{x}_{t-1} \right)^2 . \tag{2}$$

The estimated break dates $\left(\hat{T}_1, \hat{T}_2, ..., \hat{T}_K\right)$ are selected so as to satisfy

$$\left(\hat{T}_1, \hat{T}_2, ..., \hat{T}_K\right) = \arg \min_{T_1, T_2, ..., T_K} S_T(T_1, ..., T_K), \tag{3}$$

where the minimization is over all partitions such that $T_k - T_{k-1} \geq \pi T$. The trimming percentage parameter π imposes a minimum length for the time between breaks, πT. Choosing π in practice involves a trade-off between the ability to detect regimes of relatively short length and the desire to avoid overfitting the data and simply identifying 'outliers'. While πT in principle may take any value greater than or equal to the number of regressors, in practice it is best to use values significantly larger than this.[6] Given the estimated break dates $\{\hat{T}_j\}$, the estimated regression coefficients $\hat{\boldsymbol{\beta}}_k$ are

[5] We adopt the convention that $T_0 = 1$ and $T_{K+1} = T$, where T is the total number of available observations.

[6] Bai and Perron (2003) discuss computational and practical aspects of determining these design parameters.

the least squares coefficients associated with the partition comprised of the estimated break dates, i.e., $\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{k,\{\hat{T}_j\}}$. Building on previous work in Bai (1997), Bai and Perron (1998) provide results for obtaining confidence intervals for the estimated breakpoints.

2.2. *Testing for Breaks*

Several types of hypothesis tests may be of interest when multiple breaks are allowed in the return prediction model. For example, one may be interested in testing the hypothesis of no breaks versus an alternative of K breaks, or in simply testing a null hypothesis of no breaks against an alternative of at least one break. We briefly describe the idea behind these tests.

The *SupF*-type test introduced by Andrews (1993) considers the null hypothesis of no breaks versus the alternative hypothesis that there are K breaks, where K is a specified number. Given a model with K breaks, The $SupF(K)$ test statistic is simply the supremum of a set of standard F-statistics for testing the null hypothesis of no breaks, where the supremum is taken over the set of possible break fractions. Bai and Perron (1998) also suggest a related test of l versus $l+1$ breaks, denoted the $SupF(l+1|l)$ test. To perform the test, one first estimates a model with l breakpoints $\left\{\hat{T}_1, ..., \hat{T}_l\right\}$ and computes the resulting sum of squared residuals (SSR) from this model. Conditional on these breakpoints one then identifies the $l+1$-*th* breakpoint and computes the SSR for this larger model as well. By construction the SSR is always reduced as the number of breaks, K, rises. Rejection of the null only occurs if the overall minimal value of the sum of squared residuals given $l+1$ breakpoints is sufficiently smaller than the sum of squared residuals from the model with l breaks. Bai and Perron establish critical values for determining how large the reduction in the SSR needs to be for the break to be statistically significant.

Breaks may occur not simply in the regression coefficients of the prediction model (1) but also in the marginal distribution of the predictor variables, $\mathbf{x}_{t-1}$, themselves. We consider this possibility by applying a testing approach suggested by Hansen (2000). Hansen derives the large sample distributions of several test statistics for breaks allowing for structural change in the marginal distribution of the regressors and shows that the asymptotic distributions are not invariant to structural change in the regressors. Hansen suggests a 'fixed regressor bootstrap' and shows that the bootstrap is able to replicate the first-order asymptotic distributions of the test statistics. Hansen's bootstrap approach allows for heteroskedastic error processes and lagged dependent regressors but does not permit serial correlation in the regression errors. Results are developed only for the case of a single break.[7]

2.3. *The J-test for instability*

As a final test for instability we apply the J-test suggested by Elliott and Müller (2003). This test statistic alleviates the need to search over high dimensions in the case of multiple breaks and

[7]Computations related to the Hansen (2000) methodology were carried out using Gauss programs made available by Bruce Hansen.

possesses good power properties for a wide class of alternatives to stability. The model considered in Elliott and Müller is given by

$$y_t = \mathbf{X}_t'(\bar{\boldsymbol{\beta}} + \boldsymbol{\beta}_t) + \varepsilon_t \quad t = 1, ..., T \tag{4}$$

where y_t is a scalar, $\mathbf{X}_t$ is a $k \times 1$ vector (both observed), ε_t is a zero mean error, and $\bar{\boldsymbol{\beta}} + \boldsymbol{\beta}_t$ are unknown with $\boldsymbol{\beta}_1 = \mathbf{0}$ as a normalization. The hypotheses to be tested are

$$H_0 : \boldsymbol{\beta}_t = \mathbf{0} \ \forall t \text{ against } H_1 : \boldsymbol{\beta}_t \neq \mathbf{0} \text{ for some } t > 1. \tag{5}$$

Elliott and Müller's test has power against a broad class of breaking processes including specifications with rare, large breaks as well as models with small, frequent breaks. Given a specific member of this breaking class, they apply the theory of invariant tests to derive an optimal test of the null hypothesis and show that any small sample optimal test statistic (against a specific member of their class of breaking processes) of the hypothesis (5) is asymptotically equivalent to any other optimal statistic (against a different breaking process). Elliott and Müller show that all optimal test statistics converge in probability under both the null and alternative to a feasible J-test which is asymptotically optimal under fairly general assumptions concerning the error and its relationship with the regressors.

Specifics regarding the construction of the J-test are somewhat cumbersome to describe. For the special case where breaks are restricted to the intercept and where the regression errors are serially uncorrelated, the J-test is equivalent to the Most Powerful Invariant (MPI) test in a Gaussian unobserved component model as studied by Franzini and Harvey (1993) and Shively (1988). Constructing the test statistic involves creating a time series based on innovations in the standardized regression scores and conducting artificial regressions of these on a type of nonlinear time trend. The test statistic is then based on the sum of squared residuals from these regressions. We refer the reader to Elliott and Müller for further details. We emphasize that the test permits heteroskedasticity and serial correlation, as well as weakly endogenous regressors. Regressors with stochastic trends, however, are not permitted. As the test is designed to have power against a variety of alternatives, it is not well-suited for model selection based on sequential tests.

2.4. *Model Selection*

To select the number of breaks, Bai and Perron (1998) propose a method of model selection based on sequential application of the $SupF(l+1|l)$ tests. The procedure is a specific-to-general model selection strategy. The process begins with a model including a small number of breaks thought to be minimally necessary (this may be zero). Given the current number of breaks, the $SupF(l+1|l)$ test is applied and if the test results in a rejection a new break is fitted and the process repeats until the test results in no rejection or the maximum allowable number of breaks is reached, in which case the procedure stops and the terminal model is selected. Information criteria offer alternatives to the sequential method for the purposes of selecting the number of breaks and Yao (1988) suggests use of the Bayesian Information Criterion (BIC). We assess the relative merits of these approaches for our application in the next section of the paper.

3. **Finite Sample Performance of Breaks Tests**

A primary concern in our setting is the potential for 'over-fitting', i.e. spuriously finding breaks when truly none exist. The results underlying the test statistics discussed above rely on asymptotic theory. For any specific data generating process, the adequacy of the tests in small samples must be assessed via Monte Carlo simulation experiments. Since the sequential method of Bai and Perron (1998, 2003) relies on a sequence of breaks tests, finite sample size problems related to these tests generally implies finite sample overfitting problems using the sequential method.

Bai and Perron (2004) perform a series of simulation experiments and assess the size and power of the various tests for breaks under a variety of data generating processes. These range from an independent Gaussian noise process to linear processes subject to two breaks where both the regressor and the error term are distributed heterogeneously across regimes. Also considered are cases with serially dependent errors, although in these cases only intercept shifts are included. Bai and Perron (2004) find that serial correlation and/or heterogeneity in the data or errors across segments can induce significant size distortions when low values of the trimming value π are used. Thus, if these features are present in the data, π values of 15% or higher are recommended, depending on the sample size and the particular features of the data. Bai and Perron find that the sequential procedure performs better than statistical information criteria, particularly if heterogeneity across segments is present. For the processes considered by Bai and Perron (2004), the tests have reasonable power and corrections for heterogeneity and serial correlation in the residuals (when these truly exist) improve power.

While these results provide support for application of the tests and model selection method in our setting, the data considered in our study exhibit characteristics that differ significantly from the data generating mechanisms considered by Bai and Perron (2004). Specifically, at least two of the regressors in our study, the dividend yield and the short interest rate, are known to be highly persistent. It is well known that the OLS estimates of highly persistent AR(1) coefficients, while consistent, are downward biased and have sampling distributions that differ from the standard setting. More recently, Cavanagh, Elliott and Stock (1995) and Stambaugh (1999) show that when a lagged endogenous regressor follows a highly persistent AR(1) process OLS coefficient estimates follow a non-standard distribution and can be significantly biased. Hence, when financial ratios such as the dividend yield or functions of interest rates are used to predict returns the resulting least squares coefficients are biased although asymptotically consistent.[8]

[8] Diebold and Chen (1996) assess finite sample performance of asymptotic and bootrap implementations of the the *SupF* test for breaks as well as the asymptotically optimal *AveF* and *ExpF* variants suggested by Andrews and Ploberger (1994). They conclude that bootstrap methods provide a better approximation relative to the asymptotic distribution for cases with small sample sizes and/or persistent dynamics. As in Diebold and Chen, our focus is on dynamic models. The simulatation analysis in this paper may be viewed a partial extension of the Diebold and Chen study to the case of a dynamic regression in which the single regressor is a persistent variable which may be only

Many recent studies examine inference in this setting and the extent to which returns are truly predictable. Ang and Bekaert (2004), however, consider a model that includes both the dividend and earnings yields as well as the short interest rate and find that the only statistically significant regressor is the short rate and its significance is limited to short horizons.[9] An issue that, to our knowledge, has not been previously addressed in the literature is the whether and to what extent this bias might introduce size distortions in tests for structural breaks. We explore via simulation experiments the possibility of 'spurious breaks' introduced by the presence of highly persistent lagged endogenous regressors.

While persistent lagged endogenous regressors may cause size problems for the breaks tests we employ, the power of these tests is also of concern since returns are inherently very noisy and the instruments we consider explain only a small fraction of the variation in returns. The noisy nature of returns data may dilute the power of tests to detect breaks and adversely impact the finite sample performance of information criteria for model selection purposes in the presence of breaks.

3.1. *Size experiments*

If the tests are over-sized, then a true null hypothesis of no breaks will be rejected more frequently than the asymptotic theory suggests. In examining the finite-sample size properties of the breakpoint tests, we consider the following data generating process:

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + \varepsilon_{t;}\;\; \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ x_t &= \theta + \varphi x_{t-1} + \upsilon_t; \upsilon_t \sim N(0, \sigma_\nu^2);\;\; E[\varepsilon_t] = E[\upsilon_t] \;= 0;\;\; \frac{E[\varepsilon_t \upsilon_t]}{\sigma_\varepsilon \sigma_\nu} = \rho. \end{aligned} \tag{6}$$

Here y_t is generated as a linear function of lagged x_t with a Gaussian white noise error term. The variable x_t follows a first order autoregressive process with φ governing the persistency of the process. Shocks to y_t and x_t have correlation given by the parameter ρ. When $\rho = 0$ the regressor is strictly exogenous and otherwise x_{t-1} is a predetermined but not strictly exogenous regressor.

We conduct several different experiments based on the data generating process described by equation (6). First, we consider a simplified case in which we set the parameters α, β and θ equal to zero and the variances σ_ε^2 and σ_υ^2 to unity. Note that in this case y_t follows a simple Gaussian white noise process and x_t does not Granger cause y_t. With these parameters fixed, we vary the persistence parameter φ over the values 0, 0.9, 0.95 and 0.98 and the correlation parameter ρ between the values 0 and -0.9. Our interest focuses on persistence values near unity since the corresponding parameters in empirical estimates of AR(1) models for the dividend yield, short interest rate and term spread tend to range between 0.9 and 0.99 while the non-persistent case is included as a benchmark. Similarly, the value $\rho = -0.9$ roughly corresponds to the empirical correlation of the errors obtained by fitting the system described by equation (6) to data for excess returns on the

predetermined rather than strictly exogenous.

[9] Examples of other studies that examine small sample inference with lagged endogenous regressors in the context of predicting returns include Goetzmann and Jorion (1993), Hodrick (1992), Nelson and Kim (1993), Lamont (1998), Stambaugh (1999), Lewellen (2004), and Campbell and Yogo (2004).

value-weighted CRSP portfolio (as y_t) and the dividend yield (as x_t) over the sample period 1952:7 to 2003:12.[10]

In the second set of experiments, all parameters in the system described by equation (6) are tuned to the corresponding empirical estimates obtained using the value-weighted NYSE index (as y_t) and the predictor (as x_t) over the sample period 1952:7 to 2003:12. We do this in turn for the dividend yield, short interest rate, term spread and default premium. Table 1 presents empirical estimates of the parameters in equation (4) for each forecasting variable using US data.

Our final set of experiments considers 'long-horizon' regressions of cumulative returns on the lagged forecasting variables. Data continue to be generated according to equation (6), however, we now cumulate the generated returns over a specified horizon and perform the regression of the cumulated returns series on lagged x_t. For example, at the two-month horizon, the cumulated return is defined as $y_{t,2} = y_t + y_{t+1}$ and the regression is run with $y_{t,2}$ as the regressand and x_{t-1} as the regressor (a constant is also included). As is common in the literature, we run our regressions using overlapping data. Since this induces serial correlation in the cumulated return series, we apply versions of breaks tests that correct for serial correlation whenever possible.

For all experiments the sample size is set to 500, which represents a value roughly between the number of observations in our longer dataset beginning in 1952:7 and that of our shorter dataset beginning in 1970:1. Results are computed over 2000 simulations using GAUSS's random number generator.

