Regime Switches in Interest Rates

Andrew ANG
Columbia Business School and National Bureau of Economic Research, 3022 Broadway 805 Uris, New York, NY 10027 (*aa610@columbia.edu*)

Geert BEKAERT
Columbia University and National Bureau of Economic Research, 3022 Broadway 802 Uris, New York, NY 10027 (*gb241@columbia.edu*)

We examine the econometric performance of regime-switching models for interest rate data from the United States, Germany, and the United Kingdom. Regime-switching models forecast better out-of-sample than single-regime models, including an affine multifactor model, but do not always match moments very well. Regime-switching models incorporating international short-rate and term spread information forecast better, match sample moments better, and classify regimes better than univariate regime-switching models. Finally, the regimes in interest rates correspond reasonably well with business cycles, at least in the United States.

KEY WORDS: Business cycle; Forecasting; Interest rate; Regime-switching model; Term structure.

1. INTRODUCTION

The stochastic behavior of interest rates varies over time. For example, the behavior of interest rates in the 1979–1982 period in the United States or around the German reunification period seems to indicate a structural break in the time series. More generally, changes in business cycle conditions and monetary policy may affect real rates and expected inflation and cause interest rates to behave quite differently in different time periods. Regime-switching (RS) models constitute an attractive class of models to capture these changes in the stochastic behavior of interest rates within a stationary model. Many authors have built on the seminal work of Hamilton (1989) to model short rates by a model where the parameters change over time driven by a Markov state variable (assumed to be unobserved to the econometrician). For example, Hamilton (1988), Lewis (1991), Evans and Lewis (1994), Sola and Driffill (1994), Garcia and Perron (1996), Gray (1996), and Bekaert, Hodrick, and Marshall (2001) all examined empirical models of regime switches in interest rates.

Importantly, RS models accommodate regime-dependent mean reversion of interest rates. Mankiw and Miron (1986), among others, argued that the predictive power of the term spread for future short rates in the United States is very much a function of the monetary policy regime. In particular, they argued that the interest rate smoothing efforts of the Federal Reserve Bank make the U.S. short rate behave like a random walk, and this behavior causes rejections of the expectations hypothesis. When a regime-switching model is fitted to U.S. data, however, Bekaert et al. (2001) and Gray (1996) showed that such random walk behavior is only true for low interest rates, whereas high interest rates show considerable mean reversion. Several authors (Cecchetti, Lam, and Mark 1993; Garcia 1998) showed that single-regime models are econometrically rejected in favor of their RS counterparts.

Despite their economic appeal, RS models are less attractive than one-regime models from an econometric estimation perspective. Although with the work of Gray (1996) and Hamilton (1994) the likelihood construction has been simplified, estimating RS models is difficult. Often, the data do not allow clear regime classification; that is, the probability of having observed a regime ex-post may hover around a half. These problems may explain why there are few RS term structure models of interest rates (see Naik and Lee 1994; Evans 1998; Bansal and Zhou 1999).

In this article, we provide an analysis of the econometric properties of RS models, with both constant and state-dependent transition probabilities, for interest rates in the United States, Germany, and the United Kingdom. Apart from residual diagnostic tests, we use two statistical criteria to compare and rank alternative one-regime and RS models of short rates. The first criterion investigates the fit of the models with the unconditional moments of the data. One attraction of RS models is that they may accommodate some of the nonlinearities in interest rates that may show up in higher order unconditional moments (see Aït-Sahalia 1996; Stanton 1997; Ahn and Gao 2000). The dependence of mean reversion on the level of the interest rate may also induce an autocorrelogram that is difficult to match by parsimonious autoregressive moving average (ARMA) models. The second criterion concerns the forecasting power of the different models for both first and second moments. Finally, we propose a new metric to compare the performance of different RS models in identifying the regime over the sample. Our regime classification measure (RCM) uses the simple fact that the ex-post probability of observing one of the regimes ought to be close to 1 at all times when regime classification is perfect.

Given the econometric problems mentioned previously, it is not a priori clear that RS models perform well on these statistical criteria, even when they are the true data-generating process (DGP). Moreover, as Bekaert et al. (2001) stressed, the estimation may suffer from a peso problem, in that the fraction of observations drawn from one particular regime in the sample at hand may not correspond to the population frequency of that regime. In that case, the estimation is biased.

Journal of Business & Economic Statistics
April 2002, Vol. 20, No. 2

For example, it is unlikely that we could get a reliable estimate of the mean reversion at large interest rates in U.S. data without including the 1979–1982 period. Furthermore, ARMA models may generally constitute good approximations to any covariance stationary process and hence may outperform RS models in small samples if the parameter estimates of the RS models are severely biased and inefficient.

To overcome these problems, we extend the effective sample size through two channels. First, we investigate multicountry systems of interest rates. It is possible that short rates in the United States Granger-cause rates in other countries (or vice versa) and that Granger causality may be regime dependent. Whereas such relations would immediately affect the forecasting performance, we may also obtain more efficient estimates if interest rate innovations across countries are correlated. If some parameters are identical in different countries, further gains in efficiency are to be expected. The model we propose and estimate allows for correlated interest rate innovations and Granger causality between rates in some regimes. We compare the performance of several variants of the multivariate RS models to their single-regime vector-autoregressive (VAR) counterparts and to one multifactor model in the affine term structure class.

Second, we exploit information in the term structure by adding term spreads to the model. Under the null of the expectations hypothesis, spreads should forecast future short rates, so the potential for improved performance is obvious. The moments criteria here include the cross-correlations between short rates and spreads. As Pfann, Schotman, and Tschernig (1996) showed, the correlation between short rates and long rates changes with the level of the interest rate, suggesting the correlation may be informative about the regime.

Apart from a number of methodological contributions, this article offers some important empirical results. First, whereas RS models do not always outperform single-regime models in the in-sample diagnostics, they forecast very well out of sample. Second, multivariate RS models perform better than univariate models in terms of regime classification and forecasting. The best forecasting model is invariably a multivariate RS model. Hence, our results greatly expand on Gray (1996), who examined the out-of-sample forecasting power of a univariate RS model for second moments of the U.S. short rate. Third, the regime classification implied by RS models is closely related to economic business cycles and the ex-ante regime probabilities are good short-horizon predictors of the business cycle in the United States.

The article is organized as follows. Section 2 describes the data and establishes a set of stylized facts. Section 3 outlines the general empirical and econometric framework and discusses our diagnostic statistics. It presents a general multivariate RS model and considers as special cases univariate short-rate models, multicountry models of the short rate, and bivariate short-rate and term spread models for each country. A stark implication of the framework is that univariate models generally cannot be consistently estimated. Section 4 briefly discusses the empirical estimation results, and Section 5 discusses the performance of the various models. To interpret the results, we perform a Monte Carlo experiment that examines the performance of single-regime and RS models in small samples when the true DGP is an RS model. We consider the quality of regime classification and determine if the regimes are related to the business cycle in Section 6. Section 7 concludes the discussion.

2. DATA AND STYLIZED FACTS

Our empirical work uses monthly observations on 3-month short rates and 5-year long rates of zero-coupon government bonds from the United States, Germany, and the United Kingdom from January 1972 to August 1996. The dataset combines data from Jorion and Mishkin (1991) with a proprietary dataset of zero-coupon rates (see Bekaert et al. 2001). We denote the short rates as r_t^m and the spreads as z_t^m for country m. We estimate models based on an in-sample period, with forecasting done on an out-of-sample period of the last 30 months. Hence, our in-sample period has 267 observations.

Table 1 reports the first four central moments of the short rates and spread data on the in-sample period. The table also shows the autocorrelations for each country, the cross-correlations of short rates for each pair of countries, and correlations of short rates and spreads within each country. We note that the short rates for Germany and the United Kingdom do not show excess kurtosis. Short rates are very persistent, with the United Kingdom showing the least persistence. Spreads are also autocorrelated, but less so than short rates. Turning to international cross-correlations, lagged short rates of the United States are more highly correlated with current German and U.K. rates than present levels of U.S. short rates. This suggests that lagged U.S. short rates may Granger-cause short rates in Germany and the United Kingdom. The contemporaneous correlations of short rates across countries are not very high except for the U.S. and U.K. rates.

In Table 2, we determine whether the behavior of the term structure changes over the business cycle. For the United States, we use the National Bureau of Economic Research (NBER) dates for business cycle expansions and contractions, which can be found at *www.nber.org/cycles.html*; dates for Germany and the United Kingdom are from the Center for International Business Cycle Research at Columbia University (see Zarnowitz 1997). The table divides the interest rate observations into periods of expansions and contractions and performs χ^2 tests for the equality of various moments assuming independence across the cycles. As Zarnowitz (1997) noted, only the United States has a business cycle history that is "official," in the sense of being accepted by governmental authorities, and the dating of the cycles for other countries is less reliable. This means we must interpret the results for Germany and the United Kingdom with caution.

Focusing on the country with the best cycle dating, the United States, Table 2 reveals that recessions are characterized by significantly higher interest rates and somewhat more variable interest rates. The variability is, somewhat surprisingly, not significantly different across expansions and recessions. Interest rates in expansions exhibit higher kurtosis than those in recessions and they are significantly less mean reverting. Spreads are lower and more variable in recessions, but only the mean of the spread is significantly different across cycles.

Table 1. Sample Moments

Panel A: Sample central moments

	U.S.		Germany		U.K.	
Parameter	Short rate	Spread	Short rate	Spread	Short rate	Spread
Mean	7.3381	1.2198	6.9045	.4984	10.5605	.0643
	(.4449)	(.2028)	(.4197)	(.2719)	(.4268)	(.2491)
Variance	8.3103	2.0366	7.1111	3.1241	8.2388	2.7458
	(1.9390)	(.3833)	(1.3380)	(.6714)	(1.4354)	(.5292)
Skewness	.8172	−.7281	.6806	−.5410	−.1521	−.2596
	(.2167)	(.2782)	(.2515)	(.3227)	(.1797)	(.2404)
Kurtosis	3.6102	3.5921	2.6987	3.3732	2.5406	2.8086
	(.6718)	(.7179)	(.4405)	(.5768)	(.3264)	(.4071)

Panel B: Sample autocorrelations

	U.S.		Germany		U.K.	
Lag	Short rate	Spread	Short rate	Spread	Short rate	Spread
1	.9706	.8669	.9845	.9657	.9565	.9322
	(.0181)	(.0292)	(.0216)	(.0265)	(.0237)	(.0238)
2	.9295	.7663	.9583	.9207	.8948	.8776
	(.0347)	(.0497)	(.0436)	(.0507)	(.0450)	(.0425)
3	.8931	.6958	.9253	.8715	.8271	.8234
	(.0513)	(.0689)	(.0638)	(.0711)	(.0637)	(.0596)
4	.8551	.6221	.8858	.8127	.7627	.7692
	(.0653)	(.0820)	(.0812)	(.0868)	(.0784)	(.0753)
5	.8256	.5873	.8428	.7502	.7006	.7200
	(.0778)	(.0836)	(.0957)	(.0999)	(.0895)	(.0895)

Panel C: Sample cross correlations

	Short rates of countries			Short rates/spreads		
Lag	U.S./Germany	U.S./U.K.	Germany/U.K.	U.S.	Germany	U.K.
−3	.4197	.6470	.3279	−.3655	−.7929	−.6524
	(.1334)	(.0777)	(.1007)	(.1130)	(.0563)	(.0727)
−2	.4205	.6549	.3523	−.4213	−.8326	−.7016
	(.1322)	(.0725)	(.0964)	(.1091)	(.0435)	(.0607)
−1	.4120	.6521	.3696	−.4907	−.8656	−.7375
	(.1315)	(.0686)	(.0939)	(.1038)	(.0317)	(.0521)
0	.3953	.6454	.3808	−.5920	−.8804	−.7637
	(.1310)	(.0678)	(.0933)	(.0976)	(.0284)	(.0459)
1	.3756	.6139	.3782	−.5952	−.8634	−.7057
	(.1325)	(.0698)	(.0945)	(.0982)	(.0335)	(.0539)
2	.3542	.5758	.3717	−.5715	−.8389	−.6608
	(.1335)	(.0754)	(.0974)	(.1013)	(.0406)	(.0629)
3	.3294	.5485	.3650	−.5522	−.8097	−.6210
	(.1328)	(.0828)	(.1008)	(.1080)	(.0477)	(.0718)

NOTE: Sample period 1972:01–1993:02 (in-sample period). Standard errors are given in parentheses and are estimated using GMM with six Newey–West (1987) lags. In panel C, the cross-correlations are the estimates of $\mathrm{cov}(r_{t+j}^{m_1}, r_t^{m_2})/\sqrt{\mathrm{var}(r_t^{m_1})}\sqrt{\mathrm{var}(r_t^{m_2})}$ for $j = -3, -2, \ldots, +2, +3$ and country m_1 and country m_2.

