Regime Switches in Interest Rates

Andrew Ang[1] and Geert Bekaert[2]

March 25, 1998

JEL Classification Codes: C32, C53, E32, E44, G12.

Keywords: regime-switching models, interest rates, term structure of interest rates, business cycles, forecasting

[1]Graduate School of Business, Stanford University. Email: *aang@gsb.stanford.edu*

[2]Graduate School of Business, Stanford University and NBER. Email: *bekaert_geert@gsb.stanford.edu* The authors would like to thank Ken Singleton, Jun Liu and seminar participants at Stanford for comments.

Abstract

Regime-switching models are well suited to capture the non-linearities in interest rates. This paper examines the econometric performance of regime switching models for interest rate data from the US, Germany and the UK. There is strong evidence supporting the presence of regime switches but univariate models are unlikely to yield consistent estimates of the model parameters. Regime-switching models incorporating international short rate and term spread information forecast better, match sample moments better, and classify regimes better than univariate models. We show that the regimes in interest rates correspond reasonably well with business cycles, at least in the US. This may explain why regime-switching models forecast interest rates better than single regime models. Finally, the non-linear interest rate dynamics implied by regime switching models have potentially important implications for the macro-economic literature documenting the effects of monetary policy shocks on economic aggregates. Moreover, the implied volatility and drift functions are rich enough to resemble those recently estimated using non-parametric techniques.

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 U.S. 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 can change over time driven by a Markov state variable (assumed to be unobserved to the econometrician).[1]

Importantly, RS models can accomodate regime-dependent mean reversion in short rates. Mankiw and Miron (1986) among others have argued that the predictive power of the term spread for future short rates in the U.S. is very much a function of the monetary policy regime. In particular, they argue that currently the interest rate smoothing efforts of the Fed make the short rate behave like a random walk and this behavior is the cause of the rejections of the expectations hypothesis observed with recent US data. When a regime switching model is fitted to U.S. data however, Bekaert, Hodrick and Marshall (1998) and Gray (1996) show that such random walk behavior is only true for low interest rates whereas high interest rates show considerable mean reversion. As part of our analysis we show that regime-switching models which have regimes with unit root processes remain stationary as long as there is at least one strictly stationary regime. This property allows regime-switching models to capture the near unit-root persistence in interest rate data.

Despite their economic appeal, regime-switching models are less attractive than one-regime models from an econometric estimation perspective. First, it is hard to test for the presence of regimes because nuisance parameters are present under the null of a single regime model, and with some exceptions there are virtually no such tests reported in the literature.[2] Second, although with the recent work of Gray (1996) and Hamilton (1994) the likelihood construction has been simplified, estimating regime-switching models is difficult. Problems encountered include the existence of multiple maxima for the likelihood making the global hard to find and the unboundedness problem causing the conditional variance in one regime to approach zero as the other one approaches infinity.[3] Finally, often the data do not allow clear regime-classification, that is, the probability of having observed a regime ex-ante may hover around a half. These problems may explain why regime-switching models of interest rates have not enjoyed more success as a building block for term structure models.[4]

[1] Hamilton (1988), Lewis (1991), Evans and Lewis (1994), Sola and Driffill (1994), Gray (1996) and Bekaert, Hodrick and Marshall (1997b) all examine regime switches in interest rates.

[2] Hansen (1992) provides a computationally intensive test for the number of regimes but as to date it has not been applied in the literature. Lam (1990) and Cecchetti, Lam and Mark (1993) use Monte Carlo simulation to obtain an empirical likelihood ratio but in a much simpler framework than what is considered here.

[3] See Gray (1995) for more details.

[4] An interesting exception is Naik and Lee (1994) who develop a continuous-time regime switching model for bond and

In this paper, we provide an in-depth analysis of the econometric properties of regime-switching models for interest rates in the US, Germany and the UK. We start by showing that single regime models are resoundingly rejected by the data by applying the test developed by Hansen (1992) to our regime-switching model. We use two statistical criteria to compare alternative one-regime models of short rates to regime-switching models both with state-dependent and constant transition probabilities. The first criterion investigates the fit of the models with the unconditional moments of the data. One attraction of regime-switching models is that they may accommodate some of the non-linearities recently discovered in interest rates,[5] which may show up in higher order unconditional moments. The dependence of mean reversion on the level of the interest rate may also induce an autocorrelogram that is difficult to match by parsimonious ARMA models. The second criterion concerns the forecasting power of the different models, both for first and second moments.[6] Finally, we propose a new metric to compare the performance of different regime-switching models in identifying the regime over the sample. Our Regime Classification Measure (RCM) uses the simple fact that the (ex-ante or ex-post) probability of observing one of the regimes ought to be close to one at all times when regime classification is perfect.

Given the econometric problems mentioned above, it is not *a priori* clear that regime-switching models will perform well on the statistical criteria, even when they are the true data-generating process. Moreover, as stressed by Bekaert, Hodrick and Marshall (1998), 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 will be biased. 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 regime-switching models in small samples, if the parameter estimates of the regime-switching models are severely biased and inefficient.

To help overcome these problems, we extend the effective sample size through two channels. First, we investigate multi-country systems of interest rates. It is possible that short rates in the US 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

option pricing.

[5]There is a growing literature documenting the non-linearities in interest rates. For example, see Conley, Hansen, Luttmer and Scheinkman (1997), Boudoukh, Richardson, Stanton and Whitelaw (1997), Stanton (1997), Pfann, Schotman and Tschernning (1996) and Aït-Sahalia (1996).

[6]Gray (1996) examines the out-of-sample forecasting power of a regime-switching model for second moments of the U.S. short rate on weekly data and Engel (1994) examines the out-of-sample first moment performance of exchange rate regime-switching models.

regimes. We compare the performance of several variants of the multivariate regime-switching models to their single regime vector-autoregressive (VAR) counterparts.

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. We again compare the performance of several variants of the multivariate regime-switching models to their VAR counterparts. The moments criterion here include the cross-correlations between short rates and spreads. As Pfann, Schotman and Tscherning (1996) show, 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. To further analyze the non-linearities in the term structure captured by regime-switching models, we look at impulse responses following Gallant, Rossi and Tauchen (1993) and plot the drift and volatility functions implied by our model.

Apart from a number of methodological contributions, this article offers some important empirical results. First, there *are* several regimes in US, German and UK short rates. Second, RS models, despite being difficult to estimate, forecast well out of sample but do poorly at matching sample moments. Multivariate RS models perform better than univariate models in terms of regime classification and forecasting. Third, our analysis of the non-linearities implied by RS models shows that the impulse responses of shocks correspond very closely to the impulse responses from linear models when averaged over the regimes, but shocks conditional on different regimes produce impulse responses which are quite dissimilar. The non-linear conditional drifts and volatilities from RS models are similar to the drift and volatility functions estimated by a number of non-parametric studies. Finally, the regime classification implied by RS models is closely related to economic business cycles and the regime ex-ante probabilities are good short-horizon predictors of the business cycles in the US.

The paper 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. It presents a general multivariate RS model and considers as special cases univariate short rate models, multi-country 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 can generally not be consistently estimated. Section 3 also presents our diagnostic statistics. Section 4 gives the empirical estimation results and formally tests for the number of regimes present in the data. Section 5 discusses the performance of the various models. To help interpret the results we perform a Monte Carlo experiment that examines the performance of single regime and regime-switching models in small samples when the true data generating process is a regime-switching model. We consider the quality of regime classification and ask if the regimes are related to the business cycle. Section 6 explores the non-linear dynamics implied by the term structure RS models. Section 7 concludes.

2 Data and Stylized Facts

Our empirical work uses monthly observations on 3 month short rates and 5 year long rates of zero coupon bonds from the US, Germany and Great Britain from January 1972 to August 1996. The data is an updated set of the Jorion and Mishkin (1991) data series.[7] 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. This gives an in-sample period of 267 observations.

Table (1) gives 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 Great Britain do not show excess kurtosis. Short rates are very persistent, with the UK showing the least persistence. Spreads are also autocorrelated, but less so than short rates.

Turning to international cross-correlations , lagged short rates of the US are more highly correlated with current German and UK rates than present levels of US short rates. This suggests that lagged US short rates may Granger-cause movements in short rates in Germany and the UK. The contemporaneous correlations of short rates across countries are not very high except for the US and UK rates but they are significantly different from zero and 1. The correlations between spreads and short rates are highly negative but they remain significantly different from -1. This indicates that the domestic term structure is not driven by a one factor model.

In Table (2) we attempt to determine whether the behavior of the term structure depends on the business cycle.[8] 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) notes, only the US has a business cycle history which 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 UK with caution.

Focusing on the country with the best cycle dating, the US, 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 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. 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 UK. In these countries autocorrela-

[7] See Bekaert, Hodrick and Marshall (1998) for further details.

[8] The dates of NBER business cycle expansions and contractions for the US can be accessed at http://www.nber.org/cycles.html, dates for Germany and the UK are from the Center for International Business Cycle Research at Columbia University (CIBCR). We thank Risk Mishkin for providing us with the latter. For details of the methodology of CIBCR dating see Zarnowitz (1997).

tions of the short rate and spread are not significantly different across the business cycle. In Germany the patterns are similar to the US except for mean reversion which is insignificantly higher in expansions. In the UK, 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 the US, the differences across cycles are not statistically significant.

Finally, in the US and UK 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:

$$\rho(r_t^l, r_t) = \frac{\rho(z_t, r_t)}{w} + 1 \tag{1}$$

where $w = \sigma(r_t)/\sigma(r_t^l)$, which is greater 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.

For the US, the picture that emerges from our results is one where in expansions short rates are more persistent, the long rate is not as sensitive to short rate shocks and the short rate and spread are more negatively correlated. 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. In Germany and the UK, the correlations of the short rate and spread are also not significantly different, but the UK has a low p-value with the same pattern as the US.

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 short rate changing across the business cycle. Third, mean reversion in the US is significantly different across economic regimes. These patterns can potentially be accomodated in models which contain a regime variable.

3 The Empirical and Econometric Framework

3.1 A General Multivariate Regime Switching Model

We describe a general multivariate regime switching (RS) model of short rates and spreads. Let $r_t = (r_t^{us}\, r_t^{ger}\, r_t^{uk})'$, $z_t = (z_t^{us}\, z_t^{ger}\, z_t^{uk})'$ and $y_t^* = (r_t'\, z_t')'$. We assume the standard filtration is generated using only these present and lagged variables. Our most general model for a regime switching VAR of n lags is:

$$y_t^* = \mu(s_t^*) + A_1(s_t^*)y_{t-1}^* + \cdots + A_n(s_{t-n}^*)y_{t-n}^* + \Sigma_{t-1}^{\frac{1}{2}}(s_t^*)\epsilon_t \tag{2}$$

where we have Markov transition probabilities for k states at every time period t, denoted by s_t^* and $\epsilon_t \sim$ IID $N(0, I)$. To write in companion form let $y_t = (y_t^{*'}\ y_{t-1}^{*'}\ \ldots y_{t-n+1}^{*'})'$:

$$y_t = \mu(s_t) + A(s_t)y_{t-1} + \Sigma_{t-1}^{\frac{1}{2}}(s_t)u_t \tag{3}$$

where we redefine the state space so that the new state variable s_t takes on one of the $K = k^n$ values representing the k^n possible combinations for $s_t^*, s_{t-1}^*, \ldots, s_{t-n}^*$, y_t^* is a regime-switching VAR(n) in equation (2), $v(s_t) = (\mu(s_t)\, 0 \ldots 0)'$ is a $6n$x1 vector, $A(s_t)$ is the companion form given the state s_t, and $u_t = (\epsilon_t\, 0 \ldots 0)'$ is a $6n$x1 vector. By redefining the state space this way, y_t in equation (3) depends only on the current regime s_t.[9]

The Markov transition probabilities for $i = 1, \ldots, K$ states may be functions of lagged endogenous variables. For example, we can specify probabilities logistically as:[10]

$$p(s_t = i | s_{t-1} = j, \mathcal{I}_{t-1}) = \frac{e^{\alpha_{i,j} + \beta'_{i,j} y_{t-1}}}{1 + e^{\alpha_{i,j} + \beta'_{i,j} y_{t-1}}} \tag{4}$$

Let $\tilde{y}_T = (y'_T\, y'_{T-1} \ldots y'_1\, y'_0)'$ and denote the parameters of the likelihood by θ. Then we can write the conditional likelihood as:

$$f(\tilde{y}_T; \theta) = \prod_{t=1}^{T} f(y_t | \mathcal{I}_{t-1}; \theta) \tag{5}$$

Extending Gray (1996)'s methodology to multivariate conditional distributions we have:

$$\begin{aligned} f(\tilde{y}_T; \theta) &= \prod_{t=1}^{T} f(y_t | \mathcal{I}_{t-1}; \theta) && (6) \\ &= \prod_{t=1}^{T} \left(\sum_{i=1}^{K} f(y_t | \mathcal{I}_{t-1}, s_t = i; \theta) p(s_t = i | \mathcal{I}_{t-1}; \theta) \right) && (7) \end{aligned}$$

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

$$p(s_t = i | \mathcal{I}_{t-1}; \theta) = \sum_{j=1}^{K} p(s_t = i | s_{t-1} = j, \mathcal{I}_{t-1}; \theta) p(s_{t-1} = j | \mathcal{I}_{t-1}; \theta) \tag{8}$$

where the first term in the sum is the transition probability which can be state-dependent, and the second term may be decomposed by Bayes' Rule as:

$$\begin{aligned} p(s_{t-1} = j | \mathcal{I}_{t-1}; \theta) &= \frac{f(y_{t-1}, s_{t-1} = j | \mathcal{I}_{t-2}; \theta)}{f(y_{t-1} | \mathcal{I}_{t-2}; \theta)} \\ &= \frac{f(y_{t-1} | s_{t-1} = j, \mathcal{I}_{t-2}; \theta) p(s_{t-1} = j | \mathcal{I}_{t-2}; \theta)}{\sum_{m=1}^{K} f(y_{t-1} | s_{t-1} = m, \mathcal{I}_{t-2}; \theta) p(s_{t-1} = m | \mathcal{I}_{t-2}; \theta)} \end{aligned} \tag{9}$$

[9] A similar re-parameterization was proposed in Gray (1995).

[10] The specification of time-varying transition probabilities as a logistic form was first introduced by Diebold, Lee and Weinbach (1994) and is now standard. For example, see Gray (1996), and Bekaert, Hodrick and Marshall (1998).

We start the algorithm using equation (9) with $p(s_1 = i|\mathcal{I}_0)$, the stable probabilities of the system at $t = 1$ which are given by:

$$\pi_i = \frac{X_{ii}}{\sum_{j=1}^{K} X_{jj}} \tag{10}$$

where X_{ii} is the ii^{th} cofactor of the matrix $X = I - P_1$, and P_1 is the KxK 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 stable probabilities π of the transition matrix P which solve $\pi P = \pi$.

3.2 Special Cases

Since the regime-variable is unobserved to the econometrician and must be factored out of the likelihood function, it is relevant to ask under what conditions we can obtain inefficient but consistent estimates when ignoring some variables.[11] Let Z_t represent variables which do not enter into our estimation and X_t represent variables which do, so $y_t = (Z_t' \, X_t')'$. Then using conditioning arguments we can write:

$$\begin{aligned}
f(\tilde{y}_T;\theta) &= \prod_{t=1}^{T} f(y_t|\mathcal{I}_{t-1};\theta) &(11)\\
&= \prod_{t=1}^{T}\left(\sum_{i=1}^{K} f(y_t|s_t = i, \mathcal{I}_{t-1};\theta)p(s_t = i|\mathcal{I}_{t-1};\theta)\right) &(12)\\
&= \prod_{t=1}^{T}\left(\sum_{i=1}^{K} f(Z_t|X_t, s_t = i, \mathcal{I}_{t-1};\theta)f(X_t|s_t = i, \mathcal{I}_{t-1};\theta)p(s_t = i|\mathcal{I}_{t-1};\theta)\right) &(13)
\end{aligned}$$

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

$$f(Z_t|X_t, s_t = i, \mathcal{I}_{t-1};\theta) = f(Z_t|X_t, \mathcal{I}_{t-1};\theta) \tag{14}$$

We parameterize the model so that $\theta = (\theta_Z' \, \theta_X')'$ and $\{\theta_Z\} \cap \{\theta_X\} = \phi$, 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 state depends only on θ_X:

$$p(s_t = i|\mathcal{I}_{t-1};\theta) = p(s_t = i|\mathcal{I}_{t-1};\theta_X) \tag{15}$$

Then the likelihood can be written as:

$$\mathcal{L}(\tilde{y}_T;\theta) = \sum_{t=1}^{T} \ln f(Z_t|X_t, \mathcal{I}_{t-1};\theta_Z) + \sum_{t=1}^{T} \ln\left(\sum_{i=1}^{K} f(X_t|s_t = i, \mathcal{I}_{t-1};\theta_X)p(s_t = i|\mathcal{I}_{t-1};\theta_X)\right) \tag{16}$$

[11] See Gray (1995) and Bekaert and Gray (1998) for similar arguments.

Maximizing the second sum in equation (16) then yields consistent but inefficient estimates relative to full information maximum likelihood.

Estimation of the full system is infeasible, given the dimension of θ. We must 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 either real rates or expected inflation or business cycles are the source for potential regimes in nominal interest rates.[12] 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 independent across countries. It is conceivable that there is a "world business cycle"[13] driving interest rates in many countries simultaneously and in some models we consider we will 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 should suffice to capture the main empirical non-linearities shown in Section 2. Moreover, most of the past regime-switching literature has focused on two-state models.[14]

Since most of the RS literature also focuses exclusively on univariate interest rate models,[15] we start by analysing univariate short rate models for the US, Germany and UK. To consistently estimate univariate short rate RS models there must be no further information about the regime contained in the short rates or term spreads from other countries. If regimes capture business cycle effects, the different correlations in the US across cycles in Table (2) violate the assumptions needed to produce consistent estimation.

Incorporating the extra information from international and term structure data allows us to weaken the implicit assumptions but this makes estimation much more complex. In a second set of models, we add information from the short rates from other countries. In our multi-country model (below), defining the regime variable s_t becomes more involved as it will embed 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. Note that we model the term spread empirically without imposing theoretical restrictions from a pricing model as in Naik and Lee (1994). Such restrictions would probably facilitate the identification of the model parameters, but at the same time may overly constrain the model structure. It is unlikely that they capture the regime-dependent patterns in correlation and volatility we observe in the data as sucessfully as unconstrained models. Moreover, the resulting model is likely analytically intractable and very hard to estimate.

[12] Garcia and Perron (1996) and Evans and Lewis (1995) consider Markov regimes in inflation and real interest rates.

[13] See for example, Lumsdaine and Prassad (1997).

[14] Garcia and Perron (1996) and Bekaert, Hodrick and Marshall (1998) estimate three state RS models.

[15] Exceptions include Sola and Driffill (1994) and Evans (1995).

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 looking at the response of the term structure to various shocks,[16] suggests that the spread contains additional independent information which may help in the classification of regimes. For example, Eichenbaum, Evans and Marshall (1996) show 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 (1995) find that the spread is useful in predicting future activity, and the spread contains predictive information which 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 can only combine spreads and short rates in a multi-country model under severe constraints on the parameters, but we do attempt to estimate such a model. We assume independence of the states across countries and employ a cross sectional estimation. Viewing each country as an independent draw of the data generating process means we can take advantage of the increased sample size in order to draw inference from our regime switching model.

Table (3) presents a summary of the models estimated, their abbreviations used throughout the paper and the number of parameters in parentheses. We now outline each of these models briefly.

3.2.1 Univariate Models

We consider univariate regime-switching AR(1) processes because when fitting one-regime ARMA(p,q) models in Section 4 we find that the best model is an ARMA(1,0) using both AIC and BIC criteria.[17] Consequently we adopt an AR(1) conditional mean specification for our univariate RS models. For country m these are special cases of the following general model considered in Gray (1997):

$$r_t^m = \mu(s_t) + \rho(s_t) r_{t-1}^m + h^m(s_t)\epsilon_t \tag{17}$$

or equivalently:

$$\Delta r_t^m = \mu(s_t) - \beta(s_t) r_{t-1}^m + h^m(s_t)\epsilon_t \tag{18}$$

$$= f(s_t, r_{t-1}^m) + h^m(s_t)\epsilon_t \tag{19}$$

[16] These empirical papers typically investigate Impulse Responses estimated from Vector Autoregressions. Often the focus is on the effect of monetary shocks. In addition to the papers referred to in the text, see Evans and Marshall (1997) and Eichenbaum and Evans (1995).

[17] Our estimation uses conditional maximum likelihood, conditioning on the initial data points and setting the initial lagged errors to zero. Akaike's Information Criterion (AIC), and the Bayesian Information Criterion (BIC), also called the Schwarz Criterion are outlined in Judge et al. (1980) and Lütkepohl (1993).

where $\beta = 1 - \rho$, $f(s_t, r^m_{t-1})$ is the conditional drift and $h^m(s_t)$ is the conditional volatility, and the errors $\epsilon_t \sim$ IID $N(0,1)$. The conditional volatility is specified as:

$$(h^m_t(s_t))^2 = a_0(s_t) + a_1(s_t)u^2_{t-1} + b_1(s_t)(h^m_{t-1})^2 + b_2(s_t)(r^m_{t-1})^2 \tag{20}$$

$$\text{where} \qquad (h^m_t)^2 = E_{t-1}[(r^m_t)^2] - (E_{t-1}[r^m_t])^2 \tag{21}$$

$$u_t = r^m_t - E_{t-1}[r^m_t] \tag{22}$$

The regime variable s_t is either 1 or 2, and has transition probabilities

$$p(s_t = j | s_{t-1} = j) = \frac{e^{a_j + b_j r^m_{t-1}}}{1 + e^{a_j + b_j r^m_{t-1}}},\ j = 1, 2. \tag{23}$$

We will denote constant transition probabilities as P and Q for $j = 1, 2$ respectively. Denoting $p_{t,j} = p(s_t = j | \mathcal{I}_{t-1})$, we can evaluate $E_{t-1}[r^m_t]$ and $E_{t-1}[(r^m_t)^2]$ as:

$$E_{t-1}[r^m_t] = \sum_{j=1}^{2} p_{t,j}(\mu_j + \rho_j r^m_{t-1}) \tag{24}$$

$$E_{t-1}[(r^m_t)^2] = \sum_{j=1}^{2} p_{t,j} \left((\mu_j + \rho_j r^m_{t-1})^2 + (h^m_{t-1,j})^2\right) \tag{25}$$

where subscripts indicate the state $s_t = j$.