3.2. *Summary of size results*

Table 2 presents the results of the simulation experiments. All tests are evaluated at the ten percent significance level, and the tables report the percentage of cases in which the null hypothesis of no breaks is rejected when there is in fact no break in the process. For the BIC and sequential method for model selection we report the percentage of cases in which zero, one and two breaks are selected. We evaluate the size distortions of the tests and the adequacy of model selection techniques by comparing the results in Tables 2 with those predicted by the asymptotic theory. For instance, the $SupF(1)$ test applied at a 10 percent significance level rejects the null of no breaks 10 percent of the time asymptotically. We can compare this theoretical value to that obtained in the simulation analysis. Values substantially higher (lower) than 10 percent suggest that the test is over- (under-) sized.

Panel A presents results for the system described in the preceding section. This allows us to explore separately the effects of persistence and contemporaneous correlation. First, in the baseline case with no persistence and uncorrelated shocks, the tests are only slightly oversized. The BIC correctly selects a model with no breaks in nearly all cases and the sequential method performs well, selecting the true model approximately 88% of the time. As persistence is added to the system, the size distortions increase for the *SupF* and *UDMax* tests, but only marginally. The Hansen *SupF*

[10] The assumptions of Bai and Perron (1998) do not permit unit root regressors so we only consider highly persistent processes and not an actual unit root process.

test with bootstrapped p-values and the J-test continue to display excellent size properties in the presence of persistence. These results illustrate that persistence alone does not cause dramatically oversized tests or poor model selection performance using the Bai and Perron sequential method. When shocks are strongly negatively correlated but the regressor is not persistent, the size distortions are again very mild. Thus, correlation without persistence also does not result in size problems or overfitting in terms of model selection.

When we consider the case with both correlated disturbances and high persistence the distortions become much larger. In the worst case, with $\phi = 0.98$ and $\rho = -0.9$ the $SupF(1)$ test is substantially oversized, rejecting the null around 41% of the time. The $SupF(2)$ and $UDMax$ tests display even larger size distortions while the Hansen test displays smaller, but still substantial, distortions. The J-test, however, actually displays slightly under-sized behavior under both persistence and contemporaneous correlation. Thus, the behavior of the J-test fundamentally differs from that of the other breaks tests considered. It is not surprising, in light of the size distortions in the $SupF$-type tests, that the sequential method fares quite poorly in the presence of highly persistent lagged endogenous regressors. By contrast, the BIC method of model selection continues to perform well even for the most persistent processes considered. Turning to the results in Panel B, when all system parameters are tuned to empirical estimates based on US data the size distortions for the $SupF$ test are largest for the dividend yield, as expected. The $SupF$ test remains somewhat oversized for the other forecasting variables despite the fact that the contemporaneous correlations are relatively small in absolute value.

Before turning to the long-horizon size results (where we consider only the uncorrelated case), we briefly offer intuition regarding the size distortions in the $SupF$-type tests. The size distortion in the $SupF$ tests for the dividend yield is closely related to the upward bias in the estimated coefficient on the yield in univariate return regressions as studied by Stambaugh (1999), Cavanagh et al (1995) and others. Suppose for simplicity that the yield does not forecast returns so that the true, time-invariant coefficient on the yield is zero. The upward bias in the coefficient on the yield naturally translates into upward bias in the R^2-statistic for the regression in finite samples. Now consider the classic F-test for a single break with known timing. Heuristically, the F-test rejects when splitting the sample and estimating a different coefficient on each subsample results in 'too large' a reduction in the sum of squared residuals to be consistent with the null hypothesis. However, in the present case splitting the sample and estimating coefficients on each subsample increases the upward coefficient bias, and consequently the upward bias in R^2 for each subsample. There is thus a reduction in the sum of squared residuals due to the dependence of the Stambaugh bias on sample size, and this causes the F-test to be oversized. Taking the supremum over a series of F-statistics simply exacerbates this problem. The J-test performs much better precisely because it is not based on a sample-splitting approach.

Panel C presents results for the long-horizon return regressions using cumulated returns over horizons of two, four and six months. Since the use of overlapping observations induces serial correlation in the dependent variable, it is important to robustify test statistics for serial correlation wherever possible. The Hansen bootstrap procedure permits heteroskedasticity but not serial correlation. All other test statistics are robust to serial correlation. The size distortions for all the breaks tests are fairly mild for the two period horizon but increase with the horizon and are substantial when the horizon reaches six months. It is no surprise that the Hansen *SupF* procedure performs extremely poorly since it is not robust to serial correlation. The finite sample performance of the robust tests appears to be reasonably good for short horizons but degrades over longer horizons. The BIC model selection method also begins to degrade as the horizon increases. As noted by Bai and Perron (2004) a weakness of information criteria for model selection with breaks is that these criteria are not robust to serial correlation.

3.3. *Power of breaks tests and model selection when there are breaks*

The abundant noise in stock returns may hamper the detection of breaks. To assess the power of breaks tests and the adequacy of the various model selection methods, we generate data from the following process with a single breakpoint:

$$\begin{aligned} y_t &= \begin{cases} (\beta^* - \frac{\delta}{2})x_{t-1} + \varepsilon_t, & t = 1, ..., 250 \\ (\beta^* + \frac{\delta}{2})x_{t-1} + \varepsilon_t, & t = 251, ..., 500 \end{cases} \quad ; \ \varepsilon_t \sim \text{i.i.d. } N(0,1) \\ x_t &= \varphi x_{t-1} + \upsilon_t; \upsilon_t \sim N(0,1); \ E[\varepsilon_t \upsilon_t] = 0. \end{aligned} \tag{7}$$

The sample size is fixed at 500 for our experiments and the single break occurs at the midpoint of the sample. Note that this timing of the break is the most favorable possible for detection. The parameter β^* may be interpreted as the average regression coefficient and the parameter δ is the size of the break. Both shocks are normalized to have unit variance and the shocks are uncorrelated. We set the average regression coefficient β^* to be consistent with the R^2-values suggested by the empirical estimates of full-sample regressions without breaks. The Monte Carlo experiments address the power of various breaks tests and the performance of various model selection methods under a range of combinations for R^2 (and hence β^*), the size of the break expressed as a percentage of β^* and the parameter φ governing the persistence in x_t. As in the size study results are tabulated over 2000 simulations.

3.4. *Power results*

Results for the power experiments are displayed in Table 3. Panels A and B present results when R^2 is 5% and 10%, respectively. All statistical tests are conducted using empirical critical values based on 5000 simulations of the process under the null hypothesis of no break so that our results convey size-adjusted power. First consider the case where the R^2-value is 5%. While there is some variation in the power results as the persistence of x_t varies, the most dramatic variation in power occurs as the size of the break is increased from 10% to 100% of β^*. When the break is smallest (the break is 10% of β^*) the correctly sized (based on the empirical critical values under the null) $SupF(1)$ test detects the break only 9-11% of the time, depending on the persistence in x_t. The *UDMax,* Hansen and J-test statistics exhibit similar size-adjusted power. When the break is largest

(100% of β^*) the size-adjusted power of the $SupF(1)$ test is approximately 57% for the case with no persistence, and approximately 42% for the case with very high persistence. Both the Hansen test and the J-test exhibit comparable power relative to the $SupF(1)$ test. Indeed, these alternative tests exhibit slightly higher size-adjusted power when the persistence in x_t is very high. The size adjustment is important in this regard. If the tests are not size-adjusted the J-test rejects the null less frequently than the $SupF(1)$ test, particularly for persistent cases. However, in such cases the $SupF(1)$ test is oversized while the J-test is undersized. When the test is size-adjusted it is clear that there is little difference in power. Finally, the *UDMax* test exhibits lower power relative to the other tests, particularly in cases with high persistence.[11]

Turning to the model selection results, the BIC information criterion performs extremely poorly, incorrectly selecting a model with no breaks over 90% of the time, even for a break size of 100%. When the break size is small the BIC selects no breaks nearly 100% of the time. Naturally, the performance of the sequential method of Bai and Perron is closely tied to the performance of the Bai and Perron $SupF(1)$ test. When the break size is smallest the sequential method correctly selects a model with one break between 11% and 15% of the time, depending on the persistence in x_t, and in the majority of cases selects a model with no breaks. While this performance is poor, it is substantially better than the BIC, which selects no breaks essentially all of the time. When the break size is largest the sequential method correctly selects a model with one break 55% of the time when there is no persistence in x_t and 48% of the time for the most persistent case.

The degree of signal provided by x_t is increased by increasing R^2 to 10% in Panel B. As expected, the power of all of the break tests increases relative to the preceding case. It is interesting, however, that the increases are very modest for the smallest break size but go up as a function of the break size. When the break size is 100%, the size adjusted power of the $SupF(1)$ test is 70-87%, depending on the persistence in x_t. Once again, the Hansen and J-test exhibit similar size-adjusted power while the *UDMax* test is less powerful for the persistent cases. The increased power of the $SupF(1)$ test translates to better performance for the sequential method. In the best case, the sequential method correctly selects a model with one break 83% of the time, while the BIC correctly selects one break 44% of the time in this case. Further, the BIC performs extremely poorly for all break sizes under 100%. Thus, for the noisy regression models considered in this Monte Carlo study and typical for most return models, the break size must be very large for BIC to detect the break, and the sequential method appears to have superior power attributes in such cases.

[11]The break regression results discussed later in the paper suggest that breaks of 100% of the coefficient value, and even in excess of this, are empirically plausible. For example, in univariate regressions using the dividend yield (reported in Panel A of Table 6) the empirical $\frac{\delta}{\beta^*}$ associated with a break is typically on the order of 150% to 180%. The 1962 break associated with the short rate in the US exhibits a smaller empirical $\frac{\delta}{\beta^*}$ of approximately 30% so that smaller breaks are also occasionally detected empirically.

3.5. *Summary*

The preceding Monte Carlo experiments indicate several issues that plague inference regarding instability in return forecasting regressions. First, substantial size problems exist when the regressor takes the form of a persistent lagged endogenous variable. This clearly applies to the dividend yield forecasting variable considered in this study. Fortunately the *SupF*-type tests and the sequential method of Bai and Perron perform reasonably well under persistence when contemporaneous shocks are uncorrelated. The remaining forecasting variables considered in this study appear to fit this scenario, at least to an approximation. The excellent size properties of the J-test of Elliott and Müller (2003) suggest that this test can play an important role in confirming the presence of instability suggested by the Bai and Perron tests. Further, our size-adjusted power results illustrate that the J-test does not sacrifice much power relative to the *SupF*-type tests. In the empirical analysis, we point out cases where the *SupF* tests reject while the J-test does not reject. We suggest that such cases may reflect a spurious break and must be treated with caution.

The Monte Carlo experiments also reveal the limited power of tests for breaks in 'noisy' regressions. Given the extremely low R^2-values for univariate models of returns (see Table 4 below), one would expect the tests to have great difficulty in detecting any instability. The BIC model selection method performs extremely poorly when breaks are present. Put loosely, given the excessive noise in stock return regressions, unless a structural break is extremely large this information criterion will incorrectly select a model with zero breaks although one has truly occurred. For this reason, we opt to use the sequential method of Bai and Perron, despite its imperfections, as this method appears to perform better overall.

4. **Data description**

Ideally, our empirical study would examine evidence of breaks for a large number of international markets using a wide variety of forecasting variables reported in the stock return predictability literature. Since our study focuses on the possibility of occasional structural breaks affecting the relationship between stock returns and standard forecasting variables, a reasonably lengthy historical span of data is essential. Our decisions regarding the countries and sample periods examined in this study are motivated by an attempt to balance the desire for a broad and comprehensive empirical analysis with the competing desire for maximum data coverage. The empirical analysis focuses on two different datasets. The first dataset consists of monthly data for the UK and US that spans the period from July, 1952 through December, 2003. The second dataset consists of monthly data for ten OECD countries (including the UK and US) that spans the period from January, 1970 through December, 2003. The first dataset includes as much historical information as possible at the cost of including only two countries, while the second dataset provides a broader look at the international evidence at the cost of spanning a substantially shorter period.

We focus on four predictor variables that are prevalent in the empirical literature on predictability of returns. These variables are the lagged dividend yield (used, e.g., by Campbell and Shiller

(1988), Fama and French (1988), Ferson and Harvey (1991)); the short interest rate (Fama and French (1988), Fama and Schwert (1977), Ferson and Harvey (1991)); term spread (Campbell (1987), Fama and French (1988), Ferson and Harvey (1991)) and default spread (Fama and French (1988), Ferson and Harvey (1991), Keim and Stambaugh (1986)).

International data were collected primarily from Global Financial Data. Monthly total returns for 10 OECD countries were obtained along with the corresponding dividend yield series. The dividend yield is expressed as an annual rate and is constructed as the sum of dividends over the preceding 12 months, divided by the current price. For each country, a 3-month Treasury bill rate is used as a measure of the short interest rate while the yield on a long-term government bond is used as a measure of the long interest rate. Excess returns were computed as the total return on stocks in the local currency less the local short rate. This provides the local-country analog of the typical regression estimated using US data. A local country term spread is constructed as the difference between the long and short local country interest rates. The final instrument considered is the US default premium or quality spread, defined as the difference in yields between Moody's Baa and Aaa rated bonds. Since local country default premia were unavailable, the US default premium is used for each country. Data were collected for Belgium (BEL), Canada (CAN), France (FRN), Germany (GER), Italy (ITL), Japan (JPN), the Netherlands (NTH), Sweden (SWE), the United Kingdom (UK) and the United States (US) over the sample period 1970:1-2003:12 and for the UK and the US over the longer sample period spanning 1952:7-2003:12.[12]

The US return obtained from Global Financial Data is based on the S&P 500 index, and to explore robustness to this particular index we also obtained monthly value-weighted index returns and the associated dividend yields for the value-weighted NYSE-AMEX-Nasdaq (NAN) composite index and on the value-weighted NYSE index (NYSE) from the Center for Research in Security Prices (CRSP). Returns are inclusive of dividends and are calculated in excess of a one-month risk-free rate taken from the CRSP Risk Free Rates File and based on average prices.

5. **Are there breaks in stock return regressions?**

As a benchmark for the empirical analysis that follows, Table 4 presents the results of OLS regressions in the absence of structural breaks. Results are presented for both univariate models and for the multiple regression model that includes all four forecasting variables. The univariate models explain very little of the variation in excess returns. Although the maximum R^2 for a univariate model is 3.03% for the dividend yield in the UK, many of the R^2-values are less than 1%. Examination of the robust t-statistics for the univariate regressions reveals that the evidence of predictability appears stronger for the US and UK than elsewhere. Indeed, for the dividend yield these are the only two countries for which the estimated coefficient is statistically significant at conventional levels.