In recessions, there is significantly more skewness (or a lack of negative skewness) and spreads are more mean reverting.

These patterns are not perfectly replicated in Germany and the United Kingdom. In these countries, autocorrelations of the short rate and spread are not significantly different across the business cycle. In Germany, the patterns are similar to those in the United States, except for mean reversion, which is insignificantly higher in expansions. In the United Kingdom, the volatility of both spreads and interest rates is higher in expansions, although the p values are not very low. Although the point estimates of mean reversion follow the same pattern as in the United States, the differences across cycles are not statistically significant.

Finally, in the United States and the United Kingdom, the correlation between the short rate and the spread varies over the business cycle. The difference in correlations suggests that in expansions the long rate is less sensitive to short-rate shocks than in recessions. To see this, note that

$$\rho(r_t^l, r_t) = w_1[w_2\,\rho(z_t, r_t) + 1], \tag{1}$$

where $w_1 = \sigma(r_t)/\sigma(r_t^l)$, $w_2 = \sigma(z_t)/\sigma(r_t)$, which is less than 1 empirically, r_t is the short rate, z_t is the spread, r_t^l is the long rate, and $\rho(x, y)$ is the correlation between x and y. In expansions, $\rho(z_t, r_t)$ is more negative and correspondingly the correlation between short and long rates is lower.

Table 2. Interest Rate Behavior Over the Business Cycle

		U.S.			Germany			U.K.		
		Recession	Expansion	χ^2 p Value	Recession	Expansion	χ^2 p Value	Recession	Expansion	χ^2 p Value
Number of observations		50	247		149	148		128	169	
Variable	Statistic									
Short rate r	Mean	9.6466	6.5970	.0001	7.4319	5.8327	.0043	11.9695	8.6856	.0000
		(.7064)	(.3118)		(.4705)	(.3028)		(.3478)	(.4178)	
	Variance	8.4518	6.3173	.3581	8.9237	4.0253	.0055	4.7049	8.1835	.0177
		(1.8775)	(1.3677)		(1.3898)	(1.0894)		(.8396)	(1.2026)	
	Skewness	.6841	1.0360	.5199	.2782	1.3318	.0298	.3151	.4186	.7551
		(.4745)	(.2721)		(.2448)	(.4185)		(.2559)	(.2113)	
	Kurtosis	2.0077	4.5499	.0159	2.0590	5.1641	.0408	2.5627	2.1248	.3906
		(.7198)	(.7708)		(.2111)	(1.5036)		(.4227)	(.2854)	
	ρ_1	.7858	.9503	.0750	.9436	.8894	.2771	.8650	.9243	.2981
		(.0902)	(.0201)		(.0383)	(.0319)		(.0504)	(.0266)	
	ρ_2	.5657	.9061	.0276	.8878	.7641	.1754	.7674	.8353	.4504
		(.1501)	(.0368)		(.0677)	(.0613)		(.0746)	(.0503)	
Spread z	Mean	.5568	1.3835	.0474	.0247	1.2443	.0004	−.4896	.8025	.0001
		(.3903)	(.1469)		(.2688)	(.2179)		(.2128)	(.2517)	
	Variance	3.1724	1.5362	.1206	3.0657	2.1719	.2866	1.6891	2.9732	.0832
		(1.0170)	(.2776)		(.5956)	(.5906)		(.2843)	(.6845)	
	Skewness	.2928	−1.0507	.0038	−.5584	−.8092	.6436	−.3903	−.8350	.1631
		(.4090)	(.2188)		(.2996)	(.4517)		(.2207)	(.2300)	
	Kurtosis	3.4900	3.9699	.6487	2.8995	4.9111	.1210	2.2583	3.5922	.1192
		(.7742)	(.7144)		(.4963)	(1.1986)		(.3802)	(.7672)	
	ρ_1	.6461	.8630	.0362	.9010	.8590	.5751	.8487	.9123	.2889
		(.0987)	(.0313)		(.0588)	(.0465)		(.0529)	(.0284)	
	ρ_2	.3000	.7700	.0007	.8089	.6918	.3698	.7646	.8200	.5278
		(.1308)	(.0435)		(.0952)	(.0895)		(.0707)	(.0517)	
	$\rho(r,z)$	−.2580	−.6947	.0413	−.8813	−.8591	.6637	−.6272	−.7948	.0863
		(.1703)	(.0886)		(.0300)	(.0414)		(.0876)	(.0433)	

NOTE: Sample period 1972:01–1996:08 (full sample). Recessions are defined to be from the peak to the trough of the business cycle. Standard errors are in parentheses and are computed using GMM with three Newey–West lags. ρ_i denotes the ith autocorrelation, $\rho(r, z)$ denotes the correlation between short rates and spreads, and χ^2 p value denotes the p value from a χ^2 test of equality assuming independence across cycle periods.

For the United States, the picture that emerges is one in which, in expansions, short rates are more persistent, the long rate is not as sensitive to short-rate shocks, and the short rate–spread correlation is more negative. In expansions, the interest rate persistence may arise from the smoothing efforts of the monetary authorities. In recessions, long rates are more sensitive to short-rate shocks despite the lower persistence of short rates. Here, shocks to the short rate are more likely to move the whole term structure. The difference in the short rate–spread correlation across expansions and recessions is significant at the 5% level in the United States, but only significant at the 10% level in the United Kingdom and not significant in Germany. However, the pattern of the short rate–spread correlation across expansions and recessions in the United Kingdom is quantitatively similar to the pattern in the United States.

Overall, Table 2 implies the following points about the behavior of interest rates across the business cycle. First, the moments of interest rates vary from recessions to expansions; in particular, the mean is higher in recessions. Second, the spread is informative about the regime, with the spread increasing during expansions and correlations between the spread and the short rate changing across the business cycle. Third, mean reversion in the United States is significantly different across economic regimes. These patterns can potentially be accommodated in models that contain a regime variable.

3. EMPIRICAL AND ECONOMETRIC FRAMEWORK

3.1 General Multivariate Regime-Switching Model

We describe a general multivariate RS model of short rates $r_t = (r_t^{us}, r_t^{ger}, r_t^{uk})'$ and spreads $z_t = (z_t^{us}, z_t^{ger}, z_t^{uk})'$. Let $y_t = (r_t', z_t')'$. We assume that the information set $\mathcal{I}_t$ for our econometric model is composed of $[y_t', y_{t-1}', \ldots]'$. Our most general model is an RS vector autoregression (VAR):

$$y_t = \mu(s_t) + A(s_t)y_{t-1} + \Sigma_{t-1}^{1/2}(s_t)\epsilon_t, \tag{2}$$

where s_t denotes the regime realization at time t and $\epsilon_t \sim$ iid $N(0, I)$. We restrict attention to first-order VAR's because in our empirical work we usually estimate at most first-order systems.

The process s_t follows a Markov chain with K regimes and with transition probabilities that may be logistic functions of lagged endogenous variables:

$$p(s_t = i | s_{t-1} = j, \mathcal{I}_{t-1}) = \frac{e^{\alpha_{i,j} + \beta_{i,j}' y_{t-1}}}{1 + e^{\alpha_{i,j} + \beta_{i,j}' y_{t-1}}}. \tag{3}$$

Let $\tilde{y}_T = (y_T' y_{T-1}' \cdots y_1' y_0')'$ and denote the parameters of the likelihood by θ. Then, following the methodology of Hamilton (1994), we write the conditional likelihood as

$$f(\tilde{y}_T; \theta) = \prod_{t=1}^{T} \left(\sum_{i=1}^{K} f(y_t | \mathcal{I}_{t-1}, s_t = i; \theta) p(s_t = i | \mathcal{I}_{t-1}; \theta) \right). \tag{4}$$

The ex-ante probability $p_{it} = p(s_t = i|\mathcal{I}_{t-1}; \theta)$ can be written as

$$p_{it} = \sum_{j=1}^{K} p(s_t = i|s_{t-1} = j, \mathcal{I}_{t-1}; \theta) \times \left[\frac{f(y_{t-1}|s_{t-1} = j, \mathcal{I}_{t-2}; \theta)p(s_{t-1} = j|\mathcal{I}_{t-2}; \theta)}{\sum_{m=1}^{K} f(y_{t-1}|s_{t-1} = m, \mathcal{I}_{t-2}; \theta)p(s_{t-1} = m|\mathcal{I}_{t-2}; \theta)}\right], \quad (5)$$

where the first term in the sum is the transition probability, which can be state dependent, and the other terms follow from Bayes' rule.

We start the algorithm using (5) with $p(s_1 = i|\mathcal{I}_0)$ equal to the ergodic probabilities of the system at $t = 1$ given by

$$\pi_i = \frac{X_{ii}}{\sum_{j=1}^{K} X_{jj}}, \quad (6)$$

where X_{ii} is the iith cofactor of the matrix $X = I - P_1$ and P_1 is the $K \times K$ transition matrix of the system at $t = 1$, which can depend on our conditional information set $\mathcal{I}_0$. In the special case of constant transition probabilities, we start at the ergodic probabilities π of the transition matrix P that solve $\pi = P'\pi$.

3.2 Special Cases

Because the regime variable is unobserved to the econometrician and must be factored out of the likelihood function, under what conditions can we obtain inefficient but consistent estimates when ignoring some variables? Let Z_t represent variables that do not enter into our estimation and X_t represent variables that do, so $y_t = (Z_t', X_t')'$. Using conditioning arguments, we can write

$$\begin{aligned} f(\tilde{y}_T; \theta) &= \prod_{t=1}^{T} f(y_t|\mathcal{I}_{t-1}; \theta) \\ &= \prod_{t=1}^{T}\left(\sum_{i=1}^{K} f(y_t|s_t = i, \mathcal{I}_{t-1}; \theta)p(s_t = i|\mathcal{I}_{t-1}; \theta)\right) \\ &= \prod_{t=1}^{T}\left(\sum_{i=1}^{K} f(Z_t|X_t, s_t = i, \mathcal{I}_{t-1}; \theta)f(X_t|s_t = i, \mathcal{I}_{t-1}; \theta) \times p(s_t = i|\mathcal{I}_{t-1}; \theta)\right). \end{aligned} \quad (7)$$

To take $f(Z_t|X_t, s_t = i, \mathcal{I}_{t-1}; \theta)$ out of the sum, assume that the excluded variables do not depend on the regime:

$$f(Z_t|X_t, s_t = i, \mathcal{I}_{t-1}; \theta) = f(Z_t|X_t, \mathcal{I}_{t-1}; \theta). \quad (8)$$

We parameterize the model so that $\theta = (\theta_Z'\theta_X')'$ and $\{\theta_Z\} \cap \{\theta_X\} = \varnothing$, where θ_Z and θ_X affect the conditional distribution of the excluded variables and the included variables, respectively. We also assume that the ex-ante probability of being in a particular regime depends only on θ_X:

$$p(s_t = i|\mathcal{I}_{t-1}; \theta) = p(s_t = i|\mathcal{I}_{t-1}; \theta_X). \quad (9)$$

The likelihood can be written as

$$\mathcal{L}(\tilde{y}_T; \theta) = \sum_{t=1}^{T} \log f(Z_t|X_t, \mathcal{I}_{t-1}; \theta_Z) + \sum_{t=1}^{T} \log\left(\sum_{i=1}^{K} f(X_t|s_t = i, \mathcal{I}_{t-1}; \theta_X) \times p(s_t = i|\mathcal{I}_{t-1}; \theta_X)\right). \quad (10)$$

Maximizing the second sum in (10) yields consistent but inefficient estimates relative to full information maximum likelihood.