The special cases we consider involve setting $a_1 = b_1 = b_2 = 0$ (RS AR(1)), $b_2 = 0$ (RS GARCH(1,1)), $a_0 = a_1 = b_1 = 0$ (RS CIR). The last model is the RS equivalent of the discretized square root model of Cox, Ingersoll and Ross (1985).[18]

In practice, many RS models yield regimes with unit-root or near unit-root processes, and other regimes are more mean-reverting. It is important to ensure that such a process retains covariance stationarity:

Proposition 3.2.1 *Consider a univariate n-state Markov regime-switching model, with constant transition probability matrix P. The Markov chain is ergodic and the stable probabilities satisfy $\pi = P\pi$. We arbitrarily order the regimes so that the first k regimes follow unit root processes, $k < n$, and the other $n - k$ regimes follow stationary processes. The variance conditional on each regime is assumed to be constant. The stable probabilities corresponding to the unit root and stationary processes can be written as $\pi = (\pi'_k\ \pi'_{n-k})'$. Then if π_{n-k} contains a strictly positive element the overall process is (covariance) stationary.*

[18] In continuous time the CIR model has the form: $dr_t = \theta(\kappa - r_t)dt + \sigma r^{\frac{1}{2}} dB_t$. The discretization used here is standard (For example see Chan, Karolyi, Longstaff and Sanders (1992) and Pearson and Sun (1994)). We note that the discretized model allows short rates to be negative with probability 1, and this is inconsistent with the square root of the short rate appearing in the conditional volatility. However, the discretization is satisfactory for purposes of econometric estimation. As interest rates fall the upward drift tends to dominate and this property makes it hard for interest rates to go negative. In continuous time, negative interest rates are ruled out by parameter restrictions. (See Cox, Ingersoll and Ross (1985).)

Proof: See Appendix.

Intuitively, to obtain stationarity we need the unconditional autocorrelation to be strictly less than one. Although some regimes have unit roots, the presence of at least one stationary regime ensures that the unconditional autocorrelation is less than one as long as the probability of transitioning into the stationary regime is greater than zero.

3.2.2 Multi-Country Models

To motivate our RS multi-country models for $r_t = (r_t^{us}\ r_t^{ger}\ r_t^{uk})'$, we first consider one-regime VAR's. Using AIC and BIC criteria the optimal lag length is 1. Hence we consider the following general multi-country RS model:

$$\begin{pmatrix} r_t^{us} \\ r_t^{ger} \\ r_t^{uk} \end{pmatrix} = \begin{pmatrix} \alpha^{us}(s_t^{us} = i) \\ \alpha^{ger}(s_t^{ger} = j) \\ \alpha^{uk}(s_t^{uk} = k) \end{pmatrix} + A(s_t^{us} = i, s_t^{ger} = j, s_t^{uk} = k) \begin{pmatrix} r_{t-1}^{us} \\ r_{t-1}^{ger} \\ r_{t-1}^{uk} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{us} \\ \epsilon_t^{ger} \\ \epsilon_t^{uk} \end{pmatrix} \tag{26}$$

with $\epsilon_t = (\epsilon_t^{us}\ \epsilon_t^{ger}\ \epsilon_t^{uk})' \sim$ IID $N(0, \Sigma(s_t^{us} = i, s_t^{ger} = j, s_t^{uk} = k))$.

We assume that there are two states 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 parsimonious as possible, we do not consider state-dependent transition probabilities in the multi-country model. To estimate we effectively enlarge the state-space. The algorithm given here is a multivariate generalization of Gray (1996).

The Markov transition process within each country is assumed to be unaffected by the regimes in another country. Formally, for countries λ, μ, ν, with S^μ denoting the past history of states for country μ, $S_t^\mu = \{s_t^\mu, s_{t-1}^\mu, \ldots\}$:

$$p(s_t^\mu | S_t^\lambda, S_t^\mu, S_t^\nu) = p(s_t^\mu | S_t^\mu) = p(s_t^\mu | s_{t-1}^\mu) \tag{27}$$

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 2 states for 3 countries, it is possible to enlarge the state space to $2^3 = 8$ states, where the states are defined as $s_t = 1, \ldots, 8$:

s_t	US	GER	UK
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 can then calculate an 8x8 transition matrix, where for example, $p(s_t = 1|s_{t-1} = 1) = P^{us}P^{ger}P^{uk}$.

Let $r_t = (r_t^{us}\, r_t^{ger}\, r_t^{uk})'$. Then we can write equation (26) as:

$$r_t = \alpha(s_t) + A(s_t)r_{t-1} + \epsilon_t \tag{28}$$

with the states now redefined as $s_t = 1, \ldots, 8$. 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 equivalent RS VAR is infeasible. To gain efficiency we test whether some parameters are identical in the one-regime VAR. We test for Granger-causality on each country's short rates and test if parameters of the data generating process are constant across countries. We find just-significant evidence of Granger-causality of short rates of Germany and the UK by the US, but not vice versa. We also find that we cannot reject the hypothesis that ρ_i is constant across countries and we also impose this on our formulation. The tests are further detailed in Section 4. Tests of Granger-causality lead us to consider two formulations of A_i, a diagonal formulation where $A_i = \begin{pmatrix} \rho_i^{us} & 0 & 0 \\ 0 & \rho_i^{ger} & 0 \\ 0 & 0 & \rho_i^{uk} \end{pmatrix}$ and a Granger-causality formulation where $A_i = \begin{pmatrix} \rho_i^{us} & 0 & 0 \\ \zeta_i^{ger} & \rho_i^{ger} & 0 \\ \zeta_i^{uk} & 0 & \rho_i^{uk} \end{pmatrix}$. We will refer to these as A_i^1 and A_i^2 respectively.

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

$$\begin{pmatrix} \epsilon_{t,i}^{us} \\ \epsilon_{t,i}^{ger} \\ \epsilon_{t,i}^{uk} \end{pmatrix} = \begin{pmatrix} h_{t-1,i}^{us} u_t^1 \\ h_{t-1,i}^{ger} u_t^2 + \gamma_i^{ger} u_t^1 \\ h_{t-1,i}^{uk} u_t^3 + \gamma_i^{uk} u_t^1 \end{pmatrix} \tag{29}$$

where $(u_t^1\; u_t^2\; u_t^3)'$ are drawn from a IID $N(0, I(3))$ 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 US also shock the interest rates of Germany and the UK, but not vice versa. Another interpretation along the lines of a world business cycle is that there are "world" shocks which drive the dominant US economy while Germany and the UK 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 US in the world economy such a structure seems reasonable and below we test its statistical significance. The conditional covariance matrix, conditional on the state $s_t = i$ is given by:

$$\Sigma_{i,t} = E[\epsilon_t \epsilon_t' | \mathcal{I}_{t-1}, s_t = i] = \begin{pmatrix} (h_{t-1,i}^{us})^2 & \gamma_i^{ger} h_{t-1,i}^{us} & \gamma_i^{uk} h_{t-1,i}^{us} \\ \gamma_i^{ger} h_{t-1,i}^{us} & (h_{t-1,i}^{ger})^2 + (\gamma_i^{ger})^2 & \gamma_i^{ger}\gamma_i^{uk} \\ \gamma_i^{uk} h_{t-1,i}^{us} & \gamma_i^{uk}\gamma_i^{ger} & (h_{t-1,i}^{uk})^2 + (\gamma_i^{uk})^2 \end{pmatrix} \tag{30}$$

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

We can now write a distribution for r_t conditional on both the state $s_t = i$ and $\mathcal{I}_{t-1}$:

$$f(r_t|s_t = i, \mathcal{I}_{t-1}) = (2\pi)^{-\frac{3}{2}}|\Sigma_{i,t}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(r_t - \mu_{i,t})'\Sigma_{i,t}^{-1}(r_t - \mu_{i,t})\right) \tag{31}$$

where the conditional mean is given by:

$$\mu_{i,t} = \alpha(s_t = i) + A(s_t = i)r_{t-1} \tag{32}$$

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 can 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

Empirical one-regime VAR models of of $y_t^m = (r_t^m \, z_t^m)'$, the short rate and spread for country m lead us to consider a one lag RS model:

$$y_t^m = \mu(s_t) + A(s_t)y_{t-1}^m + u_t \tag{33}$$

where $u_t \sim N(0, \Sigma(s_t))$. We use 2 states, with constant transition probabilities, and also logistic state-dependent transition probabilities where:

$$p(s_t = j|s_{t-1} = j) = \frac{\exp(a_j + b_j r_{t-1}^m + c_j z_{t-1}^m)}{1 + \exp(a_j + b_j r_{t-1}^m + c_j z_{t-1}^m)} \quad j = 1, 2. \tag{34}$$

We estimate the Cholesky decomposition $R(s_t)$ of $\Sigma(s_t)$. Joint estimations of the models over all 3 countries using a cross-sectional approach following Bekaert, Hodrick and Marshall (1998) are also performed.

3.3 Model Diagnostics

To evaluate the models we consider criteria measuring the fit of the unconditional moments implied by the models to the sample estimates of the unconditional moments, out-of-sample forecast errors and the quality of the regime classification. These will be discussed in turn.

3.3.1 Unconditional Moment Comparisons

We compute the unconditional population moments of our various models using analytical expressions for one-regime VAR models and univariate one-regime GARCH and CIR processes, but using a simulation for the RS models. Because of the high persistence of the series, sample sizes of one million are needed to pin down the unconditional moments to the second decimal place.

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

$$H = (h^* - \bar{h})'\Sigma_h^{-1}(h^* - \bar{h}) \tag{35}$$

where $\bar{h}$ are sample estimates of unconditional moments, h^* are the unconditional moments from the estimated model, and Σ_h is the covariance matrix of the sample estimates of the unconditional moments. Σ_h can be obtained from a GMM estimation of the unconditional moments, and for the purposes of this paper, we use a Newey-West (1987) estimate with 6 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 $\bar{h}$ will contain the mean, variance, skewness and kurtosis; for the autocorrelogram the first 10 autocorrelations; and for cross-correlations lags from -3 to +3. We also introduce a related statistic H^*, which uses as a weighting matrix the diagonal of Σ_h. Strong correlations between the estimated moments sometimes imply that the model minimizing H does not minimize H^*.

3.3.2 Forecast Comparisons

For RS multivariate VAR models $y_t = v(s_t) + A(s_t)y_{t-1} + u_t$ with K states, we can calculate forecasts:

$$E_{t-1}(y_t) = \sum_{i=1}^{K} p_{it}(v_i + A_i y_{t-1}) \tag{36}$$

$$E_{t-1}(y_t y_t') = \sum_{i=1}^{K} p_{it}(v_i v_i' + v_i y_{t-1}' A_i' + A_i y_{t-1} V' + A_i y_{t-1} y_{t-1}' A_i' + \Sigma_{i,t}) \tag{37}$$

where p_{it} is the ex-ante probability $p(s_t = i|\mathcal{I}_{t-1})$ and $\Sigma_{i,t}$ is the covariance matrix of u_t in state i. The ex-ante probability can be determined recursively using equation (9).

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:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum(\phi_t^k - \hat{\phi}_t^{\,k})^2} \tag{38}$$

$$\text{MAD} = \frac{1}{T}\sum|\phi_t^k - \hat{\phi}_t^{\,k}| \tag{39}$$

where hatted values denote conditional forecast values. The RMSE criterion is based on Granger's (1969) result that conditional expectation is the optimal predictor under the mean squared error criterion if the underlying process is Gaussian. The MAD uses a linear penalty rather than a quadratic one. In our application we let $\phi_t = r_t$ for univariate and multi-country 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 $\phi_t = r_t z_t$.

3.3.3 Regime Classification

Previous specification tests for regime-switching models have focused on properties of residuals,[19] but here we propose a summary point statistic which captures the quality of regime classification. Define the regime classification measure, RCM, statistic for two states as:

$$RCM = 400 * \frac{1}{T} \sum_{t=1}^{T} p_t (1 - p_t) \tag{40}$$

where p_t is the ex-ante regime probability $p(s_t = 1 | \mathcal{I}_{t-1})$.[20] The constant serves to normalize the statistic to be between 0 and 100 for two states. Although the state variable s_t is unobserved, RS models can produce probability inferences about being in a particular regime through ex-ante probabilities. Weak regime inference would imply that the RS model cannot successfully distinguish between regimes from the behavior of the data and may indicate misspecification. An ideal regime switching model would classify regimes sharply so p_t would be close to one or zero; inferior models would have p_t hover close to a half. Good regime classification is associated with low RCM statistic values: a value of 0 gives perfect regime classification and a value of 100 implies that no information about the regimes is revealed.

Note that the statistic easily generalizes to multiple regimes. A general definition of the statistic for K regimes is:

$$RCM(K) = 100K^2 \frac{1}{T} \sum_{t=1}^{T} \left(\prod_{i=1}^{K} p_{i,t} \right) \tag{41}$$

where $p_{i,t} = p(s_t = i | \mathcal{I}_{t-1})$.

4 Empirical Results

In this section we summarize the results and conduct tests for the number of regimes.[21]

4.1 Are there Regimes in the Data?

4.1.1 Tests for the Number of Regimes

Since most of the RS literature focuses on univariate models and because of computational burdens, we conduct tests for the number of regimes on univariate short rate models. In testing RS models, the usual hypothesis tests (likelihood ratio, Wald, Lagrange multiplier) are not valid because they do not have the usual standard χ^2 asymptotic distribution. This results from the presence of nuisance parameters under the null.[22] For instance, in testing a one regime versus a two regime AR(1) specification, if we consider

[19] See Hamilton (1996) and Gray (1996).

[20] Alternatively we could use the smoothed probability over the entire sample, $p(s_t = 1 | \mathcal{I}_T)$.

[21] The parameter estimates for all the estimated RS models are available from the authors as an Appendix upon request.

[22] Various methods have been developed to deal with nuisance parameters in specific situations, such as those by Davies (1977, 1987), Gallant (1977) and Hansen (1996). Some of these are applied by Garcia and Perron (1996).

the null hypothesis $P = 1$ then $\gamma = (\mu_2, \rho_2, \sigma_2, Q)$ is unidentified under the null given that we start the process in the first regime, i.e. $p(s_0 = 1) = 1$. The nuisance parameters γ cause the likelihood function to be flat with respect to these parameters under the null and the Hessian (information matrix) to depend on the unidentified nuisance parameters. To test for the number of regimes, we use a test developed by Hansen (1992), which is much cited in the literature but rarely implemented because of its computational complexity. In addition we use Monte Carlo simulation to get a distribution of the empirical likelihood ratio statistic.

Hansen uses empirical process theory to bound the asymptotic distribution of a suitably standardized likelihood ratio statistic which is applicable when the assumptions of standard theory are violated. To formulate the test, let $\alpha = (\beta' \, \gamma')'$ where we wish to test the null hypothesis $\beta = 0$, and γ is the vector of nuisance parameters unidentified under the null and let θ be the parameters identified under both the null and the alternative hypothesis. To deal with the parameters θ, we concentrate them out of the sample likelihood function $L_T(\beta, \gamma, \theta) = \sum l_t$. Let

$$\hat{\theta}(\alpha) = \arg\max_{\theta} L_T(\alpha, \theta) \tag{42}$$

We define, as in Hansen (1992):

$$LR_T(\alpha) = L_T(\alpha, \hat{\theta}(\alpha)) - L_T(0, \gamma, \hat{\theta}(0, \gamma)) \tag{43}$$

$$q_t(\alpha, \hat{\theta}(\alpha)) = l_t(\alpha, \hat{\theta}(\alpha)) - l_t(0, \gamma, \hat{\theta}(0, \gamma)) - \frac{1}{T} LR_T(\alpha) \tag{44}$$

$$V_T(\alpha) = \sum_{t=1}^{T} q_t(\alpha, \hat{\theta}(\alpha))^2 \tag{45}$$

$$LR_T^* = \sup_{\alpha} \frac{LR_T(\alpha)}{\sqrt{V_T(\alpha)}} \tag{46}$$

Hansen shows that $p(LR_T^* \geq x)$ is bounded by an asymptotic distribution $p(\sup_\alpha Q_T^* \geq x) \rightarrow p(\sup_\alpha Q^* \geq x)$, where the distribution Q_T^* is defined by:

$$Q_T^* = \frac{LR_T(\alpha) - E[LR_T(\alpha)]}{\sqrt{V_T(\alpha)}} \tag{47}$$

Under the null hypothesis $E(LR_T(\alpha)) \leq 0$. Hansen assumes an empirical Central Limit Theorm holds so that $Q_T^*(\alpha) \rightarrow Q^*(\alpha)$, a Gaussian process with a known covariance function. The distribution $Q^*(\alpha)$ can be produced by simulation. The supremum itself is taken over all possible values of α. This makes the test extremely computationally intensive. In practice, the supremum must be taken over a finite grid. Most of the estimation time is spent concentrating out the likelihood function at every grid point.

In addition to Hansen's test, we employ Monte Carlo simulation to simulate a simple AR(1) model with parameters equal to the estimated parameters of each country and estimate the RS AR(1) model on each of these samples. Without the nuisance parameters γ, the likelihood ratio $2(L(\beta, \gamma, \theta) - L(0, \gamma, \theta))$ would have a χ^2 distribution with degrees of freedom equal to the number of restrictions in $\beta = 0$.

However, in the presence of nuisance parameters this will not be χ^2. The empirical distribution of the likelihood function can then be used to calculate the empirical p-value of the likelihood ratio statistic under the null hypothesis of no switching. Estimation of the RS model for every simulated sample also makes this computationally intensive.

4.1.2 Empirical Results for Tests of the Number of Regimes

We test the null $H_0 : r_t = \mu + \rho r_{t-1} + \sigma \epsilon_t$ of one regime against the alternative $H_1 : r_t = \mu(s_t) + \rho(s_t) r_{t-1} + \sigma(s_t)\epsilon_t$, with $\epsilon_t \sim$ IID $N(0,1)$, with two states. The Markov transition probabilities for remaining in state 1 (2) are denoted by P (Q). The null hypothesis is then equivalently expressed as $H_0 : P = 1$ against $H_0 : P < 1$. The algorithm we use starts the estimation in the stable probabilities of the system, $p(s_t = 1) = \frac{1-Q}{2-P-Q}$, so under the null we start in the first regime. Under the null, all parameters associated with the second regime are unidentified. For a regime switching AR(1) process in the notation of the last section we have $\beta = P - 1$, $\gamma = (\mu_2, \rho_2, \sigma_2, Q)$ and $\theta = (\mu_1, \rho_1, \sigma_1)$.

Our estimation procedure proceeds over a (coarse) finite grid as there is considerable computation time in concentrating out θ over the grid. We considered two grids for $\alpha = (\beta', \theta')'$, these being:

$$\text{grid 1} = \begin{cases} P = 0.2,\ 0.4,\ 0.6,\ 0.8 \\ \mu_2 = 0.0,\ 0.5,\ 1.0,\ 1.5 \\ \rho_2 = 0.97,\ 0.98,\ 0.99 \\ \sigma_2 = 0.2,\ 0.8,\ 1.2 \\ Q = 0.,2\ 0.4,\ 0.6,\ 0.8 \end{cases} \qquad \text{grid 2} = \begin{cases} P = 0.2,\ 0.4,\ 0.6,\ 0.8 \\ \mu_2 = 0.0,\ 0.3,\ 0.6,\ 0.9,\ 1.2,\ 1.5 \\ \rho_2 = 0.97,\ 0.98,\ 0.99 \\ \sigma_2 = 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0,\ 1.2,\ 1.4 \\ Q = 0.2,\ 0.4,\ 0.6,\ 0.8 \end{cases} \tag{48}$$

Grid 1 has 576 points, and grid 2 has 2016 points. Table (4) reports the LR_T^* estimates and the p-values using $\sup_\alpha Q_T^*$ calculated by 1000 simulations following Hansen (1992). The Table shows results that look suspiciously too good. Hansen's test unequivocally rejects the null of one regime for all countries with p-values of zero. To check the test under the null and alternative, we simulated a simple AR(1) and one RS AR(1) process both with the same sample size as the data sets for the US, Germany and UK. We would hope that under the AR(1) simulation, Hansen's test will fail to reject, and that under the RS simulation the test will reject outright. This is indeed the case, with the AR(1) producing a p-value of 0.9670 for grid 1 and 0.5880 for grid 2. The RS simulation results in outright rejections for both grids.

To our knowledge this is the first application of Hansen's test other than the original study, which focused on Hamilton's (1989) work on GDP. Our application is far more onerous, involving a non-linear autoregressive specification, with switching variances. Despite the coarse grid, the test performs extremely well and resoundingly rejects the null of one regime.

The empirical likelihood ratio test also overwhelmingly endorses regime switching. The likelihood ratio test statistics for US, Germany and UK are 215.21, 173.55 and 153.37 respectively. Figure (1) shows a plot of 1000 simulations for each country, superimposed over a χ_1^2 and χ_5^2 distribution. Without nuisance parameters we should have a χ_1^2 asymptotic distribution. With four additional nuisance parameters we also compare the empirical distributions against a χ_5^2. The solid lines in Figure (1) are the χ^2 distributions, with the top solid line being the χ_1^2. The empirical likelihood ratio is clearly not χ_1^2 distributed but is fairly similar to a χ_5^2. All countries have a similar empirical likelihood ratio distribution and the null is rejected with a zero p-value according to the empirical distribution.

4.2 Estimation Results of RS Models

Estimation of regime switching models in finite samples is plagued with the presence of multiple local maxima. To ensure that a global was found several starting values were used, and to check for the stability of the global, each of the global parameters were randomly shocked by $\pm 10\%$ to check if the same maximum was reached. Some models considered here failed to converge.

The two state regime switching models all produce one regime with a unit root and lower conditional volatility and a second regime which is stationary with higher conditional volatility. This type of estimation is found in univariate, multi-country 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 over-parameterization. Previous studies with time-varying probabilities such as Gray (1996) and Bekaert, Hodrick and Marshall (1998) have also documented this. In some of our cases the null hypothesis of constant probabilities cannot be statistically rejected. Nevertheless, the general pattern that emerges is 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, multi-country and term spread RS models in turn.

4.2.1 RS Univariate Models

In univariate models we may interpret the first regime to be a random walk with a lower mean, and lower conditional volatility and the second regime has a higher mean, higher conditional volatility and higher mean reversion.[23] The theoretical result we presented in Proposition 3.2.1 shows that despite the unit root these models are stationary.

The RS AR(1) with state-dependent probabilities for the US failed to converge. For the UK and Germany, state dependence helps ensure overall stationarity since higher interest rates eventually lead to a switch from the non-stationary first regime into the high mean reversion regime. For the UK, a likelihood ratio test fails to reject that transition probabilities are constant, suggesting the system is

[23] These results confirm those previously documented by Gray (1996) and Bekaert, Hodrick and Marshall (1998).

over-parameterized. The poor performance of state-dependent probabilities is repeated in the CIR formulations which fail to reject the restricted model of constant probabilities except for the US.

The RS GARCH model failed to converge for the UK and many of the parameters of the GARCH process for the US and Germany are insignificant. These models are likely to be over-parameterized.