[12]The specific indices to which the total return and the dividend yield series correspond are: Belgium (CBB All-Share), Canada (Toronto SE-300), France (SBF-250), Germany (CDAX), Japan (Nikko Securities Composite), Italy (BCI Global), the Netherlands(Netherlands All-Share), Sweden (Affarsvärlden Return Index), the UK (FTA All-Share), and the US (S&P 500).

Interestingly, the US default premium appears to predict returns in a number of countries (Japan, Sweden, UK) but the evidence that it predicts US excess returns is somewhat weak.

The predictive ability as measured by the in-sample R^2-value is substantially higher for the multiple regression models, ranging from 9.08% for the UK to 0.91% for Belgium, both over the shorter dataset. Again, predictability appears to be strongest in the UK and US, but the dividend yield, short interest rate, and default premium are statistically significant for many of the countries in the multivariate model.

A brief comparison of the regression results for the US S&P 500 index and the UK FTA index across the two datasets considered provides motivation for formal tests for breaks. The samples overlap from 1970:1 onward, so any differences in the regression estimates arise from the inclusion of additional data back to 1952:7. The R^2 of the multivariate regression for the UK rises from 5.14% to 9.08% for the shorter sample as the estimated coefficients on the dividend yield, short interest rate and term spread increase dramatically in this model. Results for the US multivariate model show relatively little variation. However, the univariate models suggest some instability for US data. In the case of the S&P 500 the dividend yield appears to be a stronger predictor over the longer dataset relative to that beginning in 1970:1 as the coefficient estimate drops from 0.32 to 0.23 while the R^2 of the regression falls from 0.62% to 0.33%.

5.1. *Tests for breaks and determination of the number breaks*

In our implementation we allow all coefficients to change at each break since there is no strong reason to believe that the coefficient on any of the regressors should be immune from shifts. The multivariate model is therefore

$$\begin{aligned} Ret_t &= \beta_{0k} + \beta_{1k} Div_{t-1} + \beta_{2k} Tbill_{t-1} + \beta_{3k} Spread_{t-1} + \beta_{4k} Def_{t-1} + \varepsilon_t \\ t &= T_{k-1}+1, ..., T_k;\ k = 1, ..., K+1, \end{aligned} \tag{8}$$

where Ret_t represents the (excess) return for the international index in question during month t, Div_{t-1} is the lagged dividend yield , $Tbill_{t-1}$ is the lagged local country short interest rate, $Spread_{t-1}$ is the lagged local country term spread and Def_{t-1} is the lagged US default premium. The univariate regression models take the same form as equation (8) with only a single forecasting variable included.

Table 5 presents results of various tests for structural breaks and reports the number of breaks selected by the Bai and Perron sequential method at both the 10% and 5% significance levels. The table presents test statistics and indicates those values that are statistically significant at the 5% and 10% levels. These results set the trimming percentage, π, to 15 percent of the total sample. This corresponds to a minimum window of 7 years and 8 months between breaks for our dataset beginning in 1952:7 and to a minimum window of 5 years and one month between breaks for the dataset beginning in 1970:1. We explore the implications of changing the minimum window length in Section 7.

There is abundant evidence of structural breaks in the multivariate return prediction models.

For all countries with the exception of Italy, and for both datasets examined, all of the tests for breaks conclude that breaks are present at conventional significance levels. In light of the simulation results discussed in Section 3, it is reassuring that the J-test confirms the presence of breaks in every case with the exception of Sweden. For the longer dataset the sequential method (at the 10% significance level) selects a model with two breaks for the NYSE, S&P 500 and for the UK. A model with one break is selected for the NAN. For the shorter dataset one break is selected for six of the ten countries while two breaks are selected for Germany, Sweden and the US. In most cases, the model selection results are robust to the more conservative alternative of a 5% significance level for the sequential method.

For the univariate regressions the question of stability is less clear. The $SupF(1)$ and $UDMax$ tests suggest the presence of breaks for the dividend yield regressions in most cases. However, these rejections are rarely corroborated by the J-test, and are only occasionally corroborated by the Hansen test. As noted in the summary discussion of Section 3, these results could reflect spurious breaks detected using the $SupF(1)$ and $UDMax$ tests. We conclude that the evidence regarding instability in the dividend yield regressions is fairly weak.

Since shocks to the remaining forecasting variables do not exhibit a strong contemporaneous correlation with stock returns, the $SupF$-type tests should be more reliable for these univariate regressions. For our longer dataset, there is fairly strong evidence of a break in the term spread regression for US returns. This break is not detected using the shorter dataset. This suggests that the break occurs relatively early in the sample as will be confirmed in the next section of the paper where we discuss the timing of breaks. There is also evidence of a break in the short interest rate regression in the longer dataset for both the US and UK, although in this case the J-test does not confirm the results of the other break tests. Again the break is not identified using the shorter sample. The stability results for the default premium are the most difficult to interpret as the various tests fail to agree in many cases. Exceptions include Belgium, Canada, Italy and Japan, for which all tests fail to reject, and the UK for which all tests reject the null of stability.

Finally, in a few cases where no breaks are selected using the sequential method there appears to be evidence of multiple breaks. For the short interest rate regressions in Germany and Italy, for example, the $SupF(1)$ test fails to reject the null and no breaks are selected despite the fact that the $UDMax$ test suggests that breaks are present. As Bai and Perron (2004) note, break patterns may arise for which a single break is difficult to detect while there is strong evidence of two or more breaks.[13] The authors suggest that an alternative approach is to initiate the sequential model selection process based on the $UDMax$ test against an unknown number of breaks rather than the $SupF(1)$ test. Our empirical analysis thus adopts a conservative approach by applying the $SupF(1)$ test at the initial stage of the sequential method.

[13]For intuition on this case, consider a regression with only a constant as a regressor and suppose that the sample is divided into three equally long parts. If the mean of the variable changes in the second sample but is the same in the first and third sample, a model with a single break may not pick up the change, whereas a model allowing for two breaks would identify it.

6. **Timing of breaks and changes in predictability**

The hypothesis tests discussed above suggest that breaks may be an important feature of stock return regressions. While these tests suggest that instability is statistically important, tests for stability alone cannot reveal the economic significance of breaks. Further, the timing and characterization of breaks is important both from a forecasting perspective and for attempting to connect breaks to important international events or changes in policies and institutions.

The multivariate regression models are likely to hold the greatest interest in practice, since these models explain far greater variation in returns than univariate models (see Table 4). On the other hand, with four coefficients (and a constant) in each interval, characterizing the evolution in the coefficients is more difficult than in models with a single instrument. To facilitate interpretation of the results, we therefore first examine breaks in the univariate regressions. While these univariate models are probably of limited interest to investors given their very low R^2-values, we find that examining the univariate case provides interesting insights into the source of the breaks identified for the multivariate model. They are also likely to have better power to detect breaks in the event of a partial break occurring only in a subset of the regression coefficients in the 'all' model.

6.1. *Dividend yield*

The estimated coefficients and standard errors for the dividend yield regression, along with the estimated breakdates and a 90% confidence interval for each break, are presented in Panel A of Table 6.[14] As always, caution should be exercised when interpreting the coefficient estimates on the dividend yield because of small sample bias, see, e.g., Campbell and Yogo (2004), Stambaugh (1999).[15] A single break was identified for six of the 10 OECD countries examined, including Belgium, France, Japan, Sweden, the UK and the US. A break was also identified for each of the US indices using the longer dataset. The break for the US portfolios appears to fit the run-up in stock prices in the late 1990s and subsequent reversal. Interestingly, the R^2-value of the regression increases for most countries following a break in the dividend yield. In some cases the apparent increase in predictability is substantial. For example, the R^2-value for the S&P 500 vaults from 0.6% to 11.0% after the break in our shorter dataset and from 1.2% to 7% using our longer dataset. Typically, the coefficient is statistically insignificant prior to the break. Further, for those countries where no break is identified the full-sample coefficient is insignificant in all cases.

The estimated breakpoint for the S&P 500 identified using the shorter dataset occurs during 1996, whereas for the longer dataset the estimated break occurs in 1994. A potential source of this discrepancy is the minimum window length imposed between breaks. Since this window is computed as a fraction of the total sample size, the minimum window length is longer for the 1952-

[14] All standard errors are corrected for heteroskedasticity and serial correlation using the method of Andrews and Monahan (1992).

[15] In addition to finite sample bias problems, Ang and Bekaert (2004) find strong size distortions on the dividend yield coefficient for long return horizons while the distortions are relatively small at a short horizon of one month such as the one used in our paper.

2003 sample period. With a 15% trimming percentage the window length is seven years and seven months so that a breakpoint in 1996:7 is just outside the boundary of admissible breakdates. When we reduced the trimming window to 10% for the longer dataset a break date is estimated to occur in 1996 for the NAN index but not for the NYSE or S&P 500 indices, which continue to exhibit a break in late 1994. This suggests that the minimum window length is not the sole explanation for the discrepancy in the timing of the break. In Section 7 of the paper we explore the effect of considering repurchases and a total payout yield for US data. There we find an estimated breakdate in late 1994, consistent with the longer sample results for the US portfolios.

6.2. *Short interest rate*

Panel B displays results for the short interest rate regressions with breaks. Evidence of breaks across the 10 OECD countries using the shorter dataset is limited. A single break is estimated for three countries (Belgium, France and Sweden) and for two of these three cases the estimated break occurs near the end of the sample. The post-break predictability is extremely high, however, the post-break sample is less than five years in length. Evidence of breaks appears stronger for the UK and US using the longer dataset. For each of the US indices considered, there are two breakpoints which occur in 1962 and 1974, respectively. The break dates for this model are reasonably precisely estimated although the confidence intervals for the second break are somewhat wider. The coefficients on the short rate are negative and strongly significant during the first subinterval ending with a break in 1962.[16] The coefficient estimates diminish in absolute value during the subinterval between 1962 and 1974, although they remain significant at conventional levels. Subsequent to the break in 1974, however, the estimates become insignificant and the predictive power of the short rate falls dramatically. Thus for the model based on the short interest rate our results support the finding of a breakdown in financial return prediction models.

For the UK, a single break is identified in 1974. The pattern in the coefficients is similar to that for the US, as the coefficient on the short rate is no longer significant following the break. The 1974 break in the short rate regressions for the UK and US may be related to the large macroeconomic shocks reflecting large oil price increases and the resulting break in the trend of US GDP found to have occurred around this time (Perron (1989)). One possibility is that breaks in the underlying economic fundamentals process can explain breaks in financial return models. There is no reason to expect financial return models to be immune to breaks in economic growth since these are likely to affect investors' intertemporal marginal rates of substitution and hence the process driving risk premia. It is important to note that given the minimum window length of over five years a breakpoint in 1974 is infeasible for the shorter dataset. Thus, the fact that a 1974 break is not identified more broadly using the second dataset should not be interpreted as evidence against a widespread international break.[17]

[16]The 1962 break for the US indices occurred around the time of the Cuban Missile Crisis (October 1962), generally viewed as a turning point in the cold war. It is possible that risk premia in the US were altered by this important political event.

[17]In section 7 of the paper we formally test for a contemporaneous break in the US and UK regressions using the

6.3. *Term spread and default premium*

There is only limited evidence of breaks in the univariate return models based on the local country term spread and US default premium. A break is identified in 1989 for Japan in the term spread model. The coefficient on the term spread is negative and statistically significant prior to the 1989 break, but the sign on the coefficient flips to positive (although insignificant) following the break. For the S&P 500 a break is identified in 1975 in the longer dataset. The features of this break are similar to those of the break in the short rate for the US indices above. The predictive power of the spread thus appears to be concentrated in the early part of the sample. Again, the break may be related to the oil price shock in 1974, as this lies well within the confidence interval for the breakdate.

For the regressions based on the US default premium, a single break is identified for four of the 10 countries considered, including France, Sweden, the UK and the US using the shorter dataset. The breakpoint estimated for the UK is January, 1975, which is the boundary point based on the minimum window length of five years and one month for the shorter dataset. This observation suggests that the break may in fact occur slightly earlier, and may again be related to the oil price shock. The US break is similarly timed, although the point estimate is for May 1975, which is within the bounds established by the minimum window length. The break in France also occurs at roughly the same time, while the break for Sweden is estimated to occur in 1992. For all countries the predictive ability of the default premium as measured by the regression R^2-values over different subintervals appears to be concentrated in the subinterval preceding the break.

6.4. *Multivariate regression models*

We now turn to the multivariate regression models that include the (lagged) local country dividend yield, short interest rate and term spread along with the US default premium. Table 7 presents results for the multivariate regression models with breaks. A constant is always included in the regressions but is not reported to preserve space. The breakpoint regressions reveal several interesting results. Most notably, the individual regression coefficients change substantially following a break. While some of this variation can clearly be explained by sampling variation due to sometimes large standard errors, this does not conceal the fact that breaks in the regression model appear to be sufficiently large to be of substantial economic interest. Reassuringly, however, there are not many instances where the sign of the regression coefficient estimate changes.

While the timing of breaks varies across international markets, there does appear to be some evidence of clustering in the estimated break dates. An estimated break date occurs for many of the European indices between the years 1978 and 1982, including Belgium, France, Germany, the Netherlands, and Sweden. The US does not appear to experience a break at this time while Canada does. A break is estimated for the UK in 1976 using the shorter sample of data, however, upon examining the longer sample the break appears to be timed in 1974 and the 1976 estimate may be

short interest rate.

related to power issues around the boundary based on the trimming percentage. A possible catalyst for the European breaks identified around this time was the founding of the European Monetary System in 1979. The eight original member nations agreed to hold exchange rates within certain limits. This event presumably introduced new constraints on domestic fiscal and interest rate policy. It is noteworthy that the UK does not experience a break at this time. Indeed, the UK declined to participate in the system of mutually fixed exchange rates adopted in 1979.