Estimation of the full system is infeasible given the dimension of θ, so we focus on models of subsets of the variables. Our choice here is partially based on previous literature and partially on economic reasoning. We believe that regimes in real rates, expected inflation, or business cycles are the source for potential regimes in nominal interest rates (see Garcia and Perron 1996; Evans and Lewis 1995). To obtain parsimony in modeling, we assume the existence of a two-state Markov regime variable in every country driving the entire term structure. These country-specific regime variables are assumed to be independent across countries. It is conceivable that there is a world business cycle driving interest rates in many countries simultaneously, and in some of the models we consider we allow for interdependence of various forms across countries. Nevertheless, it should be noted that the correlation between spreads and short rates within a country is typically of a higher magnitude than the correlation of short rates and spreads across countries (see Table 1), providing empirical motivation for this assumption. Although the two-regime specification may seem restrictive, it is the most the data can bear without extreme computational problems in estimation, and it suffices to capture the main empirical nonlinearities. In particular, Ang and Bekaert (2000) showed that two-state RS models can replicate the nonparametric drift and volatility functions of the short rate estimated by Aït-Sahalia (1996) and Stanton (1997). Finally, most of the past RS literature focused on two-state models, with the exception of Garcia and Perron (1996) and Bekaert et al. (2001) who estimated three-state RS models.

Because most of the RS literature focuses exclusively on univariate interest rate models, we start by analyzing univariate short-rate models for the United States, Germany, and the United Kingdom. As (8) shows, to consistently estimate univariate short-rate RS models, the distribution of the term spreads or short rates from other countries should not depend on the regime of the short rate we consider. If regimes capture business cycle effects, the different correlations in the United States across economic cycles in Table 2 violate the assumptions needed for consistent estimation.

Incorporating the extra information from international and term structure data allows us to weaken the implicit assumptions but makes estimation much more complex. In a second set of models, we add information from the short rates from other countries. In our multicountry model (discussed later), defining the regime variable s_t becomes more involved

Table 3. Summary of Models Estimated

Univariate models of short rates

One regime	Two regime equivalents: Constant probabilities	Two regime equivalents: Time-dependent probabilities
AR(1) (3)	RS1 (8)	RS2 (10)
GARCH(1,1) (5)	RS3 (12)	
CIR (3)	RS4 (8)	RS5 (10)

Multicountry models of short rates

Model	Description
VAR1u (18)	Unconstrained VAR(1)
G1 (13)	One-regime Granger-causality model; homoscedastic errors
RSG1 (16)	RS Granger causality with the same α_i, ρ_i, σ_i, P, and Q across countries; square-root errors
RSG2 (20)	RS Granger causality with the same α_i, ρ_i, P, and Q across countries, but different σ_i; square-root errors
D1 (11)	One-regime diagonal model; homoscedastic errors
RSD1 (12)	RS diagonal model with the same α_i, ρ_i, σ_i, P, and Q across countries; homoscedastic errors

Multivariate models of the term spread

One regime	Two regime equivalents: Constant probabilities	Two regime equivalents: Time-dependent probabilities
VAR(1) (9)	RSM1 (20)	RSM2 (24)
VAR(2) (13)		
ATSM (9)		

as it embeds all possible combinations of the country-specific regime variables for the three countries.

Finally, we consider models in which term spreads are added to the short rate and their dynamics remain driven by one country-specific regime variable. In most term structure models, the term spread is an exact function of a number of factors that also drive the short rate. However, the evidence from a growing literature that focuses on the response of the term structure to various shocks suggests that the spread contains additional independent information, which may help in the classification of regimes. For example, Evans and Marshall (2000) showed that monetary policy shocks have large effects on the short rate but leave the long rate unaffected, hence shrinking the spread. However, shocks from real economic activity affect the whole term structure and correspond to a level effect, increasing the interest rate but leaving the spread largely unaffected. Estrella and Mishkin (1997) found that the spread is useful in predicting future activity and that the spread contains predictive information that is not captured by other monetary policy variables. A reduced-form model where the spread and short rate have correlated innovations and different feedback rules, in which spreads help predict future regimes, may be a good representation of such a world. We estimate the short rate–spread model country by country but also consider one estimation that uses cross-country information.

Table 3 presents a summary of the models estimated, their abbreviations used throughout the article, and the number of parameters in parentheses. We now briefly outline each of these models. (Parameter estimates are available in an appendix, which is available from the authors on request.)

3.2.1 Univariate Models. For each country m, we consider special cases of the following general model considered in Gray (1996):

$$r_t^m = \mu(s_t^m) + \rho(s_t^m) r_{t-1}^m + h_{t-1}^m(s_t^m)\epsilon_t, \tag{11}$$

where $\epsilon_t \sim$ iid N(0, 1). The conditional volatility $h_{t-1}^m(s_t^m)$ is specified as

$$\left(h_{t-1}^m(s_t)\right)^2 = a_0(s_t^m) + a_1(s_t^m)\eta_{t-1}^2 + b_1(s_t^m)(h_{t-2}^m)^2 + b_2(s_t)(r_{t-1}^m), \tag{12}$$

where $(h_{t-1}^m)^2 = E_{t-1}[(r_t^m)^2] - (E_{t-1}[r_t^m])^2$ and $\eta_t = r_t^m - E_{t-1}[r_t^m]$. The regime variable s_t is either 1 or 2 and has transition probabilities:

$$p(s_t^m = j | s_{t-1}^m = j, r_{t-1}^m) = \frac{e^{a_j + b_j r_{t-1}^m}}{1 + e^{a_j + b_j r_{t-1}^m}}, \qquad j = 1, 2. \quad (13)$$

We denote constant transition probabilities as P and Q for $j = 1, 2$, respectively. We evaluate $E_{t-1}[r_t^m]$ and $E_{t-1}[(r_t^m)^2]$ as

$$E_{t-1}[r_t^m] = \sum_{j=1}^{2} p_{t,j}(\mu_j + \rho_j r_{t-1}^m),$$

$$E_{t-1}\left[(r_t^m)^2\right] = \sum_{j=1}^{2} p_{t,j}\left((\mu_j + \rho_j r_{t-1}^m)^2 + (h_{t-1,j}^m)^2\right), \quad (14)$$

where the subscripts indicate the state $s_t^m = j$.

The special cases we consider involve setting $a_1 = b_1 = b_2 = 0$ (RS AR(1)), $b_2 = 0$ (RS GARCH(1, 1)), and $a_0 = a_1 = b_1 = 0$ (RS CIR). The last model is the RS equivalent of the discretized square-root model of Cox, Ingersoll, and Ross (1985).

In practice, many interest rate RS models yield one unit-root or near unit-root regime and one more mean-reverting regime. Ang and Bekaert (1998) and Holst, Lindgren, Holst, and Thuvesholmen (1994) proved that such processes retain covariance stationarity as long as the unconditional autocorrelation is strictly less than 1. This is guaranteed by appropriate mixing of the two regimes. With constant transition probabilities, a sufficient condition is that the ergodic probability associated with the stationary regime is nonzero.

3.2.2 Multicountry Models. For $r_t = (r_t^{us} \quad r_t^{ger} \quad r_t^{uk})'$, we consider the following general multicountry RS model:

$$r_t = \begin{pmatrix} \alpha^{us}(s_t^{us}) \\ \alpha^{ger}(s_t^{ger}) \\ \alpha^{uk}(s_t^{uk}) \end{pmatrix} + A(s_t^{us}, s_t^{ger}, s_t^{uk}) r_{t-1} + \Sigma^{1/2}(s_t^{us}, s_t^{ger}, s_t^{uk})\epsilon_t, \quad (15)$$

with $\epsilon_t = (\epsilon_t^{us} \quad \epsilon_t^{ger} \quad \epsilon_t^{uk})' \sim \text{iid N}(0, I)$.

We assume that there are two regimes per country with constant probabilities, so for country m the transition matrix is $\begin{pmatrix} P^m & 1-P^m \\ 1-Q^m & Q^m \end{pmatrix}$. For computational tractability, and to keep the number of parameters as small as possible, we do not consider state-dependent transition probabilities in the multicountry model.

We assume regimes in different countries to be independent of the regimes in another country. Formally, let $S^m = \{s_t^m, s_{t-1}^m, \ldots\}$ denote the past history of regimes for country m. Then

$$p(s_t^m | S_t^{us}, S_t^{ger}, S_t^{uk}) = p(s_t^m | S_t^m) = p(s_t^m | s_{t-1}^m). \quad (16)$$

Intuitively, this means that the regime for one country is unaffected by the regime in another country. We may justify this by interpreting the regimes as arising from country-specific factors. This independence assumption can only be relaxed at considerable computational cost and proliferation of parameters. With two regimes for three countries, it is possible to enlarge the state space to $2^3 = 8$ regimes, where the regimes are defined as $s_t = 1, \ldots, 8$ (see Hamilton 1994):

s_t	U.S.	Germany	U.K.
1	1	1	1
2	2	1	1
3	1	2	1
4	2	2	1
5	1	1	2
6	2	1	2
7	1	2	2
8	2	2	2

We then calculate an 8×8 transition matrix, where, for example, $p(s_t = 1 | s_{t-1} = 1) = P^{us} P^{ger} P^{uk}$.

With the regimes now redefined as $s_t = 1, \ldots, 8$, we rewrite (15) as

$$r_t = \alpha(s_t) + A(s_t) r_{t-1} + u_t, \quad (17)$$

where $u_t \sim \text{N}(0, \Sigma(s_t))$. From hereon, subscript i's refer to the values each specific country's state comprises in the overall state i. For example, for $s_t = 4$,

$$\begin{pmatrix} \alpha_4^{us} \\ \alpha_4^{ger} \\ \alpha_4^{uk} \end{pmatrix} = \begin{pmatrix} \alpha^{us}(s_t^{us} = 2) \\ \alpha^{ger}(s_t^{ger} = 2) \\ \alpha^{uk}(s_t^{uk} = 1) \end{pmatrix}.$$

Given the number of parameters, estimation of the full model is infeasible. To gain efficiency, we test whether some parameters are identical in the one-regime VAR. In particular, we test for Granger causality on each country's short rates. These results are presented in Table 4. The table shows that a joint test for no country Granger-causing another just fails to reject (p value = .0528). Nevertheless, there is some evidence that U.S. rates Granger-cause German and U.K. rates (p value = .0029).

The results of Table 4 lead us to consider two formulations of A_i, a triangular formulationwhere $A_i = \begin{pmatrix} \rho_i^{us} & 0 & 0 \\ \zeta_i^{ger} & \rho_i^{ger} & 0 \\ \zeta_i^{uk} & 0 & \rho_i^{uk} \end{pmatrix}$,

Table 4. Granger Tests in the Multicountry VAR Model

Granger causality	*A[i, j] = 0*	*p value*
No country Granger-causes another	All off-diagonal elements = 0	.0528
U.S. Granger-causes Germany and U.K.	$A[2,1] = A[3,1] = 0$	.0029
Germany and U.K. Granger-cause U.S.	$A[1,2] = A[1,3] = 0$	.7332
Germany and U.K. Granger-cause each other	$A[2,3] = A[3,2] = 0$	.6742

NOTE: Wald tests are performed using GMM with six Newey–West lags. The notation $A[i, j]$ refers to the element in row i, column j.

which we refer to as a Granger-causality formulation, and a diagonal formulation where $A_i = \begin{pmatrix} \rho_i^{us} & 0 & 0 \\ 0 & \rho_i^{ger} & 0 \\ 0 & 0 & \rho_i^{uk} \end{pmatrix}$.