4.2.2 RS Multi-Country Models

The one-regime VAR dynamics are characterized by a companion matrix A with large diagonal elements representing the strong persistence in short rates and small, mostly insignificant off-diagonal coefficients.[24] We conduct several Wald tests on the system. These are presented in Table (5). The Table shows that a joint test for no country Granger-causing another just fails to reject (p-value = 0.0528), prompting the diagonal formulation A_i^1 for the companion matrix in the RS model. Nevertheless, there is some evidence that US rates Granger-cause German and UK rates (p-value = 0.0029). Consequently, this motivates the Granger-causality companion matrix A_i^2 of the US Granger-causing Germany and the UK in the RS model.

These results are partially consistent with the findings of Eichenbaum and Evans (1995). They show that a US monetary policy shock has a persistent effect, not only on the US interest rate but also on a number of foreign interest rates. Our results suggest that the US interest rate moves before foreign rates do. Of course, the US rate may also predict movements in the world business cycle before foreign rates do. It is striking, for example, that Granger-causality is strongest for the UK whereas Eichenbaum and Evans find that the UK rate is the only one not significantly affected by US monetary policy shocks. Of course, the US and the UK economies are very much linked and real shocks probably drive our results in this case.

We also estimate the one-regime equivalents of our RS multi-country model using the companion matrices A^1 and A^2 with our special covariance structure for the error shocks in equation (30). The Granger-causality model A^1 cannot be rejected from the unrestricted VAR using a likelihood ratio test (p-value = 0.9567). The diagonal model with A^2 also cannot be rejected, but the significance level is borderline (p-value = 0.0506). In the diagonal model we cannot reject the hypothesis that autocorrelations are the same across countries. In both one-regime formulations the coefficient on US shocks affecting the conditional volatility of Germany and UK (γ^m for country m), is insignificant for Germany but significant for the UK.

Estimation of the RS Granger-causality model is very tricky. For parsimony we initially constrained each country to have the same α_i, ρ_i, σ_i and P and Q (RSG1 and RSD1). The models with Gaussian errors were hard to estimate with unboundedness problems while the square root error models converged more easily. Consequently only the Granger-causality square root model RSG1 was successfully estimated. Constraining σ_i to be the same across countries imposes the restriction that the conditional volatility for Germany and the UK is higher than than the conditional volatility for the US. We relax this

[24]Estimations of all VAR's use a GMM estimation with 6 Newey-West lags following Bekaert and Hodrick (1992).

formulation in RSG2 and find it makes little qualitative difference.

The estimations show that Granger-causality is important only for the UK in the second mean-reverting high variance regime. Granger-causality of Germany is insignificant in both regimes. Looking at the impact of US shocks on the error terms of Germany and the UK, the Granger-causality model RSG2 has significant shock terms for Germany and the UK in the first random walk regime. The diagonal model, however, shows US shocks affecting only UK shocks in the first regime. These results point to no Granger causality in the first "normal" random walk regime, but in this regime US shocks propogate into Germany and the UK. In the second regime US short rates Granger-cause the short rates of the UK.

4.2.3 RS Term Spread Models

The one-regime benchmark is an unconstrained bivariate VAR of short rates and spreads for each country. The AIC and BIC criteria select 2 lags for the US, and 1 for Germany and the UK. Using the lag lengths of 1 and 2, we perform Granger Tests for causality of the short rate by the spread and vice versa. The results are reported in Table (6). Generally, the evidence for Granger causality is quite strong for the second-order VAR's and for the UK and Germany. For the US, the p-value for the hypothesis that spreads predict short rates is 0.0613 for the first-order VAR but only 0.1794 for the second-order VAR. Short rates do not Granger-cause spreads in the first order system and weakly predict spreads in the second-order systems. This is consistent with the evidence in Eichenbaum, Evans and Marshall (1996) who find that shocks to the short end of the yield curve have no impact on the long end. We note that the second-order VAR seems over-parameterized by its poor performance in out-of-sample forecasts and its poor matching of unconditional moments, shown in Section 5. Consequently, for RS multivariate models of the term spread we only consider first-order systems.

Let us now consider Granger-causality in the RS term spread VAR. For the US 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 US but the lower variance one for Germany). The evidence for the UK 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 US 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 that 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 5.6. The implications for the behavior of short rate shocks to the spread conditional on the regime are explored in Section 6.

In our time-varying probability formulations the transition probabilities depend on both the short rate and spread. The US model failed to converge[25], so we report a model with transition probabilities dependent only on the spread. Many of the probability coefficients are insignificant for all countries. Only the joint estimation has significant coefficients on both short rates and spreads and even here only in one regime. However, likelihood ratio tests for constant probabilities versus time varying probabilities reject for all countries. The addition of state-dependent transition probabilities does not change the results on Granger-causality and conditional correlations of the short rate and spread.

Figure (2) shows the in-sample regime classification for the RS VAR time-varying probability model for the US, Germany and UK. The solid line in the top plots are 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 $p(s_t = 1|\mathcal{I}_{t-1})$.[26] Note that the regime-classification for the UK is poor, especially for the ex-ante probabilities, and there is a high frequency of switching between regimes.[27]

5 Performance Measures

We analyze the moments and forecast performance for the univariate, multi-country and term spread models separately in Sections 5.1 to 5.3. We also specifically look at improvements when moving from univariate RS models to RS models incorporating international and term-spread information in Section 5.4. Section 5.5 summarizes the evidence and makes use of a Monte Carlo experiment to help interpret the results. Section 5.6 analyzes regime classification and examines whether the regimes are correlated with business cycle indicators. The results are reported in Tables (7) through (15). To interpret the tables the reader should refer to the nomenclature scheme in Table (3).

5.1 Univariate Performance

H-statistics for univariate models are presented in Table (7). The dismal performance of models RS1-3 for the US is partly caused by numerical problems: although theoretically stationary, the unit root in one of the regimes produces some stationarity problems in simulation.[28] For the US, the one-regime models seem to work better in matching unconditional moments than the RS models. By far the best model

[25]The same estimation problems that plague the univariate US RS time-varying probability model are shared with the RS VAR time-varying probability model.

[26]Ex-ante probabilities are calculated directly from the estimation algorithms of Hamilton (1989, 1994) or Gray (1996). Algorithms for smoothed probabilities are given by Gray (1995) and Kim (1993). Gray's is a forward looking algorithm, Kim's is a backward looking algorithm but the two are equivalent.

[27]However, there was one local maximum that did yield better behaved ex-ante probability behavior.

[28]The same problem is also observed when simulating from an AR(1) process with the autocorrelation very close to one: this model is also theoretically stationary but numerically behaves like a random walk.

however is the one-regime square root process. 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 UK, the AR(1) RS processes seem to work best with the square root processes performing more poorly. RS models with state-dependent probabilities (RS2, RS5) and GARCH (RS3) fare far less well than their constant probability counterparts.

Forecast performance for univariate models is also presented in Table (7). For the RS AR(1) models the state-dependence of the probabilities 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 UK's state dependent formulation performing better. Overall, with the exception of the UK, the GARCH models produce the best results. For the UK, the superior performance of the RS2 model, using either the RMSE or MAD criterion and for both first and second moments, is remarkable given that regime classification in the UK is rather poor. (See Figure (2)). Relative to their one-regime counterparts, RS models generally perform better. For all countries the RS AR(1) models forecast better than a simple AR(1) and the RS CIR models forecast better than the simple CIR. The one-regime GARCH model is the exception, but this may be due to over-parameterization in the RS counterpart.

5.2 Multi-Country Performance

Table (8) reports H-statistics and forecasts for each country from the multi-country models. Looking first at one-regime models, diagonal models match central moments better than the unconstrained VAR(1), which is indicative of the over-parameterization of the unconstrained VAR(1). The Granger-causality models do not perform as well as the diagonal specification.

Turning to comparisons of the RS multi-country models, with the exception of the UK, 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 US and UK but not for Germany. There is little difference in relaxing σ_i across countries in the RS Granger-causality models.

Looking at forecasts, the diagonal one-regime models out-perform the unrestricted VAR on mean forecasts and do worse for second moment forecasts only for the US, again showing over-parameterization of the unconstrained VAR. The multi-country RS diagonal model outperforms the one-regime model which is an excellent result, as we have constrained the interest rate data generating process to be the same across all countries, and shows the importance of regime shifts in forecasting.

Granger-causality seems to aid in forecasting both in one-regime and RS frameworks. The regime-switching Granger models do particularly well for the US and the UK.

5.3 Term Spread Performance

Table (9) reports the H and H* statistics for the bivariate system. The results are mixed. 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. In matching central moments, the state-dependent probability models fare better for the US and Germany than their constant probability counterparts, but for the UK this result is reversed. One-regime VAR's clearly outperform RS VAR's for central moments. The evidence is less clear for auto and cross-correlations. However, in general one-regime models produce more satisfactory fits to sample unconditional moments.

Table (10) shows forecast performance. For forecasting the first and second moments, the more parsimonious VAR(1) outperforms the VAR(2) for all countries, suggesting that the VAR(2) is over-parameterized. The RS models outperform the VAR's for forecasting the short rate, and with the exception of the UK, also for forecasting the spread. Looking at forecasts of second moments, Germany's state-dependent RS model does better than its constant probability counterpart; for the UK the state-dependent RS model also does better except for the cross-moments. For the US, the constant probability RS model clearly out-performs both one-regime VAR specifications.

5.4 Regime Switching Performance

We wish to specifically examine how incorporating extra information improves the fit of unconditional moments and forecasting of RS models. We concentrate on the H-statistics and the RMSE.

First we look at matching moments. By looking at Tables (7) and (8) we can compare the multi-country RS models with the univariate RS models. We see a dramatic improvement when incorporating multi-country information for the US but not for Germany or the UK. Comparing the univariate RS models in Table (7) with the bivariate RS term spread models in Table (9) we see the extra information allows a better match of moments only for the US, and for autocorrelations only for the UK. Overall, using the extra information from other countries or the term spread unequivocally helps the US obtain a better fit to unconditional moments, but it definitely does not help for Germany. The evidence for the UK is mixed.

Focusing now on forecasts of RS models with the RMSE criterion, the multi-country approach generally yields better forecasts than the univariate models. The RMSE statistics in Table (10) show that with the exception of univariate RS forecasts of the second moment of the short rate being better for the US, evidence favors the bivariate RS models. Generally forecasts are improved by taking a multi-country or term-spread approach.

5.5 Summary and Interpretation of Moments and Forecast Performance

In general we find that in matching sample moments one-regime models tend to perform better, despite the presence of regime-switching in the data. However, in forecasting out of sample, regime-switching models do better. Focusing on short rates, Table (11) reports the best models with the lowest H and RMSE statistics. There is no clear-cut "best" model. However, it appears that while single regime models may give lower H-statistics (for example in the case of the US), RS models forecast much better for all countries. Moreover, the best RS forecasting models incorporate information from other countries or the spread. Interestingly, RS models with state-dependent probabilities tend to forecast better than their constant probability counterparts even if they perform very poorly at matching sample moments.

How do we interpret these results? As indicated before, the RS models considered here need extremely large simulations to pin down their unconditional moments with any precision. This means that the small sample behavior of RS models may be poor. Despite the intuitive economic approach of RS models and the clear endorsement of RS models by the data, it may be that more parsimonious one-regime models produce better estimates of the sample unconditional moments than RS models in small samples. Here we run a small experiment to specifically investigate this conjecture.

Consider the following RS VAR population model of the short rate and spread, $y_t = (r_t \; z_t)'$: $y_t = \mu(s_t) + A(s_t)y_{t-1} + u_t$ where $u_t \sim N(0, \Sigma(s_t))$, $s_t = 1, 2$ with Markov state-dependent logistic transition probabilities depending on the lagged y_t. We use the parameters from the joint estimation as the population model.

Taking this model we find population moments by simulation, and then simulate a small sample of size $T + N$. We now consider several approximations to the true model and compare their unconditional moment estimates over the in-sample of size T and their forecasts over the out-sample of size N. We take T and N to be the size of our in-sample and out-sample data sets considered in the estimation and forecasts of models in this paper, 267 and 30 respectively. The models we consider are an AR(1) and a RS AR(1) on the short rates with constant probabilities, a VAR(1) and a RS VAR(1) on the bivariate short rate and spread with constant transition probabilities. We denote these as AR, RS AR, VAR, RS VAR respectively.

Unfortunately we cannot include the true model because of the problems we encountered in finding satisfactory estimates of the RS VAR with time-varying probabilities in small samples. The many convergence failures that occured even when starting from the true parameters are in itself proof of the poor small sample behavior RS model estimation may 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 gives the best (lowest) statistic value.[29] To compare unconditional forecasts, we record which model gives the lowest RMSE statistic. We repeated this for 1000 samples. Our results are listed in Table (12). The table gives 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 a RS VAR(1) with state-dependent prob-

[29]This is extremely computationally intensive and to shorten the computation time we only used sample sizes of 200,000 to estimate moments. Some experimentation showed that this should be sufficient for purposes of comparison.

abilities.

Table (12) shows that the one-regime models are good approximations in small samples to the true RS models, and that despite the true data generating process being regime-switching, 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 in forecasting the RS models perform better. 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 (13) reports the population values of the unconditional moments for the short rates and spreads and the mean values of the empirical distribution of the moments produced by the models estimated from the small sample. The table shows that the RS models tend to over-estimate the mean and under-estimate the variance of the short rate, but the population values lie within 95% confidence intervals of the small sample model moments.

5.6 Regime Classification and Regime Interpretation

To examine how well the various RS models classify the regimes, we present RCM statistics in Table (14). In univariate RS models the CIR specification produces the cleanest regime classification. For univariate models, moving from constant to state-dependent transition probabilities produces very little improvement. Multi-country estimation produces sharper regime classification for the UK and Germany at the expense of the US. Including term structure information leads to better regime classification for all countries. The results show that using more information produces better regime classification, as expected, and including the term spread uniformly decreases the RCM statistics for all countries. Our multi-country model produces less reliable classification for the US but regime-classification improves dramatically for Germany and the UK when the US is included.

The UK models classify regimes poorly, with the transition probabilities P and Q being very close to a half. In a regime switching model, if $P + Q = 1$ the model reduces to a simple switching model. In fact, using a likelihood ratio test, we are unable to reject this hypothesis for the univariate regime-switching AR(1) (p-value = 0.1202). For a pictorial representation of poor regime classification, see the UK plot in Figure (2). The high frequency of switching can be seen by the wildly fluctuating smoothed probabilities and the poor classification of ex-ante probabilities. This poor performance is reflected in the RCM value for the UK being very close to 100.

Are the regimes correlated with the business cycle? Table (15) 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.[30] The ex-ante probabilities are generated from the term structure RS model with time-varying probabilities (RSM2).[31] We report the correlations between

[30] Note the ex-ante probability $p_t = p(s_t = 1|\mathcal{I}_{t-1})$ is in the information set at time $t-1$.

[31] We use this model because it is the model with the lowest RCM statistic for the US and Germany in Table (14). Other RS models produce similar results, with those of the univariate RS models actually doing better than the results reported here.

the second regime with mean-reverting higher volatility and the economic downturns. The table shows that this regime is associated with economic recessions, while the "normal" unit root regime with lower volatility represents economic expansions. The US and Germany have significant correlations, while the correlations of the UK are insignificant.

The business cycle association of the regimes is not surprising for the US. Figure (2) shows that the ex-ante probabilities during the 1979-1982 period of monetary targeting are near zero, 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 to 1983:07), and also went through an earlier recession accompanied by high interest rates in the early 1970's (1973:09 to 1975:05). The recession brought on from re-unification, beginning in mid-1991, also saw rising interest rates but the regimes do not capture this period as sucessfully. The poor performance of the UK is not surprising given the poor regime classification of the UK model.

The last four columns of Table (15) report coefficients from a Probit regression with the recession indicator being the dependent variable, and current and lagged ex-ante probabilities being the independent variable. The Probit regressions yield significant coefficients for the US and Germany. We also list the percentage of correctly forecasted recessions in-sample from the Probit regressions. For the US, 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 UK.

Recent studies have found that the term structure can successfully predict real economic activity.[32] Estrella and Mishkin (1995) find that the spread is useful in predicting future economic activity, and Table (15) confirms their finding showing that the magnitude of correlations between recessions and the spread increases with the lag, and the Probit forecasts increase their accuracy forecasting longer future horizons. This happens across all three countries. Looking specifically at the US, the ex-ante regime probabilities have better forecast ratios for one and two month ahead predictions than the spread. While the forecast ratios increase with horizon for the spread, the forecast ratiosof the the ex-ante probabilities remain essentially flat. This evidence indicates that for the US the ex-ante regime probabilities are better contemporaneous indicators of the business cycle than the spread, and the spread is a forward looking indicator which improves its forecasting ability at longer horizons. For the other countries, the spread better predicts recession than our regime probabilities at all horizons. Given that both the regime classification and the dating the actual business cycles is less precise for these countries, this is not surprising.

[32]For example see Estrella and Mishkin (1995) and Harvey (1988).

6 Implied Short Rate and Spread Dynamics

In this section we study the short rate and spread dynamics implied by our RS term spread models along two dimensions. First, we examine impulse response (IR) functions of the bivariate RS VAR and compare these to the IR's implied by a linear one-regime VAR. Following Gallant, Rossi and Tauchen (1993) we use the key idea of IR analysis to trace the effects of a small shock through the system. We first briefly review IR's in one-regime VAR's and then extend the analysis to our non-linear RS VAR models.

Our results here may provide useful input for the rapidly growing literature on the effects of economic and policy shocks on financial variables in general and the term structure in particular.[33] Such analysis is typically constructed in a linear VAR setting. By contrasting IR's from a linear model to IR's from a non-linear framework, we may gain insights on the distortions a linear framework may introduce. The effect of short rate shocks on spreads is of independent interest since short rates are typically found to exhibit large contemporaneous effects with respect to monetary policy and other shocks.

Second, we investigate the "drift" and "volatility" functions implied by our models. There is a voluminous literature in finance on the dynamic properties of short rates and the term structure in the US. We contribute to this literature by examining the non-linearities implied by an alternative RS model and by studying the term structure in other countries as well.

6.1 Impulse Responses

6.1.1 Impulse Responses in Linear VAR's

Consider the following linear VAR: $y_t = v + Ay_{t-1} + \epsilon_t,\ \epsilon_t \sim N(0, \Sigma)$. For simplicity, we take y_t to have two elements so $y_t = (y_t^{(1)}\ y_t^{(2)})'$. The basic idea of an IR is to trace a shock δ through the system relative to a baseline. The effect of the shock one-period ahead and the baseline are given by:

$$E_t(y_{t+1}|y_t = \delta) = v + A\delta \tag{49}$$

$$E_t(y_{t+1}|y_t = 0) = v \tag{50}$$

So the first impulse response from a shock δ is:

$$IR_1 = E_t(y_{t+1}|y_t = \delta) - E_t(y_{t+1}|y_t = 0) = A\delta \tag{51}$$

Analogously the jth impulse response $IR_j = A^j\delta$.

If we consider a shock of one standard deviation of $\sqrt{\sigma_{11}}$ to $y_t^{(1)}$ as $\delta = (\sqrt{\sigma_{11}}\ 0)'$ and the covariance matrix Σ of the system is diagonal, then $\{IR_j\}_{j=1}^{\infty}$ represents the responses of the variables to a standardized shock from $y_t^{(1)}$. However, when Σ is not diagonal, then δ does not represent the typical shocks to the system because it ignores the contemporary covariances in $\Sigma = \left(\begin{smallmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{smallmatrix}\right)$. To treat this problem the literature orthogonalizes the shocks so $\widetilde{IR}_j = A^j R\delta$ where $RR' = \Sigma$. An alternative approach in

[33] For example see Eichenbaum and Evans (1995), Eichenbaum, Evans and Marshall (1995).

the spirit of Gallant, Rossi and Tauchen (1993) is to find a "typical" value of $y_t^{(2)}$ given $y_t^{(1)}$. We set the shock to $y_t^{(1)}$ at $\sqrt{\sigma_{11}}$ and then set $y_t^{(2)}$ as $E(y_t^{(2)}|y_t^{(1)} = \sqrt{\sigma_{11}})$. As y_t is assumed to be drawn from a bivariate normal $N(0, \Sigma)$, we can use the conditional normal distribution to obtain:

$$E(y_t^{(2)}|y_t^{(1)} = \sqrt{\sigma_{11}}) = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} \qquad E(y_t^{(1)}|y_t^{(2)} = \sqrt{\sigma_{22}}) = \frac{\sigma_{12}}{\sqrt{\sigma_{22}}} \tag{52}$$

So the shock $\delta = (\sqrt{\sigma_{11}}\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}})'$ can be interpreted as a shock of one standard deviation to $y_t^{(1)}$ with $y_t^{(2)}$ adjusted to be its predicted value given the movement in $y_t^{(1)}$.

6.1.2 Impulse Responses in RS VAR's

We can generalize the approach of the previous section to trace out the effect of a shock in a RS VAR: $y_t = v(s_t) + A(s_t)y_{t-1} + \epsilon_t,\ \epsilon_t \sim N(0, \Sigma(s_t))$. Assume there are two regimes so $s_t = 1, 2$ and the constant transition probability matrix is given by $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$. The columns of P are denoted $p_{.,0}$ and $p_{.,1}$. Suppose that the probabilities of being in each regime at time t are given by $\pi = (\pi_1\ \pi_2)'$. Then the first period effect of a shock δ relative to the baseline response is:

$$E_t(y_{t+1}|y_t = \delta) = (\pi_1 v_1 + \pi_2 v_2) + (\pi_1 A_1 + \pi_2 A_2)\delta \tag{53}$$

$$E_t(y_{t+1}|y_t = 0) = \pi_1 v_1 + \pi_2 v_2 \tag{54}$$

$$IR_1 = (\pi_1 A_1 + \pi_2 A_2)\delta \tag{55}$$

For the impulse responses from the RS VAR with constant transition probabilities, there exist analytical formulae.[34] Let

$$\Pi = \pi \otimes I_{2x2} = (\pi_1 I\ \pi_2 I)' \tag{56}$$

$$\mathbf{P} = \left(p_{.0} \otimes A_0 \;\middle|\; p_{.1} \otimes A_1\right) = \left(\begin{array}{c|c} p_{00}A_0 & p_{01}A_1 \\ \hline p_{10}A_0 & p_{11}A_1 \end{array}\right) \tag{57}$$

$$\mathbf{I} = \mathbf{1} \otimes I_{2x2} = (I\ I)' \tag{58}$$

where $\mathbf{1}$ is a vector of one's. The impulse responses are then given by: $IR_j = \Pi' \mathbf{P}^j \mathbf{I}\,\delta$.