The 1987 break that occurs in the US (NYSE and S&P 500 portfolios) appears to be an isolated break not experienced by other international markets. Finally, several countries experience breaks in the mid 1990s, including Germany, Japan, Sweden and the US. In the following section of the paper we discuss formal econometric tests of contemporaneous breaks across countries.

As noted earlier, previous empirical work suggest a breakdown in forecasting models in the 1990s. This work is primarily based on US data and does not attempt to time the breakpoints. Consistent with the lower coefficient estimates observed after the most recent break, the R^2-value was at or near its minimum after the most recent break for all of the US indices considered. While in-sample or ex post R^2-values in the range 0.15-0.25 are common before the first and second break, the R^2-values are near or below 0.10 after the most recent break. These values are much higher than those typically reported in studies of ex ante predictability or even for full-sample models. However, it should be recalled that, by construction, R^2-values will generally be somewhat higher because more parameters are used to minimize the sum of squared residuals than when no break is assumed. Since breaks are typically not detectable in real time, the reported R^2-values could never be exploited in an ex ante investment strategy. While one should therefore not put much into an interpretation of the observed level of these R^2-values, nevertheless broad changes in these are of considerable economic interest and are helpful in identifying patterns in ex-post predictability.

The European countries experiencing a single break between 1978 and 1981 (Belgium, France, the Netherlands) also exhibit a lower R^2-value following the break. Exceptions to the general pattern exist, though. In particular, predictability appears to have increased following the most recent break for several countries that experienced a break in the 1990s (e.g., Germany and Japan). Based on the longer dataset for the UK, the R^2-value following the second break in 1974 remained quite high at 19.1% despite the fact that this was a decrease relative to the previous subinterval.

7. Robustness and extensions

This section further considers the robustness of our evidence regarding breaks in stock return regressions, extends the previous results by considering long-horizon return regressions and tests for contemporaneous breaks across blocks of countries considered in our study.[18]

[18]We thank an anonymous referee for suggesting many of inquiries addressed in this section of the paper.

7.1. *Dividends, repurchases and total payout*

Our empirical evidence suggests a break in the dividend yield regression for US data in the mid 1990s. Earlier in the paper we discussed the potential of size distortions in tests for breaks in the dividend yield regression due to the lagged endogenous regressor bias. On the other hand, one might construct an economic argument in support of a break in the dividend yield regression arising from the gradual substitution of repurchases for dividends as a form of payout (see, e.g., Grullon and Michaely (2002)). More firms have opted for repurchases as a form of returning cash to investors and the fraction of firms paying dividends has also declined.

The top panel of Figure 1 plots a time series of the dividend yield from 1970 - 2003 and indicates our estimated breakpoint along with a 90% confidence interval for the breakdate. From the mid 1980s through the 1990s, the dividend yield appears to trend downward. Our estimated breakpoint of 1996:7 is in the midst of this period characterized by a declining dividend yield. Recent papers including Grullon and Michaely (2002) and Boudoukh, Michaely, Richardson and Roberts (2004) suggest that a measure of the total payout yield, including both dividends and repurchases, is a more sensible measure of payout for asset pricing applications. Boudoukh, Michaely, Richardson and Roberts (2004) find statistically and economically significant predictability at short and long horizons when the total payout yield is used instead of the dividend yield. They also find that while predictability diminishes after 1985 in a predictive regression using the dividend yield, predictability remains nearly constant when using the total payout yield.

To explore the robustness of our results for the dividend yield, we constructed a monthly total payout series based on repurchases data as well as dividends. We obtained year-end (annual) data for the period 1972-2003 consisting of repurchases and market capitalization values for a broad sample of Compustat firms.[19] Repurchases are defined as the total expenditure on the purchase of common and preferred stock less any reduction in the value of the net number of preferred stocks outstanding. Although repurchases data are not available at the monthly frequency, we construct a proxy for the monthly repurchases yield by scaling the year-end repurchases by the year-end market capitalization value. This provides a year-end (December) repurchases yield. In the subsequent months leading up to the next year-end update, we adjust the market capitalization value by the total return on the S&P 500 portfolio to account for market price fluctuations.[20]

The bottom panel of Figure 1 displays our constructed total payout yield. The total payout yield does not appear to trend downward as the dividend yield does; however, there does appear to be a downward shift (break) in the level of the series around 1991 or 1992. This may correspond to the recession in 1991 which reduced firms' repurchase activity. Although repurchases activity ultimately picked up again, so did share prices, so that the level of the repurchases yield remains below that observed over much of the 1980s.

To assess the sensitivity of our finding of a break in the dividend yield regression to the consid-

[19]We thank Gustavo Grullon for providing the repurchases and market capitalization data.

[20]We use the S&P 500 return as a proxy for the true monthly changes in the market capitalization value for the firms in the Compustat sample.

eration of repurchases we tested for breaks in a regression of excess returns on the S&P 500 on our total payout yield measure using the sample period 1970:1 - 2003:12.[21] The bottom panel of Figure 1 indicates our estimated breakpoint for the total payout regression. The Bai-Perron procedure selects a model with a single break where the estimated breakpoint is 1994:11, which differs slightly from the 1996:7 breakpoint identified for the dividend yield regression over the same sample period. Interestingly, however, the 1994:11 breakpoint is nearly identical to the breakpoint identified in dividend yield regressions for all US portfolios over the longer data sample period 1952:7 -2003:12 (see Table 6). Using the total payout yield thus helps resolve the slight discrepancy in the timing of the break in the yield regression across the two sample periods we examine. The behavior of the regression coefficients before and after the break are similar in the dividend yield and total payout yield regressions. The pre- and post- break coefficients on the lagged total payout yield are 0.61 and 4.70, respectively, while the corresponding values for the dividend yield regression are 0.40 and 7.46.

To summarize, our results indicate that little changes regarding evidence of breaks in yield regressions when a total payout yield is considered instead of the dividend yield. A potential source of the break is the apparent level shift in total payout yield in the early 1990s. As Figure 1 indicates, the 90% confidence interval for the estimated breakdate includes this period.

7.2. *Minimum distance between breaks*

The estimation procedure suggested by Bai and Perron (1998) and further discussed in Bai and Perron (2003) imposes a minimum number of periods between breaks of at least q, where q is the number of covariates subject to instability. In practice, Bai and Perron suggest that the minimum window should be set to a value h much greater than q. The choice of minimum window, h, is expressed as a fraction of the total sample size as the trimming percentage π and may have important implications for model selection. For a fixed number of breaks, larger values of h (equivalently of π) place more limitations on the combinations of break dates considered. For example, if the sample size is 100 and h is set to 49, then effectively the only time a break can occur is at observation 50. Increasing the window length can hence sharply reduce the combinations of breakpoints allowed. Indeed, this is an integral part of reducing the number of computations required in the efficient dynamic programming routine suggested by Bai and Perron (1998). The sequential testing method that is used to select the number of breaks in the model also relies on the trimming percentage π through its effect on the set of break partitions over which the supremum of F-statistics is computed.

To explore the robustness of the evidence of breaks in returns regressions to the choice of trimming percentage (recall that the results discussed above pertain to a trimming percentage of 15%) Table 8 presents a graphical summary of the breaks identified for different choices of π ranging from 5% to 20%. Note that in the case of a trimming percentage of 20% the maximum

[21] Since our repurchases data begin in 1972 the total payout yield for the initial months of data is simply the dividend yield. Repurchases formed a very small proportion of the total payout in 1970 and 1971, so this procedure introduces minimal error in our results and makes our sample period comparable to those used in the earlier sections.

allowable number of breaks is three. In all other cases the maximum number of breaks is set to five. Several features of Table 8 are notable. First, breaks are abundant regardless of the trimming percentage selected, indicating that the presence of breaks is not driven by a particular choice of the trimming percentage. Second, the timing of breaks is also frequently robust to the choice of trimming percentage. Exceptions tend to occur when the break occurs near the boundary point for the larger trimming percentages. For example, the breakpoint identified in 1994 for the US portfolios is robust to the choice of trimming window except when the trimming window is set to 20 percent. However, in this case a break of date of 1994 is infeasible, so a break is instead fit a bit earlier in the sample. Even if a break is technically feasible, the power of break tests declines near the boundaries, and so it is not surprising that in these cases the timing of the estimated breakdate shifts toward the interior of the sample. Overall, the main empirical results appear to be robust to the particular choice of trimming percentage.

7.3. *Misspecified functional form and breaks*

We adopt a linear functional form for the conditional expectation of excess returns. As in most applications of linear regression in economics, there is no inherent, compelling reason why the conditional expectation of excess returns need exhibit a linear form. Furthermore, the relationship between excess returns and forecasting variables implied by economic models such as that in Campbell and Cochrane (1999) is inherently nonlinear. We view the simple linear functional form as an approximation to the true, unknown, and potentially nonlinear conditional expectation. Nevertheless, a concern exists regarding whether omitted nonlinearity could give rise to spurious evidence of breaks. We note that this is fundamentally a finite sample issue, since asymptotically, under fairly general conditions, OLS parameter estimates for a misspecified linear model are consistent for the time-invariant 'pseudo-true' values that deliver the best linear approximation to the underlying nonlinear conditional expectation.[22] However, given the highly persistent nature of the forecasting variables considered in this study, it is natural to wonder whether omitted nonlinearity could give rise to spurious breaks in finite samples.

To investigate this question, we examined the robustness of our empirical evidence regarding breaks to more flexible specifications for the relation between forecasting variables and excess returns. Specifically, for the univariate forecasting models, we considered alternative models that included squared terms or both squared and cubed terms in the regression specification. The parameters on the higher order terms were assumed to be time invariant. We then tested for instability in the regression coefficients on the linear terms in the regression. If our evidence of breaks is driven by omitted nonlinearity, then we would expect to no longer find evidence of a break in the linear terms once a sufficient number of higher-order terms have been added to the specification. In fact, we find that the addition of squared and cubic terms has little impact on our evidence regarding breaks. For example, for the US portfolios we identified a break in the dividend yield regression in 1994, breaks in the short interest rate regression in 1962 and 1974 and a break in the term spread

[22] This assumes that the parameters of the true nonlinear conditional expectation function are time invariant.

regression in 1975. When we augment these univariate models with squared and cubic terms in the forecasting variable, we generally continue to find breaks at these times in the linear coefficient for the augmented model. For example, for the S&P 500 the addition of squared terms in the forecasting variable causes no change in the number of breaks identified or in the timing of the breaks, while the addition of both squared and cubed terms in the forecasting variable only affects the 1962 break for the T-bill regression.[23]

7.4. *Breaks in long-horizon return regressions*

If breaks occur in the conditional relationship between excess returns and forecasting variables, then it seems plausible that evidence of breaks should be apparent in long-horizon return regressions as well as in the single period regressions considered heretofore. The Monte Carlo results discussed in Section 3 of the paper suggest that the Bai and Perron methodology ultimately suffers from substantial size distortions as the horizon increases despite the use of tests that are robust to serial correlation. For this reason, we report results for regressions using cumulated returns only up to the six month horizon and we note that even at these shorter horizons results must be interpreted with some caution. Table 9 displays the breaks identified in regressions using cumulated returns for the multivariate regression model at horizons of two, four and six months. In all cases the trimming percentage is set to 15%, so the breaks displayed in Table 9 for cumulated returns may be compared with those displayed in the middle panel in Table 8 for the single-month horizon.

The results for the cumulated returns tend to corroborate the results for the single-month horizon, in the sense that breaks identified at the one-month horizon are frequently identified at longer horizons as well. For example, over all horizons considered a breakpoint is identified in 1994 for all US portfolios using the longer dataset. The break identified in the 1978-1982 range for many European markets also appears across all horizons considered. More breaks on average are identified for the long-horizon models. It is notable that a break is identified for all US indices using in the multivariate regression model in 1962 for the two-, four- and six-month horizons. Thus there is stronger evidence of a break in 1962 for the multivariate regression using overlapping long-horizon returns. As noted earlier, this break may be related to the Cuban Missile Crisis and appears to have affected the US market to a greater extent than the UK market. Interestingly, for all 3 US indices a break also appears in 1979-1980 in the long horizon regressions. It is not difficult to point to a potential source for this break in the US: on October 6, 1979, the Federal Reserve adopted new policy procedures that led to two recessions but also ultimately ushered in an environment of low inflation and relative economic stability.

There are two potential explanations for the appearance of additional breaks in the long-horizon regressions. First, the higher signal-to-noise ratio in longer-horizon regressions may increase the power of tests to detect breaks. Alternatively, the additional breaks may be spurious and induced by finite sample size distortions as documented in the Monte Carlo study of section 3. Based on

[23] Since the evidence regarding the timing and number of breaks is very similar to previously discussed results, we do not separately report these results in the tables to conserve space.

the Monte Carlo results, if the additional breaks were purely spurious, one would expect breaks to gradually appear as the horizon is increased. Instead, it appears that many of the additional breaks appear immediately for the two-month horizon and these are carried through to the six-month horizon.

7.5. *Tests for contemporaneous breaks across countries*

Our estimates of the timing of breakdates suggest some clustering across countries in breaks in return prediction models, i.e., that forecasting relationships undergo changes at the same or nearly the same time. In particular, the evidence based on a country-by-country analysis suggests the possibility of a contemporaneous break in multivariate return forecasting models for European indices during the period 1978-1982. Additionally, there appears to be evidence of a contemporaneous break in regressions for the US and UK, particularly for the short interest rate and term spread regressions. This section of the paper extends the previous results by providing formal tests for the presence of contemporaneous breaks.

Qu and Perron (2004) develop results that extend the estimation and testing framework of Bai and Perron (1998) to multivariate equations. The Qu and Perron (2004) methodology extends results in Bai, Lumsdaine and Stock (1998) and Bai (2000) for detecting and estimating breaks in multivariate systems. The multivariate models with breaks are estimated using quasi-maximum likelihood based on normal errors. Inference regarding the presence of breaks is based on *SupLR*-type test statistics. The testing framework is very general and permits tests for changes in the regression coefficients, tests for changes in the covariance matrix of the residuals, or both. We restrict attention to tests for changes in the regression coefficients, since the focus of this study is on instability in coefficients of linear models of returns.