To impose further structure on the error terms, we model the errors as

$$\begin{pmatrix} u_{t,i}^{us} \\ u_{t,i}^{ger} \\ u_{t,i}^{uk} \end{pmatrix} = \begin{pmatrix} h_{t-1,i}^{us}\epsilon_t^1 \\ h_{t-1,i}^{ger}\epsilon_t^2 + \gamma_i^{ger}\epsilon_t^1 \\ h_{t-1,i}^{uk}\epsilon_t^3 + \gamma_i^{uk}\epsilon_t^1 \end{pmatrix}, \tag{18}$$

where $\epsilon_t = (\epsilon_t^1, \epsilon_t^2, \epsilon_t^3)'$ are drawn from an iid $N(0, I)$ distribution and the conditional volatility of country m, $h_{t-1,i}^m$, is specified either as a constant, $h_{t-1,i}^m = \sigma_i^m$, or as a square-root process, $h_{t-1,i}^m = \sigma_i^m\sqrt{r_{t-1}^m}$. In this specification, the errors from the United States also shock the interest rates of Germany and the United Kingdom, but not vice versa. Another interpretation along the lines of a world business cycle is that there are "world" shocks, which drive the dominant U.S. economy, whereas Germany and the United Kingdom are also subject to these shocks as well as "country-specific" shocks. The extent to which these countries are exposed to the world shock depends on the state of the domestic economy. Given the dominance of the United States in the world economy, such a structure seems reasonable. The conditional covariance matrix, conditional on state $s_t = i$, is given by

$$\begin{aligned} \Sigma_t(s_t = i) &= E[u_t u_t' | \mathcal{I}_{t-1}, s_t = i] \\ &= \begin{pmatrix} (h_{t-1,i}^{us})^2 & \gamma_i^{ger} h_{t-1,i}^{us} & \gamma_i^{uk} h_{t-1,i}^{us} \\ \gamma_i^{ger} h_{t-1,i}^{us} & (h_{t-1,i}^{ger})^2 + (\gamma_i^{ger})^2 & \gamma_i^{ger}\gamma_i^{uk} \\ \gamma_i^{uk} h_{t-1,i}^{us} & \gamma_i^{uk}\gamma_i^{ger} & (h_{t-1,i}^{uk})^2 + (\gamma_i^{uk})^2 \end{pmatrix}. \end{aligned} \tag{19}$$

This specification arises because the errors $u_{t,i}^m$ inherit a multivariate normal distribution from the normality of the errors $\epsilon_{t,i}^m$. Note that German and U.K. shocks are conditionally correlated to the extent only that they correlate with the U.S. shock.

It is possible to obtain probability inferences for a particular country by summing together the relevant joint probabilities. For example, if we want the ex-ante probability $p(s_t^{us} = 1|\mathcal{I}_{t-1})$, we just sum over the probabilities $p(s_t|\mathcal{I}_{t-1})$ where $s_t^{us} = 1$. In this case, we would sum over states $s_t = 1, 3, 5, 7$.

3.2.3 Term Spread Models. For $y_t^m = (r_t^m \;\; z_t^m)'$, the short rate and spread for country m, the RS term spread model is

$$y_t^m = \mu(s_t^m) + A(s_t^m)y_{t-1}^m + \Sigma^{1/2}(s_t^m)\epsilon_t, \tag{20}$$

where $\epsilon_t \sim N(0, I)$. We use two regimes, with constant transition probabilities and logistic state-dependent transition probabilities where

$$\begin{aligned} p(s_t^m = j | s_{t-1}^m = j, y_{t-1}^m) \\ = \frac{\exp(a_j + b_j r_{t-1}^m + c_j z_{t-1}^m)}{1 + \exp(a_j + b_j r_{t-1}^m + c_j z_{t-1}^m)}, \quad j = 1, 2. \end{aligned} \tag{21}$$

We also estimate the term spread model jointly across the United States, Germany, and the United Kingdom, following Bekaert et al. (2001). This estimation views each country as an independent draw of the DGP, by assuming independence of the regimes across countries, lack of cross-country correlation, and the same parameters across countries.

We consider two classes of one-regime models as potential benchmarks. First, we estimate unconstrained VAR's of the short rate and the term spread, restricting attention to first- and second-order VAR's. Second, we consider the affine class of term structure models (see Duffie and Kan 1996). In these models, zero-coupon yields are affine (constant plus linear term) functions of the unobservable factors. This implies that we can represent $y_t^m(n)$, the yield for maturity n for country m, as an affine function of the state variables X_t^m for country m:

$$y_t^m(n) = \overline{A}(n) + \overline{B}(n)'X_t^m, \tag{22}$$

where the scalar $\overline{A}(n)$ and vector $\overline{B}(n)$ are functions of the model parameters. We represent the dynamics of X_t^m, without loss of generality, by a first-order VAR:

$$X_t^m = \phi + \Phi X_{t-1}^m + \Omega^{1/2}\epsilon_t^m, \tag{23}$$

where $\epsilon_t \sim N(0, I)$. The 1-month yield (which we do not observe) takes the form:

$$y_t^m(1) = \delta_0 + \delta_1' X_t^m, \tag{24}$$

where δ_0 is a scalar and δ_1 is a vector. The specification of a pricing kernel π_{t+1}^m, for each country m, completes the model. The pricing kernel prices all nominal bonds through the recursive relation:

$$P_t^m(n+1) = E_t\left[\pi_{t+1}^m P_{t+1}^m(n)\right], \tag{25}$$

where $P_t^m(n)$ is the zero-coupon bond price of maturity n for country m.

Different affine models make different assumptions about the state variable dynamics and the specification of the pricing kernel, in particular, the specification of the prices of risk. Standard models assume either homoscedastic state variable dynamics with constant prices of risk, for example, correlated Vasicek (1977) models, or a square-root process with time-varying prices of risk or a combination of the two. Duffee (2001) demonstrated that standard affine term structure models forecast poorly out of sample. Therefore, we consider an alternative affine model not considered by Duffee. We consider Gaussian, homoscedastic state variables, but time-varying prices of risk. More specifically, we assume that the pricing kernel has the form:

$$\pi_{t+1} = \exp\left(-\frac{1}{2}\lambda_t'\lambda_t - \delta_0 - \delta_1' X_t^m - \lambda_t'\epsilon_{t+1}\right), \tag{26}$$

where the risk premia λ_t are time varying:

$$\lambda_t = \lambda_0 + \lambda_1 X_t^m, \tag{27}$$

where λ_0 is a vector and λ_1 is a matrix.

We work with a two-factor model and for identification purposes, we impose the following parameter restrictions:

$$\phi = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Phi = \begin{pmatrix} \Phi_{11} & 0 \\ \Phi_{12} & \Phi_{22} \end{pmatrix},$$
$$\Omega = I, \qquad \lambda_0 = \begin{pmatrix} \lambda_{01} \\ 0 \end{pmatrix}, \qquad \lambda_1 = \begin{pmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \end{pmatrix}. \quad (28)$$

We call this bivariate correlated factor model the Gaussian affine term structure model (ATSM) with time-varying risk premia.

The model has a structural VAR representation in terms of the observable yields. The short rate and spread for country m can be written as

$$y_t^m \equiv \begin{pmatrix} r_t^m \\ z_t^m \end{pmatrix} = \begin{pmatrix} \overline{A}(3) \\ \overline{A}(60) - \overline{A}(3) \end{pmatrix} + \begin{pmatrix} \overline{B}(3) \\ \overline{B}(60) - \overline{B}(3) \end{pmatrix}' X_t^m \quad (29)$$

or, by appropriately defining $\overline{A}$ and $\overline{B}$, as $y_t^m = \overline{A} + \overline{B}X_t^m$. The discrete-time recursive relations determining $\overline{A}(n)$ and $\overline{B}(n)$ are derived in Ang and Piazzesi (2001). By substituting (23) into (29), it is straightforward to show that

$$y_t^m = \mu + Ay_{t-1}^m + \Sigma^{1/2}\epsilon_t, \quad (30)$$

where $\epsilon_t \sim \mathrm{N}(0, I)$, $\mu = (I - \overline{B}\Phi\overline{B}^{-1})\overline{A}$, $A = \overline{B}\Phi\overline{B}^{-1}$, and $\Sigma^{1/2} = B$. This representation makes both maximum likelihood estimation and forecasting using the observed yields easy. Clearly, the ATSM is simply a VAR model with cross-equation restrictions. Whereas the estimation of this model went smoothly for the United States, the likelihood surfaces for the United Kingdom and Germany proved very flat. Models with λ_1 restricted to 0, that is, standard correlated Vasicek (1977) models, do not converge at all for all countries. Dai and Singleton (2001) showed that a Gaussian model with affine prices of risk matches the deviations from the expectations hypothesis observed for U.S. data, but they ignore small-sample biases (see Bekaert, Hodrick, and Marshall 1997).

3.3 Model Diagnostics

We start by reporting a number of standard in-sample residual tests for our various models. Our second diagnostic more easily leads to comparisons across a large number of nonnested models of varying complexity. We measure the fit of the unconditional moments implied by the models to the sample estimates of the unconditional moments. Single-regime models may perform reasonably well along these dimensions even though they are not the true DGP. However, they are less likely to perform well over tests that exploit the changing behavior of interest rates across regimes. To easily rank the performance across all models, we focus on summary statistics for out-of-sample forecast errors. Finally, we compare different RS models, using a measure of the quality of the regime classification. We discuss these in turn.

3.3.1 Residual Tests. We report two tests on in-sample scaled residuals e_t^m of short rates of country m where $e_t^m = (r_t^m - E_{t-1}[r_t^m])/h_{t-1}^m$. The conditional volatility h_{t-1}^m is given by

$$(h_{t-1}^m)^2 = \mathrm{var}_{t-1}(r_t^m - E_{t-1}[r_t^m]) = E_{t-1}[(r_t^m)^2] - (E_{t-1}[r_t^m])^2. \quad (31)$$

For a univariate RS model, $E_{t-1}[r_t^m]$ and $E_{t-1}[(r_t^m)^2]$ are evaluated using Equation (14).

Following Bekaert and Harvey (1997), we use a generalized method of moments (GMM) test of the moment conditions on the mean of the scaled residuals:

$$E[e_t^m e_{t-j}^m] = 0 \quad \text{for} \quad j = 1, 2, \ldots, k, \quad (32)$$

which we refer to as "mean" residual tests, and a GMM test of the moments of the variance of the scaled residuals:

$$E[((e_t^m)^2 - 1)((e_{t-j}^m)^2 - 1)] = 0 \quad \text{for} \quad j = 1, 2, \ldots, k, \quad (33)$$

which we refer to as "variance" residual tests. In both tests, we choose $k = 6$ and correct for heteroscedasticity in the residuals following Andrews (1991).

3.3.2 Unconditional Moment Comparisons. We compute the unconditional population moments of our various models using analytical expressions where possible but use a simulation for the RS models with time-varying transition probabilities. Analytical formulas for moments are available only for one-regime CIR and generalized autoregressive conditional heteroscedasticity (GARCH) processes as well as for autoregressive regime-switching models with constant transition probabilities (see Timmerman 2000). Because of the high persistence of the series, we use sample sizes of one million.

To enable comparison across several models, we introduce the point statistic:

$$H = (\hat{g} - \bar{g})'\Sigma_g^{-1}(\hat{g} - \bar{g}), \quad (34)$$

where $\bar{g}$ are sample estimates of unconditional moments, $\hat{g}$ are the unconditional moments from the estimated model, and Σ_g is the covariance matrix of the sample estimates of the unconditional moments. Σ_g is estimated using a GMM estimation of the unconditional moments, and, for the purposes of this article, we use a Newey–West (1987) estimate with six lags. The point statistic assigns weights to the deviations between the unconditional moments implied by various models and the sample unconditional moments, which are inversely proportional to the error by which the sample moments are estimated.

We test for the first four central moments, the autocorrelogram, and cross-correlations. In the first case, $\bar{g}$ contains the mean, variance, skewness, and kurtosis; for the autocorrelogram, the first 10 autocorrelations; and for cross-correlations, lags from −3 to +3. Generally, the high persistence of interest rates may lead to poor estimation of unconditional moments. Therefore, there are instances where high correlation between the estimated moments leads to somewhat poorly conditioned weighting matrices. Hence, we also calculate a related statistic H^*, which uses as a weighting matrix the diagonal of Σ_g. Strong correlations between the estimated moments sometimes imply that the model minimizing H does not minimize H^*.

3.3.3 Forecast Comparisons. Our forecast methodology is to estimate only using the in-sample period and forecast without updating the parameters on the out-of-sample period. We use two point statistics for comparison of unconditional forecast errors, the root mean squared error (RMSE), and mean absolute deviation (MAD). For a time series ϕ_t, these are defined as

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum(\phi_t - \hat{\phi}_t)^2},$$
$$\text{MAD} = \frac{1}{T}\sum|\phi_t - \hat{\phi}_t|, \quad (35)$$

where the hatted values denote conditional forecast values. In our application, we let $\phi_t = r_t^k$ for univariate and multicountry models, looking at first and second moments $k = 1, 2$. In term spread models, we also consider $\phi_t = z_t$ and the cross-moment $\phi_t = r_t z_t$.

