Nonlinear Adjustment in US Bond Yields: an Empirical Analysis with Conditional Heteroskedasticity

Lucchetti, Riccardo and Palomba, Giulio

Università Politecnica delle Marche - Department of Economics

2008

Nonlinear Adjustment in US Bond Yields: an Empirical Analysis with Conditional Heteroskedasticity*

Riccardo Lucchetti† Giulio Palomba‡

Abstract

Starting from the work by Campbell and Shiller (1987), empirical analysis of interest rates has been conducted in the framework of cointegration. However, parts of this approach have been questioned recently, as the adjustment mechanism may not follow a simple linear rule; another line of criticism points out that stationarity of the spreads is difficult to maintain empirically.

In this paper, we analyse data on US bond yields by means of an augmented VAR specification which approximates a generic nonlinear adjustment model. We argue that nonlinearity captures macro information via the shape of the yield curve and thus provides an alternative explanation for some findings recently appeared in the literature.

Moreover, we show how conditional heteroskedasticity can be taken into account via GARCH specifications for the conditional variance, either univariate and multivariate.

JEL Classification: C32, C51, E43
Keywords: interest rates, cointegration, nonlinear adjustment, conditional heteroskedasticity

1 Introduction

Interest rates have been the object of extensive research in the cointegration framework in the past 20 years, stemming from the seminal paper by Campbell and Shiller (1987). A fundamental consequence of the expectation hypothesis is that the most appropriate stochastic process to represent their

*We thank, without implicating them for errors, Luca Fanelli, Michele Fratianni, Eduardo Rossi and Alberto Zazzaro for their helpful comments and suggestions.

†Dipartimento di Economia, Università Politecnica delle Marche, Piazzale Martelli n. 8, 60121 Ancona (Italy), r.lucchetti@univpm.it.

‡Dipartimento di Economia, Università Politecnica delle Marche, Piazzale Martelli n. 8, 60121 Ancona (Italy), g.palomba@univpm.it.

time-series features is some sort of $I(1)$ process. At the same time, interest rate spreads should be stationary, possibly around a non-zero mean.

Of course, this translates into very precise hypotheses on the cointegration properties of interest rates, which should cointegrate in pairs, so the cointegration rank should be $n-1$ and the cointegration vectors should be of the form $[1, 0, \ldots, -1, 0, \ldots]$. Both ideas can be incorporated in a classic Vector ECM as:

$$\Gamma(L)\Delta y_t = \mu_t + \alpha\beta' y_{t-1} + \varepsilon_t, \tag{1}$$

where $\beta' y_{t-1}$ is a vector containing the $(n-1)$ lagged spreads.

However, the above model is not guaranteed to fit the data flawlessly; in some cases, the spreads may appear non-stationary and the hypothesis that the cointegration rank is $(n-1)$ may be rejected by conventional tests. Such findings could be interpreted as an outright rejection of the expectation hypothesis; on the other hand, there is the possibility that the empirical model may have to be refined.

Several authors have pointed out the shortcomings of a plain VECM model: on one hand, Ang and Piazzesi (2003) suggest that the shape of the yield curve can be influenced by macro factors and, as a consequence, the typical persistence shown by macro data may result in substantial autocorrelation in the spreads, to the point that there are even doubts on their stationarity (see Giese, 2006).

On the other hand, there is some evidence that the adjustment mechanism implicit in a cointegration model may follow a nonlinear dynamic in the case of bond yields. In most cases, this effect is modelled via a threshold model à la Balke and Fomby (1997). Hansen and Seo (2002) argue that adjustment follows two regimes, and is noticeable in one but not in the other. A similar argument is put forward in Krishnakumar and Neto (2005), where the authors argue that the adjustment is brought about by the monetary authority's interventions, and therefore occurs sporadically. A serious drawback of this class of models is that inference is rather complex, and the issues arising when modelling more than two series are quite difficult to handle.

An additional complication may arise because interest rates, like any other financial variable, show considerable changes in volatility if sampled at a monthly frequency or higher. This empirical regularity is widely acknowledged and has spurred the development of the gigantic literature on conditionally heteroskedastic processes, from Engle (1982) onwards. In this context, highly heteroskedastic innovations may have a dramatic impact on standard inferential procedures: estimator efficiency is an obvious issue, but there may also be robustness concerns.

In this article, we propose an empirical analysis that combines nonlinear effects in the conditional mean with conditional heteroskedasticity. The paper is structured as follows: section 2 describes our dataset and provides some preliminary evidence to motivate our preferred models, which are presented in section 3, while section 4 contains the estimates, their economic interpretation and an out-of-sample comparison of the forecasts obtained with our models with some of the alternatives. Section 5 concludes.

2 Integration and cointegration properties

We have used three weekly time series for US government bonds selected for different maturities: the variables in the model are the US Treasury constant maturities 3-month (*short*, r_t^s), the US Treasury constant maturities 2-year (*medium*, r_t^m) and the US Treasury constant maturities 10-year (*long*, r_t^l). The data source is DATASTREAM[1]. The sample period goes from 1982/10/08 through 2008/01/25 and includes 1321 observations for each series; time series plots are shown in Fig. 1. The two spreads sm_t and sl_t are defined as $(r_t^m - r_t^s)$ and $(r_t^l - r_t^s)$, respectively.

The choice of modelling weekly data basically depends on the fact that monthly frequencies would not allow us to capture the adjustments occurring during the period. On the other hand, using daily data may raise other concerns, due to the fact that information arrival is not uniform through time[2]. For these reasons we assume the week as the "natural" timeframe for adjustments.

In order to ensure that monetary policy rules are broadly consistent within the sample period, the sample period starts at 1982/10/07, when the FOMC announcement was made of the switch from M1 to a target rate as the main objective, as per Thornton (2005). Moreover, in order to evaluate the out-of-sample predicting properties of our model for a reasonable time span, we kept the last 52 observation out of the sample used for estimation. These choices yield a sample size of 1269 observations, that we deem adequate for our purpose.

As a preliminary step, we ran a battery of unit-root tests, reported in Table A-1 in the appendix: the classic ADF test[3], which is reported here for completeness, although it is known to suffer from power problems; the DF-GLS test by Elliott, Rothenberg, and Stock (1996) and the KPSS test by Kwiatkowski, Phillips, Schmidt, and Shin (1992).

The results appear to support the conventional view only to some extent: while it is safe to characterise the rates as nonstationary processes, the evidence for the spreads is more mixed. While for sm conventional tests seem to favour stationarity, this happens to a lesser extent for sl. However,

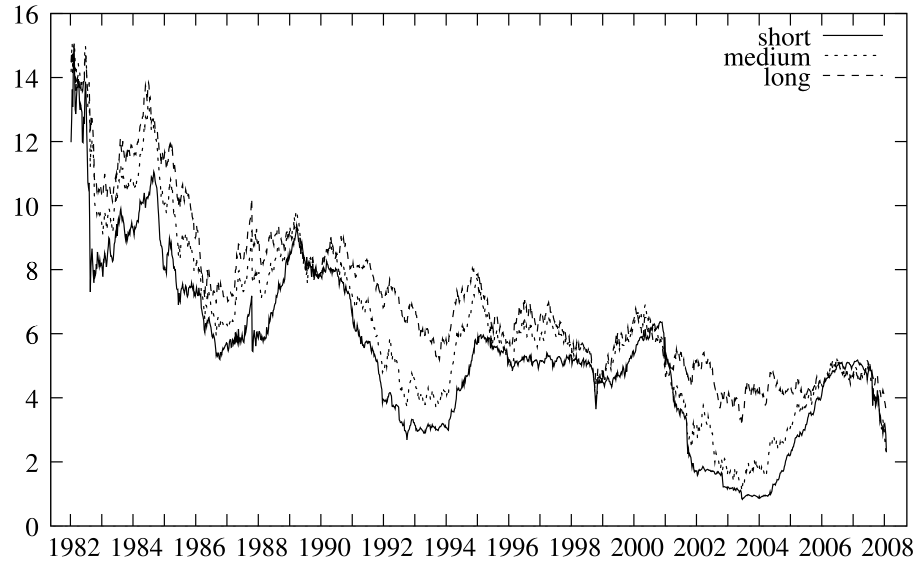
[1] The codes corresponding to the available series are FRTCM3M, FRTCM2Y and FRTCM10.