We test for a single contemporaneous break for the US and UK returns in the full regression model (including all forecasting variables for each country) over the sample period 1952:7 - 2003:12. The *SupLR* test statistic is 41.7 which is highly significant at the 5% level. Thus, we find evidence of a contemporaneous break for the US and UK. The estimated timing of the break is November, 1974 and a 90 percent confidence interval for the break extends from December, 1973 through March, 1975, so the breakpoint is quite precisely estimated. The estimated breakpoint coincides closely with the oil price shock of 1974-5 following the Yom Kippur War of 1973.

We additionally tested for a contemporaneous break across all markets in our study using the sample period 1970:1 - 2003:12 and separately for just the subset including the European markets for which a break was identified in the 1978-1982 range in the country-by-country analysis (Belgium, France, Germany, the Netherlands, Sweden). In neither case was the null hypothesis of no joint break rejected.[24]

While the inability of the joint test to detect a common break for the European nations may seem surprising, it should be noted that the extent of predictability, as measured by the full-sample R^2-value for the regression with all predictor variables (see Table 4) is quite weak for most of the

[24] Tests for contemporaneous breaks in the univariate models also did not result in rejections of the null.

European nations that exhibit a break in the 1978-1982 period. Indeed, over the 1970-2003 period the highest R^2-values are for the UK and US, respectively. This may help explain why we are able to identify a common break for the US and UK but not for the European group of nations, despite the closely clustered breakdates as per the single-country regressions.

In moving from a single-country regression to the multi-country regression setting, the number of parameters increases substantially. If some of the regression parameters do not change substantially following a break, then this may dilute the ability of the multivariate test to detect a common break. Looking at the single-country regression results with breaks (Table 7), it is clear that not all coefficients exhibit substantial change following a break. While one might attempt to specify *a priori* which coefficients are likely to exhibit breaks, it is difficult to provide a firm economic justification for such restrictions.

8. **Conclusion**

This study presents systematic empirical evidence of structural breaks in models of predictable components in international stock returns based on the lagged dividend yield, short interest rate, term spread and default premium. We find evidence of breaks for the vast majority of countries in multivariate regression models for excess returns. Further, our results indicate that the relationship between particular state variables and stock returns may change substantially following a break. Empirical evidence of predictability is not uniform over time and is concentrated in certain periods. For a number of the countries examined in our study ex post predictability appears to be substantially weaker after the most recent break, although a few exceptions exist. Using a longer historical dataset for the UK and US we find evidence of a common break around 1974-1975, which we relate to the oil price shock. Additionally, there is some evidence of a common break experienced by a number of European stock markets during the period 1978-1982. We suggest that this break may be related to the introduction of the European Monetary System in 1979 and the associated constraints imposed on monetary and fiscal policy in member nations.

The presence of structural breaks in predictive return regressions raises several economically interesting issues. First, how should conditional expected returns be estimated in the presence of breaks? One possibility is to use data after the most recently identified break date. Pesaran and Timmermann (2002) propose a procedure that reverses the ordering of the data in the prediction model and estimates the model parameters using only post-break data. This method determines the break date using a Cusum-squared test and directly addresses the question of how much historical data to use. Such an approach is, however, unlikely to work well if the data sample after the most recent break is very short. For this case, Pesaran and Timmermann (2004) prove that it can be optimal to use pre-break data provided that the break is not very large (so the bias in the predictive return regression does not get too large). They propose viewing the length of the data window used in estimation of the conditional mean as a separate parameter that is optimally chosen to trade-off (squared) bias against reduction in parameter estimation error resulting from using pre-break data.

A very active strand in the finance literature considers asset allocation problems in the presence of predictable asset returns (see, e.g., Ait-Sahalia and Brandt (2001), Barberis (2000), Brandt (1999), Campbell and Viceira (1998) and Kandel and Stambaugh (1996)). The predominant approach in the literature is to presume a time-invariant relationship between forecasting variables and expected returns. Clearly, the possibility of breaks and time-variation in the conditional mean function for asset returns complicates the portfolio allocation problem. Thus, a second question is how to adjust portfolio weights in the presence of breaks to the return forecasting equation. Answering this question requires a full-blown model for the underlying breakpoint process since the predictive density of future stock returns now becomes a mixture of the return distribution conditional on no future breaks (i.e., remaining in the current regime) and the return distribution given that a break occurs. Hence the probability of a future break must be computed and the parameters of the return prediction model after a break must be drawn from some 'meta distribution' characterizing the parameters across the various break segments. This is best accomplished using a Bayesian approach, but relies on distributional assumptions about the regressor variables, error terms and the underlying breakpoint process that go beyond the present paper. These issues are addressed by Pettenuzzo and Timmermann (2005) who find that structural breaks can have significant effects on the optimal asset allocation.

References

[1] Ait-Sahalia, Y. and M. Brandt, 2001, Variable selection for portfolio choice. Journal of Finance 56, 1297-1350.

[2] Andreou, E. and E. Ghysels, 2002, Detecting Multiple Breaks in Financial Market Volatility Dynamics. Journal of Applied Econometrics 17, 579-600.

[3] Andreou, E. and E. Ghysels, 2003, Tests for Breaks in the Conditional co-movements of asset returns. Statistica Sinica 13, 1045-1074.

[4] Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821-856.

[5] Andrews, D.W.K. and J.C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953-966.

[6] Andrews, D.W.K. and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

[7] Ang, A. and G. Bekaert, 2004, Stock return predictability: Is it there? Manuscript, Columbia University and NBER.

[8] Avramov, D. and T. Chordia, 2002, Stock Returns are Predictable Ex Ante: A Portfolio Evaluation Perspective. Manuscript, University of Maryland.

[9] Bai, J., 1997, Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551-563.

[10] Bai, J., 2000, Vector autoregressive model with structural changes in regression coefficients and in variance-covariance matrices. Annals of Economics and Finance 1, 303-339.

[11] Bai, J., R.L. Lumsdaine and J.H. Stock, 1998, Testing for and dating breaks in multivariate time series. Review of Economic Studies 65, 395-432.

[12] Bai, J. and P. Perron, 1998, Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

[13] Bai, J. and P. Perron, 2003, Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1-22.

[14] Bai, J. and P. Perron, 2004, Multiple structural change models: A simulation analysis. Forthcoming in Econometric Essays, D. Corbea, S. Durlauf and B.E. Hansen (eds.), Cambridge University Press.

[15] Barberis, N., 2000, Investing for the long run when returns are predictable. Journal of Finance 55, 225-264.

[16] Bekaert, G., 2001, Editors Foreword. Journal of Empirical Finance 8, 451-457.

[17] Bekaert, G. and R.J. Hodrick, 1992, Characterizing predictable components in excess returns in equity and foreign exchange markets. Journal of Finance 47, 467-509.

[18] Bossaerts, P. and P. Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What do we Learn?. Review of Financial Studies 12, 405-428.

[19] Boudoukh, J., R. Michaely, M. Richardson and M.R. Roberts, 2004, On the implications of measuring payout yield: Implications for empirical asset pricing. NBER Working paper No. 10651.

[20] Brandt, M.W., 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach. Journal of Finance 54, 1609-1646.

[21] Campbell, J.Y., 1987, Stock returns and the term structure. Journal of Financial Economics 18, 373-399.

[22] Campbell, J.Y., 2003, Consumption-based asset pricing, in: Constantinides, G.M., Harris, M., and Stulz, R. eds., Handbook of the Economics of Finance, Vol. 1B (North-Holland, Amsterdam) 803-888.

[23] Campbell, J.Y. and J.H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205-251.

[24] Campbell, J.Y. and R.J. Shiller, 1988, Stock prices, earnings, and expected dividends. Journal of Finance 43, 661-676.

[25] Campbell, J.Y. and L. Viceira, 1998, Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114, 433-495.

[26] Campbell, J.Y. and M. Yogo, 2004, Efficient tests of stock return predictability. Working paper, Harvard University.

[27] Cavanagh, C.L., G. Elliott and J.H. Stock, 1995, Inference in models with nearly integrated regressors. Econometric Theory 11, 1131-1147.

[28] Christopherson, J.A., W. Ferson and D.A. Glassman, 1998, Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance. Review of Financial Studies 11, 111-142.

[29] Cochrane, J., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations. Journal of Finance 46, 209-235.

[30] Cooper, M., R.C. Gutierrez Jr. and W. Marcum, 2005, On the Predictability of Stock Returns in Real Time. Journal of Business 78, No 2.

[31] Diebold, F.X. and C. Chen, 1996, Testing structural stability with endogenous break point: a size comparison of analytic and bootstrap procedures. Journal of Econometrics 70, 221–241.

[32] Dimson, E. and P. Marsh, 1999, Murphy's law and market anomalies. Journal of Portfolio Management 25(2), 53-69.

[33] Elliott, G. and U. Müller, 2003, Optimally Testing General Breaking Processes in Linear Time Series Models. Working paper, Department of Economics, University of California, San Diego.

[34] Fama, E.F. and G.W. Schwert, 1977, Asset returns and inflation. Journal of Financial Economics 5, 115-146.

[35] Fama, E.F. and K.R. French, 1988, Dividend yields and expected stock returns. Journal of Financial Economics 22, 3-25.

[36] Ferson, W.E. and R.W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, Journal of Finance 51, 425-461.

[37] Ferson, W. and C.R. Harvey, 1991, The variation of economic risk premiums. Journal of Political Economy 99, 385-415.

[38] Franzini, L. and A. Harvey, 1993, Testing for deterministic trend and seasonal components in time series models, Biometrika 70, 673-682.

[39] French, K.R., G.W. Schwert and R.F. Stambaugh, 1987, Expected stock returns and volatility. Journal of Financial Economics 19, 3-29.

[40] Ghysels, E., 1998, On Stable Factor Structures in the Pricing of Risk: Do Time-varying Betas Help or Hurt? Journal of Finance 53, 549-574.

[41] Goetzmann, W.N. and P. Jorion, 1993, Testing the predictive power of dividend yields. Journal of Finance 48, 663-679.

[42] Goyal, A. and I. Welch, 2003, Predicting the equity premium with dividend ratios. Management Science 49, 639-654.

[43] Grullon, G. and R. Michaely, 2002, Dividends, share repurchases, and the substitution hypothesis. Journal of Finance, 57, 1649-1684.

[44] Hansen, B., 2000, Testing for structural change in conditional models. Journal of Econometrics 97, 93-115.

[45] Hjalmarsson, E., 2004, On the predictability of global stock returns. Working Paper, Yale University.

[46] Hodrick, R.J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5, 357-386.

[47] Inoue, A. and L. Killian, 2004, In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? Econometric Reviews 23, 371-402.

[48] Kandel, S. and R. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective. Journal of Finance 51, 385-424.

[49] Keim, D.B. and R.F. Stambaugh, 1986, Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357-390.

[50] Kim, C., J.C. Morley and C. Nelson, 2000, Is there a structural break in the equity premium? Manuscript, University of Washington.

[51] Lamont, O., 1998, Earnings and expected returns. Journal of Finance 53, 1563-1587.

[52] Lettau, M. and S. Ludvigsson, 2001, Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-849.

[53] Lewellen, J., 2004, Predicting returns with financial ratios. Journal of Financial Economics 74, 209-235.

[54] Marquering, W. and M. Verbeek, 2005, The economic value of predicting stock index returns and volatility. Forthcoming in Journal of Financial and Quantitative Analysis.

[55] Nelson, C.R. M.J. Kim, 1993, Predictable stock returns: The role of small sample bias. Journal of Finance 48, 641-661.

[56] Pastor, L. and R. Stambaugh, 2001, The equity premium and structural breaks. Journal of Finance 56, 1207-1245.

[57] Perez-Quiros, G. and A. Timmermann, 2000, Firm size and cyclical variations in stock returns. Journal of Finance 55, 1229-1262.

[58] Perron, P., 1989, The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361-1401.

[59] Pesaran, M.H. and A. Timmermann, 1995, Predictability of stock returns: Robustness and economic significance. Journal of Finance 50, 1201-1228.

[60] Pesaran, M.H. and A. Timmermann, 2002, Market Timing and Return Prediction under Model Instability. Journal of Empirical Finance 9, 495-510.

[61] Pesaran, M.H. and A. Timmermann, 2004, Selection of Estimation Window in the Presence of Breaks. Manuscript, University of Cambridge and UCSD.

[62] Pettenuzzo, D. and A. Timmermann, 2005, Predictability of Stock Returns and Asset Allocation under Structural Breaks. Mimeo UCSD.

[63] Qu, Z. and P. Perron, 2004, Estimating and testing structural changes in multivariate regressions. Working paper, Boston University.

[64] Rapach, D.E. and M.E. Wohar, 2005, Structural breaks and predictive regression models of aggregate U.S. stock returns. Manuscript, University of Nebraska-Omaha.

[65] Rapach, D.E., Wohar, M.E. and J. Rangvid, 2002, Macro variables and international stock return predictability. Manuscript, University of Nebraska-Omaha.

[66] Shively, T., 1988, An analysis of tests for regression coefficient stability. Journal of Econometrics, 39, 367-386.

[67] Sowell, F., 1996, Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework. Econometrica 64, 1085-1107.

[68] Stambaugh, R.F., 1999, Predictive regressions. Journal of Financial Economics 54, 375-421.

[69] Stock, J.H. and M.W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11-30.

[70] Sullivan, R., A. Timmermann, and H. White, 1999, Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. Journal of Finance 54, 1647-1691.

[71] Whitelaw, R.F., 1994, Time variations and covariations in the expectation and volatility of stock market returns. Journal of Finance 49, 515-541.

[72] Valkanov, R., 2003, Long horizon regressions: Theoretical results and applications, Journal of Financial Economics, 68, 201-232.

[73] Yao, Y-C., 1988, Estimating the number of change-points via Schwarz' criterion. Statistics and Probability Letters 6, 181-189.

Table 1. Calibration of parameters for Monte Carlo simulations. The table presents estimation results for the system

$$
\begin{aligned}
y_t &= \alpha + \beta x_{t-1} + \varepsilon_t \\
x_t &= \theta + \varphi x_{t-1} + \upsilon_t ,
\end{aligned}
$$

based on excess returns for the NYSE and the US forecasting variables. These parameters are subsequently used in simulations discussed in section X of the paper. For each equation in the system the table presents the estimated coefficients as well as the estimated standard error for the shock to the equation. The estimated contemporaneous correlation between shocks, denoted ρ, is also presented. Panels A, B, C and D present results for the dividend yield, short interest rate, term spread and default premium, respectively. The sample period is monthly 1952:7-2003:12.