3.3.4 Regime Classification. Previous specification tests for RS models have focused on properties of residuals (Gray 1996) or scores (Hamilton 1996), but here we propose a summary point statistic that captures the quality of regime classification. An RS model assumes that at each point of time the data are drawn from one of the regimes that is observed by agents in the economy but not by the econometrician. To conduct inference about the regime, most articles focus on the smoothed (ex-post) regime probabilities, $p(s_t = 1|\mathcal{J}_T)$, which we denote as p_t. Weak regime inference implies that the RS model cannot successfully distinguish between regimes from the behavior of the data and may indicate misspecification. An ideal RS model should classify regimes sharply so that p_t is close to 1 or 0; in inferior models, p_t may hover close to a half.

To measure the quality of regime classification, we therefore propose the regime classification measure (RCM), defined for two states as

$$\text{RCM} = 400 \times \frac{1}{T}\sum_{t=1}^{T} p_t(1-p_t). \quad (36)$$

The constant serves to normalize the statistic to be between 0 and 100. Good regime classification is associated with low RCM statistic values: A value of 0 means perfect regime classification and a value of 100 implies that no information about the regimes is revealed. Because the true regime is a Bernoulli random variable, the RCM statistic is essentially a sample estimate of its variance.

The statistic easily generalizes to multiple regimes. A general definition of the statistic for K regimes is

$$\text{RCM}(K) = 100K^2\frac{1}{T}\sum_{t=1}^{T}\left(\prod_{i=1}^{K} p_{i,t}\right), \quad (37)$$

where $p_{i,t} = p(s_t = i|\mathcal{J}_T)$.

4. EMPIRICAL RESULTS

4.1 Estimation Results of the Regime-Switching Models

Estimation of regime-switching models in finite samples is plagued by the presence of multiple local maxima. To ensure that we identify the global maximum for the 31 RS models we estimate, we use the following procedure. First, we obtain estimates for a large set of starting values and select a candidate global maximum. Second, to check for stability of the global, we reestimate using starting values randomly chosen in a $\pm 10\%$ interval around the parameters of the provisional global maximum. When models have trouble converging to a well-behaved global using this procedure, we either dropped the model or simplified it, rather than continuing the numerical search toward poorly identified models.

The RS models all produce one regime with a unit root and lower conditional volatility and a second regime that is stationary with higher conditional volatility. This type of estimation is found in univariate, multicountry, and term spread models. Economically, the first regime corresponds to "normal" periods where monetary policy smoothing makes interest rates behave like a random walk. When extraordinary shocks occur, interest rates are driven up, volatility becomes higher, and interest rates become more mean reverting.

In general, models with time-varying transition probabilities have many insignificant coefficients in the probability terms, which suggests overparameterization. Previous studies with time-varying probabilities such as Gray (1996) also document this. For some of our models, we fail to reject the null hypothesis of constant probabilities. Nevertheless, the general pattern that emerges in the majority of cases is as expected: Higher short rates (and spreads) increase the probability of switching to the high-volatility regime.

To highlight the features of specific models, we discuss univariate, multicountry, and term spread RS models in turn. Recall that Table 3 presents the nomenclature scheme of the models.

4.1.1 Univariate Models. As Table 3 shows, we consider three different conditional volatility specifications. We retain constant transition probability models for all countries for all the formulations, except for the UK GARCH model. We do not estimate state-dependent models for the GARCH formulation, because the constant probability models are already overparameterized. In estimating models with state-dependent transition probabilities, we only find significant state dependence for the US CIR model and the German RS AR(1) model. We drop the RS AR(1) model with state-dependent transition probabilities for the United States.

4.1.2 Multicountry Models. RSD1 is a diagonal model with the same parameters (α_i, ρ_i, σ_i) across countries and homoscedastic within-regime errors. The RSG1 model is identical but has square-root errors. Constraining σ_i to be the same across countries imposes the restriction that the conditional volatility for Germany and the United Kingdom is higher than the conditional volatility for the United States. We relax this formulation in the RSG2 model and find that it makes little qualitative difference.

The estimation results show that Granger causality by U.S. shocks is important only for the United Kingdom in the second mean-reverting high-variance regime. Granger causality of Germany is insignificant in both regimes. Looking at the impact of U.S. shocks on the error terms of Germany and the United Kingdom, the Granger-causality model RSG2 has significant shock terms for Germany and the United Kingdom in the first random walk regime. The diagonal model, however, shows U.S. shocks affecting only U.K. shocks in the first regime. To summarize, in the "normal" random walk regime U.S. shocks propagate into Germany and the United Kingdom, whereas in the second regime U.S. short rates Granger-cause only U.K. short rates.

4.1.3 Term Spread Models. In the RS term spread models, we find that Granger causality is model dependent. For the United States and Germany, one regime produces a significant $A_i[1,2]$ term, so the spread Granger-causes the short rate in only one regime (the higher variance regime for the United States but the lower variance one for Germany). The evidence for the United Kingdom is less clear as the coefficient is just insignificant in one regime but very insignificant in the other. Similarly, the short rates Granger-cause spreads only in one regime but these may not be the same regimes where spreads Granger-cause short rates. In the United States, these are in opposite regimes, but for Germany these regimes are the same. In the joint estimation where we assume independence and the same parameters across countries, short rates and spreads Granger-cause each other in the same regime (the lower conditional variance regime).

The correlation between short rates and spreads differs markedly across regimes. The high-variance less persistent regime has more negative correlation than the low-variance regime. Wald tests for equality across the regimes reject with zero p value for all countries. Short rates and spreads seem less correlated in the first regime, which corresponds to "normal" periods. However, note from Table 2 that the correlation between the short rate and spread is more negative in expansions, which is the opposite to what the regime-switching models imply. Nevertheless, the high-mean, high-variance second regime does correspond to economic recessions. We examine this further in Section 6.

For our time-varying probability formulations, the transition probabilities depend on both the short rate and the spread, except for the United States where we use a model with transition probabilities dependent only on the spread. Likelihood ratio tests for constant transition probabilities versus time-varying probabilities reject for all countries. The results on Granger causality and regime-dependent correlations hold for both the constant and the time-varying transition probability models.

5. PERFORMANCE MEASURES

Section 5.1 analyzes residual tests and the moment performance, and Section 5.2 analyzes forecast performance. Section 5.3 summarizes the evidence and makes use of a Monte Carlo experiment to help interpret the results. The results are reported in Tables 5 through 10.

5.1 In-Sample Tests

5.1.1 Residual Tests. Table 5 lists the results of the mean and variance residual tests. Turning first to the U.S. results, the benchmark single-regime models perform well, passing both the mean and the variance residual tests. However, in each of the univariate, multicountry, and term spread models, the variance residual test has a p value of only slightly larger than 5%. The only models that comfortably pass both the mean and the variance residual tests incorporate term spread information in an RS model (RSM1 and RSM2). The single-regime or RS (RS3) univariate GARCH models and the CIR models fail

Table 5. Residual Tests on Short Rates

	U.S.		Germany		U.K.	
Model	*Mean*	*Variance*	*Mean*	*Variance*	*Mean*	*Variance*
			Univariate models			
AR1	.3170	.0523	.0212*	.4826	.4855	.8401
RS1	.0101*	.4980	.0003**	.2782	.4868	.6243
RS2	—	—	.0001**	.0094**	.0000**	.0248*
GARCH	.0085**	.1529	.0000**	.5841	.4543	.8894
RS3	.0153	.5452	.0011**	.4612	—	—
CIR	.0071**	.0000**	.0088**	.1321	.4410	.7602
RS4	.0039**	.5368	.0000**	.1489	.4519	.8656
RS5	.0030**	.6816	.0000**	.0592	.0000**	.2700
			Multicountry models			
VAR1u	.3728	.0508	.0137*	.2463	.5197	.6841
G1	.8767	.0000**	.0246*	.1189	.7495	.6441
RSG1	.0483*	.1081	.0002**	.5369	.7702	.5769
RSG2	.0311*	.1478	.0002**	.5476	.7753	.6668
D1	.2497	.0513	.0211*	.4820	.4885	.8407
RSD2	.0016**	.0000**	.0000**	.0000**	.6210	.4704
			Term spread models			
VAR1	.5087	.0547	.0197*	.4432	.4598	.4517
VAR2	.8715	.0185*	.3146	.1572	.4568	.4146
ATSM	.0120*	.0911	.0233*	.4902	.0329*	.4132
RSM1	.3831	.1753	.0011**	.3581	.5929	.8920
RSM2	.2988	.1434	.0013**	.3682	.4338	.9694

NOTE: The table reports p values from mean and variance residual tests. p values significant at the 5% (1%) level are labeled (*) (**).

Table 6. Moments of Univariate and Multicountry Models

Panel A: Univariate models

	Statistic	RS1	RS2	RS3	RS4	RS5	AR(1)	GARCH	CIR
					U.S.				
Central moments	*H*	330.99	—	327.56	113.72	63.03	30.32	—	3.11*
	*H**	112.15	—	72.92	36.78	46.24	15.26	—	1.71*
Autocorrelogram	*H*	10.01	—	6.67	8.82	5.23	3.88*	—	3.91
	*H**	20.84	—	7.81	16.31	6.19	1.30*	—	5.34
					Germany				
Central moments	*H*	67.03	4,563.76	5,211.55	100.78	17.03*	165.53	—	34.11
	*H**	17.80	153.80	4,088.45	27.78	9.61	7.98	—	6.54*
Autocorrelogram	*H*	6.82	9.21	5.08*	6.14	7.67	6.91	—	5.96
	*H**	13.01	22.30	3.07*	12.46	20.59	13.74	—	9.59
					U.K.				
Central moments	*H*	3.49*	4.00	—	29.02	36.90	5.81	6.82	25.11
	*H**	4.38	4.19	—	18.34	34.27	2.83*	4.09	19.13
Autocorrelogram	*H*	7.43	7.75	—	7.98	7.36*	9.51	8.84	8.43
	*H**	11.06*	13.07	—	14.01	14.61	20.24	17.55	15.58

Panel B: Multicountry models

	Statistic	VAR1u	D1	RSD1	G1	RSG1	RSG2
				U.S.			
Central moments	*H*	30.83	21.73	13.38*	30.31	28.66	32.66
	*H**	15.25	15.76	11.10*	15.26	17.26	22.64
Autocorrelogram	*H*	3.43	3.34*	8.06	3.87	9.77	11.70
	*H**	.97	.25*	13.44	1.29	19.84	27.34
				Germany			
Central moments	*H*	174.98	166.62	54.91	207.78	26.09*	26.47
	*H**	7.90*	7.98	15.50	8.20	10.09	11.09
Autocorrelogram	*H*	6.12*	6.91	7.19	6.99	7.96	9.12
	*H**	12.43*	13.82	13.21	15.10	16.55	21.84
				U.K.			
Central moments	*H*	6.40	6.06*	64.62	7.94	146.54	287.66
	*H**	2.76	2.80	55.29	2.71*	81.77	123.56
Autocorrelogram	*H*	10.08	9.63*	26.13	11.67	31.97	34.04
	*H**	21.90	20.69*	81.71	26.34	106.81	113.68

NOTE: Lowest statistic values are denoted with an asterisk.

to pass the mean residual tests. The multicountry RS models generally do poorer than their single-regime counterparts.

The mean residual tests for Germany reject all the models, with the exception of a second-order VAR, despite a first-order VAR being the optimal Akaike information criterion (AIC) and Bayesian information criterion (BIC) choice. Several RS specifications (RS2 and RSD2) do less well than their single-regime counterparts, with the variance residual test also rejecting them. In comparison, almost all the models pass the residual tests on U.K. data, with univariate RS state-dependent transition probability specifications (RS2 and RS5) and the ATSM being the exception.

The Gaussian ATSM's reject the mean residual test at a 5% level across all countries. The implied factors from affine models are severely biased, which leads to the poor in-sample performance, but the ATSM's manage to pass the variance residual tests. This confirms evidence in Ghysels and Ng (1998), who rejected the conditional mean specification of affine models, but also found less evidence of misspecification with second moments.

In summary, no single model passes all the residual tests for all countries. For the United States and the United Kingdom, RS term spread models comfortably pass the residual tests, whereas almost all models fail to pass the residual tests on German data.

5.1.2 Matching Sample Moments. We present *H* statistics for univariate models in panel A of Table 6. For the United States, the one-regime models seem to work better in matching unconditional moments than the RS models. The dismal performance of models RS1–3 for the United States is partly caused by the unit root in one of the regimes, although the models are theoretically stationary. For Germany, RS2 and RS3 do poorly because they produce large values for kurtosis. The best fits for the moments for Germany are for the one-regime and RS CIR models. For the United Kingdom, the AR(1) RS processes work best with the square-root processes performing more poorly. RS models with state-dependent probabilities (RS2, RS5) and GARCH errors (RS3) fare less well than the constant probability models, RS1 and RS4.