[2] It is well known (see Ghysels, Harvey, and Renault, 1996) that weekend effects, quote arrivals, dividend announcements or market closures can represent some examples for this evidence.

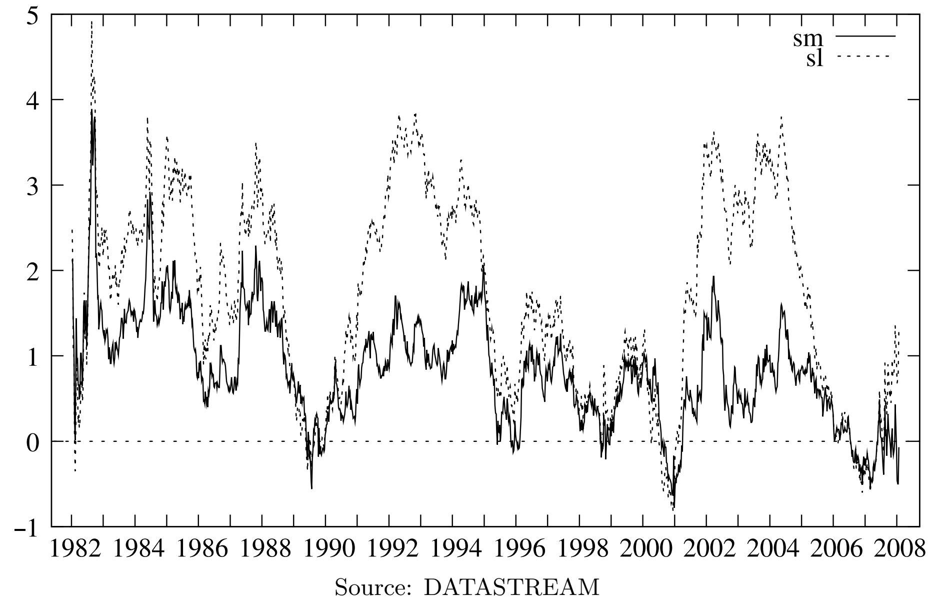
[3] Following Hall (1994), the number of lags for the ADF tests was chosen via a general-to-specific approach.

Figure 1: The data

(a) Interest rates



(b) Spreads with respect to r_t^s



Source: DATASTREAM

if we employ the KSS test (see Kapetanios, Shin, and Snell, 2003), which is specially tailored to have higher power against nonlinear alternatives, the null hypothesis of a unit root in sl is strongly rejected.

A similar picture comes from a joint analysis on the whole sample, undertaken by using the Johansen procedure (see Johansen, 1996). The results are shown in Table A-2 and can be briefly summarised by saying that the expected outcomes of the base model (1) can be found (with very wide confidence intervals) only if a cointegration rank of 2 is imposed; however, this hypothesis is rejected rather strongly by both the trace and λ-max tests.

This is consistent with the findings by Giese (2006), who argues that cointegration holds between spreads: under this hypothesis, in a trivariate system such as ours, the cointegration vector should be of the form $[1, -\beta, (\beta - 1)]$, where β is the cointegration parameter between spreads. Hence, the three elements of the cointegration vector should sum to 0. However, Table A-2 shows that this hypothesis was tested by means of an LR test and rejected.

All in all, there is no clear evidence for the feature that should most typically characterise cointegrated systems, that is mean reversion of the disequilibrium series which, in our case, are represented by the spreads. In other words, if we confine ourselves to linear models, there seems to be too little tendency in the interest rates to move in the direction needed to bring back the spreads to their long-run equilibrium value.

In the next section, we argue that part of the problem may come from an adjustment mechanism that cannot be adequately captured by the models we have used so far. However, there is also reason to believe that unmodelled features of the innovation process may be important. The most obvious candidate is clearly high persistence in variance, which is a typical feature of high-frequency financial data. There is some evidence (Ling, Li, and McAleer, 2003) that conditional heteroskedasticity may have a dramatic impact on conventional unit-root tests.

In our opinion, the model put forward in the following section offers a more convincing alternative.

3 A nonlinear adjustment model

The empirical framework outlined in the previous section delivers results which are not fully consistent with economic theory; clearly, this may be a spurious outcome of the base model being mis-specified. There are several ways in which the base model could be extended: for example, structural breaks or long-memory effects could be accounted for. In this section, we argue that the fundamental element to take into account is nonlinearity in the adjustment process.

A more general adjustment model can be represented as

$$\Phi(L)\Delta y_t = g(z_{t-1}) + \varepsilon_t,$$

where the function $g(\cdot)$ is a function for which there exists at least one value z^e such that $g(z^e) = 0$. The disequilibrium term z_t is a linear combination $\beta' y_t$, where β is the cointegration matrix, assumed known[4].

Special cases include the ordinary linear VECM, in which $g(z_t) = \mu + \alpha z_t$ and $z^e = E(z_t)$, or a threshold model, where the function $g(\cdot)$ is a piecewise linear function (a variation of a Band-TAR model, in Balke and Fomby's terminology):

$$g(z_t) = \begin{cases} \varphi_1(1-\rho_1) + \rho_1 z_t & \text{if} \quad z_t > \varphi_1 \\ 0 & \text{if} \quad \varphi_2 < z_t < \varphi_1 \\ \varphi_2(1-\rho_2) + \rho_2 z_t & \text{if} \quad z_t < \varphi_2, \end{cases} \tag{2}$$

Threshold models have received some attention for modelling interest rates. However, such models are unsuitable for the purpose of the present work for two reasons: first, the numerical issues in the computation of the estimates are far from obvious[5], mainly due to the necessity of applying numerical optimisation methods to a function which presents the discontinuities inherent in (2). Moreover, a threshold cointegration model is not straightforward to generalise to cases, such as ours, when the number of cointegrated processes is greater than two. The issues involved are explored in Lo and Zivot (2001). One serious problem with multivariate threshold model is the multidimensional generalisation of the "no-adjustment" region, which is a segment in the univariate case: clearly, there is no *a priori* reason for assuming that it should be a parallelogram or a circle or an ellipse, and any choice could only be justified on the grounds of analytical or computational convenience.

Another possibility is put forward in Escribano (2004): drawing on earlier work by Escribano and Navarro (2002), the author proposes to approximate the function $g(\cdot)$ by Padé polynomials in a model for money demand. This solution is remarkably elegant, but again is difficult to generalise to a multivariate setting.

As a consequence, we decided to work with a third-order Taylor expansion of the unknown adjustment function, which allows for the needed flexibility in the adjustment function while keeping the number of parameters reasonable.

[4]A considerable body of literature has been developed in which the general properties of nonlinear autoregressive models are analysed. The obligatory reference here is Meyn and Tweedie (1993) which is, however, much more general and much more technical than what would be needed here; an excellent recent paper which is closer to our present setting is Saikkonen (2005).

[5]See Hansen and Seo (2002) for an example. In order to overcome some of the difficulties, a Bayesian approach is advocated in Balcombe (2006).

We assume here that the cointegration rank is $(n-1)$ and the disequilibrium terms are given by the spread variables, which are assumed to possess finite moments of all orders. In this case, our approximate adjustment function can be written as

$$g(s_t) = \mu + \alpha' s_t + \theta'(s_t \otimes s_t)_+ + \lambda'(s_t \otimes s_t \otimes s_t)_+ \quad (3)$$

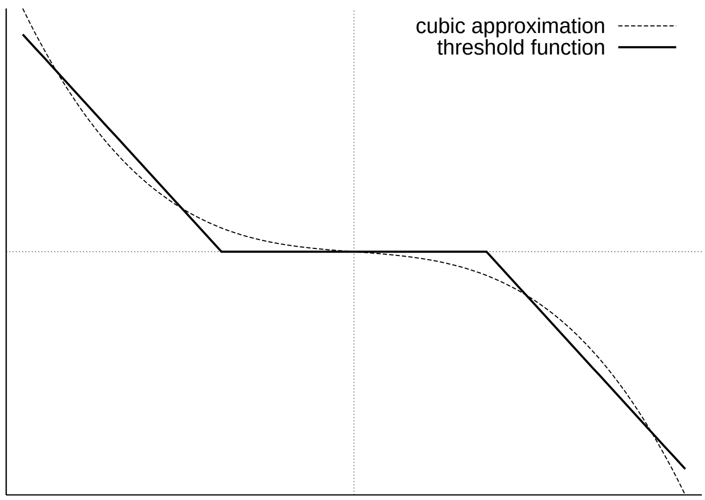
where the $()_+$ operator is understood to remove all duplicate elements[6].