A: Dividend Yield

	Intercept	*Div. Yld.(t-1)*	*Std. Error*	ρ
NYSE(t)	-0.00466	0.31742	0.04131	-0.92175
Div. Yld.(t)	*Intercept*	*Div. Yld.(t-1)*	*Std. Error*	
	0.00032	0.98846	0.00156	

B: Short Interest Rate

	Intercept	*T-bill(t-1)*	*Std. Error*	ρ
NYSE(t)	0.01428	-2.05721	0.04118	-0.01917
T-bill(t)	*Intercept*	*T-bill(t-1)*	*Std. Error*	
	0.00010	0.97534	0.00052	

C: Term Spread

	Intercept	*Trm. Spd.(t-1)*	*Std. Error*	ρ
NYSE(t)	0.00276	6.26948	0.04124	0.08834
Trm. Spd.(t)	*Intercept*	*Trm. Spd.(t-1)*	*Std. Error*	
	0.00003	0.94091	0.00023	

D: Default Premium

	Intercept	*Def. Prm.(t-1)*	*Std. Error*	ρ
NYSE(t)	-0.00010	7.48732	0.04137	0.04674
Def. Prm.(t)	*Intercept*	*Def. Prm.(t-1)*	*Std. Error*	
	0.00002	0.97213	0.00008	

Table 2. Monte Carlo simulations results: size. The simulation experiments assess the size properties of tests for structural breaks in stock return regressions and the adequacy of various model selection procedures under the null of no breaks. In Panel A, the dependent variable is generated as a zero mean Gaussian white noise process. The regressor is generated as a zero mean stationary AR(1) process with persistence parameter Φ. In both cases the shocks are normalized to have unit variance and the innovations to the dependent variable and to the regressor have contemporaneous correlation ρ of either zero or -0.9. In Panel B, the system described by equation (6) of the paper is calibrated to each regressor in turn, based on US data over the period 1952:7 - 2003:12 using the parameters in Table 1. Panel C presents size results for long-horizon regressions for the system described by equation (6). The nominal size of the tests is 10%. In all cases the sample size is 500, and results are reported for 2000 Monte Carlo replications.

A: Size Effects of Persistence and Contemporaneous Correlation

ρ	Φ	*SupF(1)*	*SupF(2)*	*UDMax*	*Hans. SupF*	*J-Stat*	*BIC(0)*	*BIC(1)*	*BIC(2)*	*SEQ(0)*	*SEQ(1)*	*SEQ(2)*
0	0	12.5%	15.2%	14.5%	11.3%	11.1%	99.7%	0.3%	0.0%	87.6%	12.0%	0.5%
0	0.9	14.5%	18.9%	17.9%	10.5%	10.6%	99.8%	0.3%	0.0%	85.5%	13.9%	0.7%
0	0.95	16.6%	21.1%	21.1%	11.3%	10.3%	99.7%	0.3%	0.0%	83.4%	15.5%	1.2%
0	0.98	16.5%	22.5%	20.8%	11.0%	8.2%	99.7%	0.4%	0.0%	83.6%	15.7%	0.8%
-0.9	0	13.3%	16.0%	15.3%	11.1%	11.0%	99.8%	0.2%	0.0%	86.8%	12.7%	0.5%
-0.9	0.9	21.3%	33.6%	31.1%	17.9%	8.3%	99.6%	0.5%	0.0%	78.8%	19.1%	2.2%
-0.9	0.95	27.6%	50.4%	44.4%	22.3%	5.9%	98.8%	1.2%	0.0%	72.4%	22.3%	4.9%
-0.9	0.98	41.2%	68.2%	63.7%	33.0%	6.1%	97.3%	2.7%	0.0%	58.9%	30.5%	8.6%

B: Size Results for Process Tuned to Emprical Estimates

Variable	*SupF(1)*	*SupF(2)*	*UDMax*	*Hans. SupF*	*J-Stat*	*BIC(0)*	*BIC(1)*	*BIC(2)*	*SEQ(0)*	*SEQ(1)*	*SEQ(2)*
Div. Yld.	56.1%	81.6%	81.0%	42.6%	6.4%	96.3%	3.4%	0.3%	43.9%	35.6%	17.7%
Short Rate	45.0%	72.7%	69.2%	32.6%	5.1%	98.0%	2.0%	0.1%	55.1%	32.1%	10.9%
Term Spread	28.6%	50.3%	44.8%	20.1%	6.7%	99.2%	0.8%	0.1%	71.4%	24.2%	4.1%
Default Premium	42.2%	70.6%	67.1%	31.0%	4.7%	98.1%	1.9%	0.1%	57.9%	30.8%	10.1%

C: Size Results for Long-Horizon Return Regressions

Horizon	Φ	*SupF(1)*	*SupF(2)*	*UDMax*	*Hans. SupF*	*J-Stat*	*BIC(0)*	*BIC(1)*	*BIC(2)*	*SEQ(0)*	*SEQ(1)*	*SEQ(2)*
2	0	16.8%	21.8%	22.2%	35.1%	12.9%	94.8%	4.7%	0.6%	83.3%	15.5%	1.3%
2	0.9	23.1%	38.1%	36.9%	52.9%	18.2%	88.1%	9.2%	2.5%	77.0%	20.7%	2.3%
2	0.95	25.1%	41.2%	39.2%	53.0%	14.5%	86.9%	10.2%	2.8%	75.0%	21.8%	3.0%
2	0.98	26.9%	44.6%	45.0%	55.7%	12.9%	85.5%	11.2%	3.1%	73.2%	23.6%	3.0%
4	0	21.2%	31.5%	36.0%	70.0%	17.7%	63.3%	20.1%	12.9%	78.8%	18.9%	2.2%
4	0.9	32.0%	55.6%	58.4%	89.0%	24.2%	34.6%	22.3%	26.8%	68.0%	26.5%	4.9%
4	0.95	34.7%	59.0%	61.9%	90.9%	20.2%	32.1%	22.7%	28.5%	65.3%	26.9%	6.9%
4	0.98	38.6%	63.6%	68.5%	91.7%	19.1%	28.5%	22.1%	29.6%	61.5%	29.7%	7.7%
6	0	24.9%	41.6%	48.1%	85.9%	17.8%	33.8%	21.0%	28.7%	75.1%	22.0%	2.6%
6	0.9	37.9%	64.6%	70.7%	97.3%	25.2%	10.6%	14.5%	30.4%	62.2%	29.9%	7.0%
6	0.95	40.5%	66.4%	74.4%	97.5%	22.6%	9.0%	13.3%	31.0%	59.6%	29.7%	8.9%
6	0.98	45.8%	74.8%	82.0%	97.7%	21.4%	6.8%	12.0%	29.0%	54.3%	31.8%	11.4%

Table 3. Monte Carlo simulation results: size-adjusted power. This table reports the results of 2,000 simulation experiments with a single break. The dependent variable is generated as a function of a single driving variable, which itself follows an AR(1) process with persistence Φ. The specific data generating process is described in the equations below. In all cases the sample size is 500 observations and the break occurs at the midway point of the sample. The second column of the tables indicates the break size as a percentage of the average coefficient, which is tuned to the corresponding R^2-value. The table reports the frequency of cases that various break tests reject at the 10% significance level and reports model selection outcomes using both the BIC model selection criteria and the sequential method of Bai and Perron (1998) at the 10% level of significance. All tests are conducted using empirical critical values based on 5,000 simulations under the null of no break so the results indicate the size-adjusted power of the tests.

$$y_t = \left(\beta^* - \frac{\delta}{2}\right)x_{t-1} + \varepsilon_t;\ t = 1,\ldots,250$$

$$y_t = \left(\beta^* + \frac{\delta}{2}\right)x_{t-1} + \varepsilon_t;\ t = 251,\ldots,500$$

$$x_t = \varphi x_{t-1} + \upsilon_t;\ \upsilon_t \sim N(0,1);\ \varepsilon_t \sim N(0,1);\ \mathrm{E}[\varepsilon_t \upsilon_t] = 0.$$

A: $R^2 = 5\%$

Φ	$\frac{\delta}{\beta^*}\times 100$	*SupF(1)*	*UDMax*	*Hans. SupF*	*J-Stat*	*BIC(0)*	*BIC(1)*	*BIC(2)*	*SEQ(0)*	*SEQ(1)*	*SEQ(2)*
0	10	9.7%	10.7%	9.6%	11.0%	99.8%	0.2%	0.0%	87.9%	11.4%	0.6%
0	25	13.8%	13.6%	12.8%	14.3%	99.5%	0.5%	0.0%	84.6%	14.6%	0.8%
0	50	20.6%	20.5%	21.4%	20.8%	98.8%	1.2%	0.0%	76.8%	21.9%	1.4%
0	100	57.1%	56.1%	60.9%	56.8%	89.3%	10.7%	0.1%	41.0%	55.2%	3.9%
0.95	10	10.1%	11.6%	10.1%	9.7%	99.7%	0.3%	0.0%	84.2%	14.3%	1.4%
0.95	25	12.1%	12.3%	12.2%	11.9%	99.7%	0.4%	0.0%	81.6%	17.0%	1.5%
0.95	50	17.5%	18.2%	18.4%	19.3%	98.9%	1.2%	0.0%	73.6%	24.6%	1.9%
0.95	100	50.2%	50.1%	53.3%	51.3%	90.3%	9.7%	0.0%	44.8%	50.8%	4.4%
0.98	10	10.6%	11.4%	11.3%	11.4%	99.5%	0.6%	0.0%	83.6%	15.2%	1.2%
0.98	25	11.5%	11.9%	12.5%	12.4%	99.8%	0.2%	0.0%	81.6%	17.0%	1.5%
0.98	50	15.7%	15.7%	17.3%	17.1%	98.8%	1.2%	0.0%	74.2%	24.0%	1.9%
0.98	100	41.8%	39.5%	45.5%	45.4%	91.3%	8.8%	0.0%	48.1%	47.9%	4.1%

B: $R^2 = 10\%$

Φ	$\frac{\delta}{\beta^*}\times 100$	*SupF(1)*	*UDMax*	*Hans. SupF*	*J-Stat*	*BIC(0)*	*BIC(1)*	*BIC(2)*	*SEQ(0)*	*SEQ(1)*	*SEQ(2)*
0	10	10.4%	10.3%	10.8%	10.6%	99.3%	0.7%	0.0%	87.3%	11.8%	1.0%
0	25	14.8%	13.8%	15.6%	15.2%	99.5%	0.6%	0.0%	82.6%	16.8%	0.7%
0	50	33.8%	31.4%	35.5%	34.3%	95.7%	4.4%	0.0%	60.1%	37.1%	2.8%
0	100	87.3%	84.7%	89.0%	85.0%	55.9%	44.1%	0.1%	12.5%	82.6%	4.9%
0.95	10	11.2%	11.5%	11.4%	11.1%	99.5%	0.6%	0.0%	83.3%	15.5%	1.3%
0.95	25	16.3%	16.0%	16.5%	17.1%	99.5%	0.6%	0.0%	77.3%	21.2%	1.6%
0.95	50	32.9%	30.4%	33.9%	33.0%	97.0%	3.0%	0.1%	64.6%	32.4%	3.0%
0.95	100	79.5%	66.0%	82.9%	80.1%	62.8%	37.1%	0.1%	16.5%	76.2%	7.2%
0.98	10	11.8%	11.9%	11.6%	10.3%	99.7%	0.4%	0.0%	82.8%	15.8%	1.4%
0.98	25	14.4%	14.5%	14.6%	14.3%	99.2%	0.9%	0.0%	77.0%	21.6%	1.4%
0.98	50	26.7%	24.9%	28.9%	28.3%	96.3%	3.7%	0.1%	63.6%	33.7%	2.7%
0.98	100	70.3%	67.1%	72.7%	70.0%	71.1%	28.8%	0.2%	23.7%	69.7%	6.4%

Table 4. Full-sample regression results with no breaks. The dependent variable is local country returns in excess of the local Treasury bill rate. For each country the forecasting variables are local country measures lagged one-month with the exception of the default premium, which is the US default premium for all countries. Standard errors to the right of each coefficient are computed using standard errors corrected for heteroskedasticity and serial correlation. R^2-statistics are also provided for each regression. In the top portion of each panel, results are reported over the sample period 1952:6 - 2003:12 for the NYSE, NYSE/AMEX/NASDAQ, S&P 500 and UK (FTA) portfolios. In the bottom portion of each panel results are reported over the shorter sample period 1970:1 - 2003:12 for 10 OECD portfolios.