We can take the probabilities of being in each regime π at time t to be $(1\,0)'$ to start in regime 1, $(0\,1)'$ to start in regime 2 or to be the stable probabilities of the system. A shock δ will also take into account the contemporaneous correlation of $y_t^{(1)}$ and $y_t^{(2)}$ conditional on the regime. The conditional covariance in regime 1 is Σ_1 so we can use $\delta = (\sqrt{\sigma_{11,1}}\ \frac{\sigma_{12,1}}{\sqrt{\sigma_{11,1}}})'$ to represent a shock of 1 standard deviation to $y_t^{(1)}$ conditional on regime 1, adjusting $y_t^{(2)}$ to take into account the contemporaneous movement in $y_t^{(1)}$ in regime 1.

[34] For RS VAR's with time-varying probabilities there are no corresponding analytical formulae. These must be obtained by simulation: we did not do these because of the computational requirements.

6.1.3 Empirical Results

Figure (3) presents the impulse responses from shocks of one standard deviation. Solid lines represent the IR from the one-regime VAR and non-solid lines the IR's from the regime-switching model.[35] The numbers on the horizontal axis are months. A first remarkable fact about the figure is the similarity of the dynamics across countries. It is also remarkable that the impulse responses starting from the stable probabilities of the regime-switching model mimic almost exactly the impulse responses from the linear VAR models. However, the effects of a shock conditional on a regime are quite different.

Let us first consider the effects from a one standard deviation shock to short rates, in the first column in Figure (3). The standard deviation is small in regime 1, the unit root regime, and has little initial effect. The standard deviation is larger in regime 2, the stationary regime, and has a much greater initial effect. In both cases short rates are very persistent but the shocks conditional on regime 2 dissipate to approach the shocks conditional on regime 1 and the stable probabilities. Only in Germany is this convergence not complete after 30 months. The positive shock to short rates is associated with a negative shock to the spread. In regime 1, shocks generate very little effect, but in regime 2, where volatility is much higher and spreads are much more negatively correlated with short rates, the term spread narrows considerably more.

In the second column in Figure (3) we see the effects of conditional shocks to the spread. The effects are similar to what we had before. Nevertheless, spreads are generally less persistent than short rates so that the shocks die out sooner. Germany's shocks are the slowest to die out, reflecting the high persistence of its spreads and short rates. A typical positive shock to the spread immediately reduces the short rate in our framework due to the negative contemporaneous correlation structure for the shocks. In the first "normal" regime, shocks to the spreads have very little effect on short rates and the response line is almost flat. In the second regime, higher volatility and a more negative contemporaneous correlation drives the short rate much further down as the spread is shocked upwards. The short rate is pulled back up much faster because of higher mean-reversion in the second regime.

A similar picture emerges from graphs where we look at unit shocks (Figure (4)) rather than one standard deviation shocks. To give an example, consider a 100 basis point (bp) increase in the short rate (induced by monetary policy, say). The effect on the short rate dies out slowly to a level of about 40 bp over 30 months in a linear VAR. However, in the high variance regime it reaches 40 bp after barely 10 months. These IR's dramatically illustrate the potential importance of regimes in policy analysis - suppose that the short rate shocks correspond to monetary policy shocks. Clearly, the IR dynamics seem very much dependent on the regime the economy is in at the time of the shock.

[35] We use a VAR length of 2 for the US, and 1 for Germany and the UK, which is the optimal lag by AIC and BIC criteria (see Section 4.2.3).

6.2 The Drift and Volatility Functions

There is now a large literature documenting empirical non-linearities in interest rates. Aït-Sahalia (1996) parametrically specifies the drift and volatility functions for US 7-day Eurodollar spot rate changes using non-linear functions. Aït-Sahalia finds a highly non-linear drift with strong mean-reversion at very low and high interest rates but the drift is essentially zero in the middle region. The volatility function assumes a J-shape so the spot rate is more volatile outside the middle region, with the highest volatility occurring at very high interest rates. Conley et al (1997)'s drift estimations on overnight Fed funds rate changes look very similar to Aït-Sahalia's plots, but without the strong mean reversion at high interest rates. In their formulation, stationarity at high interest rates is induced by increasing volatility. Stanton (1997)'s non-parametrically estimated drift on daily 3 month T-bill rate changes is zero until high interest rates where the drift becomes very negative. Stanton's non-parametrically estimated volatility looks very similar to Aït-Sahalia's, with volatility increasing at higher levels of interest rates.

These findings suggest that interest rates exhibit strong non-linear drifts, with the drift being zero over much of the support of the data, but strongly mean-reverting at low or high interest rates. The volatility of interest rates generally increases with the level of the interest rate with the lowest volatility appearing in the middle of the support.

In this section we specifically look at the drift and volatility functions implied by the RS models. To obtain more information about interest rates at very high and very low levels, we pool the information from the US, Germany and UK to estimate a joint RS process. This was done for a univariate constant transition probability RS model, and bivariate term spread RS models with constant and time-varying probabilities.[36] Section 6.2.1 reviews the definitions of drift and volatilities for linear models, Section 6.2.2 outlines how the drift and volatilities for RS models are obtained, and Section 6.2.3 presents the empirical results.

6.2.1 Linear Drift and Volatilities

The conditional drift and volatility functions for a multivariate linear process $y_t = v + Ay_{t-1} + \Sigma_{t-1}^{\frac{1}{2}}\epsilon_t$, with $\epsilon_t \sim$ IID N(0,I) are given by:

$$\text{drift} = E_{t-1}(\Delta y_t) = v - (I - A)y_{t-1} \tag{59}$$

$$\text{vol} = \text{diag}\Sigma_{t-1}^{\frac{1}{2}} \tag{60}$$

For univariate processes y_t, we may plot the drift and volatility against y_{t-1}. For a simple mean-reverting process of short rates rates $y_t = r_t$, the drift will be a downward sloping line, showing positive drift at low levels of r_{t-1} and negative drift at high levels. In this way interest rates are pulled back toward their long-term mean at low and high levels. For a CIR model the volatility is $\sigma\sqrt{r_{t-1}}$ which is increasing with the interest rate level.

[36] A joint estimation of a univariate RS model with time-varying probabilities was attempted but failed to converge.

Suppose in our bivariate VAR $y_t = (r_t\, z_t)'$ with constant conditional covariance Σ, we wish to obtain the drift of r_t. The conditional mean of r_t is given by:

$$E_{t-1}(r_t) = \mu_1 + A_{11}r_{t-1} + A_{12}z_{t-1} \tag{61}$$

where subscripts indicate the appropriate element in the parameter matrix A. To obtain $E(r_t|r_{t-1})$ we need to integrate out z_{t-1}:

$$E(r_t|r_{t-1}) = \mu_1 + A_{11}r_{t-1} + A_{12}E(z_{t-1}|r_{t-1}) \tag{62}$$

Since the system $y_t = (r_t\, z_t)'$ is jointly normal, we can use the conditional distribution to evaluate $E(z_{t-1}|r_{t-1})$. The system y_t is unconditionally normally distributed with mean $\tilde{\mu}$ and variance $\tilde{\Sigma}$ where $\tilde{\mu} = (I - A)^{-1}\mu$ and $\text{vec}\tilde{\Sigma} = (1 - A \otimes A)^{-1}\text{vec}\Sigma$. Then the conditional mean $E(z_t|r_t)$ is given by:

$$E(z_t|r_t) = \tilde{\mu}_2 + \frac{\tilde{\Sigma}_{12}}{\tilde{\Sigma}_{22}}(r_t - \tilde{\mu}_1) \tag{63}$$

Hence we can obtain the drift of r_t from the bivariate system $E(\Delta r_t|r_{t-1}) = E(r_t|r_{t-1}) - r_{t-1}$.

6.2.2 Drifts and Volatilities from RS VAR's

For a RS VAR of $y_t = (r_t\, z_t)'$, $y_t = \mu(s_t) + A(s_t)y_{t-1} + \Sigma(s_t)^{\frac{1}{2}}\epsilon_t$, we can obtain drifts and volatilities of y_t by integrating out s_t. To obtain the drift of r_t we need to integrate out both s_t and z_t. This can be done numerically by the following procedure for both constant and time-varying transition probabilities.

First, simulate out the system y_t. Record the drift and volatility of y_t for every observation, where drift $= v_i - (I - A_i)y_{t-1}$ and vol $= \text{diag}\Sigma_i^{\frac{1}{2}}$ and subscripts indicate the state $s_t = i$. Divide interest rates into bins of width 25 basis points, and then within each bin calculate the average drift and the average conditional volatility. We use the mid-points of the bins to plot an appropriate drift and volatility function. From the simulation we record the regime realizations to enable us to estimate $E(s_t|r_{t-1})$. Large simulations (upwards of 500,000) are necessary to obtain smooth plots for our bin size.

6.2.3 Empirical Results

We only report the drift and volatility functions from the joint processes, as these give us the most information possible, especially at very low and high interest rates.[37]

Figure (5) presents the estimated drifts from the joint RS models. The top panel shows the drift function for the short rate from a bivariate RS term-spread VAR with constant transition probabilities. The dashed lines correspond to the drift functions conditional on each regime and are linear. The first regime is a near-unit root regime (almost zero drift) and the second regime is strongly mean-reverting

[37] Plots of the drifts and volatilities from individual countries exhibit similar patterns. The conditional volatilities from the UK, however, are much flatter than the plots given here. This is due to the simple-switching nature of the UK RS process. The conditional volatility can only vary as much as the expected state at each interest rate level varies. With a simple switching process the expected state is always a half.

(downward sloping line). The drift function for the RS process is a weighted average of the linear drift functions in each regime, with the weights determined by the different amounts of time spent in each regime at different interest rate levels. The drift from this model retains a fairly linear shape.

Moving to the bottom plot in Figure (5), we see the drift functions for the RS VAR with constant and time-varying probabilities and the univariate RS model with constant transition probabilities. The univariate RS model with constant probabilities looks very similar to the bivariate RS model with constant probabilities. However, the drift in the model with state-depndent transition probabilities closely resembles the drift presented in Stanton (1997), with a very flat drift close to zero until the middle of the support and then turning negative at higher interest rates. Stanton's drift starts turning negative around 14%, while our drift starts turning negative around 10%. The shape is also similar to Aït-Sahalia (1996), but Aït-Sahalia's drift starts turning negative at 18%. The difference in the ranges can be attributed to using different data sets, but the important observation is that the state-dependent probability model can reproduce the shape of the non-parametric estimations.

The kinked shape from the state-dependent probability model results from a much faster transition into the stronger mean-reverting regime at higher interest rates and so at higher interest rates more time is spent in the second regime. This places more weight on the drift of the second regime at higher interest rates than in the RS models with constant transition probabilities. Similarly, at lower interest rates the transition from the unit root regime is slower. Note that at very low interest rates the drift increases slightly.

Figure (6) presents plots of the conditional volatility functions of the short rate. The top plot shows the volatility from the constant probability bivariate RS model. The conditional volatility varies only because the expected state varies with the level of the short rate. (In each regime the conditional volatility is constant.) The plot shows that at very high short rates the process is likely to be in the second higher conditional volatility regime. The bottom plot of Figure (6) shows the conditional volatility of the short rate for the three joint RS models. The constant probability univariate and bivariate models yield similar shapes for conditional volatility and bear a strong resemblance to Aït-Sahalia (1996)'s J-shaped estimations. The volatility implied by the bivariate RS model with time-varying probabilities loses much of its upturn at lower interest rates but then increases rapidly with rising interest rates. It looks very much like Stanton (1997)'s non-parametric estimation. The main source driving the volatility increase is the increasing probability of remaining in the high variance regime, which decreases at very high interest rates. Overall, the volatilities are increasing in the level of the interest rate, with some upturn at lower interest rates.

The bivariate RS models allow us to look at the drifts and volatilities for spreads. We present these in Figure (7). The top plot shows that the drift function of the spread is very linear and the addition of time-varying probabilities does not change the shape of the drift function. The middle plot shows how the conditional volatility of the spread varies with the expected state corresponding to each spread level. The conditional volatility is lowest for "normal" levels of the spread between 0 and 1. The conditional volatility of the spread for both the constant and time-varying bivariate RS model is shown in the bottom plot of Figure (7). The addition of state-dependent transition probabilities now makes the spread volatility less symmetric, as the probability of staying in the higher variance regime now increases with higher spreads. Whereas before, both unusually high and low spreads are associated with the high variance regime, state dependence makes it more likely to remain in the low variance regime when the spread increases.

7 Conclusions

This paper demonstrates theoretically and empirically that univariate regime-switching models can capture the non-linear mean reversion observed in interest rates in an economically appealing and stationary model. Moreover, there is overwhelming evidence for multiple regimes in the data generating process of short rates.

Given the well-known econometric problems estimating regime-switching models in small samples, we compare their econometric performance relative to their one-regime counterparts. First, the moments implied by regime-switching models do not fit the sample moments as well as simpler models do because of the difficulties in estimating regime-switching models in small samples. A Monte Carlo experiment confirms this happens even when the regime-switching model is the true data generating process. Second, regime-switching models tend to forecast better than one-regime models.

To improve the econometric performance of regime-switching models it is important to incorporate additional information. In fact, we show that univariate regime-switching models will typically yield inconsistent estimates as soon as the omitted variables contain information on the regime. We compare the performance of univariate versus multi-country and term spread approaches. In particular, US short rates improve both the regime classification and the statistical performance for German and UK short rates (but not vice versa). Furthermore, inclusion of term spread information leads to dramatic superior performance in regime inference and general improvements over univariate models in forecasting. However, the inclusion of extra information did not always improve the fit of the unconditional moments.

The regimes correspond well with business cycle expansions and contractions. For the US using Probit regressions, the ex-ante probabilities of regimes forecast future recessions better than the term spread for short horizons (less than 2 months ahead), while the spread shows increasing accuracy for longer horizons (6 months ahead).

The behavior of the term structure varies dramatically with the regime. For example, correlations between short rates and spreads are significantly different across regimes. We examine the non-linearities implied by regime-switching models by looking at their impulse responses and their drift and volatility functions. When averaged over each state, the impulse responses correspond almost exactly to the impulse responses implied by linear Vector Autoregressions. However, conditional on a regime, impulse responses behave very differently. As the impulse responses are so dissimilar in each regime this has implications for the policy analysis of shocks.

The drift and volatility functions of regime-switching models correspond closely to the empirical drift and volatilities estimated in recent literature using non-parametric techniques. In particular, regime-switching models with time-varying probabilities can produce highly non-linear drifts with unit root behavior for most of the support of the data, with strong mean-reversion at high interest rates. The conditional volatilities from regime-switching models also match the empirical estimations in the literature. Hence a simple parametric model is sufficient to match the rich non-linear dynamics of short rates.

Some interesting extensions would be to further improve regime classification by imposing theoretical restrictions from a term structure model. Attempting to trace the macro-economic or policy sources of the shocks driving the impulse responses is another promising area for further research. This paper has also shown the importance of endogenizing regime switches in future economic models.

References

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

[2] Bekaert, G., and Gray, S.F., 1998, "Target Zones and Exchange Rates: An Empirical Investigation", forthcoming in the *Journal of International Economics*.

[3] Bekaert, G., R.J. Hodrick, and D.A. Marshall, 1998, " "Peso Problem" Explanations for Term Structure Anomalies", NBER working paper.

[4] Bekaert, G., and R.J. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets", *The Journal of Finance*, 47, 2, 467-509.

[5] Boudoukh, J., Richardson, M., Stanton, R., and Whitelaw, R., 1997, "The Stochastic Behavior of Interest Rates: Implications from a Nonlinear, Continuous-Time, Multifactor Model", working paper.

[6] Cecchetti, S.J., P. Lam, and N.C. Mark, 1993, "The equity premium and the risk-free rate: matching the moments", *Journal of Monetary Economics*, 31, 21-46.

[7] Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-term Interest Rate", *Journal of Finance*, 47, 1209-1227.

[8] Conley, T. G., Hansen, L. P., Luttmer, E. G. J., Scheinkman, J. A., 1997, "Short-Term Interest Rates as Subordinated Diffusions", *Review of Financial Studies*, 10, 3, 525-577.

[9] Cox, J., J. Ingersoll, and S. Ross, 1985, "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, 385-408.

[10] Davies, R.B., 1977, "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, 64, 247-254.

[11] Davies, R.B., 1987, "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, 74, 1, 33-43.

[12] Diebold, F.X., Lee, J.H., and G. C. Weinbach, 1994, "Regime switching with time-varying transition probabilities", in Hargreaves, C., ed., *Time Series Analysis and Cointegration*, Oxford University Press.

[13] Eichenbaum, M., and C. L. Evans, 1995, "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates", *Quarterly Journal of Economics*, 110, 4, 975-1009.

[14] Eichenbaum, M., Evans, C., and D. Marshall, 1996, "Economic Determinants of the Term Structure of Nominal Interest Rates", working paper.

[15] Engel, C., 1994, "Can the Markov switching model forecast exchange rates?", *Journal of International Economics*, 36, 151-165.

[16] Estrella, A., and F. S. Mishkin, 1995, "The Term Structure of Interest Rates and its Role in Monetary Policy for the European Central Bank", NBER working paper 5279.

[17] Evans, C. L., and D. A. Marshall, 1997, "Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory", working paper.

[18] Evans, M. D., 1995, "Peso Problems: Their Theoretical and Empirical Implications", working paper, Stern School of Business, NYU.

[19] Gallant, A.R., 1977, "Testing a nonlinear regression specification: a nonregular case", *Journal of the American Statistical Association*, 72, 359, 523-530.

[20] Gallant, A.R., Rossi, P.E., and Tauchen, G., 1993, "Nonlinear Dynamic Structures", *Econometrica*, 61, 4, 871-907.

[21] Garcia, R. and P. Perron, 1996, "An analysis of the real interest rate under regime shifts", *The Review of Economics and Statistics*, 78, 1, 111-125.

[22] Granger, C.W., 1969, "Prediction with a Generalized Cost of Error Function", *Operational Research Quarterly*, 20, 199-207.

[23] Gray, S.F., 1995, "An Analysis of Conditional Regime-Switching Models", manuscript, Graduate School of Business, Stanford University.

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

[25] 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 and Control*, 12, 385-423.

[26] Hamilton, J.D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

[27] Hamilton J.D., 1994, *Time Series Analysis*, Princeton University Press, New Jersey.

[28] Hamilton, J.D., 1996, "Specification Testing in Markov-Switching Time Series Models", *Journal of Econometrics*, 70, 127-157.

[29] Hansen, B.E., 1992, "The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP", *Journal of Applied Econometrics*, 7, 61-82.

[30] Hansen, B.E., 1996, "Inference when a nuisance parameter is not identified under the null hypothesis", *Econometrica*, 64,2,413-430.

[31] Harvey, C., 1988, "The Real Term Structure and Consumption Growth", *Journal of Financial Economics*, 22, 2, 305-33.

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

[33] Judge, G.G., W.E. Griffiths, R.C. Hill, and T.C. Lee, 1980, *The Theory and Practice of Econometrics*, Wiley, New York.

[34] Kim, C.J., 1993, "Dynamic linear models with Markov-switching", *Journal of Econometrics*, 60, 1-22.

[35] Lam, P., 1990, "The Hamilton Model with a General Autoregressive Component", *Journal of Monetary Economics*, 26, 409-432.

[36] Lewis, K.K., 1991, "Was There a "Peso Problem" in the U.S. Term Structure of Interest Rates: 1979-1982?", *International Economic Review*, 32, 159-173.

[37] Lumsdaine, R.L., and S.E. Prassad, 1997, "Identifying the Common Component in International Business Fluctuations", working paper.

[38] Lütkepohl, H., 1993, *Introduction to Multiple Time Series*, 2nd ed, Springer-Verlag, Berlin.

[39] Mankiw, N.G., and J.A. Miron, 1986, "The Changing Behavior of the Term Structure of Interest Rates", *Quarterly Journal of Economics*, 101, 211-228.

[40] Naik, V., and Lee, M. H., 1994, "The Yield Curve and Bond Option Prices with Discrete Shifts in Economic Regimes", working paper.

[41] Nerlove, M., D. Grether, and J.L. Carvalho, 1995, *Analysis of Economic Time Series: A Synthesis*, revised ed, Academic Press, San Diego.

[42] Newey, W. K., and K. D. West, 1987, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-8.

[43] Pearson, N., and Sun, T.S., 1994, "Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll and Ross model", *Journal of Finance*, 54, 1279-1304.

[44] Pfann, G.A., Schotman, P.C., and Tschernig, R., 1996, "Nonlinear interest rate dynamics and implications for the term sturcture", *Journal of Econometrics*, 74, 149-176.

[45] Prasolov, V.V., 1991, *Problems and Theorems in Linear Algebra*, Translations of Mathematical Monographs, vol 134, American Mathematical Society.

[46] Sola, M., and J. Driffill, 1994, "Testing the Term Structure of Interest Rates Using a Vector Autoregressive with Regime Switching", *Journal of Economic Dynamics and Control*, 18, 601-628.

[47] 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.

[48] Zarnowitz, V., 1997, "Some new lessons from international cyclical indicators", in Oppenländer, K., ed., *Business Cycle Indicators*, Ashgate Publishing, p175-204.

Appendix A: Proof of Proposition 3.2.1

We set up the following model:

$$y_t = \mu + \nu s_t + (1 - \beta s_t) y_{t-1} + \epsilon_t \tag{A-1}$$

with $\epsilon_t \sim$ IID $N(0, \sigma_1^2 + \alpha s_t)$, $\sigma_1^2 + \alpha > 0$ and $0 < \beta < 1$. The state $s_t = 0, 1$ and is independent of ϵ_t so when $s_t = 0$ we have a random walk, and when $s_t = 1$ the regime is mean-reverting. Without loss of generality we may consider only these two regimes since we may group all unit root regimes into regime 0, and we need only consider an AR(1) process for regime 1 as an example of a stationary process. Longer AR(p) processes can be handled similarly in the manner presented here, and MA(q) components can be treated by expanding the state space in a suitable manner to remove the state-dependence of the MA terms.

We denote the Markov transition probabilities as $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$. We only require that the probability of entering the stationary regime is non-zero ($p_{01} > 0$) and the probability of staying in the stationary regime is non-zero (ie $p_{11} > 0$).

We can recursively substitute to get:

$$y_t = \sum_{j=0}^{t} \left(\left[\prod_{i=0}^{j-1} (1 - \beta s_{t-i}) \right] (\mu + \nu s_{t-j}) \right) + \left[\prod_{i=0}^{t-1} (1 - \beta s_{t-i}) \right] y_0 + \sum_{j=0}^{t} \left(\left[\prod_{i=0}^{j-1} (1 - \beta s_{t-i}) \right] \epsilon_{t-j} \right) \tag{A-2}$$

where the product term is understood to give 1 when the index is negative.