Table 7. Unconditional Moments of Term Spread Models

		Statistic	VAR1	VAR2	ATSM	RSM1	RSM2
					U.S.		
Central moments	r_t	H	31.51	29.99	29.62*	193.20	141.95
		H^*	15.26*	15.27	15.35	97.05	54.83
	z_t	H	10.09	10.07*	10.88	119.70	30.96
		H^*	7.62*	7.62	7.87	86.30	21.77
Autocorrelations	r_t	H	2.46*	890.48	52.25	4.32	5.13
		H^*	.84*	17.35	249.45	1.33	12.11
	z_t	H	21.70	5,724.77	5.30*	16.82	10.58
		H^*	43.26	68.67	5.67*	69.27	14.52
Cross-correlation	$r_t z_t$	H	86.99	444.51	73.71	12.73	2.05*
		H^*	16.24	8.82	260.11	31.29	.19*
					Germany		
Central moments	r_t	H	232.01	157.55*	250.97	374.27	268.22
		H^*	8.24	8.08*	10.42	20.11	11.40
	z_t	H	6.43	6.03*	17.11	38.98	18.69
		H^*	3.29*	3.69	7.49	10.04	5.97
Autocorrelations	r_t	H	7.41	2,941.41	4.85*	6.65	6.04
		H^*	15.28	23.09	3.24*	13.90	10.63
	z_t	H	8.50	316.92	5.70*	15.57	14.66
		H^*	17.39	34.19	8.68*	51.67	47.65
Cross-correlation	$r_t z_t$	H	6.96*	142.92	1,718.14	17.30	10.80
		H^*	8.89	7.71	4,228.46	10.91	4.21*
					U.K.		
Central moments	r_t	H	4.84	4.93	4.33*	23.51	32.49
		H^*	3.03	3.00*	3.10	4.33	5.60
	z_t	H	2.25	2.17*	77.39	9.80	11.09
		H^*	1.42	1.40*	46.42	7.63	9.15
Autocorrelations	r_t	H	8.04	50.26	7.24*	8.69	8.98
		H^*	16.26	60.28	13.31*	19.00	21.69
	z_t	H	2.82*	119.41	8.61	2.99	3.09
		H^*	.38*	21.27	16.61	2.00	2.42
Cross-correlation	$r_t z_t$	H	7.87*	199.73	397.72	17.00	11.36
		H^*	11.44	10.93	1,233.65	9.34	2.09*

NOTE: Lowest statistic values are denoted with an asterisk.

Panel B of Table 6 reports H statistics for the multicountry models. Among the one-regime models, diagonal models match central moments better than the unconstrained VAR(1) or Granger-causality models, suggesting overparameterization in these models. With the exception of the United Kingdom, the RS diagonal model performs better than its one-regime diagonal counterpart. This is quite an achievement, considering that this model constrains each country to have the same parameters. The RS Granger-causality models perform more poorly than the RS diagonal models for the United States and the United Kingdom but not for Germany. There is little difference when we no longer constrain σ_i to be equal across countries in the RS Granger-causality models.

Table 7 reports the H and H^* statistics for the bivariate short rate–spread system. The one-regime models (VAR(1), VAR(2), and ATSM) generally outperform the RS models (RSM1 and RSM2) at matching unconditional moments. For one-regime models, the more parsimonious VAR(1) definitely does better at matching autocorrelations than VAR(2), with comparable results for the central moments. For the United States, the ATSM performs almost as well as VAR(1) and VAR(2) in matching central moments, but this is not the case for the United Kingdom and Germany. In matching autocorrelations, the ATSM performs best across the board in Germany, performs best for the short-rate aucorrelations in the United Kingdom, and also performs best for spread correlations in the United States. However, the ATSM's perform extremely poorly in all countries matching the short rate–spread cross-correlation. This is because the off-diagonal term in the companion matrix of the factors [Φ_{12} in (28)] is near 0 for Germany and the United Kingdom. Turning to the RS models, the state-dependent probability models fare better for the United States and Germany than their constant probability counterparts, but for the United Kingdom this result is reversed. One-regime models clearly outperform RS VAR's for central moments and autocorrelograms. Only for cross-correlations does RSM2 provide a good fit.

Does incorporating extra information improve the performance of RS models? By looking across panels A and B of Table 6, we compare the univariate RS models with the multicountry RS models. We see a dramatic improvement when incorporating multicountry information for the United States but not for Germany or the United Kingdom. Comparing the univariate RS models in Table 6, with the bivariate RS term spread models in Table 7, the term spread information leads to a better match of moments only for the United States, and for autocorrelations only for the United Kingdom. Overall, using the extra information from other countries or the term spread unequivocally helps the United States obtain a better

Table 8. Forecasts of Univariate and Multicountry Models

Panel A: Univariate models

	Statistic	RS1	RS2	RS3	RS4	RS5	AR(1)	GARCH	CIR
					U.S.				
r_t	MAD	.1488	—	.1458	.1458	.1487	.1483	.1387*	.1664
	RMSE	.1956	—	.1943	.1945	.1968	.1888*	.1999	.1999
r_t^2	MAD	1.5161	—	1.4696	1.4771	1.5048	1.6540	1.3410*	1.9874
	RMSE	1.9421	—	1.9167*	1.9277	1.9525	2.0335	1.9207	2.3042
					Germany				
r_t	MAD	.1307	.1299	.1329	.1285	.1327	.1501	.1207*	.1615
	RMSE	.1732	.1732	.1716	.1694	.1732	.1900	.1627*	.2006
r_t^2	MAD	1.2097	1.1979	1.1822	1.1407	1.1824	1.4568	1.0895*	1.4985
	RMSE	1.5936	1.5943	1.5351	1.5114	1.5423	1.8174	1.4736*	1.8637
					U.K.				
r_t	MAD	.2509	.2137*	—	.2449	.2288	.2419	.2555	.2539
	RMSE	.2890	.2668*	—	.2819	.2771	.2772	.2910	.2893
r_t^2	MAD	3.5109	2.9807*	—	3.2783	3.0626	3.3666	3.4550	3.3090
	RMSE	4.0030	3.5617*	—	3.7319	3.6206	3.8180	3.9192	3.7569

Panel B: Multicountry models

	Statistic	VAR1u	D1	RSD1	G1	RSG1	RSG2
				U.S.			
r_t	MAD	.1619	.1499	.1378	.1483	.1160*	.1174
	RMSE	.2002	.1891	.1841	.1888	.1625*	.1626
r_t^2	MAD	1.5550	1.7159	1.2139	1.3992	.9949*	1.0388
	RMSE	1.8065	2.0771	1.4980	1.6453	1.1930*	1.2146
				Germany			
r_t	MAD	.1580	.1500	.1429	.1327*	.1451	.1466
	RMSE	.1959	.1899	.1868	.1704*	.2035	.2062
r_t^2	MAD	1.6591	1.4557	1.2822	1.4632	1.1957	1.2436*
	RMSE	1.8537	1.8164	1.5706	1.6303	1.5206*	1.5899
				U.K.			
r_t	MAD	.2747	.2410	.1429	.2668	.1124*	.1142
	RMSE	.3116	.2762	.2017	.3055	.1766*	.1833
r_t^2	MAD	2.2897	3.3541	1.6389	2.1274	1.2960	1.2872*
	RMSE	2.0020	3.8040	2.1859	1.8801	1.8465*	1.8554

NOTE: Lowest statistic values are denoted with an asterisk.

fit to unconditional moments, but it definitely does not help Germany. The evidence for the United Kingdom is mixed.

5.2 Out-of-Sample Tests

Tables 8 and 9 list the forecast performance results. Focusing first on univariate models in panel A of Table 8, the state dependence of the probabilities in RS AR(1) models produces superior forecasts, even though many of the estimated coefficients are insignificant and the performance in matching the sample moments is poor. However, this result is not shared by the RS CIR model, with only the United Kingdom's state-dependent formulation performing better. Overall, with the exception of the United Kingdom, the GARCH models produce the best results. For the United Kingdom, the superior performance of the RS2 model, using either the RMSE or the MAD criterion for both first and second moments, is remarkable, given that regime classification in the United Kingdom is rather poor (see Fig. 1). Relative to their one-regime counterparts, RS models generally perform better. For all countries, with the exception of the one-regime GARCH model, the RS AR(1) models forecast better than a simple AR(1) and the RS CIR models forecast better than the single-regime CIR models.

Panel B of Table 8 presents the forecasting results for the multicountry models. The diagonal one-regime models outperform the unrestricted VAR on mean forecasts and do worse for second-moment forecasts only for the United States, again showing overparameterization of the unconstrained VAR. The multi-country RS diagonal model outperforms the one-regime model despite having the interest rate DGP constrained to be the same across all countries. This is a strong endorsement of the importance of regime shifts in forecasting. Granger causality seems to aid in forecasting in both one-regime and RS frameworks. The RS Granger models do particularly well for the United States and the United Kingdom.

Table 9. Forecasts of Term Spread Models

Statistic	Model														
	VAR1			*VAR2*			*ATSM*			*RSM1*			*RSM2*		
							U.S.								
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	.1885	.2117		.1918	.2186		.1224*	.2134		.1531	.2070*		.1588	.2072	
RMSE	.2312	.2760		.2490	.2735		.1641*	.2825		.1948	.2653*		.2025	.2650	
	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$
MAD	2.1183	.7141	1.2672	2.1267	.7422	1.2122	1.2065*	.8501	1.0970*	1.5907	.6267*	1.1044	1.6529	.6280	1.0964
RMSE	2.5226	.9181	1.5309	2.6521	.9534	1.4835	1.5005*	1.0292	1.4816	1.9941	.8292*	1.1391*	2.0755	.8287	1.3870
							Germany								
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	.1359	.2191		.1796	.2395		.1212	.2176*		.1197	.2214		.1074*	.2186	
RMSE	.1765	.2796		.2219	.2826		.1632	.2734*		.1617	.2821		.1471*	.2755	
	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$
MAD	1.3464	.7084*	.9882	1.6568	.8016	1.0049	1.2121	.7507	.8999*	1.1333	.7168	1.0243	.9772*	.7225	1.0092
RMSE	1.7109	.8757*	1.3273	2.1278	.9241	1.2957	1.5805	.8999	1.2992	1.5495	.8909	1.3346	1.3531*	.8962	1.2937*
							U.K.								
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	.2172	.2455*		.2607	.2603		.4510	.3274		.2211	.2491		.1768*	.2590	
RMSE	.2529	.3031		.3112	.3384		.4867	.4308		.2180*	.3078		.2414	.3121	
	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$	r_t^2	z_t^2	$r_t z_t$
MAD	3.0729	1.0989*	1.2995*	3.5870	1.1921	1.3297	6.2222	1.2128	1.5194	3.1016	1.1168	1.3197	2.4471*	1.1370	1.3864
RMSE	3.4981	1.4991*	1.5292*	4.1074	1.6290	1.6097	6.6166	1.9078	1.9809	3.5814	1.5343	1.5416	3.1786*	1.5423	1.5746

NOTE: Lowest statistic values are denoted with an asterisk.

Table 9 reports forecast performance in the term spread models. In forecasting the first and second moments, the more parsimonious VAR(1) outperforms the VAR(2) for all countries, suggesting that the VAR(2) is overparameterized. In the United States, the ATSM provides better forecasts of the short rate than unrestricted VAR's, which confirms the results in Ang and Piazzesi (2001). This finding is repeated for Germany but not for the United Kingdom, where the ATSM fails to beat the VAR specifications. For the United States, the ATSM outperforms all the other bivariate specifications for forecasting short rates and second moments of short rates. Duffee (2001) noted that affine models with constant risk premia forecast very poorly, but he did not consider forecasts of affine models with time-varying risk premia as in our ATSM specification. In contrast to the U.S. results, in Germany and the United Kingdom RS models outperform the one-regime models for forecasting the level and square of short rates. The results of forecasts of spreads and cross-moments are mixed. Whereas the RS models outperform the one-regime specifications in the United States, the ATSM and VAR specifications provide better forecasts in Germany and the United Kingdom. The lowest RMSE statistics for cross-moment forecasts belong to the RS models for the United States and Germany; the best cross-moment forecast for the United Kingdom is VAR(1).