A: Univariate Regression Models

	Dividend Yield			*Short Interest Rate*			*Term Spread*			*US Default Premium*		
	R^2	*Beta*	*SE*	R^2	*Beta*	*SE*	R^2	*Beta*	*SE*	R^2	*Beta*	*SE*
NYSE (52:7 -)	0.71%	0.33	0.16	1.00%	-1.77	0.68	1.15%	4.62	1.90	0.36%	7.18	5.96
NYSE/AMEX/NASDAQ	0.71%	0.34	0.18	1.06%	-1.90	0.71	1.22%	4.95	2.03	0.37%	7.58	6.18
S&P 500	0.62%	0.32	0.17	1.09%	-1.88	0.68	1.11%	4.64	1.96	0.20%	5.45	5.81
UK (FTA)	2.79%	0.70	0.35	0.14%	-0.70	0.86	0.39%	1.93	1.64	0.68%	12.77	6.55
BEL (70:1 -)	0.02%	0.04	0.17	0.05%	-0.43	1.12	0.08%	1.67	2.71	0.30%	7.55	7.66
CAN	0.05%	0.09	0.27	0.97%	-1.59	0.95	0.54%	2.78	2.03	0.15%	5.36	9.96
FRN	0.07%	0.08	0.17	0.27%	-1.07	1.09	0.69%	4.05	2.62	0.03%	3.09	8.48
GER	0.14%	0.20	0.26	0.42%	-2.05	1.62	0.15%	2.03	2.40	0.29%	8.33	6.87
ITL	0.23%	-0.31	0.32	0.06%	-0.43	0.95	0.15%	2.03	2.37	0.00%	-0.48	12.50
JPN	0.63%	0.55	0.41	0.00%	0.02	1.60	0.14%	2.32	3.34	1.18%	16.04	6.32
NTH	0.21%	0.13	0.16	0.44%	-1.62	1.06	0.31%	2.02	1.47	0.41%	9.30	7.34
SWE	0.20%	0.20	0.22	0.19%	0.94	1.35	0.00%	0.22	2.17	1.06%	18.71	12.19
UK	3.03%	0.76	0.47	0.00%	0.11	1.05	0.37%	1.73	1.75	1.59%	21.25	8.21
US (S&P 500)	0.33%	0.23	0.20	0.58%	-1.50	0.86	1.27%	4.81	2.24	0.65%	10.41	7.27

B: Multiple Regression Model

	R^2	*Div. Yld.*	*SE*	*Int. Rate*	*SE*	*Term Spd.*	*SE*	*Def. Prm.*	*SE*
NYSE	4.26%	0.46	0.18	-4.05	1.07	0.40	1.96	17.33	7.90
NYSE/AMEX/NASDAQ	4.43%	0.48	0.20	-4.31	1.12	0.48	2.09	18.38	8.37
S&P 500	3.91%	0.47	0.19	-3.88	1.07	0.85	2.03	14.42	7.97
UK (FTA)	5.14%	0.84	0.32	-3.94	1.52	-1.97	1.77	26.14	9.69
BEL	0.91%	0.16	0.26	-2.82	2.24	-1.21	3.68	14.22	10.05
CAN	3.07%	0.44	0.32	-4.88	1.72	-3.39	3.33	18.07	10.46
FRN	1.15%	0.35	0.36	-3.11	2.88	-0.03	4.84	6.47	13.17
GER	1.69%	0.60	0.37	-5.66	2.83	-2.28	3.31	7.52	9.21
ITL	0.37%	-0.30	0.32	-0.20	1.19	1.66	3.03	1.93	12.81
JPN	2.17%	0.67	0.46	-2.37	1.97	-2.02	3.55	20.82	8.45
NTH	2.29%	0.53	0.27	-7.43	3.26	-6.24	3.94	9.16	9.99
SWE	1.29%	-0.05	0.28	0.62	1.95	2.95	3.15	20.88	16.24
UK	9.08%	2.44	0.98	-12.04	4.10	-9.15	3.34	20.58	11.34
US	3.90%	0.83	0.37	-6.02	1.97	-2.08	3.23	14.14	8.68

Table 5. Tests for breaks and model selection. The table presents test statistics for various hypothesis tests regarding the occurrence of breaks in the regression model for U.S. and international excess stock returns. The test statistics reported include the SupF(1) test, the UDMax test of Bai and Perron (1998), the SupF(1) test with bootstrap critical values corrected for heteroskedasticity as suggested by Hansen (2002) and the J-test of Elliott and Muller (2003). The Bai and Perron and Hansen SupF(1) test statistics differ generally since the former is computed using a HAC covariance estimator. See the appendix for further details regarding the various tests. The trimming percentage for the *SupF* tests is set at 15% of the sample size. The table also reports the number of breaks selected based on the sequential method suggested by Bai and Perron (1998, 2003) at both the 10% and 5% significance levels.

		A: Sample period 1952:7 - 2003:12				B: Sample period 1970:1 - 2003:12									
		NYSE	NAN	S&P 500	UK (FTA)	BEL	CAN	FRN	GER	ITL	JPN	NTH	SWE	UK	US
Div. Yield	SupF(1)	25.31 [a]	22.74 [a]	25.28 [a]	4.85	10.56 [b]	8.45	11.69 [a]	8.29	6.15	11.06 [b]	9.04	10.58 [b]	14.42 [a]	15.98 [a]
	UDMax	25.31 [a]	22.74 [a]	25.28 [a]	23.14 [a]	10.56 [b]	8.45	11.69 [b]	8.35	9.16	11.06 [b]	14.67 [a]	10.58 [b]	14.42 [a]	15.98 [a]
	Hansen SupF	10.91 [b]	11.80 [a]	15.04 [a]	19.72 [b]	14.13 [a]	6.51	7.77	5.06	5.44	15.63 [a]	8.19	9.25	7.82	11.29 [b]
	J-Stat	-9.53	-8.97	-13.73 [b]	-8.31	-11.08	-5.21	-8.59	-9.95	-9.18	-10.54	-12.58	-11.18	-9.92	-11.55
	# Breaks (10%)	1	1	1	0	1	0	1	0	0	1	0	1	1	1
	# Breaks (5%)	1	1	1	0	0	0	1	0	0	0	0	0	1	1
T-bill rate	SupF(1)	13.66 [a]	13.01 [a]	14.92 [a]	17.95 [a]	12.65 [a]	4.87	12.14 [a]	8.61	8.06	9.25	9.51	17.47 [a]	2.88	9.07
	UDMax	14.52 [a]	14.08 [a]	14.92 [a]	21.94 [a]	12.65 [a]	9.92	12.14 [a]	10.50 [b]	12.45 [b]	9.25	9.51	17.47 [a]	6.97	9.52
	Hansen SupF	19.40 [a]	18.26 [a]	12.48 [a]	30.59 [a]	8.33	6.02	9.33	11.70 [a]	7.80	16.38 [b]	10.77 [b]	15.82 [a]	11.42	13.09
	J-Stat	-8.48	-8.15	-10.39	-11.45	-11.92	-6.50	-10.84	-8.96	-11.05	-12.31	-11.10	-11.63	-7.92	-9.62
	# Breaks (10%)	2	2	2	1	1	0	1	0	0	0	0	1	0	0
	# Breaks (5%)	1	1	2	1	1	0	1	0	0	0	0	1	0	0
Term Spd.	SupF(1)	13.93 [a]	12.70 [a]	16.59 [a]	8.60	7.64	3.00	6.43	8.38	6.99	10.76 [b]	9.38	6.38	9.63	9.80
	UDMax	13.93 [a]	12.70 [a]	16.59 [a]	8.60	10.33	5.15	6.46	8.38	8.58	10.76 [b]	9.38	6.53	9.63	9.80
	Hansen SupF	14.75 [a]	13.18 [a]	15.04 [a]	4.94	5.09	3.64	5.56	8.82	6.86	12.54 [b]	8.17	11.22 [b]	23.62 [b]	14.37 [b]
	J-Stat	-13.06 [b]	-11.74	-13.72 [b]	-6.33	-10.25	-5.63	-11.13	-8.71	-12.29	-11.94	-13.48 [b]	-13.31 [b]	-6.79	-15.15 [a]
	# Breaks (10%)	1	1	1	0	0	0	0	0	0	1	0	0	0	0
	# Breaks (5%)	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Def. Prm.	SupF(1)	6.38	6.46	6.21	9.80	5.44	6.62	9.96 [b]	7.57	5.49	4.18	8.22	10.47 [b]	12.31 [a]	16.91 [a]
	UDMax	11.52 [b]	11.58 [b]	11.35 [b]	10.46 [b]	9.65	8.69	9.96	7.57	8.14	5.80	8.22	10.47 [b]	12.31 [a]	16.91 [a]
	Hansen SupF	5.79	5.86	5.27	-10.03	7.21	9.27	11.07	9.94	6.53	4.96	12.45	11.87	41.72 [a]	11.12 [b]
	J-Stat	-12.90 [b]	-12.72	-13.58 [b]	-17.74 [a]	-12.63	-8.02	-11.00	-13.01 [b]	-10.96	-10.40	-17.46 [a]	-13.18 [b]	-18.64 [a]	-9.89
	# Breaks (10%)	0	0	0	0	0	0	1	0	0	0	0	1	1	1
	# Breaks (5%)	0	0	0	0	0	0	0	0	0	0	0	0	1	1
All Reg.	SupF(1)	22.30 [a]	20.03 [a]	24.31 [a]	44.11 [a]	27.09 [a]	29.39 [a]	26.89 [a]	24.49 [a]	11.46	22.55 [a]	41.73 [a]	20.63 [a]	20.40	28.57 [a]
	UDMax	27.13 [a]	27.95 [a]	27.12 [a]	46.40 [a]	27.09 [a]	29.39 [a]	26.89 [a]	24.49 [a]	18.17 [b]	22.55 [a]	41.73 [a]	29.09 [a]	22.04	32.85 [a]
	Hansen SupF	31.64 [a]	29.11 [a]	35.11 [a]	97.53 [a]	33.31 [a]	30.03 [a]	23.14 [a]	23.87 [a]	14.31	23.88 [b]	35.93 [a]	27.66 [a]	37.82 [a]	33.61 [a]
	J-Stat	-34.12 [a]	-33.67 [a]	-33.544 [a]	-39.09 [a]	-38.24 [a]	-30.54 [b]	-30.79 [a]	-37.13 [a]	-22.17	-30.69 [a]	-50.25 [a]	-26.53	-38.00 [a]	-38.01 [a]
	# Breaks (10%)	2	1	2	2	1	1	1	2	0	1	1	2	1	2
	# Breaks (5%)	1	1	1	2	1	1	1	2	0	1	1	2	1	2

Note: Superscripts [a] and [b] represent statistical significance at the 5% and 10% levels, respectively.

Table 6. Univariate regression models with breaks. This table presents the estimated coefficients and associated heteroskedasticity and autocorrelation consistent standard errors during each subinterval identified using the sequential breakpoint method of Bai and Perron (1998) for regressions of US and international excess stock returns on each univariate predictor variable in turn. We only display results for the portfolios where at least one breakpoint was identified. The minimum window length was set to 15% of the total sample size in all cases. For each breakpoint identified, the squared brackets present the estimated breakdate as well as the lower and upper bounds of a 90% confidence interval for this estimate.

A: Dividend Yield

	Subinterval 1			*Breakpoint*	*Subinterval 2*		
Portfolio	R^2	*Beta*	*S.E.*		R^2	*Beta*	*S.E.*
NYSE (52:7 -)	1.5%	0.60[a]	*0.22*	[92:8 94:11 95:4]	6.7%	2.88[a]	*0.74*
NYSE/AMEX/NASDAQ	1.5%	0.61[a]	*0.23*	[92:8 94:12 95:4]	6.5%	3.38[a]	*0.98*
S&P 500	1.2%	0.55[a]	*0.21*	[92:10 94:12 95:3]	7.5%	3.55[a]	*0.90*
UK (FTA)	2.8%	0.70	*0.35*				
BEL (70:1 -)	1.2%	-0.30	*0.19*	[70:1 81:10 82:6]	2.9%	0.71[a]	*0.30*
CAN	0.0%	0.09	*0.27*				
FRN	1.6%	0.56[b]	*0.33*	[78:8 82:11 86:8]	1.4%	0.60[a]	*0.25*
GER	0.1%	0.20	*0.26*				
ITL	0.2%	-0.31	*0.32*				
JPN	0.1%	-0.18	*0.42*	[82:9 89:11 90:12]	3.6%	4.49[a]	*1.85*
NTH	0.2%	0.13	*0.16*				
SWE	0.2%	0.23	*0.32*	[70:1 80:8 81:10]	2.2%	0.96[a]	*0.33*
UK	4.2%	1.09	*0.78*	[91:4 92:7 99:2]	8.4%	1.71[a]	*0.43*
US	0.6%	0.40	*0.28*	[96:6 96:7 96:8]	11.0%	7.46[a]	*1.93*

B: Short Interest Rate

	Subinterval 1			*Breakpoint*	*Subinterval 2*			*Breakpoint*	*Subinterval 3*		
Portfolio	R^2	*Beta*	*S.E.*		R^2	*Beta*	*S.E.*		R^2	*Beta*	*S.E.*
NYSE (52:7 -)	11.6%	-16.79[a]	*3.61*	[61:3 62:9 64:11]	13.8%	-11.76[a]	*2.97*	[73:3 74:8 81:7]	0.3%	-0.94	*0.98*
NYSE/AMEX/NASDAQ	11.6%	-16.79[a]	*3.60*	[61:1 62:9 64:10]	13.9%	-12.05[a]	*3.06*	[73:2 74:8 81:10]	0.3%	-1.11	*1.06*
S&P 500	12.1%	-17.82[a]	*4.24*	[61:7 62:9 64:12]	14.8%	-11.78[a]	*2.88*	[73:4 74:8 81:1]	0.3%	-0.97	*0.98*
UK (FTA)	12.5%	-9.12[a]	*2.19*	[73:1 74:10 82:8]	0.4%	1.25	*1.08*				
BEL (70:1 -)	5.2%	-5.67[a]	*1.74*	[74:10 81:8 83:12]	0.2%	0.97	*1.27*				
CAN	1.0%	-1.59[b]	*0.95*								
FRN	0.7%	-1.97	*1.25*	[98:4 98:5 98:7]	9.5%	-30.43[a]	*8.91*				
GER	0.4%	-2.05	*1.62*								
ITL	0.1%	-0.43	*0.95*								
JPN	0.0%	0.02	*1.60*								
NTH	0.4%	-1.62	*1.06*								
SWE	0.1%	0.65	*1.34*	[98:5 98:6 98:7]	14.2%	-69.00[a]	*16.55*				
UK	0.0%	0.11	*1.05*								
US	0.6%	-1.50[b]	*0.86*								

Note: The superscripts [a] and [b] indicate statistical significance at the 5% and 10% levels.

Table 6. Univariate regression models with breaks (continued).