For a stochastic process $\{x_t\}$ defined by $x_t = \sum_{j=0}^{\infty} \xi_{t-j}$ to be (covariance) stationary a sufficient condition is that uniformly:

$$\lim_{t \to \infty} \sum_{j=0}^{t} E(\xi_{t-j}) = E(x_t) \tag{A-3}$$

$$\lim_{t \to \infty} \frac{1}{t} E \left(\sum_{j=0}^{t} \xi_{t-j} \right)^2 = 0 \tag{A-4}$$

See Nerlove, Grether and Carvalho (1995).

We want the second term in equation (A-2) to converge to zero in mean square so we can ignore this term in the limit, i.e. we would like:

$$E \left[\prod_{i=0}^{t-1} (1 - \beta s_{t-i}) \right]^2 \underset{t \to \infty}{\longrightarrow} 0 \tag{A-5}$$

We need the following Lemmas:

Lemma 1 *If the sum of the absolute values of the elements of every row of a square matrix is less than 1, then all the eigenvalues have modulus less than 1.*

Proof: See Theorem 11.7.2 of Prasolov (1991).

Lemma 2 *If P is a real square matrix with eigenvalues which have all modulus less than 1, then:*

1. $P^j \underset{j\to\infty}{\longrightarrow} 0$
2. $\sum_{j=0}^{\infty} P^j = (I-P)^{-1}$ *exists.*

Proof: See Theorem A.9.1 of Lütkepohl (1993).

We introduce the following notation:

$$\mathbf{1} = (1\,1)', \text{ vector of 1's} \tag{A-6}$$

$$\pi = (\pi_0\,\pi_1)', \text{ the stable probabilities of P} \tag{A-7}$$

The stable probabilities π satisfy $P\pi = \pi$. We assume a unique solution to this exists, or equivalently, we assume that the Markov chain is ergodic.

We also define:

$$\hat{P} = \begin{pmatrix} p_{00} & p_{01}(1-\beta) \\ p_{10} & p_{11}(1-\beta) \end{pmatrix} \tag{A-8}$$

$$\hat{\hat{P}} = \begin{pmatrix} p_{00} & p_{01}(1-\beta)^2 \\ p_{10} & p_{11}(1-\beta)^2 \end{pmatrix} \tag{A-9}$$

$$\bar{P} = \begin{pmatrix} p_{00}\mu & p_{01}(\mu+\nu) \\ p_{10}\mu & p_{11}(\mu+\nu) \end{pmatrix} \tag{A-10}$$

$$\bar{\bar{P}} = \begin{pmatrix} p_{00}\mu^2 & p_{01}(\mu+\nu)^2 \\ p_{10}\mu^2 & p_{11}(\mu+\nu)^2 \end{pmatrix} \tag{A-11}$$

$$\hat{\bar{P}} = \begin{pmatrix} p_{00}\mu & p_{01}(\mu+\nu)(1-\beta) \\ p_{10}\mu & p_{11}(\mu+\nu)(1-\beta) \end{pmatrix} \tag{A-12}$$

$$\hat{\pi} = \begin{pmatrix} \pi_0 \\ (1-\beta)\pi_1 \end{pmatrix} \tag{A-13}$$

$$\hat{\hat{\pi}} = \begin{pmatrix} \pi_0 \\ (1-\beta)^2\pi_1 \end{pmatrix} \tag{A-14}$$

$$\hat{\bar{\pi}} = \begin{pmatrix} \pi_0\mu \\ \pi_1(\mu+\nu)(1-\beta) \end{pmatrix} \tag{A-15}$$

where a bar represents the action of the "constant term", and a hat represents the action of the "mean reversion term" on the probability weights.

Using induction it is easy to show that:

$$E\left[\prod_{i=0}^{t-1}(1-\beta s_{t-i})\right]^2 = (\hat{\hat{\pi}})'(\hat{\hat{P}})^{t-1}\mathbf{1} \tag{A-16}$$

And so using Lemma 1 and 2 on $\hat{\hat{P}}$, $(\hat{\hat{P}})^t \to 0$, so the second term in equation (A-2) converges to zero in mean square.

Taking the first term in equation (A-2) we can show that the conditions for (covariance) stationarity are met. Using Lemmas 1 and 2 applied now to $\hat{P}$ we have:

$$\begin{aligned}\sum_{j=0}^{t} E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\} &= \hat{\pi}'(I+\hat{P}+\hat{P}^2+\cdots+\hat{P}^{t-1})\bar{P}\mathbf{1} \\ &\underset{t\to\infty}{\longrightarrow} \hat{\pi}'(I-\hat{P})^{-1}\bar{P}\mathbf{1} = E(y_t)\end{aligned} \tag{A-17}$$

The second condition (A-4) is a little trickier:

$$\begin{aligned}&\frac{1}{t}E\left\{\sum_{j=0}^{t}\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\}^2 \\ &= \frac{1}{t}\sum_{j=0}^{t} E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\}^2 \\ &\quad + \frac{2}{t}\sum_{k=0}^{t-1}\sum_{j=0}^{k} E\left\{\left[\prod_{i=0}^{k-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-k})\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\}\end{aligned} \tag{A-18}$$

We see that for a simple switching model with s_t independent across time, the second term disappears and the first term in the sum can be factored out by independence, which clearly converges to zero as $t \to \infty$. However, for a regime-switching model with Markov dependence, we have to bound both the first and second terms. First observe that:

$$\begin{aligned}\frac{1}{t}\sum_{j=0}^{t} E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\}^2 &= \frac{1}{t}\hat{\hat{\pi}}(I+\hat{\hat{P}}+\cdots+\hat{\hat{P}}^{t-1})\bar{\bar{P}}\mathbf{1} \\ &\leq \frac{1}{t}\hat{\hat{\pi}}(I-\hat{\hat{P}})^{-1}\bar{\bar{P}}\mathbf{1} \\ &\underset{t\to\infty}{\longrightarrow} 0\end{aligned} \tag{A-19}$$

applying Lemmas 1 and 2 on $\hat{\hat{P}}$.

Now taking the sum to $t=\infty$ in the second term in equation (A-18) we can expand this out as:

$$\begin{aligned}&\sum_{k=0}^{\infty}\sum_{j=0}^{k} E\left\{\left[\prod_{i=0}^{k-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-k})\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right](\mu+\nu s_{t-j})\right\} \\ &= \quad (\hat{\bar{\pi}})'\bar{P}\mathbf{1} \\ &\quad + (\hat{\bar{\pi}})'\hat{P}\bar{P}\mathbf{1} + \ (\hat{\bar{\pi}})'\hat{\bar{P}}\bar{P}\mathbf{1} \\ &\quad + (\hat{\bar{\pi}})'\hat{P}^2\bar{P}\mathbf{1} + (\hat{\bar{\pi}})'\hat{\bar{P}}\hat{P}\bar{P}\mathbf{1} + (\hat{\bar{\pi}})'\hat{\hat{P}}\hat{\bar{P}}\bar{P}\mathbf{1} \\ &\quad + (\hat{\bar{\pi}})'\hat{P}^3\bar{P}\mathbf{1} + (\hat{\bar{\pi}})'\hat{\bar{P}}\hat{P}^2\bar{P}\mathbf{1} + (\hat{\bar{\pi}})'\hat{\hat{P}}\hat{\bar{P}}\hat{P}\bar{P}\mathbf{1} + (\hat{\bar{\pi}})'(\hat{\hat{P}})^2\hat{\bar{P}}\bar{P}\mathbf{1} \\ &\quad + \ldots\end{aligned} \tag{A-20}$$

We can sum each column of the triangular array by applying Lemmas 1 and 2 to $\hat{P}$:

$$
\begin{aligned}
&= (\hat{\bar{\pi}})'(I-\hat{P})^{-1}\bar{P}\mathbf{1} + (\hat{\hat{\pi}})'\hat{\bar{P}}(I-\hat{P})^{-1}\bar{P}\mathbf{1} + (\hat{\hat{\pi}})'\hat{\hat{P}}\hat{\bar{P}}(I-\hat{P})^{-1}\bar{P}\mathbf{1} + (\hat{\hat{\pi}})'(\hat{\hat{P}})^2\hat{\bar{P}}(I-\hat{P})^{-1}\bar{P}\mathbf{1} + \dots \\
&= (\hat{\bar{\pi}})'(I-\hat{P})^{-1}\bar{P}\mathbf{1} + (\hat{\hat{\pi}})'\left[I + \hat{\hat{P}} + (\hat{\hat{P}})^2 + \dots\right]\hat{\bar{P}}(I-\hat{P})^{-1}\bar{P}\mathbf{1} \\
&= (\hat{\bar{\pi}})'(I-\hat{P})^{-1}\bar{P}\mathbf{1} + (\hat{\hat{\pi}})'(I-\hat{\hat{P}})^{-1}\hat{\bar{P}}(I-\hat{P})^{-1}\bar{P}\mathbf{1} \\
&= \left[(\hat{\bar{\pi}})' + (\hat{\hat{\pi}})'(I-\hat{\hat{P}})^{-1}\hat{\bar{P}}\right](I-\hat{P})^{-1}\bar{P}\mathbf{1}
\end{aligned}
\tag{A-21}
$$

Where the penultimate equality results from applying lemmas 1 and 2 to $\hat{\hat{P}}$. Hence it is clear that when divided by t the second term in equation (A-18) converges to zero as $t \to \infty$.

We obtain similar expressions for the third term in equation (A-2):

$$
\sum_{j=0}^{t} E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right]\epsilon_{t-j}\right\} = 0 \tag{A-22}
$$

We use the same trick as above to show the second condition (A-4) is satisfied:

$$
\begin{aligned}
\frac{1}{t}E\left\{\sum_{j=0}^{t}\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right]\epsilon_{t-j}\right\}^2 &= \frac{1}{t}\sum_{j=0}^{t}E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right]\epsilon_{t-j}\right\}^2 \\
&\quad + \frac{2}{t}\sum_{k=0}^{t-1}\sum_{j=0}^{k}E\left\{\left[\prod_{i=0}^{k-1}(1-\beta s_{t-i})\right]\epsilon_{t-k}\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right]\epsilon_{t-j}\right\}
\end{aligned}
\tag{A-23}
$$

The first term in equation (A-23) we can write as:

$$
\begin{aligned}
\frac{1}{t}\sum_{j=0}^{t}E\left\{\left[\prod_{i=0}^{j-1}(1-\beta s_{t-i})\right]\epsilon_{t-j}\right\}^2 &= \frac{1}{t}\hat{\hat{\pi}}(I + \hat{\hat{P}} + \dots + \hat{\hat{P}}^{t-1})\bar{\bar{P}}V \\
&\le \frac{1}{t}\hat{\hat{\pi}}(I - \hat{\hat{P}})^{-1}\bar{\bar{P}}V \\
&\underset{t\to\infty}{\longrightarrow} 0
\end{aligned}
\tag{A-24}
$$

where we define $V = \begin{pmatrix} \sigma_1^2 \\ \sigma_1^2 + \alpha \end{pmatrix}$. The second term in equation (A-23) is zero by the independent sampling across time of the error terms.

Hence the process $\{y_t\}$ in equation (A-2) is composed of two (covariance) stationary terms and so the sum of these is also stationary, thus completing the proof.

We note that strict stationarity will also follow because of the strict stationarity of ϵ_t and s_t.

Table 1: Sample Moments

Panel A: Sample Central Moments						
Parameter	US		GER		UK	
	short rate	spread	short rate	spread	short rate	spread
mean	7.3381	1.2198	6.9045	0.4984	10.5605	0.0643
	(0.4449)	(0.2028)	(0.4197)	(0.2719)	(0.4268)	(0.2491)
variance	8.3103	2.0366	7.1111	3.1241	8.2388	2.7458
	(1.9390)	(0.3833)	(1.3380)	(0.6714)	(1.4354)	(0.5292)
skewness	0.8172	-0.7281	0.6806	-0.5410	-0.1521	-0.2596
	(0.2167)	(0.2782)	(0.2515)	(0.3227)	(0.1797)	(0.2404)
kurtosis	3.6102	3.5921	2.6987	3.3732	2.5406	2.8086
	(0.6718)	(0.7179)	(0.4405)	(0.5768)	(0.3264)	(0.4071)
Panel B: Sample Autocorrelations						
Lag	US		GER		UK	
	short rate	spread	short rate	spread	short rate	spread
1	0.9706	0.8669	0.9845	0.9657	0.9565	0.9322
	(0.0181)	(0.0292)	(0.0216)	(0.0265)	(0.0237)	(0.0238)
2	0.9295	0.7663	0.9583	0.9207	0.8948	0.8776
	(0.0347)	(0.0497)	(0.0436)	(0.0507)	(0.0450)	(0.0425)
3	0.8931	0.6958	0.9253	0.8715	0.8271	0.8234
	(0.0513)	(0.0689)	(0.0638)	(0.0711)	(0.0637)	(0.0596)
4	0.8551	0.6221	0.8858	0.8127	0.7627	0.7692
	(0.0653)	(0.0820)	(0.0812)	(0.0868)	(0.0784)	(0.0753)
5	0.8256	0.5873	0.8428	0.7502	0.7006	0.7200
	(0.0778)	(0.0836)	(0.0957)	(0.0999)	(0.0895)	(0.0895)
6	0.7975	0.5501	0.7943	0.6839	0.6392	0.6689
	(0.0857)	(0.0866)	(0.1071)	(0.1097)	(0.0970)	(0.1016)
7	0.7771	0.5113	0.7423	0.6167	0.5771	0.6119
	(0.0916)	(0.0828)	(0.1166)	(0.1186)	(0.1015)	(0.1139)
8	0.7642	0.5083	0.6888	0.5490	0.5118	0.5553
	(0.0973)	(0.0732)	(0.1246)	(0.1267)	(0.1034)	(0.1205)
9	0.7425	0.4739	0.6363	0.4824	0.4526	0.5164
	(0.0983)	(0.0742)	(0.1319)	(0.1348)	(0.1044)	(0.1245)
10	0.7163	0.4611	0.5858	0.4217	0.3951	0.4711
	(0.0992)	(0.0802)	(0.1381)	(0.1427)	(0.1036)	(0.1291)
Panel C: Sample Cross Correlations						
	Short rates of countries			Short rates/Spreads		
Lag	US/DEM	US/UK	DEM/UK	US	GER	UK
-3	0.4197	0.6470	0.3279	-0.3655	-0.7929	-0.6524
	(0.1334)	(0.0777)	(0.1007)	(0.1130)	(0.0563)	(0.0727)
-2	0.4205	0.6549	0.3523	-0.4213	-0.8326	-0.7016
	(0.1322)	(0.0725)	(0.0964)	(0.1091)	(0.0435)	(0.0607)
-1	0.4120	0.6521	0.3696	-0.4907	-0.8656	-0.7375
	(0.1315)	(0.0686)	(0.0939)	(0.1038)	(0.0317)	(0.0521)
0	0.3953	0.6454	0.3808	-0.5920	-0.8804	-0.7637
	(0.1310)	(0.0678)	(0.0933)	(0.0976)	(0.0284)	(0.0459)
1	0.3756	0.6139	0.3782	-0.5952	-0.8634	-0.7057
	(0.1325)	(0.0698)	(0.0945)	(0.0982)	(0.0335)	(0.0539)
2	0.3542	0.5758	0.3717	-0.5715	-0.8389	-0.6608
	(0.1335)	(0.0754)	(0.0974)	(0.1013)	(0.0406)	(0.0629)
3	0.3294	0.5485	0.3650	-0.5522	-0.8097	-0.6210
	(0.1328)	(0.0828)	(0.1008)	(0.1080)	(0.0477)	(0.0718)

Sample period 1972:01 to 1993:02 (in-sample period). Standard errors are in parentheses and are estimated using Generalized Method of Moments with 6 Newey-West lags. The standard errors are calculated setting up moment conditions for each country separately for each of the central moments and auto-correlations in Panels A and B. In Panel C, the cross-correlations are the estimates of $\frac{\mathrm{cov}(r^{us}_{t+j}, r^{ger}_t)}{\sqrt{\mathrm{var}(r^{us}_t)}\sqrt{\mathrm{var}(r^{ger}_t)}}$ for $j = -3, -2, \ldots, +2, +3$, where each pair of countries is now used to construct the moment conditions.

Table 2: Interest Rate Behavior over the Business Cycle

		US			Germany			UK		
		recession	expansion	χ^2 p-val	recession	expansion	χ^2 p-val	recession	expansion	χ^2 p-val
Number of observations		50	247		149	148		128	169	
	mean	9.6466	6.5970	0.0001	7.4319	5.8327	0.0043	11.9695	8.6856	0.0000
		(0.7064)	(0.3118)		(0.4705)	(0.3028)		(0.3478)	(0.4178)	
	variance	8.4518	6.3173	0.3581	8.9237	4.0253	0.0055	4.7049	8.1835	0.0177
		(1.8775)	(1.3677)		(1.3898)	(1.0894)		(0.8396)	(1.2026)	
short	skewness	0.6841	1.0360	0.5199	0.2782	1.3318	0.0298	0.3151	0.4186	0.7551
rate		(0.4745)	(0.2721)		(0.2448)	(0.4185)		(0.2559)	(0.2113)	
r	kurtosis	2.0077	4.5499	0.0159	2.0590	5.1641	0.0408	2.5627	2.1248	0.3906
		(0.7198)	(0.7708)		(0.2111)	(1.5036)		(0.4227)	(0.2854)	
	ρ_1	0.7858	0.9503	0.0750	0.9436	0.8894	0.2771	0.8650	0.9243	0.2981
		(0.0902)	(0.0201)		(0.0383)	(0.0319)		(0.0504)	(0.0266)	
	ρ_2	0.5657	0.9061	0.0276	0.8878	0.7641	0.1754	0.7674	0.8353	0.4504
		(0.1501)	(0.0368)		(0.0677)	(0.0613)		(0.0746)	(0.0503)	
	mean	0.5568	1.3835	0.0474	0.0247	1.2443	0.0004	-0.4896	0.8025	0.0001
		(0.3903)	(0.1469)		(0.2688)	(0.2179)		(0.2128)	(0.2517)	
	variance	3.1724	1.5362	0.1206	3.0657	2.1719	0.2866	1.6891	2.9732	0.0832
		(1.0170)	(0.2776)		(0.5956)	(0.5906)		(0.2843)	(0.6845)	
	skewness	0.2928	-1.0507	0.0038	-0.5584	-0.8092	0.6436	-0.3903	-0.8350	0.1631
		(0.4090)	(0.2188)		(0.2996)	(0.4517)		(0.2207)	(0.2300)	
spread	kurtosis	3.4900	3.9699	0.6487	2.8995	4.9111	0.1210	2.2583	3.5922	0.1192
z		(0.7742)	(0.7144)		(0.4963)	(1.1986)		(0.3802)	(0.7672)	
	ρ_1	0.6461	0.8630	0.0362	0.9010	0.8590	0.5751	0.8487	0.9123	0.2889
		(0.0987)	(0.0313)		(0.0588)	(0.0465)		(0.0529)	(0.0284)	
	ρ_2	0.3000	0.7700	0.0007	0.8089	0.6918	0.3698	0.7646	0.8200	0.5278
		(0.1308)	(0.0435)		(0.0952)	(0.0895)		(0.0707)	(0.0517)	
	$\rho(r,z)$	-0.2580	-0.6947	0.0413	-0.8813	-0.8591	0.6637	-0.6272	-0.7948	0.0863
		(0.1703)	(0.0886)		(0.0300)	(0.0414)		(0.0876)	(0.0433)	

Sample period: 1972:01 to 1996:08 (full sample). Recessions are defined to be from the peak to the trough of the business cycle. Data for expansions and recessions are from NBER for the US and the Center for International Business Cycle Research at Columbia University (CIBCR) for Germany and the UK. The CIBCR dates end at 1995:12, where Germany is still in recession from post-reunification. Our sample periods ends in 1996:08, and we assume that Germany remains in recession over this period. The results do not change if this assumption is modified. Moments are calculated over expansions and recessions using Generalized Method of Moments with 3 Newey-West lags, with standard errors in parentheses. The symbols ρ_i represents the ith autocorrelation, and $\rho(r,z)$ is the correlation between short rates and spreads. The column denoted χ^2 p-val denotes the p-value from a χ^2 test of equality assuming independence across cycle periods.

Table 3: Summary of Models Estimated

Univariate Models of short rates

One-regime	Two-regime equivalents	
	const probs	time-dep probs
AR(1)	RS1	RS2
(3)	(8)	(10)
GARCH(1,1)	RS3	
(5)	(12)	
CIR	RS4	RS5
(3)	(8)	(10)

Multi-Country Models of short rates

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

Multivariate Models of the Term Spread

One-regime	Two-regime equivalents	
	const probs	time-dep probs
VAR(1)	RSM1	RSM2
(9)	(20)	(24)
VAR(2)		
(13)		

RSM2 for the US has state-dependent probabilities depending only on the spread, and contains 22 parameters. The full model failed to converge.

Table 4: LR_T^* statistics for RS AR(1) model

Grid 1 Results

	LR_T^*	$\sup_\alpha Q_T^*$ distn min	max	mean	median	stdev	p-value
US	7.7534	0.3665	4.5122	1.8606	1.7920	0.6219	0.0000
GER	4.4794	0.2976	4.1095	1.6177	1.5553	0.6146	0.0000
UK	4.4902	0.6444	4.4995	1.9027	1.8570	0.5990	0.0010
AR(1)	1.0213	0.6487	4.8505	1.9771	1.8880	0.6116	0.9670
RS	5.9467	0.5537	4.5376	1.8743	1.8492	0.5782	0.0000

Grid 2 Results

	LR_T^*	$\sup_\alpha Q_T^*$ distn min	max	mean	median	stdev	p-value
US	7.9284	0.3386	4.5492	2.0491	2.0055	0.6321	0.0000
GER	5.8593	0.4622	3.8087	1.8876	1.8401	0.6199	0.0000
UK	6.9021	0.4544	4.3885	2.0577	2.0031	0.6286	0.0000
AR(1)	2.0561	0.9541	5.4106	2.2614	2.1858	0.6125	0.5880
RS	7.0000	0.6197	4.7178	2.0737	2.0069	0.5911	0.0000

LR_T^* is Hansen (1992)'s supremum test statistic for one versus two regimes. To obtain the LR_T^* statistic, the supremum is taken over a finite grid: grid 1 contains 576 points, and grid 2 contains 2016 points. In Hansen's test $p(LR_T^* \geq x)$ is bounded by an asymptotic distribution $p(\sup Q_T^* \geq x)$ which is produced by 1000 simulations. The remaining columns give the distribution of the bounding distribution, with empirical p-values in the last column. The AR(1) row gives the results after simulating a single AR(1) process with parameters ($\rho = 0.98$, $\mu = 0.20$, $\sigma^2 = 0.46$). The RS row gives the results after simulating a single RS AR(1) process with parameters ($\mu_1 = 0.02$, $\mu_2 = 0.30$, $\beta_1 = 0.0025$, $\beta_2 = 0.0100$, $\sigma_1 = 0.25$, $\sigma_2 = 0.80$, $P = 0.9526$, $Q = 0.8808$). For both simulations we used $r_0 = 6.00$.