For compactness, define

$$s_t = [\ sm_t \quad sl_t\ ]' \quad (4)$$
$$q_t = (s_t \otimes s_t)_+ = [\ sm_t^2 \quad sm_t \cdot sl_t \quad sl_t^2\ ]' \quad (5)$$
$$c_t = (s_t \otimes s_t \otimes s_t)_+ = [\ sm_t^3 \quad sm_t^2 \cdot sl_t \quad sm_t \cdot sl_t^2 \quad sl_t^3\ ]' \quad (6)$$

so s_t, q_t and c_t are, respectively, the first-, second- and third-order terms of the Taylor expansion.

Figure 2: Cubic approximation



The terms q_t and c_t do not have, *per se*, a direct economic interpretation: they only represent corrections to the linear term which are necessary to capture the threshold effect adequately. The main idea is that a cubic function

[6] For example, for the second-order term

$$(s_t \otimes s_t)_+ = D_n^+(s_t \otimes s_t)$$

where D_n^+ is the Moore-Penrose inverse of the duplication matrix, which, for $n = 2$, equals

$$D_n^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

See Magnus and Neudecker (1988).

can provide a good approximation to an adjustment function for the empirically relevant range of values, while remaining linear in the parameters, so estimation of equation (3) is relatively straightforward. Since it is reasonable to assume that the adjustment function should be non-increasing, it is obvious that a quadratic term alone would not suffice and a cubic term is needed. For example, the approximation to a piecewise linear function such as that of a threshold model is depicted in Figure 2.

It must be stressed that this should be considered a *local* approximation to the unknown adjustment function, whose main virtue is computational simplicity. We do not claim in any way that equation (3) is the actual Data Generating Process, but merely a local approximation to an unknown DGP, which we assume to satisfy stability conditions such as condition (iv) in Escribano (2004) (p. 80) or condition (ii) in Saikkonen (2005) (p. 72), even if the approximating cubic function does not.

4 Estimation results

The choice of the sample to use for building our empirical model is a crucial one. On one hand, the data in the sample should be as homogeneous as possible, since the nonlinear adjustment mechanism that we aim to quantify is a stylised representation of occasional events, most likely monetary policy interventions. As policy rules change, so does their representation, leading to structural breaks that inevitably jeopardise the entire statistical analysis.

In the following subsections, we propose three different alternatives to model the nonlinear adjustment, which take the form of an augmented version of a VAR model.

4.1 Single-equation models

A first set of models we propose which incorporate the above mechanism is a battery of univariate ECM models in which the quadratic and cubic terms are added to take the nonlinear adjustment effect into account.

We therefore estimate three equations, one for each rate, whose general form is

$$\Delta r_t^i = d_t + A_i(L)' \Delta R_{t-1} + \alpha_i' s_{t-1} + \theta_i' q_{t-1} + \lambda_i' c_{t-1} + \varepsilon_t \qquad (7)$$

where $i = s, m, l$, while R_t is a column vector containing the three rates and s_t, q_t and c_t are defined in equations (4)–(6). $A_i(L)$ is a row vector of lag polynomials and d_t is a vector of deterministic terms, which include a constant and dummy variables for the "Black Monday" of October 19^{th}, 1987 and the WTC attack of September 11^{th} 2001.

We first estimate the three equations (7) unrestricted, with the order of $A_i(L)$ equal to 6. We then proceed to restrict the base models via a general-to-specific approach so to obtain a parsimonious representation. The usual battery of diagnostic tests is then run against the restricted models. The results are shown in Table A-3.

The customary array of diagnostic checks show no sign of misspecification, except for a very substantial evidence of ARCH effects in each equation; this is hardly surprising, given the nature of the data. In order to take this fact into account and as a robustness check, we ran the same specification for the conditional mean with a GARCH(1,1) specification for the conditional variance. The estimate for the GARCH parameters are reported in Table A-4 in the appendix, together with the general-to-specific and nonlinearity tests.

Table 1: Single-equation ECM estimates — adjustment parameters

	Dep. variable: Δr_t^s			Dep. variable: Δr_t^m			Dep. variable: Δr_t^l		
Variable	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat
sm_{t-1}	0.107	0.044	2.405						
sl_{t-1}	-0.025	0.032	-0.777						
sm_{t-1}^2	-0.148	0.058	-2.563	-0.063	0.077	-0.820	-0.043	0.071	-0.604
sl_{t-1}^2	-0.011	0.022	-0.482	-0.013	0.023	-0.566	-0.009	0.022	-0.392
$sm \cdot sl_{t-1}$	0.063	0.056	1.141	0.076	0.080	0.951	0.060	0.077	0.773
sm_{t-1}^3	0.004	0.030	0.139	-0.118	0.046	-2.557	-0.092	0.041	2.271
sl_{t-1}^3	0.006	0.007	0.925	0.013	0.010	1.374	0.009	0.009	1.006
$sm_{t-1}^2 \cdot sl_{t-1}$	0.049	0.052	0.938	0.178	0.072	2.483	0.135	0.068	1.985
$sm_{t-1} \cdot sl_{t-1}^2$	-0.032	0.029	-1.098	-0.091	0.043	-2.140	-0.070	0.041	-1.684
Wald test for total nonlinearity:									
	Robust F = 1.984			Robust F = 2.176			Robust F = 3.054		
	p-value = 0.054			p-value = 0.033			p-value = 0.003		

Sample: 1982/10/08-2007/01/26 (1269 observations)
Newey-West HAC standard errors (window size = 8)

The estimates for the coefficients α, θ and λ are reported in Table 1, together with a robust Wald test for overall significance. The important point to note here is that the combined effect of the non-linear terms is significant at the 5% level in two of the three equations and in all three at 10%. Combined with the previous results, this indicates that a nonlinear adjustment mechanism is visible in the data and at the same time clears the model for the conditional mean of any misspecification.

4.2 A multivariate model

Instead of estimating the equations (7) as an array of univariate models, it may be preferable to use a full multivariate model, in which the adjustment mechanism is captured via the lagged spreads plus their squares and cubes. The deterministic part (constant and dummy variables) is the same as equation (7). The multivariate equivalent to equation (7) then becomes

$$\Phi(L)\Delta R_t = \gamma' d_t + \alpha_i' s_{t-1} + \theta_i' q_{t-1} + \lambda_i' c_{t-1} + \varepsilon_t \qquad (8)$$

In this case, however, the model to be used to take conditional heteroskedasticity into account becomes an issue. In order to close the model, a law of motion for the conditional covariance matrix Ω_t is needed. Several choices are available from the wide literature about multivariate GARCH models: the first attempt to model multivariate conditional covariances is the Vech Model introduced by Bollerslev, Engle, and Wooldridge (1988) together with its restricted formulation known as Diagonal GARCH. Other relevant contributions are Factor GARCH by Engle and Ng (1993) and the Dynamic Conditional Correlations (DCC model) by Engle (2002). More recently, models like O-GARCH (Alexander and Chibumba, 1996), GO-GARCH (Van der Weide, 2002) or the Generalized Orthogonal Factor GARCH (Lanne and Saikkonen, 2007), based on principal components have been suggested to solve the problem of estimation in presence of a great number of time series. A fairly comprehensive survey of the literature is provided in Laurent, Bauwens, and Rombouts (2006).