C: Term Spread

	Subinterval 1			Breakpoint	Subinterval 2		
Portfolio	R^2	*Beta*	*S.E.*		R^2	*Beta*	*S.E.*
NYSE (52:7 -)	8.1%	18.49 [a]	*4.22*	[72:10 75:5 86:3]	0.2%	1.75	*2.31*
NYSE/AMEX/NASDAQ	8.0%	18.61 [a]	*4.36*	[72:4 75:5 87:1]	0.3%	2.19	*2.52*
S&P 500	8.7%	18.91 [a]	*3.97*	[72:10 75:5 84:1]	0.2%	1.77	*2.36*
UK (FTA)	0.4%	1.93	*1.64*				
BEL (70:1 -)	0.1%	1.67	*2.71*				
CAN	0.5%	2.78	*2.03*				
FRN	0.7%	4.05	*2.62*				
GER	0.1%	2.03	*2.40*				
ITL	0.2%	2.03	*2.37*				
JPN	1.5%	-6.22 [a]	*3.07*	[83:8 89:11 94:12]	0.6%	8.12	*7.72*
NTH	0.4%	2.01	*1.87*				
SWE	0.0%	0.22	*2.17*				
UK	0.4%	1.73	*1.75*				
US	1.3%	4.81 [a]	*2.24*				

D: US Default Premium

	Subinterval 1			Breakpoint	Subinterval 2		
Portfolio	R^2	*Beta*	*S.E.*		R^2	*Beta*	*S.E.*
NYSE (52:7 -)	0.4%	7.18	*5.96*				
NYSE/AMEX/NASDAQ	0.4%	7.58	*6.18*				
S&P 500	0.7%	12.77 [b]	*6.55*				
UK (FTA)	0.2%	5.45	*5.81*				
BEL (70:1 -)	0.3%	7.55	*7.66*				
CAN	0.1%	5.36	*9.96*				
FRN	8.4%	67.55	*25.57*	[74:12 77:3 82:4]	0.1%	-4.52	*8.66*
GER	0.3%	8.33	*6.87*				
ITL	0.0%	-0.48	*12.50*				
JPN	1.2%	16.04 [z]	*6.32*				
NTH	0.4%	9.30	*7.34*				
SWE	5.4%	39.38 [z]	*12.63*	[88:9 92:9 96:10]	1.3%	-42.47	*37.53*
UK	28.2%	226.81 [z]	*76.88*	74:11 75:1 79:10]	0.3%	8.04	*5.55*
US	18.6%	92.27 [z]	*20.43*	[74:1 75:5 77:11]	0.1%	3.53	*6.622*

Note: The superscripts [a] and [b] indicate statistical significance at the 5% and 10% levels.

Table 7. Multivariate regression models with breaks. This table presents the estimated coefficients and standard errors for the multivariate return model including the dividend yield, short interest rate, term spread and default premium regressors during each subinterval identified using the sequential breakpoint method of Bai and Perron (1998). Standard errors are heteroskedasticity and autocorrelation consistent. The minimum window length was set equal to 15% of the total sample size. For each breakpoint identified, the squared brackets present the estimated breakdate as well as the lower and upper bounds of a 90% confidence interval for this estimate.

	Subinterval 1				*Subinterval 2*				*Subinterval 3*		
Portfolio	*R^2*	*Beta*	*S.E.*	*Breakpoint*	*R^2*	*Beta*	*S.E.*	*Breakpoint*	*R^2*	*Beta*	*S.E.*
NYSE	9.5%			[87:6 87:7 87:8]	20.5%			[94:12 95:3 96:10]	9.5%		
Div. Yield		0.60 [a]	*0.25*			9.04 [a]	*3.16*			3.69 [a]	*1.07*
T-Bill Rate		-5.81 [a]	*1.22*			-26.52 [a]	*12.26*			-12.22	*7.79*
Spread		1.25 [a]	*2.68*			-17.54	*13.84*			-16.03 [b]	*9.26*
Def. Prem.		29.03 [b]	*9.24*			-105.50	*64.74*			-36.63	*38.78*
NYSE/AMEX/NASDAQ	9.6%			[87:6 87:7 87:9]	7.3%						
Div. Yield		0.62 [a]	*0.26*			3.04 [a]	*1.13*				
T-Bill Rate		-6.06 [a]	*1.29*			-19.02 [a]	*8.18*				
Spread		0.81	*2.73*			-23.27 [a]	*10.75*				
Def. Prem.		30.60 [a]	*9.62*			-40.15 [b]	*22.31*				
S&P 500	9.5%			[87:6 87:7 87:8]	21.3%			[94:12 95:3 96:9]	9.9%		
Div. Yield		0.56 [a]	*0.23*			9.75 [a]	*3.52*			4.16 [a]	*1.19*
T-Bill Rate		-5.71 [a]	*1.16*			-28.06 [a]	*13.68*			-13.79	*9.06*
Spread		1.64	*2.57*			-18.95	*15.15*			-16.15	*11.08*
Def. Prem.		27.29 [a]	*8.67*			-123.11 [b]	*69.22*			-49.16	*44.70*
UK (FTA)	7.7%			[65:10 67:1 67:6]	29.1%			[73:5 74:11 75:1]	19.1%		
Div. Yield		0.63	*0.43*			0.73 [a]	*0.33*			4.31 [a]	*1.10*
T-Bill Rate		-8.44 [a]	*4.59*			-20.61 [a]	*2.82*			-16.80 [a]	*4.93*
Spread		-0.57	*9.74*			-24.18 [a]	*5.27*			-13.87 [a]	*4.38*
Def. Prem.		15.12	*21.34*			95.15 [a]	*36.24*			3.18	*9.00*
BEL	9.8%			[79:10 81:10 81:12]	7.3%						
Div. Yield		1.07 [b]	*0.57*			2.63 [a]	*0.77*				
T-Bill Rate		-15.23 [a]	*5.15*			-10.59 [a]	*4.09*				
Spread		-1.52	*13.05*			-10.72 [b]	*5.62*				
Def. Prem.		15.11	*14.95*			-16.40	*14.15*				
CAN	14.9%			[77:4 78:1 78:9]	8.2%						
Div. Yield		1.67	*1.27*			2.39 [a]	*0.60*				
T-Bill Rate		-15.19	*10.28*			-11.36 [a]	*2.75*				
Spread		-3.64	*16.17*			-9.15 [a]	*3.78*				
Def. Prem.		35.14	*26.23*			9.80	*9.72*				
FRN	18.2%			[77:4 78:1 79:6]	2.4%						
Div. Yield		3.09 [a]	*0.82*			1.12 [a]	*0.54*				
T-Bill Rate		-33.94 [a]	*8.64*			-5.23	*3.73*				
Spread		-30.76 [a]	*8.69*			-6.39	*6.14*				
Def. Prem.		30.04	*27.32*			-12.79	*14.23*				

Note: The superscripts [a] and [b] indicate statistical significance at the 5% and 10% levels.

Table 7. Multiple regression model with breaks (continued).

	Subinterval 1				*Subinterval 2*				*Subinterval 3*		
Portfolio	R^2	*Beta*	*S.E.*	*Breakpoint*	R^2	*Beta*	*S.E.*	*Breakpoint*	R^2	*Beta*	*S.E.*
GER	11.3%			[81:6 82:12 83:2]	9.2%			[92:11 94:7 94:9]	11.9%		
Div. Yield		1.00 ^{b}	*0.51*			2.72 ^{a}	*1.02*			1.73	*1.78*
T-Bill Rate		-3.70	*2.66*			-16.14 ^{a}	*6.75*			-37.77	*11.16*
Spread		8.50 ^{a}	*4.23*			-28.65 ^{a}	*11.63*			4.29	*16.09*
Def. Prem.		25.03 ^{a}	*8.95*			53.53 ^{a}	*21.83*			-127.87 ^{a}	*47.43*
ITL	0.4%										
Div. Yield		-0.30	*0.32*								
T-Bill Rate		-0.20	*1.19*								
Spread		1.66	*3.03*								
Def. Prem.		1.93	*12.81*								
JPN	5.6%			[96:4 96:5 96:6]	12.7%						
Div. Yield		1.17 ^{a}	*0.49*			17.74 ^{a}	*5.77*				
T-Bill Rate		-10.22 ^{a}	*3.46*			-119.59 ^{a}	*50.97*				
Spread		-7.32 ^{a}	*3.62*			70.70 ^{a}	*26.38*				
Def. Prem.		30.88 ^{a}	*8.74*			-79.72 ^{b}	*46.44*				
NTH	22.4%			[79:6 81:8 82:1]	5.4%						
Div. Yield		1.51 ^{a}	*0.40*			1.71 ^{a}	*0.55*				
T-Bill Rate		-28.11 ^{a}	*5.37*			-5.31	*4.23*				
Spread		-27.82 ^{a}	*6.26*			3.43	*5.57*				
Def. Prem.		52.88 ^{a}	*12.59*			-43.59 ^{a}	*16.58*				
SWE	25.6%			[78:9 78:12 79:7]	7.0%			[90:5 92:9 93:2]	9.8%		
Div. Yield		6.14 ^{a}	*1.04*			-0.07	*0.33*			4.21 ^{a}	*1.67*
T-Bill Rate		-35.39 ^{a}	*6.80*			1.52	*7.54*			2.93	*4.07*
Spread		-20.32 ^{a}	*6.61*			0.04	*9.94*			6.19	*7.60*
Def. Prem.		56.50 ^{a}	*17.21*			44.24 ^{b}	*24.34*			-106.57 ^{a}	*53.46*
UK	29.1%			[76:4 76:9 79:7]	6.7%						
Div. Yield		2.75	*1.76*			2.65 ^{a}	*0.94*				
T-Bill Rate		-26.93 ^{a}	*9.58*			-9.18 ^{a}	*4.34*				
Spread		-29.21 ^{a}	*12.65*			-7.09 ^{b}	*3.98*				
Def. Prem.		147.72 ^{a}	*46.30*			-0.85	*7.97*				
USA	14.2%			[87:6 87:7 87:10]	26.9%			[94:2 94:5 96:6]	9.8%		
Div. Yield		1.416 ^{a}	0.489			11.539 ^{a}	3.521			3.789 ^{a}	*1.05*
T-Bill Rate		-8.499 ^{a}	2.077			-38.469 ^{a}	14.974			-16.851 ^{a}	*7.75*
Spread		-0.46	3.267			-27.991 ^{b}	16.292			-20.542 ^{a}	*8.92*
Def. Prem.		33.009 ^{a}	9.634			-90.25	62.684			-51.941	*43.43*

Note: The superscripts ^{a} and ^{b} indicate statistical significance at the 5% and 10% levels.

Table 8. Robustness of breaks to choice of trimming percentage. The table assesses the robustness of the results of the sequential breaks model selection procedure of Bai and Perron (1998) to different choices of the trimming percentage. The table displays the estimated breakpoints for regressions of excess returns on various portfolios on a constant and lagged values of the (local country) dividend yield, Treasury bill rate, term spread and US default premium. An "X" is placed in each column where the midpoint of an estimated break falls. The bin labeled "71 to 75" contains an "X" if an estimated break date occurs between 1971:1 and 1975:12. An "XX" indicates two estimated breaks within the period. Other bin labels should be interpreted similarly.

Portfolio	Trimming Percentage 10%										Trimming Percentage 15%										Trimming percentage 20%									
	52 to 55	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*95 to 99*	*52 to 55*	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 99*	*52 to 55*	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 99*
NYSE (52:7 -)								X	X									X	X				X					X		
NYSE/AMEX/NASDAQ								X	X									X					X					X		
S&P 500								X	X									X	X									X		
UK (FTA)		X		X	X								X		X								X		X					

Portfolio	Trimming Percentage 10%						Trimming Percentage 15%						Trimming Percentage 20%					
	71 to 75	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*
BEL (70:1 -)	X		X						X						X			
CAN		X				XX		X						X				
FRN	X	X						X										
GER						X			X		X				X		X	
ITL																		
JPN						X						X						X
NTH			X						X						X			
SWE		X	X		X			X			X			X			X	
UK	X		X					X						X				
US										X	X					X	X	

Table 9. Long-horizon returns regressions with overlapping observations. The table indicates the estimated breakpoints for multiple regression models for horizons of two, four and six months. Results are reported for the model selected by the sequential breaks model selection procedure of Bai and Perron (1998) at the 10% significance level. In all cases the trimming percentage is set to 15%. An "X" is placed in each column where the midpoint of an estimated break falls. The bin labeled "71 to 75" contains an "X" if an estimated break date occurs between 1971:1 and 1975:12. An "XX" indicates two estimated breaks within the period. Other bin labels should be interpreted similarly. The SupF-type tests used in the model selection procedure are corrected for serial correlation and heteroskedasticity.

	Two-Month Horizon										**Four-Month Horizon**										**Six-Month Horizon**									
Portfolio	*52 to 55*	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*95 to 99*	*52 to 55*	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 99*	*52 to 55*	*56 to 59*	*60 to 65*	*66 to 70*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 99*
NYSE (52:7 -)			X			X		X					X			X			X				X			X			X	
NAN			X			X			X				X			X			X				X			X			X	
S&P 500			X			X		X					X				X		X				X			X			X	
UK (FTA)				X	X		X							X	X		X							X	X					

	Two-Month Horizon						**Four-Month Horizon**						**Six-Month Horizon**					
Portfolio	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*	*71 to 75*	*76 to 80*	*81 to 85*	*86 to 90*	*91 to 95*	*96 to 00*
BEL (70:1 -)			X		X		X		X		X		X		X	X	X	
CAN		X						X	X		X	X		X				X
FRN		X		X		X		X	X	X		X		X	X		X	
GER		X			X			X		X	X	X		X		X		X
ITL	X		XX			X	X		XX			X	X	X	X		X	X
JPN	X			X	X		X		X	X		X	X		X	X		X
NTH	X	X		X			X		X		X		X		X	X	X	
SWE		X	X		X	X		X	X		X	X		X	X		X	X
UK		X	X					X	X	X				X	X	X		
US					X			X			X		X		X	X	X	X

Figure 1. **Dividend yield versus total payout yield**. The top panel of the figure plots the monthly dividend yield series for the US over the sample period 1969:12 – 2003:12 and indicates the estimated breakpoint in a regression of excess returns for the S&P 500 on the lagged dividend yield. The bottom panel plots the monthly total payout yield constructed as described in Section 7 of the paper, along with the estimated breakpoint in a regression of excess returns for the S&P 500 on the lagged total payout yield. In both panels, the estimated breakpoint is indicated by a solid line, while dashed lines indicate a 90% confidence interval for the breakpoint. The vertical axis is measured in percentage points.