Table 5: Granger Tests in the Multi-Country VAR Model

Granger-causality	$A[i,j] = 0$	χ^2	p-value
no country Granger-causes another	all off-diagonal elements = 0	12.4435	0.0528
US Granger-causes Germany and UK	$A[2,1] = A[3,1] = 0$	11.6562	0.0029
Germany and UK Granger-cause US	$A[1,2] = A[1,3] = 0$	0.6206	0.7332
Germany and UK Granger-cause each other	$A[2,3] = A[3,2] = 0$	0.7885	0.6742

With $r_t = (r_t^{us}\ r_t^{ger}\ r_t^{uk})'$ we estimate $r_t = v + Ar_{t-1} + u_t$. The Wald tests are performed using a Generalized Methods of Moments estimation of the parameters with 6 Newey-West lags. The notation $A[i,j]$ refers to the element in row i, column j.

Table 6: Granger Tests in the Term Spread VAR Model

	Spreads Granger-cause short rates				Short rates Granger-cause spreads			
	lags=1		lags=2		lags=1		lags=2	
	χ^2 value	pval	χ^2 value	pval	χ^2 value	pval	χ^2 value	pval
US	3.5026	0.0613	3.4868	0.1794	0.0227	0.8802	5.0037	0.0819
GER	1.1257	0.2887	15.4510	0.0004	0.3804	0.5374	21.7171	0.0000
UK	2.2611	0.1327	15.7548	0.0004	0.5111	0.4747	13.1207	0.0014

We estimate a VAR(p) $y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p}$, of $y_t = (r_t\, z_t)'$ short rates and spreads with p the number of lags. We conduct a joint test of $A_i[1,2] = 0$, the entry in row 1 column 2 of A_i, $i = 1,2$ to test if the spread Granger-causes the short rate. We test $A_i^[2,1] = 0$ to test if the short-rate Granger-causes the spread. The Wald tests are performed using a Generalized Methods of Moments estimation of the parameters with 6 Newey-West lags.

Table 7: Univariate models: Unconditional moments and Forecasts

		RS1	RS2	RS3	RS4	RS5	AR(1)	GARCH	CIR
US									
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
Forecasts r_t	MAD	0.1488	-	0.1458	0.1458	0.1487	0.1483	0.1387*	0.1664
	RMSE	0.1956	-	0.1943	0.1945	0.1968	0.1888*	0.1999	0.1999
Forecasts 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
GER									
		RS1	RS2	RS3	RS4	RS5	AR(1)	GARCH	CIR
Central moments	H	67.03	4563.76	5211.55	100.78	17.03*	165.53	-	34.11
	H*	17.80	153.80	4088.45	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
Forecasts r_t	MAD	0.1307	0.1299	0.1329	0.1285	0.1327	0.1501	0.1207*	0.1615
	RMSE	0.1732	0.1732	0.1716	0.1694	0.1732	0.1900	0.1627*	0.2006
Forecasts 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
UK									
		RS1	RS2	RS3	RS4	RS5	AR(1)	GARCH	CIR
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
Forecasts r_t	MAD	0.2509	0.2137*	-	0.2449	0.2288	0.2419	0.2555	0.2539
	RMSE	0.2890	0.2668*	-	0.2819	0.2771	0.2772	0.2910	0.2893
Forecasts 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

RS1 refers to the constant probability RS AR(1) model.
RS2 refers to the state-dependent probability RS AR(1) model.
RS3 refers to the RS GARCH(1,1) model with constant probabilities.
RS4 refers to the RS square root CIR model with constant probabilities.
RS5 is the CIR model with time-varying probabilities.
Dashes mean that the model could not be satisfactorily estimated for RS2 and RS3, and non-stationarity for GARCH(1,1).
Asterixed values are the lowest statistic values.

Table 8: Multi-Country: Unconditional Moments and Forecasts

				US			
		VAR1u	D1	RSD1	G1	RSG1	RSG2
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*	0.97	0.25*	13.44	1.29	19.84	27.34
Forecasts r_t	MAD	0.1619	0.1499	0.1378	0.1483	0.1160*	0.1174
	RMSE	0.2002	0.1891	0.1841	0.1888	0.1625*	0.1626
Forecasts r_t^2	MAD	1.5550	1.7159	1.2139	1.3992	0.9949*	1.0388
	RMSE	1.8065	2.0771	1.4980	1.6453	1.1930*	1.2146
				GER			
		VAR1u	D1	RSD1	G1	RSG1	RSG2
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
Forecasts r_t	MAD	0.1580	0.1500	0.1429	0.1327*	0.1451	0.1466
	RMSE	0.1959	0.1899	0.1868	0.1704*	0.2035	0.2062
Forecasts 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
				UK			
		VAR1u	D1	RSD1	G1	RSG1	RSG2
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
Forecasts r_t	MAD	0.2747	0.2410	0.1429	0.2668	0.1124*	0.1142
	RMSE	0.3116	0.2762	0.2017	0.3055	0.1766*	0.1833
Forecasts 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

VAR1u refers to the unconstrained VAR of lag length 1.
D1 is the one-regime model with different AR(1) processes for each country.
RSD1 is the RS diagonal model with the α_i, ρ_i, σ_i, P, Q across countries.
G1 is the one-regime Granger-causality model for UK and Germany from the US.
RSG1 is the RS Granger-causality model with the same α_i, ρ_i, σ_i, P, Q across countries.
Asterixed values are the lowest statistic values.

Table 9: Term Spread Models: Unconditional Moments

US

			VAR1	VAR2	RSM1	RSM2
Central Moments	r_t	H	31.51	29.99*	193.20	141.95
		H*	15.26*	15.27*	97.05	54.83
	z_t	H	10.09	10.07*	119.70	30.96
		H*	7.62*	7.62	86.30	21.77
Autocorrelations	r_t	H	2.46*	890.48	4.32	5.13
		H*	0.84*	17.35	1.33	12.11
	z_t	H	21.70	5724.77	16.82	10.58*
		H*	43.26	68.67	69.27	14.52*
Crosscorrelation	$r_t z_t$	H	86.99	444.51	12.73	2.05*
		H*	16.24	8.82	31.29	0.19*

Germany

			VAR1	VAR2	RSM1	RSM2
Central Moments	r_t	H	232.01	157.55*	374.27	268.22
		H*	8.24	8.08*	20.11	11.40
	z_t	H	6.43	6.03*	38.98	18.69
		H*	3.29*	3.69	10.04	5.97
Autocorrelations	r_t	H	7.41	2941.41	6.65	6.04*
		H*	15.28	23.09	13.90	10.63*
	z_t	H	8.50*	316.92	15.57	14.66
		H*	17.39*	34.19	51.67	47.65
Crosscorrelation	$r_t z_t$	H	6.96*	142.92	17.30	10.80
		H*	8.89	7.71	10.91	4.21*

UK

			VAR1	VAR2	RSM1	RSM2
Central Moments	r_t	H	4.84*	4.93	23.51	32.49
		H*	3.03	3.00*	4.33	5.60
	z_t	H	2.25	2.17*	9.80	11.09
		H*	1.42	1.40*	7.63	9.15
Autocorrelations	r_t	H	8.04*	50.26	8.69	8.98
		H*	16.26*	60.28	19.00	21.69
	z_t	H	2.82*	119.41	2.99	3.09
		H*	0.38*	21.27	2.00	2.42
Crosscorrelation	$r_t z_t$	H	7.87*	199.73	17.00	11.36
		H*	11.44	10.93	9.34	2.09*

The variables r_t and z_t refer to the short rate and spread, respectively.
VAR1 refers to a VAR model of lag length 1, VAR2 to a VAR model of lag length 2, RSM1 to a bivariate regime switching model with constant probabilities, RSM2 to the state-dependent probability model. RSM2 for the US contains state-dependent probabilities depending only on the spread, as the full model failed to converge.
Asterixed values are the lowest statistic values.

Table 10: Term Structure: First and Second moment forecasts

	VAR1			VAR2			RSM1			RSM2		
US												
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	0.1885	0.2117		0.1918	0.2186		0.1531*	0.2070*		0.1588	0.2072	
RMSE	0.2312	0.2760		0.2490	0.2735		0.1948*	0.2653*		0.2025	0.2650	
	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	$r_t z_t$
MAD	2.1183	0.7141	1.2672	2.1267	0.7422	1.2122	1.5907*	0.6267*	1.1044*	1.6529	0.6280	1.0964
RMSE	2.5226	0.9181	1.5309	2.6521	0.9534	1.4835	1.9941*	0.8292*	1.1391*	2.0755	0.8287	1.3870
Germany												
	VAR1			VAR2			RSM1			RSM2		
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	0.1359	0.2191		0.1796	0.2395		0.1197	0.2214		0.1074*	0.2186*	
RMSE	0.1765	0.2796		0.2219	0.2826		0.1617	0.2821		0.1471*	0.2755*	
	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	$r_t z_t$
MAD	1.3464	0.7084*	0.9882*	1.6568	0.8016	1.0049	1.1333	0.7168	1.0243	0.9772*	0.7225	1.0092
RMSE	1.7109	0.8757*	1.3273	2.1278	0.9241	1.2957	1.5495	0.8909	1.3346	1.3531*	0.8962	1.2937*
UK												
	VAR1			VAR2			RSM1			RSM2		
	r_t	z_t		r_t	z_t		r_t	z_t		r_t	z_t	
MAD	0.2172	0.2455*		0.2607	0.2603		0.2211	0.2491		0.1768*	0.2590	
RMSE	0.2529	0.3031		0.3112	0.3384		0.2180*	0.3078		0.2414	0.3121	
	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	$r_t z_t$
MAD	3.0729	1.0989*	1.2995*	3.5870	1.1921	1.3297	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	3.5814	1.5343	1.5416	3.1786*	1.5423	1.5746

The variables r_t and z_t refer to the short rate and spread, respectively.
VAR1 refers to a VAR model of lag length 1, VAR2 to a VAR model of lag length 2, RSM1 to a bivariate regime switching model with constant probabilities, RSM2 to the state-dependent probability model. RSM2 for the US contains state-dependent probabilities depending only on the spread, as the full model failed to converge.
Asterixed values are lowest statistic values.

Table 11: Overall Moments and Forecast Comparisons for short rates

Best H-statistics

	US	Ger	UK
Central moments	CIR	RS5	RS1
Autocorrelogram	VAR1	RS3	VAR2

Best RMSE-statistics

	US	Ger	UK
r_t	RSG1	RSM2	RSG1
r_t^2	RSG1	RSM2	RSG1

See Table (3) for nomenclature of the models estimated.

Table 12: Small Sample Experiment: % 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%

We simulate data of length 297 from the joint estimation from Table (A-10) of a bivariate system of the short rate r_t and spread z_t. 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. We record which model gives the lowest H and RMSE statistics. The table lists the percentage times of which model performed the best in small the small sample. We conducted 1000 simulations.

For unconditional moments "r_t central" refers to the H-statistic for the mean, variance, skewness and kurtosis of r_t over the in-sample period of 267, $\rho(r_t)$ refers to the H-statistic for the first 10 autocorrelations of the r_t, while $\rho(r_t, z_t)$ refers to the cross-correlations from lags -3 to +3 of the r_t and z_t.

The forecasts use RMSE over an out-sample size of 30.

Table 13: Small Sample Distribution of Moments

Short Rates

	Population	AR	RSAR	VAR	RSVAR
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	0.5750		0.2032		0.1185
			(0.1700)		(0.3087)
Kurtosis	3.0639		3.1360		3.2287
			(0.3263)		(3.3094)

Spreads

	Population	AR	RSAR	VAR	RSVAR
Mean	0.8642			0.8509	0.3410
				(0.3903)	(0.4304)
Variance	1.5460			1.4306	1.0500
				(0.5161)	(0.2705)
Skewness	-0.1815				-0.0790
					(0.2812)
Kurtosis	3.0084				3.2709
					(1.8155)

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 (12). The skewness and kurtosis for the AR and VAR models was not recorded because these are theoretically 0 and 3 respectively.

Table 14: RCM Statistics

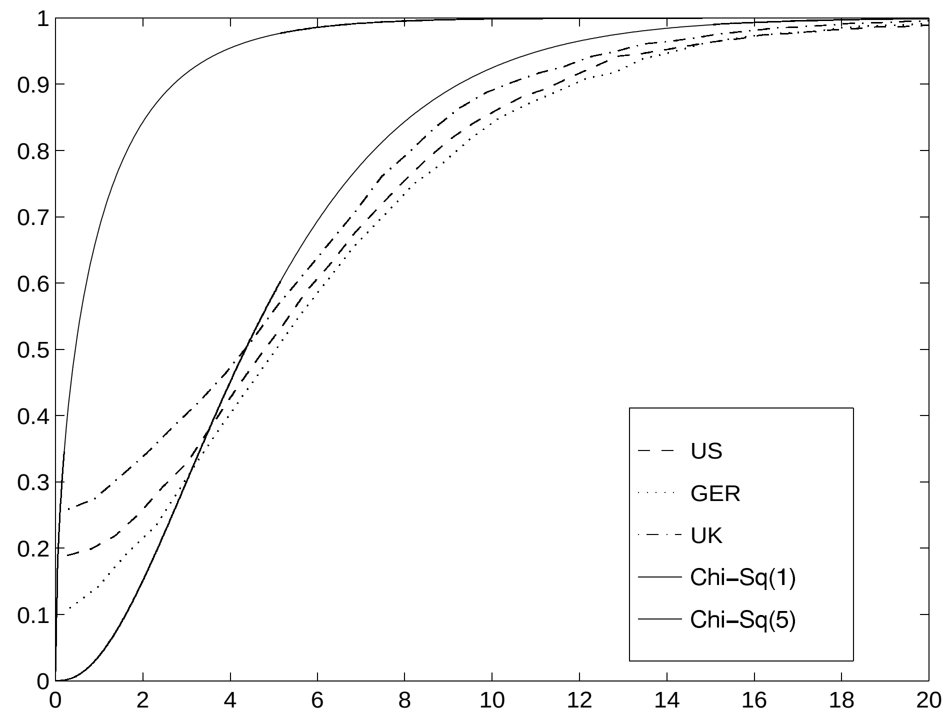
	US	Ger	UK
RS1	28.15	55.29	96.14
RS2	-	51.84	95.12
RS3	37.44	75.30	-
RS4	29.61	46.87	96.56
RS5	23.29	46.13	94.21
RSD1	40.73	30.41	49.75
RSG1	51.21	41.28	52.54
RSG2	50.90	44.07	51.04
RSM1	24.73	44.38	96.12
RSM2	22.19	26.34	86.28

See Table (3) for nomenclature of the models estimated.
Blanks in the table mean the RS model could not be estimated. RSM2 for the US contains state-dependent probabilities depending only on the spread, rather than both the short rate and the spread, as the full mode failed to converge.

Table 15: Markov Regimes and Business Cycles

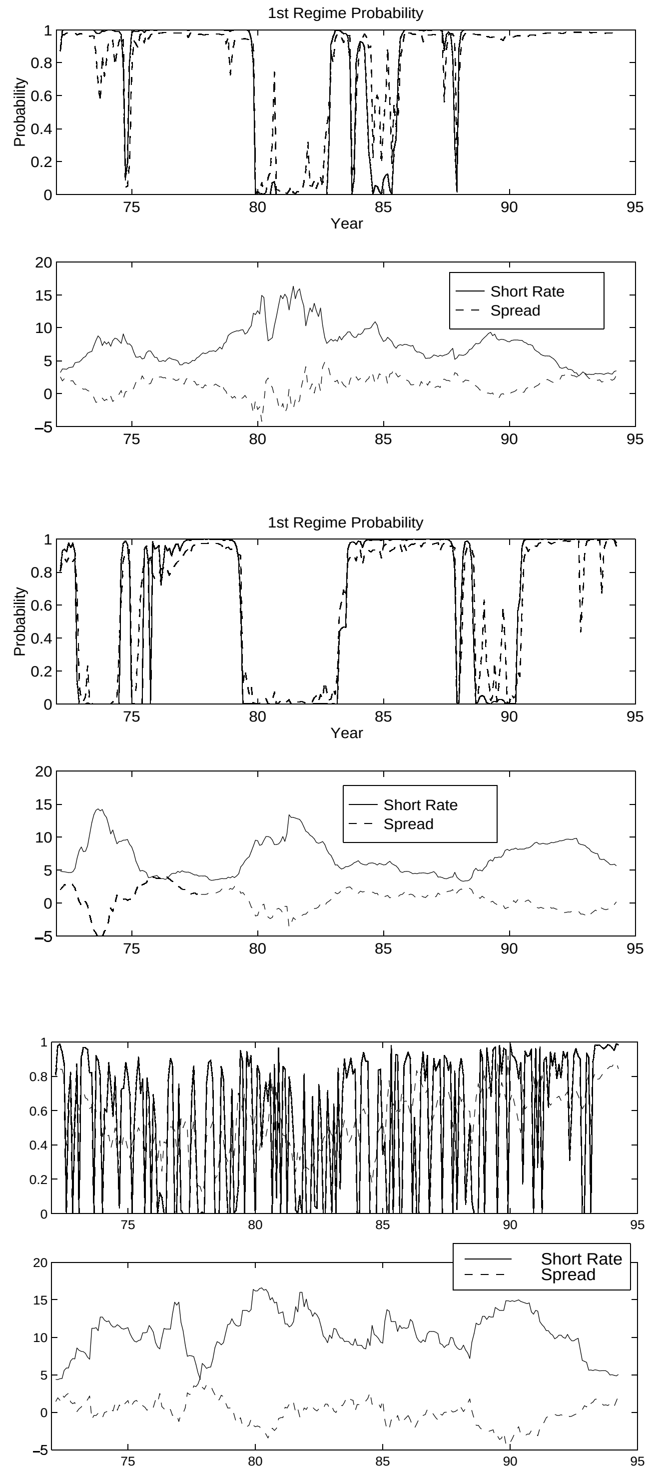
			US			
	Correlations			Probit Forecasting		
mths ahead j	$\rho(1-p_{t-j+1}, \text{rec}_t)$	$\rho(z_{t-j}, \text{rec}_t)$	$\beta(1-p_{t-j+1})$	%forecast	$\beta(z_{t-j})$	%forecast
1	0.4264	-0.3047	1.6203	83.8	-0.2811	80.8
	(0.1153)	(0.1104)	(0.2569)		(0.0605)	
2	0.4618	-0.3989	1.7537	84.2	-0.3847	82.3
	(0.1149)	(0.1028)	(0.2603)		(0.0645)	
4	0.4840	-0.5096	1.8428	84.4	-0.5611	86.7
	(0.1123)	(0.0851)	(0.2640)		(0.0760)	
6	0.4122	-0.5296	1.5569	85.1	-0.5750	87.0
	(0.1126)	(0.0820)	(0.2584)		(0.0745)	
			Germany			
	Correlations			Probit Forecasting		
mths ahead j	$\rho(1-p_{t-j+1}, \text{rec}_t)$	$\rho(z_{t-j}, \text{rec}_t)$	$\beta(1-p_{t-j+1})$	%forecast	$\beta(z_{t-j})$	%forecast
1	0.1892	-0.5276	0.5789	60.2	-0.4903	75.2
	(0.1109)	(0.0719)	(0.1879)		(0.0601)	
2	0.2162	-0.5830	0.6632	61.5	-0.6073	75.8
	(0.1107)	(0.0615)	(0.1890)		(0.0696)	
4	0.2472	-0.6590	0.7615	63.9	-0.8474	77.9
	(0.1101)	(0.0508)	(0.1908)		(0.0927)	
6	0.2392	-0.6811	0.7366	63.6	-0.9400	81.6
	(0.1106)	(0.0483)	(0.1915)		(0.1024)	
			UK			
	Correlations			Probit Forecasting		
mths ahead j	$\rho(1-p_{t-j+1}, \text{rec}_t)$	$\rho(z_{t-j}, \text{rec}_t)$	$\beta(1-p_{t-j+1})$	%forecast	$\beta(z_{t-j})$	%forecast
1	0.0911	-0.3439	0.6856	54.1	-0.2821	67.3
	(0.1066)	(0.0999)	(0.4590)		(0.0506)	
2	0.0779	-0.3828	0.5864	53.6	-0.3218	69.4
	(0.1067)	(0.0962)	(0.4601)		(0.0522)	
4	0.0098	-0.4508	0.0740	51.3	-0.4018	74.1
	(0.1077)	(0.0899)	(0.4646)		(0.0564)	
6	-0.0230	-0.4680	-0.1756	49.0	-0.4274	72.4
	(0.1063)	(0.0837)	(0.4710)		(0.0584)	

Recessions are coded as a 1, expansions as 0, and are obtained from NBER and the Center for International Business Cycle Research at Columbia University (CIBCR). 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). (Note that p_t is in the information set at time $t-1$.) The first regime corresponds to the unit root regime with lower conditional volatility, so $1-p_t$ is the probability of being in the second higher conditional volatility mean-reverting regime. The first two columns 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 3 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 $F(.)$ 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.



χ_1^2 distribution, the bottom solid line is a χ_5^2.