In this paper, we used a BEKK model (Engle and Kroner, 1995), so to achieve a reasonable level of generality. Hence, the conditional covariance matrix is assumed to be

$$\Omega_t = CC' + A\varepsilon_{t-1}\varepsilon'_{t-1}A' + B\Omega_{t-1}B' \tag{9}$$

In equation (9), C is a lower-triangular matrix, whose diagonal elements are constrained to be non-negative, while A and B are full-rank square matrices.

In our case, the BEKK model is probably the best tool to use because we are modelling a small number of series, as it combines high generality with relative parsimony and has the property that the conditional covariance matrices Ω_t are positive definite by construction under very mild conditions. As is well known, the BEKK model includes a relatively high number of parameters, which makes it unsuitable for large-scale models; this problem is mitigated by the use of the analytical score (see Lucchetti, 2002), which we also employ here. For larger models, it would be wiser to model the persistence in variance by a more parsimonious approach.

It should be noted that a dynamic conditional forecast of the covariance matrix can be useful for many purposes. In fact, if the forecast is a tool to be employed in some financial activity, such as portfolio allocation or risk hedging, the forecast of the conditional covariances is likely to be the main object of interest.

For lack of a better term, we will refer to this model as the NECH (Nonlinear Error Correction with Heteroskedasticity) model.

We chose to estimate our model with six lags following the indications provided by several information criteria[7] and LR tests; these criteria failed

[7]In particular we used AIC, BIC, HQC (Hannan and Quinn, 1979) and LWZ (Liu, Wu, and Zidek, 1997).

Table 2: Estimates for equation (8) - adjustment parameters

Equation for Δr_t^s				Equation for Δr_t^m				Equation for Δr_t^l			
Par.	Coeff.	S.E.	z-stat	Par.	Coeff.	S.E.	z-stat	Par.	Coeff.	S.E.	z-stat
$\alpha_{1,1}$	0.203	0.057	3.528	$\alpha_{2,1}$	0.032	0.054	0.601	$\alpha_{3,1}$	0.006	0.045	0.134
$\alpha_{1,2}$	-0.071	0.040	-1.797	$\alpha_{2,2}$	0.013	0.038	0.339	$\alpha_{3,2}$	0.001	0.032	0.034
$\theta_{1,1}$	-0.251	0.082	-3.071	$\theta_{2,1}$	-0.124	0.094	-1.325	$\theta_{3,1}$	-0.062	0.090	-0.693
$\theta_{1,2}$	0.014	0.028	0.497	$\theta_{2,2}$	-0.019	0.033	-0.573	$\theta_{3,2}$	-0.001	0.031	-0.033
$\theta_{1,3}$	0.074	0.068	1.090	$\theta_{2,3}$	0.081	0.092	0.886	$\theta_{3,3}$	0.042	0.093	0.452
$\lambda_{1,1}$	0.006	0.026	0.220	$\lambda_{2,1}$	-0.136	0.038	-3.603	$\lambda_{3,1}$	-0.111	0.040	-2.789
$\lambda_{1,2}$	0.002	0.007	0.224	$\lambda_{2,2}$	0.013	0.010	1.270	$\lambda_{3,2}$	0.008	0.010	0.729
$\lambda_{1,3}$	0.070	0.050	1.394	$\lambda_{2,3}$	0.207	0.070	2.946	$\lambda_{3,3}$	0.165	0.070	2.358
$\lambda_{1,4}$	-0.037	0.031	-1.197	$\lambda_{2,4}$	-0.095	0.044	-2.144	$\lambda_{3,4}$	-0.074	0.044	-1.664

to yield a clear-cut indication of the optimum number of lags. We therefore decided to sacrifice parsimony in exchange for robustness; our results, however, do not change qualitatively with other choices.

Table 3: NECH model - Wald tests for the adjustment parameters

series	Total	Linear	Nonlinear
	$\alpha_{1,\cdot} = 0,\ \theta_{1,\cdot} = 0,\ \lambda_{1,\cdot} = 0$	$\alpha_{1,\cdot} = 0$	$\theta_{1,\cdot} = 0,\ \lambda_{1,\cdot} = 0$
Δr_t^s	27.0182	14.3615	11.3839
	(0.0014)	(0.0008)	(0.1227)
	$\alpha_{2,\cdot} = 0,\ \theta_{2,\cdot} = 0,\ \lambda_{2,\cdot} = 0$	$\alpha_{2,\cdot} = 0$	$\theta_{2,\cdot} = 0,\ \lambda_{2,\cdot} = 0$
Δr_t^m	21.4705	8.8296	16.9247
	(0.0107)	(0.0121)	(0.0179)
	$\alpha_{3,\cdot} = 0,\ \theta_{3,\cdot} = 0,\ \lambda_{3,\cdot} = 0$	$\alpha_{3,\cdot} = 0$	$\theta_{3,\cdot} = 0,\ \lambda_{3,\cdot} = 0$
Δr_t^l	15.7207	2.1794	15.4139
	(0.0729)	(0.3363)	(0.0310)
Joint test for total nonlinearity:			
$(\theta = 0,\ \lambda = 0)$:			
	W=33.5727	p-val=0.0402	
Test for the same nonlinearity between *medium* and *long*:			
$(\theta_{2,\cdot} = \theta_{3,\cdot}$ and $\lambda_{2,\cdot} = \lambda_{3,\cdot})$			
	W=5.6026	p-val = 0.5868	

The estimates for the parameters of the linear terms of the conditional mean and the conditional variance have little interest for interpretation; therefore, they are reported in the appendix as Tables A-5 and A-6. Suffice it to say that the estimates for the A and B matrices leave no room for considering the innovations homoskedastic. Moreover, the hypothesis of diagonality of the matrices A and B is strongly rejected, which confirms that a full multivariate model is arguably preferable.

The adjustment parameters α, θ and λ are displayed in Table 2, while Table 3 reports a battery of robust Wald tests for several hypotheses of interest[8].

As can be seen by comparing tables 1 and 2, the estimates are very similar in sign and in magnitude. Again, despite the fact that several coefficients are not individually significant, joint tests indicate that the adjustment effect is detectable for all rates and the nonlinear effect is highly significant when we consider all the equations jointly. The hypothesis of no nonlinear adjustment is only accepted for r_t^s. These tests together indicate that some error correction operates, but it cannot be described by a simple linear rule, as would be the case had we estimated an ordinary VECM model.

Moreover, the estimates for the parameter vectors θ and λ in the r_t^m and r_t^l equations are very similar in sign and magnitude. In fact, this prompts the intriguing hypothesis that the nonlinear effects may in fact be the same for the two rates. Accepting this idea would lead one to think that if nonlinearity effectively captures monetary policy interventions, then these operate in a parallel fashion on the long-term end of the yield curve; in other words, interventions operate on the 2-year and the 10-year rates by altering their level, but not the spread between them. This hypothesis was tested via a robust Wald test: the test statistic for the hypothesis of a common adjustment for the two longer-term bond is 5.0626, with a p-value of 0.5868, which leads to accepting this hypothesis too. Further investigation on this kind of parallelism seems very promising: the possibility of generalising the idea to a wider span of the yield curve looks particularly interesting and will be analysed in future work.

Since nonlinearity seems to exert a lesser effect on the short rate, a restricted model was also estimated, combining the restriction on the short rate equation $\theta_1 = 0, \lambda_1 = 0$ with the cross-equation restrictions $\theta_2 = \theta_3, \lambda_2 = \lambda_3$. However, the restriction was rejected quite strongly (LR = 67.007, with p-value 6.69E-09).