Adding information from other countries or term spreads to the estimation uniformly improves forecasts. Focusing on the RMSE criterion, Table 8 shows that the multicountry approach generally yields better forecasts than the univariate models. Table 9 shows that adding term spreads improves forecasts, with the RS spread models beating univariate forecasts with the exception of the United States, where the ATSM dominates.

5.3 Summary and Interpretation

In general, we find that, in matching sample moments, RS models do not systematically outperform one-regime models. However, in forecasting out of sample, RS models almost invariably do better. Focusing on short rates, Table 10 reports the best models with the lowest H, RMSE, and MAD statistics. There is no clear-cut "best" model. However, it appears that, whereas single-regime models may produce lower H statistics (e.g., in the U.S. case), RS models forecast better for all countries. We note that, for the United States, the ATSM comes very close to giving the best forecast for the short rate. Moreover, the best RS forecasting models incorporate information from other countries or the spread. Interestingly, RS models with state-dependent transition probabilities tend to forecast better than their constant probability counterparts, even though they perform very poorly at matching sample moments.

How do we interpret these results? As indicated before, the RS models considered here need large simulations to pin

Table 10. Overall Moments and Forecast Comparisons for Short Rates

	Best H statistics		
	U.S.	*Germany*	*U.K.*
Central moments	CIR	RS5	RS1
Autocorrelogram	VAR1	ATSM	ATSM

	Best RMSE statistics			Best MAD statistics		
	U.S.	*Germany*	*U.K.*	*U.S.*	*Germany*	*U.K.*
r_t	RSG1	RSM2	RSG1	RSG1	RSM2	RSG1
r_t^2	RSG1	RSM2	RSG1	RSG1	RSM2	RSG2

Table 11. Small-Sample Experiment: Percentage of Time Models Do Best

	Unconditional moments					Forecasts			
	AR	RS AR	VAR	RS VAR		AR	RS AR	VAR	RS VAR
r_t central	15.9%	59.9%	14.8%	9.4%	r_t	30.6%	16.3%	24.5%	28.6%
$\rho(r_t)$	43.4%	3.3%	43.7%	9.6%	r_t^2	29.4%	18.0%	20.5%	32.1%
z_t central			90.1%	9.9%	z_t			45.8%	54.2%
$\rho(z_t)$			36.3%	63.7%	z_t^2			46.1%	53.9%
$\rho(r_t, z_t)$			88.9%	11.1%	Cross			44.2%	55.8%

NOTE: We simulate data of length 297 from the joint estimation across the U.S.–Germany–U.K. of a bivariate system of the short rate r_t and spread z_t with time-varying probabilities (model RSM2). We then estimate an AR(1), a regime-switching AR(1), a VAR, and a regime-switching VAR, denoted AR, RS AR, VAR, and RS VAR, respectively and record which model gives the lowest H and RMSE statistics. The table lists the percentage times the model performed the best in each small sample. We conducted 1,000 simulations.

down their unconditional moments with any precision. This means that the small-sample behavior of RS models may be poor. In other words, it is conceivable that more parsimonious one-regime models produce better estimates of the sample unconditional moments than RS models in small samples, even though an RS model is the true DGP. Here we run a Monte Carlo experiment to specifically investigate this conjecture.

Consider the following RS VAR population model of the short rate and spread, $y_t = (r_t\, z_t)'$: $y_t = \mu(s_t) + A(s_t)y_{t-1} + \Sigma^{1/2}(s_t)\epsilon_t$ where $\epsilon_t \sim \mathrm{N}(0, I)$, $s_t = 1, 2$ with Markov state-dependent logistic transition probabilities depending on lagged y_t. We use the parameters from the joint estimation as the population model and find the true population moments of this model using a very long simulation. Then, we simulate a small sample of size $T+N$ and compute unconditional moment estimates over the in-sample of size T and RMSE forecast statistics over the out-sample of size N for several aproximations to the true model. We set T and N to be the size of our in-sample and out-sample datasets in this article, 267 and 30, respectively. The models we consider are an AR(1) and an RS AR(1) on the short rates with constant probabilities and a VAR(1) and an RS VAR(1) on the bivariate short rate and spread with constant transition probabilities. We denote these as AR, RS AR, VAR, and RS VAR, respectively.

Unfortunately, we cannot include the true model because of the problems we encounter in finding satisfactory estimates of the RS VAR with time-varying probabilities in small samples. The many convergence failures that occur, even when starting from the true parameters, are themselves proof of the small-sample problems RS models face.

To compare the unconditional moment estimators, we calculate H statistics with the mean, standard deviation, skewness, and kurtosis and then record which of the four models yields the best (lowest) statistic value for each simulated sample. To compare out-of-sample forecasts, we record which model produces the lowest RMSE statistic. We use 1,000 Monte Carlo replications. Table 11 reports the percentage times each model best fit the population moments or produced the best forecasts. For example, for the simulations performed, in 15.9% of cases the AR(1) model gave the best fit to the population moments as measured by the H statistic, even though the true model was an RS VAR(1) with state-dependent probabilities.

Table 11 shows that the one-regime models are good approximations of the true RS models in small samples and that, despite the true DGP being an RS model, parsimonious one-regime models may perform better at matching moments and forecasting. It is notable that RS models perform quite poorly in matching unconditional moments, but perform better in forecasting. These results parallel our findings for the actual RS models estimated on real data.

We also examine the empirical distribution of the moments produced by the models in small samples. Table 12 reports the population values of the unconditional moments for the short rates and spreads. The table also lists the mean values and standard deviations of the small-sample distribution of the moments produced by the various models. RS models tend to overestimate the mean and underestimate the variance of the short rate, but the population values lie within 95% confidence intervals of the small-sample model moments. However, the AR and VAR single-regime models produce close to unbiased estimates of the mean and variance. This result may help justify the popularity of VAR-type models to test unconditional term structure hypotheses, such as the expectations

Table 12. Small-Sample Distribution of Moments

Parameter	Population	AR	RS AR	VAR	RS VAR
Short rates					
Mean	7.3289	7.3905	8.5011	7.4066	8.8526
		(1.3454)	(1.4462)	(1.3802)	(1.7742)
Variance	11.2885	10.9206	7.8944	11.0027	8.9975
		(3.8646)	(2.2026)	(4.3127)	(2.6317)
Skewness	.5750		.2032		.1185
			(.1700)		(.3087)
Kurtosis	3.0639		3.1360		3.2287
			(.3263)		(3.3094)
Spreads					
Mean	.8642			.8509	.3410
				(.3903)	(.4304)
Variance	1.5460			1.4306	1.0500
				(.5161)	(.2705)
Skewness	−.1815				−.0790
					(.2812)
Kurtosis	3.0084				3.2709
					(1.8155)

NOTE: These are the means, with standard errors in parentheses, of the moments of the estimated models in a small sample of 267 in the experiment of Table 11. Skewness and kurtosis for the AR and VAR models are theoretically 0 and 3, respectively.

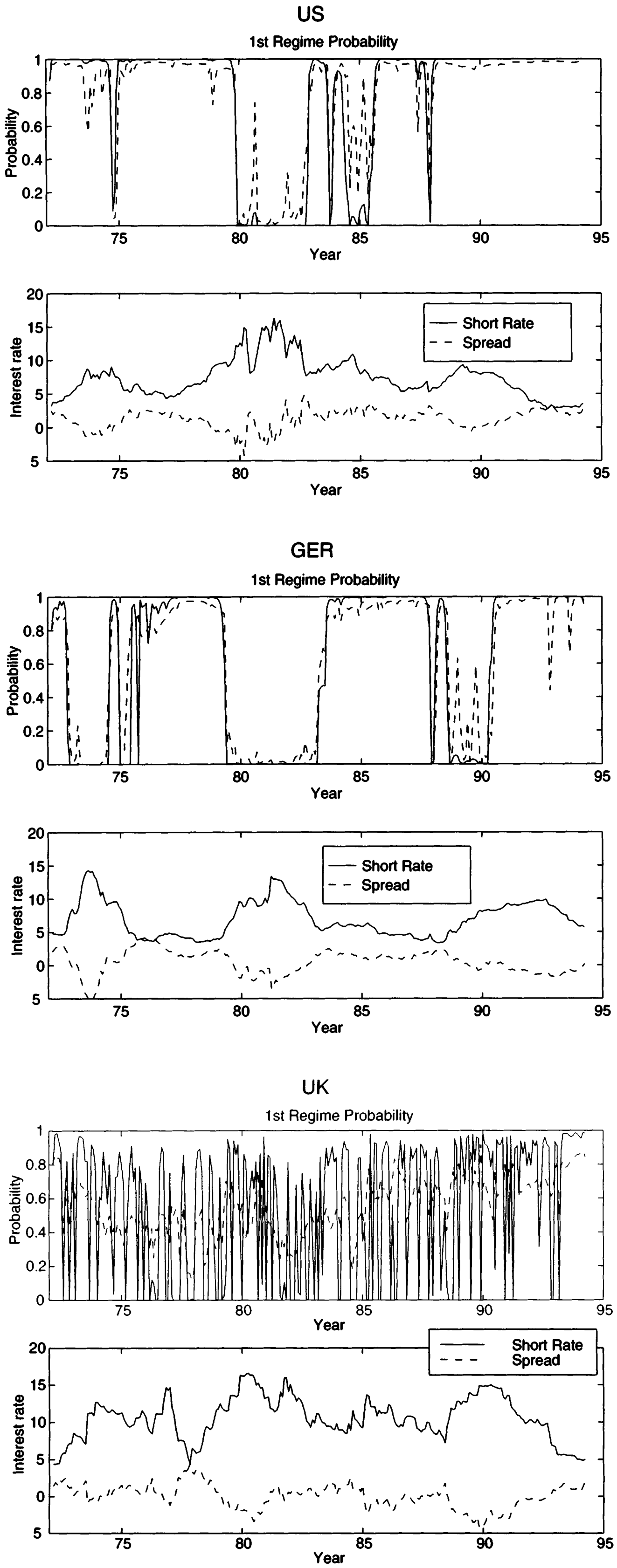

Figure 1. Regime Probabilities. The figure shows the ex-ante probabilities $p(s_t = 1|\mathcal{I}_{t-1})$ (dotted line) and smoothed probabilities $p(s_t = 1|\mathcal{I}_T)$ (solid line) in the top subplots for each country, and the short rate and spread for each country in the bottom subplots.

hypothesis, even in the presence of significant nonlinearities in the data.

6. REGIME CLASSIFICATION AND REGIME INTERPRETATION

Figure 1 displays the regime probabilities for the RS VAR state-dependent transition probability model for the United States, Germany, and the United Kingdom. The solid line in the top plots represents smoothed probabilities $p(s_t = 1|\mathcal{I}_T)$ using information over the full sample of size T and the broken line represents ex-ante probabilities $p(s_t = 1|\mathcal{I}_{t-1})$. Plots of ex-ante and smoothed probabilities for the other models look similar. For the United Kingdom, there is a high frequency of switching between regimes because the transition probabilities P and Q are very close to a half. In a regime-switching model, if $P+Q=1$ the model reduces to a simple switching model. For the U.K. models, we often cannot reject this hypothesis and the regime classification also appears poor because the smoothed regime probability often is far away from 1 or 0.

For a more quantitative examination of regime classification, we present RCM statistics in Table 13. In univariate models, the RS AR(1) model produces the sharpest regime classification for the United States, whereas RS CIR models produce the sharpest regime classification for Germany and the United Kingdom. For univariate models, moving from constant to state-dependent transition probabilities produces very little improvement. Our multicountry model produces sharper regime classification for the United Kingdom and Germany at the expense of the United States. In particular, there is a large improvement in regime classification for the United Kingdom by adding U.S. information. Including term spread information leads to lower RCM statistics for all countries.

Are the regimes correlated with the business cycle? Table 14 attempts to answer this question. The table first presents correlations between various lags j of the ex-ante probabilities p_{t-j+1} and a recession indicator for the business cycles of each country. The ex-ante probabilities are generated from the term spread RS model with time-varying probabilities (RSM2). We use this model because it is the model with the lowest RCM statistic for the United States in Table 13. We report the correlations between the second regime with mean-reverting higher volatility and the economic downturns. The table shows that this regime is associated with economic recessions, whereas

Table 13. RCM Statistics

Model	*U.S.*	*Germany*	*U.K.*
RS1	10.44	22.57	43.14
RS2	—	23.69	41.54
RS3	19.04	52.53	—
RS4	11.53	21.02	42.29
RS5	12.88	20.45	40.64
RSD1	18.11	13.44	27.23
RSG1	21.16	19.72	28.17
RSG2	21.94	22.25	24.48
RSM1	7.67	14.60	38.70
RSM2	6.68	16.12	34.90

Table 14. Markov Regimes and Business Cycles

U.S.