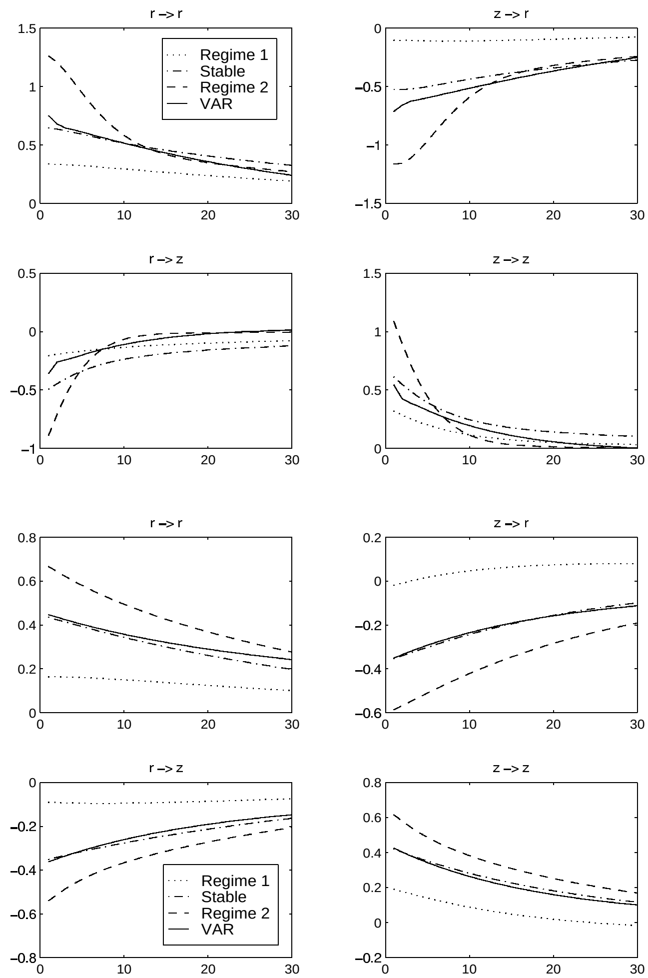
Figure 1: Empirical Distribution of the Likelihood Ratio Test



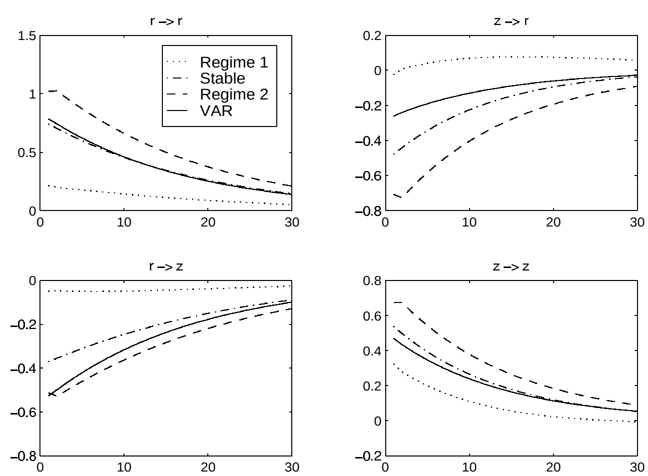
$p(s_t = 1|\mathcal{I}_{t-1})$ (dotted line) and smoothed probabilities $p(s_t = 1|\mathcal{I}_T)$ (solid line). The bottom panels show the short rate and spread.

Figure 2: VAR RS model: time-varying probs

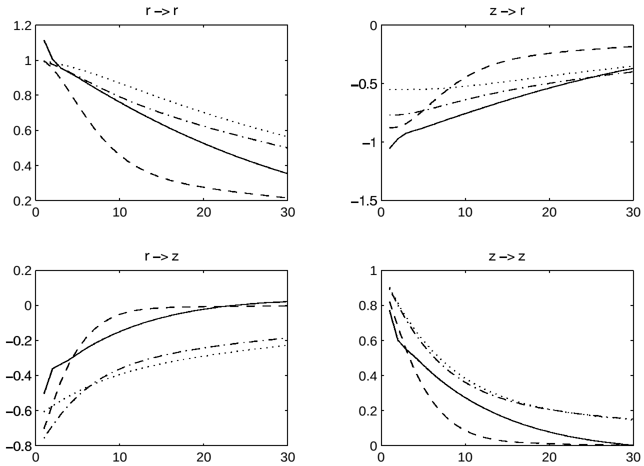
US


r –> r
z –> r
r –> z
z –> z
Regime 1
Stable
Regime 2
VAR

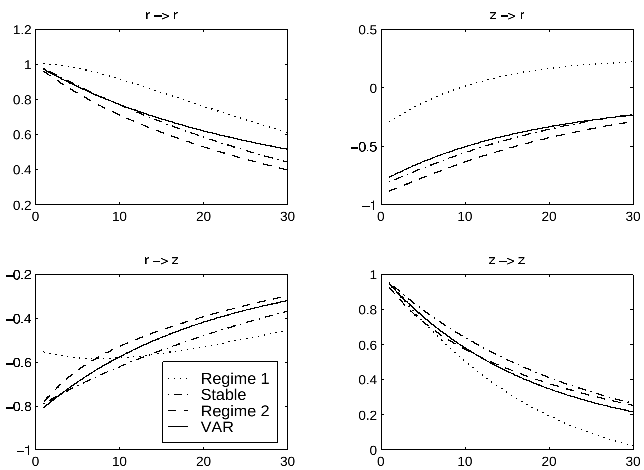
UK


r –> r
z –> r
r –> z
z –> z
Regime 1
Stable
Regime 2
VAR

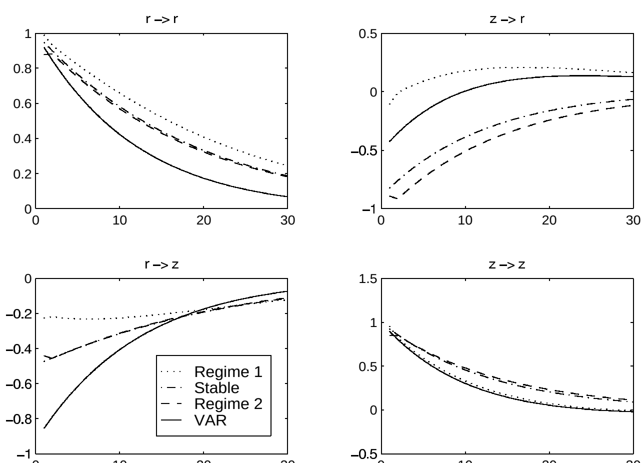
US – unit shock

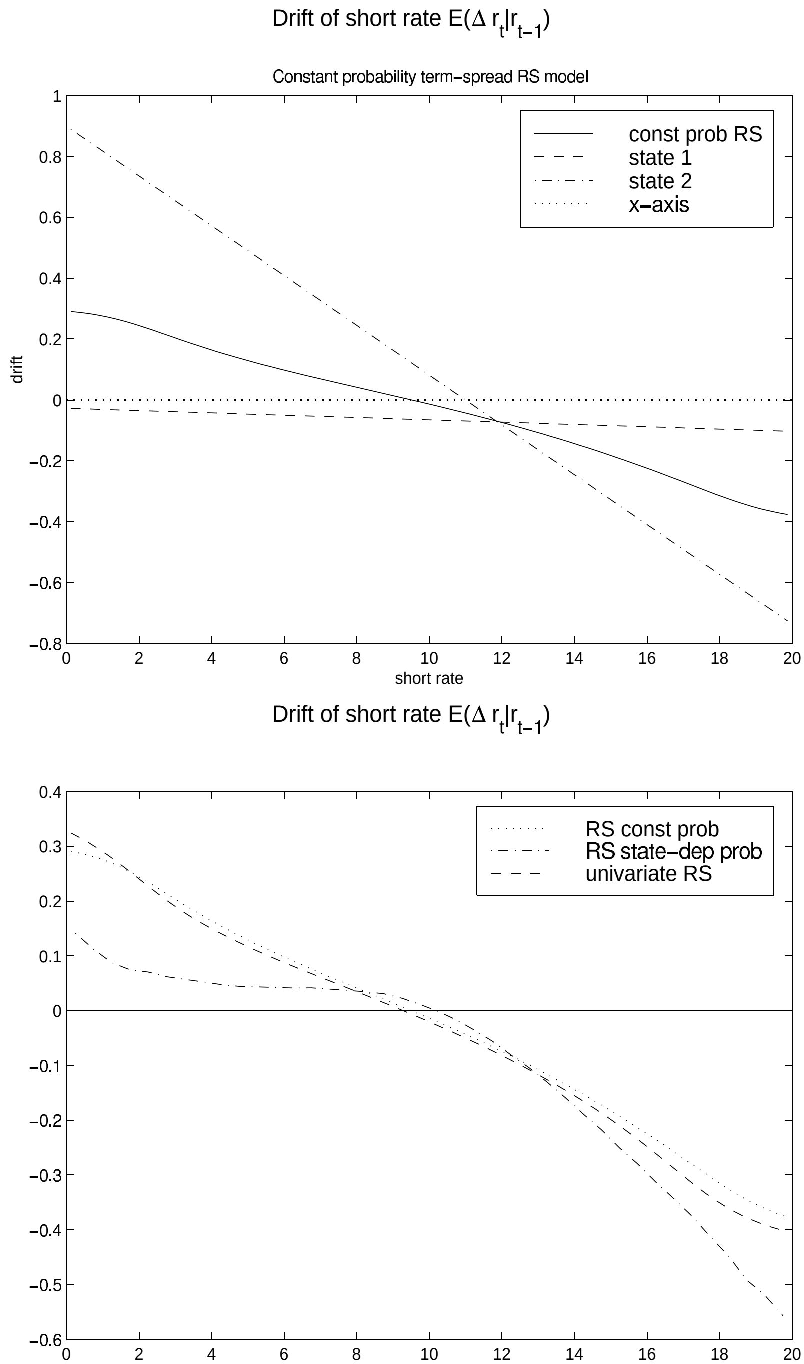


GER – unit shock



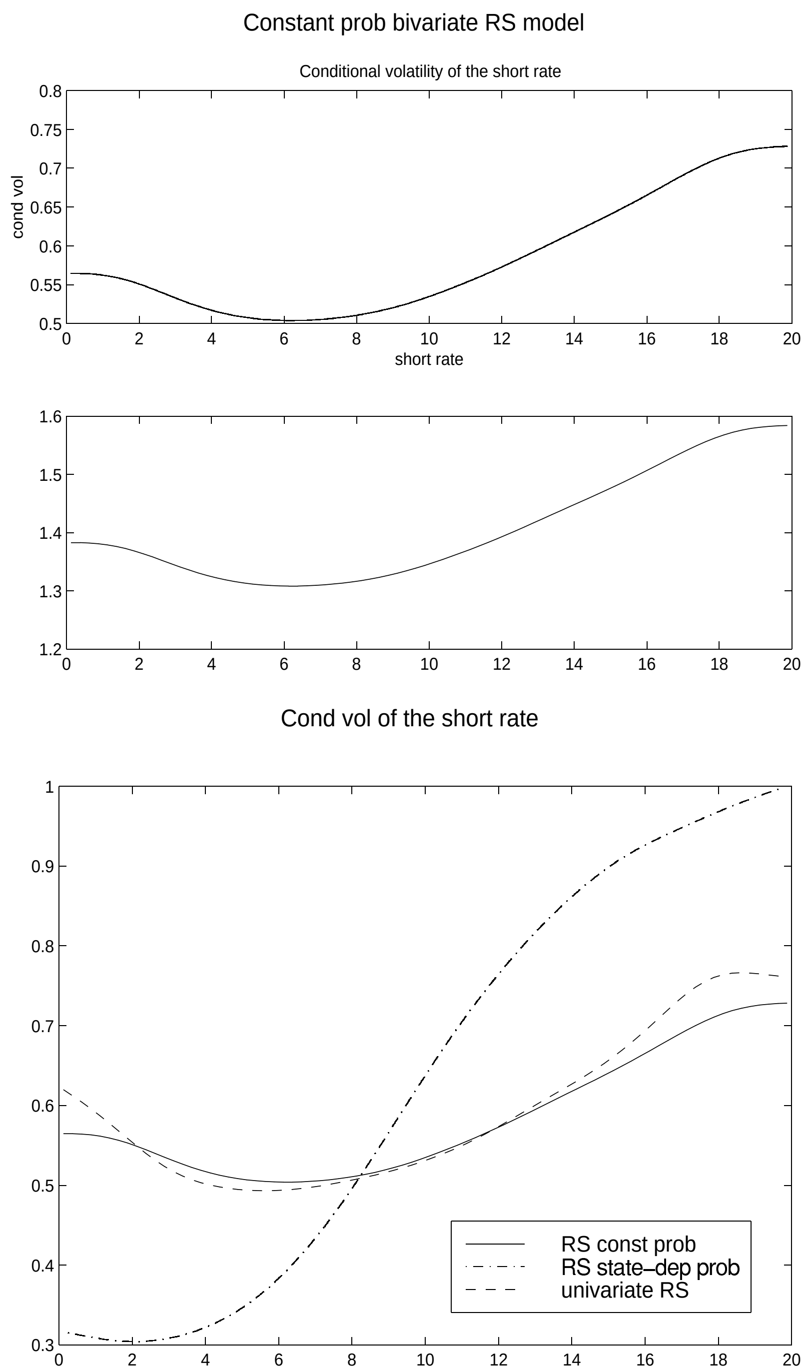
UK – unit shock





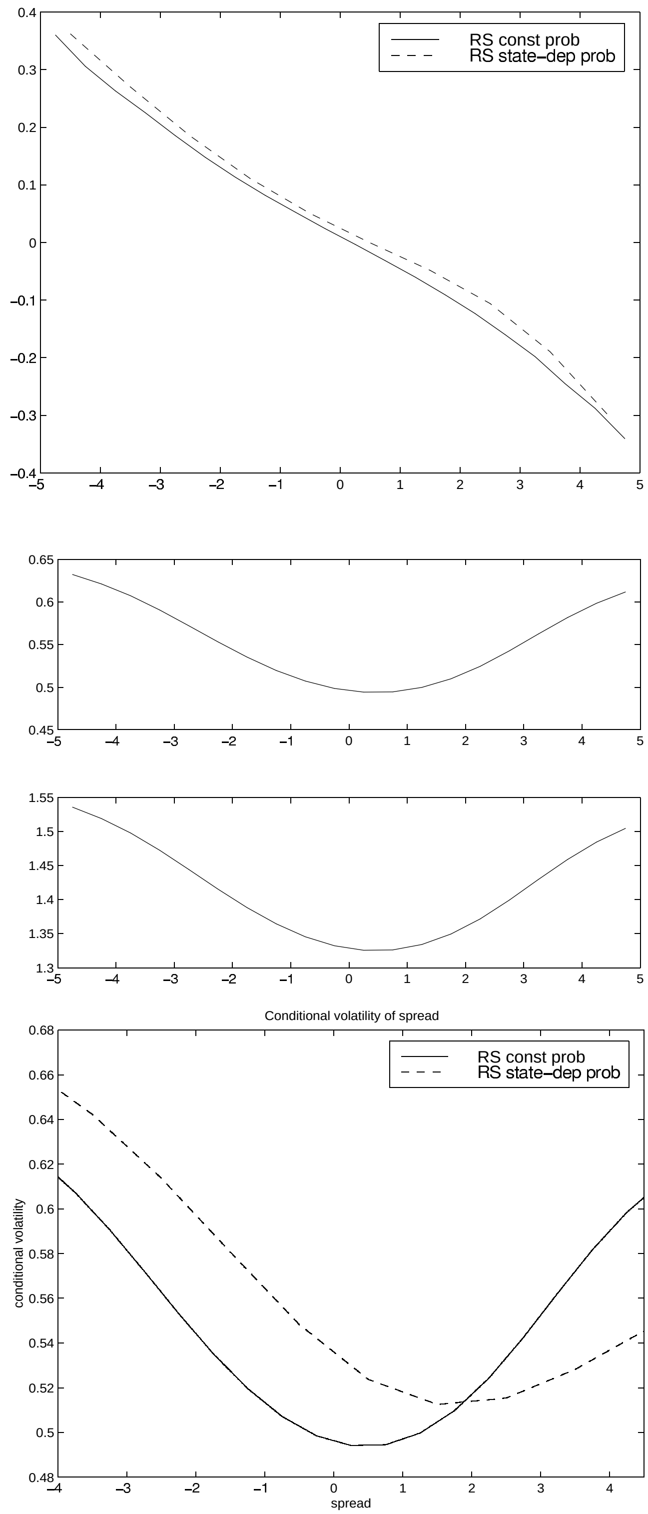
r_t from the bivariate term spread RS VAR with constant transition probabilities, after integrating out the spread z_t and the state s_t. The dashed lines represent the the drifts conditional on each state. The second plot gives the drifts of the short rate from 3 RS models: the RS VAR with constant transition probabilities, the RS VAR with time-varying transition probabilities, and the univariate RS short rate model with constant transition probabilities. All RS models are estimated jointly over the US, Germany and UK.

Figure 5: Drift functions of RS models



r_t from the bivariate term spread RS VAR with constant transition probabilities, after integrating out the spread z_t and the state s_t. The top panel gives the conditional volatility itself while the bottome panel shows the expected state associated with each level of the short rate. The second plot gives the conditional volatilities of the short rate from 3 RS models: the RS VAR with constant transition probabilities, the RS VAR with time-varying transition probabilities, and the univariate RS short rate model with constant transition probabilities. All RS models are estimated jointly over the US, Germany and UK.

Figure 6: Conditional Volatility of RS models



z_t from the bivariate term spread RS VAR with constant transition probabilities and the bivariate term spread RS VAR with time-varying probabilities after integrating out the spread r_t and the state s_t. The middle plot shows the the conditional volatility along with the expected state from the bivariate term spread RS VAR with constant transition probabilities. The bottom plot shows the conditional volatiltiy of the spread for both RS models.

Figure 7: Conditional Drift and Volatility of the Term Spread

Appendix: Tables of Estimated Models

Table A-1: Univariate RS AR(1) model: Constant Transition Probabilities

	United States			Germany			Gt Britain			Joint		
	Param	Std error	pvalue	Param	Std error	pvalue	Param	Std error	pvalue	Param	Std error	pvalue
μ_1	0.0284	0.0228	0.2135	-0.0089	0.0464	0.8475	-0.0786	0.0721	0.2760	0.0567	0.0311	0.0680
μ_2	0.6887	0.5934	0.2457	0.3990	0.2398	0.0961	1.3380	0.4677	0.0042	0.7541	0.2338	0.0013
β_1	0.0000	0.0000	0.9999	0.0007	0.0053	0.8859	0.0023	0.0047	0.6259	0.0120	0.0005	0.0000
β_2	0.0739	0.0383	0.0537	0.0452	0.0193	0.0190	0.1127	0.0297	0.0002	0.0702	0.0059	0.0000
σ_1	0.3002	0.0180	0.0000	0.1681	0.0130	0.0000	0.2101	0.0189	0.0000	0.2382	0.0137	0.0000
σ_2	1.2725	0.1215	0.0000	0.7178	0.0585	0.0000	1.1567	0.0836	0.0000	1.0528	0.0525	0.0000
P	0.9757	0.0680	0.0000	0.9198	0.0305	0.0000	0.6272	0.0606	0.0000	0.8960	0.0239	0.0000
Q	0.9199	0.0548	0.0000	0.8584	0.0560	0.0000	0.5410	0.0779	0.0000	0.8142	0.0436	0.0000

A likelihood test for P+Q=1 for the UK gives a χ^2 value of 2.4152, with p-value 0.1202.

Estimated model:

$$\Delta r_t = \mu(s_t) - \beta(s_t) r_{t-1} + \sigma(s_t)\epsilon_t$$
$$\epsilon_t \sim \text{IID}\, N(0,1)$$
$$P = p(s_t = 1 | s_{t-1} = 1)$$
$$Q = p(s_t = 2 | s_{t-1} = 2)$$

Table A-2: Univariate RS models: Time-varying Transition Probabilities

	Germany			UK		
Parameter	Est	Std error	pvalue	Est	Std error	pvalue
μ_1	-0.0078	0.0157	0.6175	-0.0707	0.0716	0.3236
μ_2	0.3851	0.2407	0.1096	1.3385	0.4636	0.0039
β_1	0.0000	0.0003	1.0000	0.0032	0.0283	0.9098
β_2	0.0438	0.0585	0.4540	0.1122	0.0718	0.1179
σ_1	0.1647	0.0128	0.0000	0.2091	0.0185	0.0000
σ_2	0.7123	0.0607	0.0000	1.1545	0.0836	0.0000
a_1	2.7918	1.2205	0.0222	1.8502	0.9368	0.0483
b_1	-0.0804	0.1885	0.6698	-0.1357	0.0932	0.1452
a_2	-1.0862	1.8266	0.5521	1.3594	1.2152	0.2633
b_2	0.3726	0.2845	0.1904	-0.1189	0.1120	0.2885
LR test	χ^2=6.12, pvalue = 0.0468			χ^2=2.55, pvalue = 0.2800		

LR refers to a likelihood ratio test for $b_1, b_2 = 0$.

US model and joint estimation failed to converge.

Estimated model:

$$\Delta r_t = \mu(s_t) - \beta(s_t) r_{t-1} + \sigma(s_t)\epsilon_t, \quad \epsilon_t \sim \text{IID}\, N(0,1)$$
$$p(s_t = i | s_{t-1} = i) = \frac{e^{a_i + b_i r_{t-1}}}{1 + e^{a_i + b_i r_{t-1}}}, \qquad i = 1, 2.$$

Table A-3: Univariate RS GARCH(1,1) model

	US			GER		
Parameter	Est	Std error	pvalue	Est	Std error	pvalue
μ_1	0.0323	0.4066	0.9366	0.0200	0.0174	0.2503
μ_2	0.6927	0.5548	0.2119	0.3668	0.1179	0.0019
β_1	0.0032	0.0781	0.9672	0.0000	0.0000	1.0000
β_2	0.0700	0.0470	0.1365	0.0814	0.0175	0.0000
a_{01}	0.0503	0.0160	0.0016	0.0120	0.0033	0.0003
a_{02}	0.5123	1.7291	0.7670	0.0714	0.0332	0.0317
a_{11}	0.0284	0.4464	0.9500	0.8470	0.2125	0.0001
a_{11}	0.0393	0.2756	0.8865	0.8371	0.2840	0.0032
b_{01}	0.2112	0.3000	0.4814	0.0000	0.0000	1.0000
b_{11}	0.8213	1.5084	0.5861	0.0050	0.0893	0.9551
P	0.9731	0.0326	0.0000	0.9051	0.0400	0.0000
Q	0.9296	0.0377	0.0000	0.9063	0.0582	0.0000

UK model failed to converge.

Estimated model:

$$\Delta r_t = \mu(s_t) - \beta(s_t) r_{t-1} + \sqrt{h_t(s_t)} u_t,\ u_t \sim \text{IID}\, N(0,1)$$
$$h_t(s_t) = a_0(s_t) + a_1(s_t)\epsilon_{t-1}^2 + b_1(s_t) h_{t-1}$$
$$h_t = E_{t-1}[r_t^2] - (E_{t-1}[r_t])^2$$
$$\epsilon_t = r_t - E_{t-1}[r_t]$$
$$P = p(s_t = 1 | s_{t-1} = 1),\ Q = p(s_t = 2 | s_{t-1} = 2)$$

Table A-4: Univariate RS CIR - constant probs

	US			GER			UK		
Parameter	Est	Std error	pvalue	Est	Std error	pvalue	Est	Std error	pvalue
μ_1	0.0471	0.0670	0.4817	0.0701	0.0446	0.1157	-0.0145	0.0663	0.8273
μ_2	0.5273	0.5076	0.2990	0.2883	0.2223	0.1947	1.2283	0.4378	0.0050
β_1	0.0044	0.0008	0.0000	0.0182	0.0010	0.0000	0.0086	0.0006	0.0000
β_2	0.0564	0.0120	0.0000	0.0235	0.0046	0.0000	0.1023	0.0135	0.0000
σ_1	0.1239	0.0076	0.0000	0.0800	0.0055	0.0000	0.0679	0.0059	0.0000
σ_2	0.3785	0.0372	0.0000	0.2608	0.0238	0.0000	0.3764	0.0277	0.0000
P	0.9806	0.0120	0.0000	0.9509	0.0273	0.0000	0.6362	0.0591	0.0000
Q	0.9338	0.0366	0.0000	0.8922	0.0645	0.0000	0.5170	0.0709	0.0000

A Wald test for P+Q=1 for the UK gives a χ^2 value of 2.7541, with p-value 0.0970.