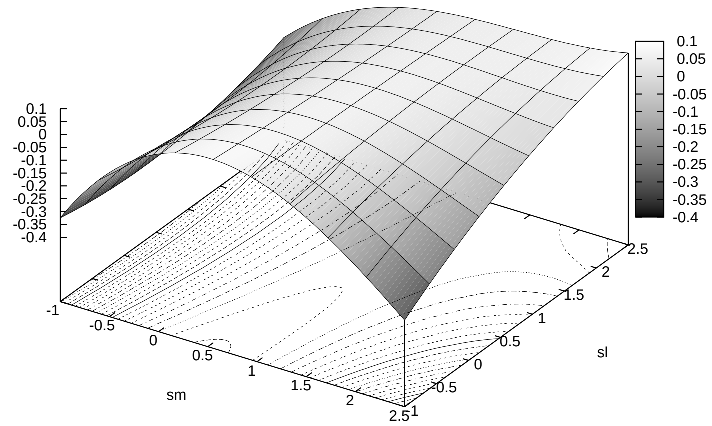
4.3 An interpretation of the adjustment process

In order to visualise how changes in rates are driven by the spreads, Figure 3 displays a plot of the adjustment function of the three rates in response to the two lagged spreads sm_{t-1} and sl_{t-1}. The nonlinearity is evident: if the adjustment process had followed the standard linear VECM mechanism, the contour lines of the adjustment surface would have been parallel, equidistant lines. In contrast, notice that there are ample regions of the $\{sm_t, sl_t\}$ plane where the adjustment mechanism is idle, whereas outside those regions nonlinearity operates and adjustment occurs much more effectively. An alternative way to express this result is to consider the attractor of the long-run relationships not as a single line, but rather as a region of the three-dimensional space.

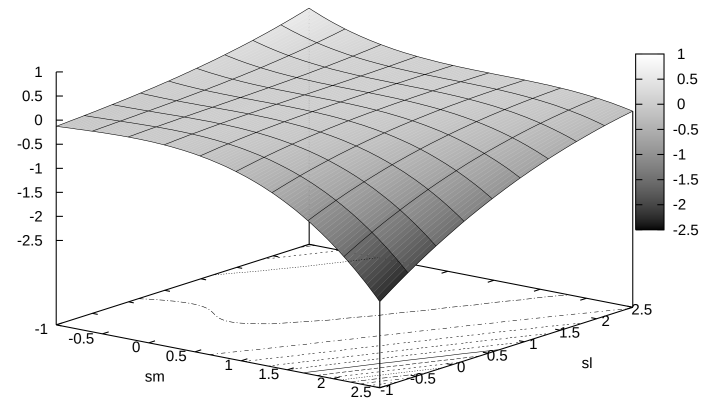
[8] All standard errors and test statistics are computed by using the Bollerslev and Wooldridge (1992) robust variance estimator.

Figure 3: Adjustment surfaces

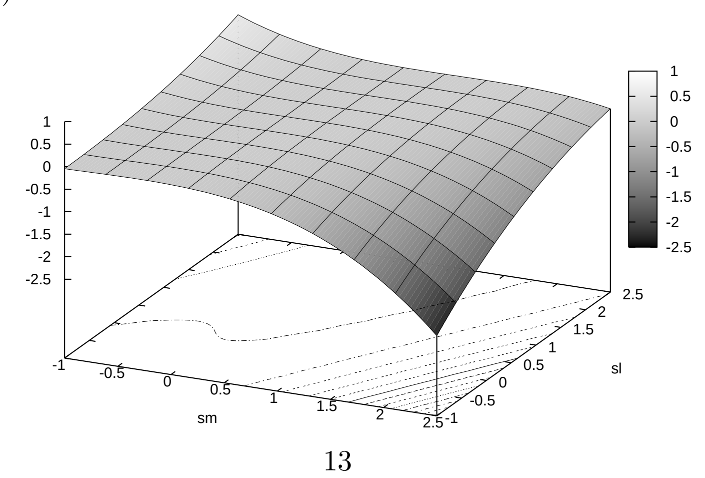
(a) short (r_t^s)



(b) medium (r_t^m)

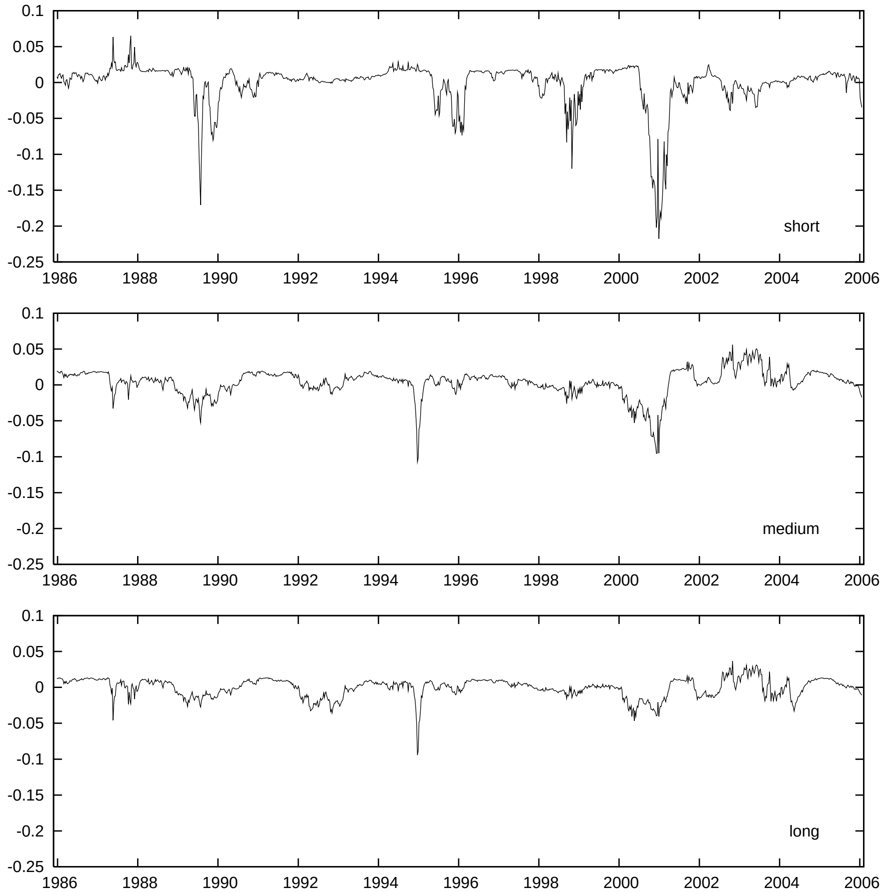


(c) long (r_t^l)



In other words, the response of bonds yields to movements in the short-term rate depends on the shape of the yield curve. If slope and curvature remain "standard", then little adjustment occurs, if any at all. On the contrary, when the curve shape becomes "unusual" adjustment is triggered. This is the effect captured by the nonlinear part of our model, which operates in a way comparable to the threshold models cited above.

Figure 4: Historical adjustments



It is also interesting to evaluate graphically when the error-correction part of the model played a significant role in driving the dependent variables. Figure 4 reports the combined estimated effects on the bond rates of the adjustment parameters reported in Table 2.

Most of the spikes in the real data coincide with relevant shocks to the bond market and consequent policy interventions. A few examples are given by the Brady Plan implementation in July 1989 (see Ünal, Demirgüç-Kunt, and Leung (1993) for a detailed account of the timeline), the Mexican Peso crisis of December 1994 and the "dot-com" bubble burst that occurred in early 2000.

4.4 Out-of-sample performance

In order to assess the predictive ability of our estimated model, 52 observations were kept out of sample. This type of check is particularly important because nonlinear models often overfit the data and show poor out-of-sample performance. Hence, the aim of this section is to perform an overall robustness check, rather than to advocate our models as forecasting tools.[9]

The one-step-ahead forecasts for our univariate specifications (ECM for the homeoskedastic variant and GARCH for the conditionally heteroskedastic one) plus the multivariate model (NECH) were computed. For comparison purposes, we did the same with the following array of competing models:

RW The random walk model. In this model, the forecast error is simply the first difference of the series. This is equivalent to an ARIMA(0,1,0) model.

LVAR A vector autoregressive model of order 6 on the rates in levels. The order of the model was chosen by considering the BIC and the HQC information criteria.

C2 Same as above, in VECM form with cointegration rank set to 2. No restrictions are imposed on the cointegration vectors.

C1 Same as above, with cointegration rank set to 1.

DVAR Same as above, with cointegration rank set to 0, that is, an unrestricted VAR in differences.

BEKK Same as above, with a BEKK specification for the conditional covariance. This model is included because it is customary in applied financial analysis to set up multivariate conditional heteroskedasticity models on the log-returns, discarding the information supplied by the series in levels.