	Correlations		Probit forecasting			
Months Ahead j	$\rho(1-p_{t-j+1}, rec_t)$	$\rho(z_{t-j}, rec_t)$	$\beta(1-p_{t-j+1})$	% Forecast	$\beta(z_{t-j})$	% Forecast
1	.4264	−.3047	1.6203	83.8	−.2811	80.8
	(.1153)	(.1104)	(.2569)		(.0605)	
2	.4618	−.3989	1.7537	84.2	−.3847	82.3
	(.1149)	(.1028)	(.2603)		(.0645)	
4	.4840	−.5096	1.8428	84.4	−.5611	86.7
	(.1123)	(.0851)	(.2640)		(.0760)	
6	.4122	−.5296	1.5569	85.1	−.5750	87.0
	(.1126)	(.0820)	(.2584)		(.0745)	

Germany

	Correlations		Probit forecasting			
Months Ahead j	$\rho(1-p_{t-j+1}, rec_t)$	$\rho(z_{t-j}, rec_t)$	$\beta(1-p_{t-j+1})$	% Forecast	$\beta(z_{t-j})$	% Forecast
1	.1892	−.5276	.5789	60.2	−.4903	75.2
	(.1109)	(.0719)	(.1879)		(.0601)	
2	.2162	−.5830	.6632	61.5	−.6073	75.8
	(.1107)	(.0615)	(.1890)		(.0696)	
4	.2472	−.6590	.7615	63.9	−.8474	77.9
	(.1101)	(.0508)	(.1908)		(.0927)	
6	.2392	−.6811	.7366	63.6	−.9400	81.6
	(.1106)	(.0483)	(.1915)		(.1024)	

U.K.

	Correlations		Probit forecasting			
Months Ahead j	$\rho(1-p_{t-j+1}, rec_t)$	$\rho(z_{t-j}, rec_t)$	$\beta(1-p_{t-j+1})$	% Forecast	$\beta(z_{t-j})$	% Forecast
1	.0911	−.3439	.6856	54.1	−.2821	67.3
	(.1066)	(.0999)	(.4590)		(.0506)	
2	.0779	−.3828	.5864	53.6	−.3218	69.4
	(.1067)	(.0962)	(.4601)		(.0522)	
4	.0098	−.4508	.0740	51.3	−.4018	74.1
	(.1077)	(.0899)	(.4646)		(.0564)	
6	−.0230	−.4680	−.1756	49.0	−.4274	72.4
	(.1063)	(.0837)	(.4710)		(.0584)	

NOTE: Recessions are coded as a 1, expansions as 0. The symbol p_t represents the ex-ante probabilities $p(s_t = 1|\mathcal{I}_{t-1})$ of the first regime from the term spread RS model with time-varying transition probabilities (RSM2). Columns 2 and 3 give the correlation of the recession indicator (rec) with the ex-ante probability of the second regime and the spread z_t. Standard errors are calculated using GMM with three Newey–West lags. The last four columns show results from fitting the Probit model $p(\text{rec}_t = 1) = F(\alpha + \beta(.)a_{t-j})$, where $F(\cdot)$ is the normal cumulative distribution function, β is the coefficient corresponding to the variable a_{t-j}, and we let a_{t-j} be current and lagged values of $1-p_{t-j}$ and z_{t-j-1}. Lags are in months. The % Forecast column is the percentage of correctly forecasted (in-sample) values from the Probit regression.

the "normal" unit-root regime with lower volatility represents economic expansions. The United States and Germany have significant correlations, whereas the correlations of the United Kingdom are insignificant.

The business cycle association of the regimes is not surprising for the United States. Figure 1 shows that the ex-ante probabilities during the 1979–1982 period of monetary targeting are near 0, placing this period in the second regime. During this period, high variable interest rates were accompanied by a large recession. Germany also experienced a similar episode around the same time (1980:03–1983:07) and also went through an earlier recession accompanied by high interest rates in the early 1970s (1973:09–1975:05). The recession brought on by the re-unification, beginning in mid-1991, also saw rising interest rates, but the regimes do not capture this period as successfully. The poor results for the United Kingdom are not surprising, given the poor regime classification of the U.K. model.

The last four columns of Table 14 report coefficients from a Probit regression with the recession indicator being the dependent variable and current and lagged ex-ante probabilities being the independent variables. The Probit regressions yield significant coefficients for the United States and Germany. We also list the percentage of correctly forecasted recessions in-sample from the Probit regressions. For the United States, the ex-ante probabilities successfully predict 84% of recessions one-month ahead, with the success ratio slightly increasing as we try to predict further into the future. The success ratio is around 60% for Germany and, not surprisingly, only 50% for the United Kingdom.

Harvey (1988) and Estrella and Mishkin (1997) found that term spreads successfully predict real economic activity. Table 14 confirms their findings, showing that the magnitude of correlations between recessions and the spread increases with the lag and that the accuracy of the Probit forecasts increases with the forecast horizon. This happens in all three countries. Looking specifically at the United States, the ex-ante regime probabilities have better forecast ratios for one- and two-month-ahead predictions than the spread. Whereas the forecast ratios increase with horizon for the spread, the forecast ratios of the ex-ante probabilities remain essentially flat. This evidence indicates that, for the United States, the ex-ante regime probabilities are better contemporaneous indicators of the business cycle than the spread, and the spread is a forward-looking indicator with greater forecasting ability at longer horizons. For the other countries, the spread better predicts recessions than our regime probabilities at all horizons. Given that both the regime classification and the dating of the actual business cycles is less precise for these countries, this is not surprising.

7. CONCLUSIONS

We compare the econometric performance of regime-switching (RS) models relative to their one-regime counterparts in several ways. First, residual tests show that RS models often perform worse than single-regime models. For the United States, only RS models with term spread information comfortably pass the residual tests. Second, the moments implied by RS models do not always fit the sample moments as well as simpler models do because of the difficulties in estimating RS models in small samples. A Monte Carlo experiment confirms that this happens, even when the RS model is the true data-generating process. Finally, RS models invariably forecast better than one-regime models, although a parsimonious multifactor affine term structure model with time-varying prices of risk performs almost as well for U.S. short rates.

To improve the econometric performance of RS models, it is important to incorporate additional information. In fact, univariate RS models yield inconsistent estimates when the omitted variables contain information on the regime. We compare the performance of univariate models with multicountry short-rate models and models incorporating term spreads. In particular, U.S. short rates improve both the regime classification and the statistical performance for German and U.K. short rates (but not vice versa). Furthermore, inclusion of term spread information leads to general improvements over univariate models in forecasting and to dramatic superior performance in regime inference. The inclusion of additional cross-sectional country short rates or term spreads does not always improve the fit of the unconditional moments. However, the regimes correspond well with business cycle expansions and contractions.

ACKNOWLEDGMENTS

The authors thank Rob Engle, Eric Ghysels, Ken Singleton, Jun Liu, Liuren Wu, two anonymous referees, and seminar participants at Stanford University and the 1999 Econometric Society meetings for many useful comments. Geert Bekaert acknowledges financial support from a National Science Foundation grant.

[Received June 2000. Revised July 2001.]

REFERENCES

Ahn, D. H., and Gao, B. (2000), "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, 12 (4), 721–762.

Aït-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9 (2), 385–426.

Andrews, D. (1991), "Heteroskedasticity and Autocorrelation Consistent Matrix Estimation," *Econometrica*, 59, 817–858.

Ang, A., and Bekaert, G. (1998), "Regime Switches in Interest Rates," Working Paper 6508, National Bureau of Economic Research.

——— (2000), "Short Rate Nonlinearities and Regime Switches," forthcoming *Journal of Economic Dynamics & Control*.

Ang, A., and Piazzesi, M. (2001), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics With Macroeconomic and Latent Factors," Working Paper 8363, National Bureau of Economic Research.

Bansal, R., and Zhou, H. (1999), "Term Structure of Interest Rates With Regime Shifts," mimeo, Duke University.

Bekaert, G., and Harvey, C. R. (1997), "Emerging Equity Market Volatility," *Journal of Financial Economics*, 43 (1), 29–77.

Bekaert, G., Hodrick, R. J., and Marshall, D. A. (1997), "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, 44 (3), 309–348.

——— (2001), "'Peso Problem' Explanations for Term Structure Anomalies," *Journal of Monetary Economics*, 48 (2), 241–270.

Cecchetti, S. J., Lam, P., and Mark, N. C. (1993), "The Equity Premium and the Risk-Free Rate: Matching the Moments," *Journal of Monetary Economics*, 31, 21–46.

Cox, J., Ingersoll, J., and Ross, S. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–408.

Dai, Q., and Singleton, K. (2001), "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," forthcoming *Journal of Financial Economics*.

Duffee, G. R. (2001), "Forecasting Future Interest Rates: Are Affine Models Failures?" unpublished manuscript submitted to *Journal of Finance*.

Duffie, D., and Kan, R. (1996), "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6 (4), 379–406.

Estrella, A., and Mishkin, F. S. (1997), "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, 41 (7), 1375–1401.

Evans, C., and Marshall, D. (2000), "Economic Determinants of the Term Structure of Nominal Interest Rates," mimeo, Chicago Federal Reserve.

Evans, M. D. (1998), "Regime Shifts, Risk and the Term Structure," mimeo, Georgetown University.

Evans, M, D., and Lewis, K. K. (1995), "Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation?" *Journal of Finance*, 50 (1), 225–253.

Garcia, R. (1998), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39 (3), 763–788.

Garcia, R., and Perron, P. (1996), "An Analysis of the Real Interest Rate Under Regime Shifts," *Review of Economics and Statistics*, 78 (1), 111–125.

Ghysels, E., and Ng, S. (1998), "A Semiparametric Factor Model of Interest Rates and Tests of the Affine Term Structure," *Review of Economics and Statistics*, 80 (4), 535–548.

Gray, S. F. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Hamilton, J. D. (1988), "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics & Control*, 12, 385–423.

——— (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

——— (1994), *Time Series Analysis*, Princeton, Princeton University Press, NJ:

——— (1996), "Specification Testing in Markov-Switching Time Series Models," *Journal of Econometrics*, 70, 127–157.

Harvey, C. R. (1988), "The Real Term Structure and Consumption Growth," *Journal of Financial Economics*, 22 (2), 305–333.

Holst, U., Lindgren, G., Holst, J., and Thuvesholmen, M. (1994), "Recursive Estimation in Switching Autoregressions With a Markov Regime," *Journal of Time Series Analysis*, 15 (5), 489–506.

Jorion, P., and Mishkin, F. (1991), "A Multicountry Comparison of Term-Structure Forecasts at Long Horizons," *Journal of Financial Economics*, 29, 59–80.

Lewis, K. K. (1991), "Was There a 'Peso Problem' in the US Term Structure of Interest Rates: 1979–1982?" *International Economic Review*, 32, 159–173.

Mankiw, N. G., and Miron, J. A. (1986), "The Changing Behavior of the Term Structure of Interest Rates," *The Quarterly Journal of Economics*, 101, 211–228.

Naik, V., and Lee, M. H. (1994), "The Yield Curve and Bond Option Prices With Discrete Shifts in Economic Regimes," mimeo, University of British Columbia.

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

Pfann, G. A., Schotman, P. C., and Tschernig, R. (1996), "Nonlinear Interest Rate Dynamics and Implications for the Term Structure," *Journal of Econometrics*, 74, 149–176.

Sola, M., and Driffill, J. (1994), "Testing the Term Structure of Interest Rates Using a Vector Autoregression With Regime Switching," *Journal of Economic Dynamics & Control*, 18, 601–628.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52 (5), 1973–2002.

Timmerman, A. (2000), "Moments of Markov Switching Models," *Journal of Econometrics*, 96, 75–111.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Zarnowitz, V. (1997), "Some New Lessons From International Cyclical Indicators," in *Business Cycle Indicators*, ed. K. Oppenländer Ashgate Publishing, pp. 175–204.