Estimated model:

$$\Delta r_t = \mu(s_t) - \beta(s_t) r_{t-1} + \sigma(s_t) r_{t-1}^{\frac{1}{2}} \epsilon_t$$
$$\epsilon_t \sim \text{IID}\, N(0,1)$$
$$P = p(s_t = 1 | s_{t-1} = 1),\ Q = p(s_t = 2 | s_{t-1} = 2)$$

Table A-5: Univariate RS CIR - time varying probs

	US			GER			UK		
Parameter	Est	Std error	p value	Est	Std error	p value	Est	Std error	p value
μ_1	0.0330	0.0219	0.1311	0.0706	0.0449	0.1163	-0.0152	0.0616	0.8054
μ_2	0.4183	0.5942	0.4814	0.3010	0.2188	0.1688	1.2448	0.4379	0.0045
β_1	0.0000	0.0000	0.0000	0.0183	0.0199	0.3595	0.0088	0.0205	0.6694
β_2	0.0497	0.1205	0.6799	0.0262	0.0714	0.7143	0.1037	0.0746	0.1646
σ_1	0.1225	0.0081	0.0000	-0.0783	0.0063	0.0000	-0.0680	0.0057	0.0000
σ_2	0.3894	0.0441	0.0000	0.2559	0.0215	0.0000	0.3763	0.0277	0.0000
a_1	6.4688	3.1058	0.0373	1.4738	1.4260	0.3014	1.6577	0.9211	0.0719
b_1	-0.4288	0.3620	0.2362	0.2398	0.2614	0.3589	-0.1078	0.0885	0.2231
a_2	-7.5211	5.1933	0.1476	-0.0445	1.5631	0.9773	1.9314	1.2441	0.1206
b_2	1.1684	0.7200	0.1046	0.3056	0.2105	0.1466	-0.1806	0.1143	0.1141
		χ^2	pvalue		χ^2	pvalue		χ^2	pvalue
Wald Test		4.47	0.1068		0.46	0.7948		3.08	0.2147
LR Test		14.30	0.0008		2.03	0.3627		3.16	0.2061

Wald and likelihood ratio (LR) test refer to b_1, $b_2 = 0$.

Estimated model: $\Delta r_t = \mu(s_t) - \beta(s_t) r_{t-1} + \sigma(s_t) r_{t-1}^{\frac{1}{2}} \epsilon_t \quad \epsilon_t \sim$ IID $N(0,1)$

$$p(s_t = i | s_{t-1} = i) = \frac{e^{a_i + b_i r_{t-1}}}{1 + e^{a_i + b_i r_{t-1}}}, \qquad i = 1, 2.$$

Table A-6: Multi-country one-regime restricted models

Granger causality: G1

Parameter	Estimate	Std	p value
α_{us}	0.2162	0.1161	0.0626
α_{ger}	-0.0067	0.0906	0.9410
α_{uk}	0.4202	0.1878	0.0253
ρ_{us}	0.9706	0.0147	0.0000
ρ_{ger}	0.9735	0.0114	0.0000
ρ_{uk}	0.9114	0.0223	0.0000
ζ_{ger}	0.0261	0.0106	0.0138
ζ_{uk}	0.0706	0.0222	0.0015
γ_{ger}	0.0066	0.0279	0.8139
γ_{uk}	0.1062	0.0488	0.0295
σ_{us}	0.6893	0.0299	0.0000
σ_{ger}	0.4557	0.0198	0.0000
σ_{uk}	0.7917	0.0343	0.0000

Likelihood ratio test of this model vs unrestricted VAR $\chi^2 = 1.0708$ with p-value = 0.9567.

Likelihood ratio test of $\rho_{us} = \rho_{ger} = \rho_{uk} = \rho$ gives $\chi^2 = 6.3702$ with p-value = 0.0414.

Estimated model: $r_t = \alpha + Ar_{t-1} + \epsilon_t$ where $A = \begin{pmatrix} \rho_{us} & 0 & 0 \\ \zeta_{ger} & \rho_{ger} & 0 \\ \zeta_{uk} & 0 & \rho_{uk} \end{pmatrix}$. Covariance is homoskedastic given by the one-regime equivalent of eqn (30).

Diagonal Model: D1

Parameter	Estimate	Std	p value
α_{us}	0.2563	0.1177	0.0294
α_{ger}	0.1082	0.0786	0.1685
α_{uk}	0.4595	0.1909	0.0161
ρ_{us}	0.9652	0.0149	0.0000
ρ_{ger}	0.9846	0.0106	0.0000
ρ_{uk}	0.9568	0.0174	0.0000
γ_{ger}	0.0089	0.0286	0.7557
γ_{uk}	0.1119	0.0502	0.0258
σ_{us}	0.6895	0.0299	0.0000
σ_{ger}	0.4608	0.0120	0.0000
σ_{uk}	0.8064	0.0350	0.0000

Likelihood ratio test of this model vs unrestricted VAR $\chi^2 = 16.8828$ with p-value = 0.0506.

Likelihood ratio test of $\rho_{us} = \rho_{ger} = \rho_{uk} = \rho$ gives $\chi^2 = 2.2204$ with p-value = 0.3295.

Estimated model: $R_t = \alpha + \text{diag}(\rho) R_{t-1} + \epsilon_t$. Covariance is homoskedastic given by the one-regime equivalent of eqn (30).

Table A-7: Multicountry RS Models

Granger-causality model: RSG1
(Restrict α_i, ρ_i, σ_i, P, Q across countries)

Parameter	Estimate	Std error	pvalue
α_1	0.0110	0.0346	0.7500
α_2	0.4851	0.1470	0.0010
ρ_1	0.9956	0.0056	0.0000
ρ_2	0.9336	0.0187	0.0000
ζ_{11}	0.0022	0.0037	0.5498
ζ_{12}	0.0046	0.0131	0.7247
ζ_{21}	-0.0022	0.0048	0.6483
ζ_{22}	0.0455	0.0179	0.0112
γ_1^{ger}	-0.0105	0.0066	0.1103
γ_2^{ger}	0.0221	0.0330	0.5038
γ_1^{uk}	0.0210	0.0063	0.0009
γ_2^{uk}	0.0387	0.0629	0.5387
σ_1	-0.0483	0.0047	0.0000
σ_2	0.3210	0.0173	0.0000
P	3.7133	0.3602	0.0000
Q	1.8963	0.1541	0.0000

Estimated model: $r_t = \alpha(s_t) + A(s_t)r_{t-1} + \epsilon_t$ where $A = \begin{pmatrix} \rho(s_t^{us}) & 0 & 0 \\ \zeta_1(s_t^{ger}) & \rho(s_t^{ger}) & 0 \\ \zeta_2(s_t^{uk}) & 0 & \rho(s_t^{uk}) \end{pmatrix}$.
Covariance is homoskedastic given by eqn (30) with the same σ_i across countries.
Probabilities are in logit form, so the actual transition probability for $p(s_t = 1|s_{t-1} = 1) = \frac{e^P}{1+e^P}$ and $p(s_t = 2|s_{t-1} = 2) = \frac{e^Q}{1+e^Q}$.

Diagonal model: RSD1
(Restrict α_i, ρ_i, σ_i, P, Q across countries)

Parameter	Estimate	Std error	pvalue
α_1	0.0617	0.0250	0.0136
α_2	0.6255	0.2401	0.0092
ρ_1	0.9909	0.0031	0.0000
ρ_2	0.9349	0.0240	0.0000
γ_1^{ger}	0.0048	0.0179	0.7888
γ_2^{ger}	0.0307	0.0754	0.6842
γ_1^{uk}	0.0562	0.0274	0.0400
γ_2^{uk}	0.1450	0.1469	0.3236
σ_1	0.1655	0.0096	0.0000
σ_2	1.0990	0.0532	0.0000
P	3.9549	0.3599	0.0000
Q	1.8607	0.1480	0.0000

Model: Each country i has the process $r_t^i = \alpha(s_t) - \rho(s_t)r_{t-1}^i + \epsilon_t^i$ with the covariance given by eqn (30) with the same σ_i across countries.
Probabilities are in logit form, so the actual transition probability for $p(s_t = 1|s_{t-1} = 1) = \frac{e^P}{1+e^P}$ and $p(s_t = 2|s_{t-1} = 2) = \frac{e^Q}{1+e^Q}$.

Table A-8: Multicountry RS Models

Granger-causality model: RSG2
(Restrict α_i, ρ_i, P, Q across countries)

Parameter	Estimate	Std error	p-value
α_1	0.0026	0.0190	0.8923
α_2	0.4747	0.1508	0.0016
ρ_1	0.9967	0.0028	0.0000
ρ_2	0.9365	0.0171	0.0000
ζ_{11}	0.0021	0.0019	0.2830
ζ_{12}	0.0035	0.0129	0.7879
ζ_{21}	-0.0017	0.0024	0.4919
ζ_{22}	0.0414	0.0189	0.0288
γ_1^{ger}	-0.0154	0.0034	0.0000
γ_2^{ger}	0.0197	0.0266	0.4608
γ_1^{uk}	0.0252	0.0040	0.0000
γ_2^{uk}	0.0378	0.0428	0.3773
σ_{11}	0.0586	0.0034	0.0000
σ_{12}	0.3069	0.0185	0.0000
σ_{21}	0.0455	0.0025	0.0000
σ_{22}	0.2583	0.0193	0.0000
σ_{31}	0.0369	0.0026	0.0000
σ_{32}	0.3712	0.0257	0.0000
P	3.9463	0.7485	0.0000
Q	1.9252	0.2347	0.0000

Estimated model: $r_t = \alpha(s_t) + A(s_t)r_{t-1} + \epsilon_t$ where $A = \begin{pmatrix} \rho(s_t^{us}) & 0 & 0 \\ \zeta_1(s_t^{ger}) & \rho(s_t^{ger}) & 0 \\ \zeta_2(s_t^{uk}) & 0 & \rho(s_t^{uk}) \end{pmatrix}$.

Covariance is homoskedastic given by eqn (30).

Probabilities are in logit form, so the actual transition probability for $p(s_t = 1|s_{t-1} = 1) = \frac{e^P}{1+e^P}$ and $p(s_t = 2|s_{t-1} = 2) = \frac{e^Q}{1+e^Q}$.

Table A-9: Term Structure RS VAR Coefficients: constant probs

	US			GER			UK			Joint		
Parameter	Estimate	Std error	p value	Estimate	Std error	p value	Estimate	Std error	p value	Estimate	Std error	p value
μ_{11}	0.0222	0.1551	0.8861	-0.3454	0.0792	0.0000	0.0560	0.1220	0.6461	-0.0275	0.0650	0.6723
μ_{12}	0.4797	0.1948	0.0138	0.3875	0.0975	0.0001	0.4610	0.1840	0.0122	0.1842	0.0699	0.0084
μ_{21}	2.4747	1.1253	0.0279	0.6137	0.5420	0.2575	0.6305	0.7183	0.3800	0.8995	0.3293	0.0063
μ_{22}	-1.1338	1.1255	0.3137	0.1161	0.5125	0.8208	0.1578	0.5114	0.7577	-0.1902	0.2845	0.5038
$A_1[1,1]$	0.9993	0.0190	0.0000	1.0432	0.0109	0.0000	0.9846	0.0120	0.0000	0.9962	0.0077	0.0000
$A_1[1,2]$	0.0060	0.0332	0.8564	0.0715	0.0156	0.0000	-0.0291	0.0222	0.1904	0.0275	0.0145	0.0585
$A_1[2,1]$	-0.0547	0.0243	0.0240	-0.0548	0.0135	0.0000	-0.0349	0.0178	0.0501	-0.0176	0.0082	0.0321
$A_1[2,2]$	0.8708	0.0404	0.0000	0.9288	0.0189	0.0000	0.9534	0.0318	0.0000	0.9416	0.0156	0.0000
$A_2[1,1]$	0.7845	0.0932	0.0000	0.9291	0.0672	0.0000	0.9507	0.0642	0.0000	0.9182	0.0305	0.0000
$A_2[1,2]$	-0.2371	0.1155	0.0402	-0.0376	0.0998	0.7064	0.1423	0.1040	0.1713	-0.0314	0.0470	0.5049
$A_2[2,1]$	0.1210	0.0938	0.1970	-0.0174	0.0636	0.7843	-0.0221	0.0456	0.6282	0.0195	0.0266	0.4626
$A_2[2,2]$	0.9207	0.1168	0.0000	0.9097	0.0948	0.0000	0.8256	0.0720	0.0000	0.9014	0.0415	0.0000
$R_1[1,1]$	0.3397	0.0190	0.0000	0.1637	0.0116	0.0000	0.2159	0.0184	0.0000	0.2532	0.0139	0.0000
$R_1[1,2]$	-0.2150	0.0255	0.0000	-0.0878	0.0170	0.0000	-0.0429	0.0383	0.2618	-0.1301	0.0174	0.0000
$R_1[2,2]$	0.2918	0.0157	0.0000	0.1834	0.0113	0.0000	0.3362	0.0331	0.0000	0.2482	0.0113	0.0000
$R_2[1,1]$	1.2660	0.1247	0.0000	0.6940	0.0511	0.0000	1.1604	0.0841	0.0000	1.0665	0.0544	0.0000
$R_2[1,2]$	-1.1312	0.1446	0.0000	-0.5806	0.0538	0.0000	-0.5898	0.0658	0.0000	-0.7218	0.0518	0.0000
$R_2[2,2]$	0.6888	0.0662	0.0000	0.3256	0.0234	0.0000	0.5296	0.0421	0.0000	0.5977	0.0291	0.0000
P	0.9738	0.0123	0.0000	0.9421	0.0230	0.0000	0.6449	0.0550	0.0000	0.7761	0.0414	0.0000
Q	0.9000	0.0440	0.0000	0.9049	0.0378	0.0000	0.5382	0.0748	0.0406	0.8812	0.0205	0.0000
ρ_1	-0.5931	0.0496	0.0000	-0.4318	0.0709	0.0000	-0.1267	0.1151	0.0000	-0.4643	0.0500	0.0000
ρ_2	-0.8541	0.0359	0.0000	-0.8722	0.0240	0.0000	-0.7441	0.0450	0.0000	-0.7702	0.0248	0.0000
Eigen-	$\lvert\lambda_j\rvert$ of A_1= 0.9967, 0.8734			$\lvert\lambda_j\rvert$ of A_1= 0.9863, 0.9863			$\lvert\lambda_j\rvert$ of A_1= 1.0044, 0.9335			$\lvert\lambda_j\rvert$ of A_1= 0.9850, 0.9528		
values	$\lvert\lambda_j\rvert$ of A_2= 0.8666, 0.8666			$\lvert\lambda_j\rvert$ of A_2= 0.9468, 0.8920			$\lvert\lambda_j\rvert$ of A_2= 0.9158, 0.8605			$\lvert\lambda_j\rvert$ of A_2= 0.9101, 0.9101		

Bivariate RS VAR with constant probabilities of the short rate and spread. The joint coefficients are from a joint-estimation of the regime switching VAR with the same coefficients impoStd errord over all countries and treating each country as independent, following Bekaert, Hodrick and Marshall (1997b)
Model is $Y_t = A(s_t)Y_{t-1} + U_t, U_t \sim$ IID $N(0, \Sigma(s_t))$ with $Y_t = (r_t\, z_t)'$, the short rate and spread, and $R(s_t) = \text{chol}(\Sigma(s_t))$, $s_t = 1, 2$. We have constant probabilities $P = p(s_t = 1|s_{t-1} = 1)$ and $Q = p(s_t = 2|s_{t-1} = 2)$.
Wald Test for P+Q=1 for the UK yields $\chi^2 = 3.89$, pvalue = 0.0486.
$|\lambda_j|$ are the modulus of the eigenvalues of the companion matrix.
ρ_i is the correlation between short rates and spread conditional on regime i. Wald tests for equality across regimes all yield p-values of 0.0000 for all countries.

Table A-10: Term Structure RS VAR Coefficients: State dependent probs

	US			GER			UK			Joint		
Parameter	Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value
μ_{11}	0.0420	0.1584	0.7910	-0.3587	0.0827	0.0000	0.0460	0.1160	0.6918	-0.0112	0.0785	0.8861
μ_{12}	0.5119	0.1725	0.0030	0.4160	0.0947	0.0000	0.4215	0.1752	0.0161	0.1841	0.0773	0.0172
μ_{21}	2.4729	1.1555	0.0323	0.7477	0.4402	0.0894	0.7232	0.7321	0.3232	1.1460	0.4033	0.0045
μ_{22}	-1.0562	1.1574	0.3615	0.0071	0.3943	0.9856	0.1560	0.5182	0.7634	-0.3445	0.3371	0.3069
$A_1[1,1]$	0.9975	0.0196	0.0000	1.0424	0.0114	0.0000	0.9857	0.0115	0.0000	0.9940	0.0094	0.0000
$A_1[1,2]$	0.0008	0.0333	0.9819	0.0797	0.0162	0.0000	-0.0261	0.0214	0.2230	0.0264	0.0165	0.1107
$A_1[2,1]$	-0.0592	0.0217	0.0063	-0.0575	0.0131	0.0000	-0.0309	0.0174	0.0762	-0.0165	0.0093	0.0766
$A_1[2,2]$	0.8647	0.0350	0.0000	0.9211	0.0185	0.0000	0.9535	0.0319	0.0000	0.9385	0.0167	0.0000
$A_2[1,1]$	0.7845	0.0956	0.0000	0.9140	0.0552	0.0000	0.9411	0.0645	0.0000	0.8985	0.0360	0.0000
$A_2[1,2]$	-0.2383	0.1187	0.0447	-0.0614	0.0859	0.4746	0.1230	0.1056	0.2440	-0.0492	0.0521	0.3448
$A_2[2,1]$	0.1152	0.0963	0.2315	-0.0048	0.0496	0.9231	-0.0215	0.0456	0.6379	0.0313	0.0305	0.3049
$A_2[2,2]$	0.9198	0.1203	0.0000	0.9277	0.0789	0.0000	0.8388	0.0736	0.0000	0.9120	0.0451	0.0000
$R_1[1,1]$	0.3424	0.0197	0.0000	0.1785	0.0103	0.0000	0.2108	0.0168	0.0000	0.2651	0.0166	0.0000
$R_1[1,2]$	-0.2183	0.0262	0.0000	-0.0961	0.0157	0.0000	-0.0390	0.0358	0.2756	-0.1512	0.0196	0.0000
$R_1[2,2]$	0.2920	0.0156	0.0000	0.1819	0.0106	0.0000	0.3289	0.0288	0.0000	0.2520	0.0103	0.0000
$R_2[1,1]$	1.2796	0.1277	0.0000	0.6868	0.0490	0.0000	1.1562	0.0823	0.0000	1.1006	0.0570	0.0000
$R_2[1,2]$	-1.1432	0.1476	0.0000	-0.5813	0.0524	0.0000	-0.5923	0.0651	0.0000	-0.7406	0.0549	0.0000
$R_2[2,2]$	0.6949	0.0680	0.0000	0.3251	0.0230	0.0000	0.5334	0.0396	0.0000	0.6202	0.0322	0.0000
a_1	3.2219	0.5747	0.0000	6.5026	2.9188	0.0259	4.2030	1.5538	0.0068	7.0987	1.2995	0.0000
b_1				-0.3257	0.3793	0.3905	-0.3568	0.1506	0.0178	-0.6026	0.1344	0.0000
c_1	0.3978	0.4068	0.3282	-1.0273	0.5479	0.0608	-0.4671	0.2450	0.0566	-0.5943	0.2018	0.0032
a_2	3.3823	1.1917	0.0045	8.4219	4.5860	0.0663	-2.1590	1.8461	0.2422	-1.2325	1.1693	0.2919
b_2				-0.5629	0.5269	0.2854	0.1922	0.1644	0.2423	0.1998	0.1110	0.0718
c_2	-0.6385	0.3995	0.1100	-2.2325	0.8988	0.0130	0.7450	0.2799	0.0078	0.2500	0.1507	0.0971
ρ_1	-0.5988	0.0500	0.0000	-0.4672	0.0630	0.0000	-0.1178	0.1093	0.0000	-0.5145	0.0489	0.0000
ρ_2	-0.8545	0.0361	0.0000	-0.8728	0.0236	0.0000	-0.7431	0.0435	0.0000	-0.7667	0.0261	0.0000
Eigen-	$\lvert\lambda_j\rvert$ of A_1= 0.9972, 0.8651			$\lvert\lambda_j\rvert$ of A_1= 0.9823, 0.9823			$\lvert\lambda_j\rvert$ of A_1= 1.0022, 0.9369			$\lvert\lambda_j\rvert$ of A_1= 0.9846, 0.9479		
values	$\lvert\lambda_j\rvert$ of A_2= 0.8655, 0.8655			$\lvert\lambda_j\rvert$ of A_2= 0.9024, 0.9393			$\lvert\lambda_j\rvert$ of A_2= 0.8900, 0.8900			$\lvert\lambda_j\rvert$ of A_2= 0.9060, 0.9060		

Model is $Y_t = A(s_t)Y_{t-1} + U_t$, $U_t \sim$ IID $N(0, \Sigma(s_t))$ with $Y_t = (r_t\ z_t)'$, the short rate and spread, and $R(s_t) = \text{chol}(\Sigma(s_t))$. Time varying probabilities are $p(s_t = i|s_{t-1} = i) = \frac{\exp(a_i+b_i r_{t-1}+c_i z_{t-1})}{1+\exp(1+a_i+b_i r_{t-1}+c_i z_{t-1})} \quad i = 1, 2.$

The full US model failed to converge, so we report a restricted estimation with probabilities dependent only on the lagged spread. The joint system refers to a cross-Std errorctional estimation across all countries assuming the independence of countries following Bekaert, Hodrick and Marshall (1997b).

Likelihood ratio test for constant vs time varying probabilities gives $\chi^2 = 6.46$, p-value = 0.0396 for the US, $\chi^2 = 14.54$, p-value = 0.0058 for Germany and $\chi^2 = 20.51$ with p-value 0.0000 for the UK. For the joint system $\chi^2 = 55.56$ with p-value 0.0000.

$|\lambda_j|$ are the modulus of the eigenvalues of the companion matrix.

ρ_i is the correlation between short rates and spread conditional on regime i. Wald tests for equality across regimes all yield p-values of 0.0000 for all countries.