All the above models were estimated over the same sample as in section 4 and both statistics were calculated up to the 25^{th} of January, 2008. In all the estimated models the dummy variables for the "Black Monday" and the WTC attack are used.

The RMSE and the Diebold and Mariano (1995) test (DM) were used to evaluate their predictive ability; for the DM test, we used a V-shaped loss

[9] As (Clements, 2005, pp. 39–40) argues, "... [I]t is often argued that non-linear models will be better in some states than others[...] If those occasions which favour the non-linear model are relatively infrequent, then the good performance at those times may be diluted by averaging squared forecast errors over all periods".

function, while the long-run variance was computed by using pre-whitening as in Andrews and Monahan (1992) and a Bartlett window size of 3.

Tables 4 and 5 summarise the results; the null hypothesis of the DM test is that of equal predictive accuracy between our proposed models (NECH, GARCH, ECM) and all the alternative models.

Table 4: Out-of-sample RMSE

Model	r_t^s	r_t^m	r_t^l
RW	0.2066	0.1502	0.1077
LVAR	0.4019	0.3014	0.2162
C2	0.4015	0.3008	0.2151
C1	1.1727	0.3006	0.5929
DVAR	0.2116	0.1522	0.1054
BEKK	0.2092	0.1606	0.1147
ECM	0.1951	0.1505	0.1075
GARCH	0.1947	0.1530	0.1098
NECH	0.2046	0.1493	0.1093

In terms of RMSE, our proposed models are in most cases superior to all the competing models; in the worst cases, they are comparable with the best ones. Moreover, it is noteworthy that among the competing models the random walk model is consistently among the best, if not the best overall[10]. For the short-term rate r_t^s, the models containing a nonlinear part exhibit the lowest RMSE of all; for the other two rates, the results are less strong, but similar. The DM tests carry a similar message: apart from one case (the RW for the short rate), the nonlinear models always outperform their linear alternatives when the test rejects the null. It should be noted, however, that even the superiority displayed in this case by the RW for the short rate is contrasted by a higher value of the RMSE.

It should also be noted that our out-of-sample period includes the subprime mortgage crisis, which could potentially have disruptive effects on forecasts if our models suffered from in-sample overfitting.

The overall message from the out-of-sample forecasting exercise is that no evidence of misspecification and overfitting by the nonlinear models is visible; on the contrary, nonlinearity seem to help forecasts, albeit marginally.

[10]The fact that no model clearly outperforms the random walk model is not surprising. It is a well-established fact that, on the very short run, financial variables are essentially impossible to predict. The considerations in Kilian and Taylor (2003) on the difficulties of beating the random walk model when important nonlinearities are present in the data generating process also apply here.

Table 5: Diebold and Mariano (1995) test versus the nonlinear adjustment models

Model	r_t^s		r_t^m		r_t^l	
			versus NECH			
RW	**-3.3228**	***	**1.7928**	*	**2.3043**	**
LVAR	**3.6649**	***	**2.1834**	**	**2.2304**	**
C2	**3.6332**	***	**2.2158**	**	**2.2767**	**
C1	1.3252		0.1118		0.2603	
DVAR	-0.1026		1.0800		1.2505	
BEKK	0.1933		0.9623		0.2712	
			versus GARCH			
RW	**-4.0512**	***	1.6007		**2.4417**	**
LVAR	**3.0511**	***	**2.1486**	**	**2.3417**	**
C2	**3.0341**	***	**2.1791**	**	**2.3920**	**
C1	1.0576		0.2537		0.2010	
DVAR	0.5284		0.6207		0.6930	
BEKK	0.7288		0.1671		-0.8878	
			versus ECM			
RW	**-4.0837**	***	1.6245		**2.4991**	**
LVAR	**3.1638**	***	**2.1671**	**	**2.3520**	**
C2	**3.1443**	***	**2.1974**	**	**2.4019**	**
C1	1.1294		0.2734		0.1976	
DVAR	0.9298		0.8118		**1.8468**	*
BEKK	0.9566		0.3646		-0.0095	

Two-tailed tests: positive test values indicate superior forecasting accuracy for the nonlinear model; negative test values indicate superior forecasting accuracy for the linear model.

5 Summary and conclusions

The main message of our paper is clear: nonlinear adjustment is an important empirical feature in US bond rates. From an economic point of view, this is an interesting result because it suggests that the notion of long-run equilibrium should be broadened to include the concept that equilibrium could be attained in a region of the state space, rather than a single point (or a collection of isolated points). From a statistical point of view, failure to include nonlinearity into an empirical model leads to mis-specification and may hamper predictive ability.

Our models for US bonds approximate a nonlinear adjustment mechanism via a simple variable addition to an otherwise ordinary VAR model. Moreover, incorporating conditional heteroskedasticity can be done via standard methods. Hence, they are much less complex to estimate, from a computational point of view, than multivariate threshold models and can also be used when the number of time series is greater than two.

In our empirical application, we provide a description of the data that reconciles the findings of different strands of applied work in this area. The most prominent features of our dataset are encompassed, identifying nonlinear adjustment effects in particular periods such as the Brady Plan introduction, the Mexican Peso crisis, the Russian crisis or the “dot-com” bubble burst. In addition, our out-of-sample analysis shows that the 2007 subprime mortgage crisis is handled satisfactorily.

Finally, the description of policy shocks transmission to long-term bonds that our model offers prompts some interesting considerations: out of the three interest rates considered, the short-term rate is the one displaying the least compelling evidence for nonlinearity and for which the improvements in forecasting power are least obvious. This may suggest that the adjustment mechanism we have studied is particularly important for longer-term bond rates. This aspect may be investigated in more detail by considering a wider spectrum of maturities. This point will be the object of future research.

References

Alexander, C. and A. Chibumba (1996): “Multivariate orthogonal factor GARCH,” *Discussion Paper in mathematics, University of Sussex*.

Andrews, D. W. K. and J. C. Monahan (1992): “An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator,” *Econometrica*, 60, 953–966.

Ang, A. and M. Piazzesi (2003): “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables,” *Journal of Monetary Economics*, 50.

Balcombe, K. G. (2006): "Bayesian estimation of cointegrating thresholds in the term structure of interest rates," *Empirical Economics*, 31, 277–289.

Balke, N. and T. Fomby (1997): "Threshold cointegration," *International Economic Review*, 38, 627–645.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): "A capital asset pricing model with time varying covariances," *Journal of Political Economy*, 96, 116–131.

Bollerslev, T. and J. M. Wooldridge (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances," *Economic Review*, 11, 143–172.

Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062–1088.

Clements, M. P. (2005): *Evaluating Econometric Forecasts Of Economic And Financial Variables*, Palgrave Macmillan.

Diebold, F. X. and R. M. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Elliott, G., T. Rothenberg, and J. Stock (1996): "Efficient tests for an autoregressive unit root," *Econometrica*, 64, 813–836.

Engle, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the U.K. inflation," *Econometrica*, 50, 987–1008.

Engle, R. F. (2002): "Dynamic conditional correlation - A simple class of multivariate GARCH models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F. and K. F. Kroner (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150.

Engle, R. F. and V. K. Ng (1993): "Measuring and testing the the impact of news on volatility," *Journal of Finance*, 48, 1749–1778.

Escribano, A. (2004): "Nonlinear Error Correction: the Case of Money Demand in the United Kingdom (1878–2000)," *Macroeconomic Dynamics*, 8, 76–116.

Escribano, A. and S. M. Navarro (2002): "Nonlinear Error Correction Models," *Journal of Time Series Analysis*, 23, 509–522.

Ghysels, E., A. C. Harvey, and E. Renault (1996): *Stochastic volatility*, in G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, vol. 14, Statistical Methods in Finance, Elsevier, North Holland.

Giese, J. (2006): "Level, slope, curvature: The yield curve's derivatives and their relations to macro variables," Paper presented at the European Economic Association and the Econometric Society European Meeting (EEA-ESEM), Vienna, August.

Hall, A. (1994): "Testing for a unit root in time series with pretest data-based model selection," *Journal of Business and Economic Statistics*, 12, 461–469.

Hannan, E. J. and B. G. Quinn (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society (B)*, 41, 190–195.

Hansen, B. and B. Seo (2002): "Testing for two-regime threshold cointegration in vector error correction models," *Journal of Econometrics*, 110, 293–318.

Johansen, S. (1996): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Kapetanios, G., Y. Shin, and A. Snell (2003): "Testing for a unit root in the nonlinear STAR framework," *Journal of Econometrics*, 112, 359–379.

Kilian, L. and M. P. Taylor (2003): "Why is it so difficult to beat the random walk forecast of exchange rates?" *Journal of International Economics*, 60, 85–107.

Krishnakumar, J. and D. Neto (2005): "Testing unit root in threshold cointegration," *Cahiers du Département d'Économétrie, Faculté des sciences économiques et Sociales, Université de Genève*, 4.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54, 159–178.

Lanne, M. and P. Saikkonen (2007): "A multivariate generalized orthogonal factor GARCH model," *Journal of Business and Economic Statistic*, 25, 61–75.

Laurent, S., L. Bauwens, and J. V. K. Rombouts (2006): "Multivariate GARCH models: a survey," *Journal of Applied Econometrics*, 21, 79–109.

Ling, S., W. Li, and M. McAleer (2003): "Estimation and testing for unit root processes with GARCH (1, 1) errors: Theory and monte carlo evidence," *Econometric Reviews*, 22, 179–202.

Liu, J., S. Wu, and J. V. Zidek (1997): "On segmented multivariate regressions," *Statistica Sinica*, 7, 497–525.

Lo, M. C. and E. Zivot (2001): "Threshold cointegration and nonlinear adjustment to the law of one price," *Macroeconomic Dynamics*, 5, 533–76.

Lucchetti, R. (2002): "Analytical score for multivariate GARCH models," *Computational Economics*, 19, 133–143.

MacKinnon, J. G. (1996): "Numerical distribution functions for unit root and cointegration tests," *Journal of Applied Econometrics*, 11, 601–618.

Magnus, J. R. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Meyn, S. P. and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability*, Springer.

Saikkonen, P. (2005): "Stability results for nonlinear error correction models," *Journal of Econometrics*, 127, 69–81.

Thornton, D. L. (2005): "When did the FOMC begin targeting the federal funds rate? what the verbatim transcripts tell us," Working Papers 2004–015, Federal Reserve Bank of St. Louis.

Ünal, H., A. Demirgüç-Kunt, and K.-W. Leung (1993): "The Brady Plan, 1989 Mexican debt-reduction agreement, and bank stock returns in United States and Japan," *Journal of Money, Credit and Banking*, 25, 410–29.

Van der Weide, R. (2002): "GO-GARCH: a multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17, 549–564.

Appendix

Table A-1: Stationarity tests for rates and spreads

Sample: 1982/10/08 - 2007/01/26 (1269 observations)								
series	ADF	DFGLS		KPSS		KSS	
r_t^s	-1.4800	-0.1638		2.8847	***		
r_t^m	-2.0591	0.0317		3.3792	***		
r_t^l	-1.8446	0.4898		4.0010	***		
sm_t	-2.1617	-1.8720	*	0.1518		-2.1564	
sl_t	-2.3812	-1.6406	*	0.4103	*	-2.9957	**

Stars indicate rejection of the null hypothesis at the 10%, 5% and 1% level (the KPSS test has stationarity as its null hypothesis). The p-values for the ADF tests were computed via the algorithm by MacKinnon (1996). The critical values for the DF-GLS test are from Elliott et al. (1996). The KPSS test is carried out with a window size of 26; critical values are from Kwiatkowski et al. (1992). The KSS test was carried out with a contant and no trend; lag selection via Hannan and Quinn's (1979) criterion. For the critical values see Kapetanios et al. (2003).

Table A-2: Johansen test and cointegrating vectors

Time series

Sample: 1982/10/08-2007/01/26 (1269 observations)

Restricted constant - Selected number of lags: 2

Rank	Eigenvalue	Trace test	p-value	λ-max test	p-value
0	0.0629	89.6430	0.0000	82.4690	0.0000
1	0.0036	7.1737	0.8803	4.5282	0.9139
2	0.0021	2.6456	0.6540	2.6456	0.6529
1 cointegrating vector:		r_{t-1}^s	r_{t-1}^m	r_{t-1}^l	const
	β'	1.0000	-1.4501	0.59036	-0.57064
		(0.0000)	(0.0562)	(0.0615)	(0.1539)

LR Test for $\beta_1 + \beta_2 + \beta_3 = 0$: 28.0490 , p-value = 1.18e-07

Optimal number of lags is selected via Hannan and Quinn (1979) information criterion.

Table A-3: Single-equation ECM estimates

	Dep. variable: Δr_t^s			Dep. variable: Δr_t^m			Dep. variable: Δr_t^l		
Variable	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat
const	-0.011	0.008	-1.328	-0.006	0.008	-0.715	-0.006	0.008	-0.784
wcc_t	-1.752	0.020	-87.678	-1.485	0.030	-50.169	-1.217	0.023	-52.617
wtc_t	-0.595	0.014	-42.510	-0.657	0.015	-42.861	-0.228	0.014	-16.753
wtc_{t-1}	-0.358	0.038	-9.444	0.033	0.038	0.853	0.121	0.035	3.409
Δr_{t-1}^s	-0.145	0.069	-2.109	-0.044	0.057	-0.772	-0.121	0.046	-2.642
Δr_{t-2}^s	0.027	0.052	0.513	0.128	0.055	2.340			
Δr_{t-3}^s	0.020	0.040	0.488	0.102	0.042	2.404			
Δr_{t-4}^s	0.092	0.035	2.636	0.067	0.034	1.998			
Δr_{t-5}^s	0.070	0.032	2.170						
Δr_{t-1}^m	0.166	0.048	3.460	0.032	0.066	0.479	0.136	0.062	2.172
Δr_{t-2}^m				-0.120	0.072	-1.659			
Δr_{t-1}^l	-0.088	0.049	-1.805	0.006	0.061	0.092	-0.073	0.066	-1.109
Δr_{t-2}^l				0.138	0.071	1.959			
sm_{t-1}	0.107	0.044	2.405						
sl_{t-1}	-0.025	0.032	-0.777						
sm_{t-1}^2	-0.148	0.058	-2.563	-0.063	0.077	-0.820	-0.043	0.071	-0.604
sl_{t-1}^2	-0.011	0.022	-0.482	-0.013	0.023	-0.566	-0.009	0.022	-0.392
$sm \cdot sl_{t-1}$	0.063	0.056	1.141	0.076	0.080	0.951	0.060	0.077	0.773
sm_{t-1}^3	0.004	0.030	0.139	-0.118	0.046	-2.557	-0.092	0.041	2.271
sl_{t-1}^3	0.006	0.007	0.925	0.013	0.010	1.374	0.009	0.009	1.006
$sm_{t-1}^2 \cdot sl_{t-1}$	0.049	0.052	0.938	0.178	0.072	2.483	0.135	0.068	1.985
$sm_{t-1} \cdot sl_{t-1}^2$	-0.032	0.029	-1.098	-0.091	0.043	-2.140	-0.070	0.041	-1.684
General-to-specific test:									
	F(11, 1238) = 1.493			F(12, 1238) = 1.326			F(17, 1238) = 0.931		
	p-value = 0.128			p-value = 0.197			p-value = 0.537		
RESET test with squares and cubes:									
	F(2. 1248) = 1.407			F(2. 1249) = 1.548			F(2. 1254) = 1.281		
	p-value = 0.245			p-value = 0.213			p-value = 0.278		
Breusch-Godfrey test for first-order autocorrelation:									
	LMF stat = 1.102			LMF stat = 3.315			LMF = 0.983		
	p-value = 0.294			p-value = 0.069			p-value = 0.322		
	TR^2 = 1.120			TR^2 = 3.359			TR^2 = 0.994		
	p-value = 0.290			p-value = 0.067			p-value = 0.319		
ARCH(2) test:									
	LM stat = 163.362			LM stat = 15.681			LM stat = 18.290		
	p-value = 0.000			p-value = 0.000			p-value = 0.000		
CUSUM test for stability of parameters (Harvey-Collier):									
	t(1248) = 1.807			t(1250) = -0.471			t(1255) = 0.764		
	p-value = 0.071			p-value = 0.638			p-value = 0.445		
Wald test for total nonlinearity:									
	Robust F = 1.984			Robust F = 2.176			Robust F = 3.054		
	p-value = 0.054			p-value = 0.033			p-value = 0.003		

Sample: 1982/10/08-2007/01/26 (1269 observations)
Newey-West HAC standard errors (window size = 8)

Table A-4: Univariate GARCH models

	Dep. variable: Δr_t^s			Dep. variable: Δr_t^m			Dep. variable: Δr_t^l		
Variable	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat	Coeff.	S.E.	z-Stat
const	0.000	0.000	1.433	0.001	0.001	1.759	0.000	0.001	0.416
ARCH	0.142	0.057	2.486	0.088	0.031	2.817	0.044	0.048	0.920
GARCH	0.858	0.050	17.097	0.859	0.055	15.703	0.943	0.077	12.179
General-to-specific test:									
	W = 12.763			W = 9.149			W = 13.334		
	p-value = 0.309			p-value = 0.690			p-value = 0.714		
Wald test for total nonlinearity:									
	W = 13.860			W = 24.589			W = 18.585		
	p-value = 0.054			p-value = 0.001			p-value = 0.010		

QML standard errors; complete estimates are available upon request.

Table A-5: Estimates for equation (8) - short-run parameters

	Equation for Δr_t^s				Equation for Δr_t^m				Equation for Δr_t^l		
Par.	Coeff.	S.E.	z-stat	Par.	Coeff.	S.E.	z-stat	Par.	Coeff.	S.E.	z-stat
μ_1	-0.037	0.014	-2.725	μ_2	-0.019	0.011	-1.719	μ_3	-0.012	0.010	-1.208
$\gamma_{1,1}$	-1.829	0.027	-68.560	$\gamma_{2,1}$	-1.516	0.030	-50.400	$\gamma_{3,1}$	-1.184	0.029	-40.380
$\gamma_{1,2}$	-0.570	0.015	-38.980	$\gamma_{2,2}$	-0.667	0.017	-39.370	$\gamma_{3,2}$	-0.222	0.016	-13.690
$\gamma_{1,3}$	-0.420	0.236	-1.783	$\gamma_{2,3}$	-0.056	0.125	-0.452	$\gamma_{3,3}$	0.141	0.056	2.510
$\phi_{1,1,1}$	-0.057	0.044	-1.298	$\phi_{1,2,1}$	0.009	0.050	0.183	$\phi_{1,3,1}$	-0.074	0.042	-1.749
$\phi_{1,1,2}$	0.068	0.043	1.573	$\phi_{1,2,2}$	0.007	0.066	0.113	$\phi_{1,3,2}$	0.163	0.061	2.649
$\phi_{1,1,3}$	0.009	0.039	0.229	$\phi_{1,2,3}$	0.064	0.056	1.138	$\phi_{1,3,3}$	-0.086	0.057	-1.509
$\phi_{2,1,1}$	0.024	0.038	0.629	$\phi_{2,2,1}$	0.052	0.043	1.208	$\phi_{2,3,1}$	-0.023	0.041	-0.556
$\phi_{2,1,2}$	-0.018	0.046	-0.390	$\phi_{2,2,2}$	-0.012	0.066	-0.181	$\phi_{2,3,2}$	0.028	0.064	0.433
$\phi_{2,1,3}$	-0.005	0.040	-0.114	$\phi_{2,2,3}$	0.027	0.060	0.444	$\phi_{2,3,3}$	-0.002	0.061	-0.033
$\phi_{3,1,1}$	0.017	0.037	0.443	$\phi_{3,2,1}$	0.056	0.045	1.258	$\phi_{3,3,1}$	-0.047	0.040	-1.183
$\phi_{3,1,2}$	-0.050	0.045	-1.107	$\phi_{3,2,2}$	-0.050	0.067	-0.758	$\phi_{3,3,2}$	0.046	0.064	0.715
$\phi_{3,1,3}$	0.053	0.042	1.276	$\phi_{3,2,3}$	0.083	0.063	1.331	$\phi_{3,3,3}$	0.009	0.061	0.144
$\phi_{4,1,1}$	0.145	0.038	3.846	$\phi_{4,2,1}$	0.074	0.042	1.746	$\phi_{4,3,1}$	0.014	0.040	0.344
$\phi_{4,1,2}$	-0.051	0.045	-1.130	$\phi_{4,2,2}$	0.016	0.069	0.232	$\phi_{4,3,2}$	-0.001	0.059	-0.015
$\phi_{4,1,3}$	0.029	0.041	0.709	$\phi_{4,2,3}$	0.021	0.064	0.328	$\phi_{4,3,3}$	0.023	0.058	0.396
$\phi_{5,1,1}$	0.008	0.037	0.219	$\phi_{5,2,1}$	-0.048	0.040	-1.223	$\phi_{5,3,1}$	-0.091	0.040	-2.289
$\phi_{5,1,2}$	-0.000	0.041	-0.010	$\phi_{5,2,2}$	0.066	0.060	1.103	$\phi_{5,3,2}$	0.102	0.061	1.671
$\phi_{5,1,3}$	0.005	0.033	0.160	$\phi_{5,2,3}$	0.019	0.055	0.343	$\phi_{5,3,3}$	-0.026	0.056	-0.466
$\phi_{6,1,1}$	0.004	0.033	0.124	$\phi_{6,2,1}$	0.045	0.036	1.236	$\phi_{6,3,1}$	0.010	0.036	0.269
$\phi_{6,1,2}$	0.053	0.038	1.401	$\phi_{6,2,2}$	-0.021	0.056	-0.380	$\phi_{6,3,2}$	0.023	0.056	0.409
$\phi_{6,1,3}$	-0.042	0.033	-1.284	$\phi_{6,2,3}$	0.006	0.053	0.113	$\phi_{6,3,3}$	-0.039	0.055	-0.713

Table A-6: Estimates for equation (9)

Par.	Coeff.	S.E.	z-stat	Par.	Coeff.	S.E.	z-stat
$C_{1,1}$	0.0113	0.0038	2.9470	$A_{1,3}$	-0.0016	0.1115	-0.0147
$C_{2,1}$	0.0000	0.0088	0.0008	$A_{2,3}$	-0.1081	0.0864	-1.2510
$C_{3,1}$	-0.0030	0.0062	-0.4865	$A_{3,3}$	0.1217	0.0655	1.8600
$C_{2,2}$	0.0260	0.0054	4.7740	$B_{1,1}$	0.9331	0.0305	30.5700
$C_{3,2}$	0.0079	0.0060	1.3230	$B_{2,1}$	0.0254	0.0230	1.1060
$C_{3,3}$	-0.0000	0.0000	-0.8786	$B_{3,1}$	-0.0106	0.0166	-0.6359
$A_{1,1}$	0.3594	0.0805	4.4620	$B_{1,2}$	0.0129	0.0453	0.2836
$A_{2,1}$	0.0258	0.0652	0.3959	$B_{2,2}$	0.9182	0.0360	25.4900
$A_{3,1}$	0.0685	0.0488	1.4050	$B_{3,2}$	0.0265	0.0296	0.8971
$A_{1,2}$	-0.0226	0.1126	-0.2003	$B_{1,3}$	-0.0054	0.0395	-0.1361
$A_{2,2}$	0.2755	0.0836	3.2970	$B_{2,3}$	0.0361	0.0256	1.4100
$A_{3,2}$	-0.0170	0.0740	-0.2292	$B_{3,3}$	0.9685	0.0198	49.0300

Moduli of the eigenvalues of $(A \otimes A + B \otimes B)$:

0.9995 0.9774 0.9625 0.9614 0.9516 0.9459 0.9161 0.9066 0.9024

Wald test for diagonal BEKK: $W = 46.0351$, p-value = 6.8e-6