ELSEVIER

Journal of Banking & Finance 21 (1997) 685–720

Journal of BANKING & FINANCE

The information content in bond model residuals: An empirical study on the Belgian bond market

Piet Sercu [a,*], Xueping Wu [b,1]

[a] Katholieke Universiteit, Naamsestraat 69, B-3000 Leuven, Belgium
[b] City University of Hong Kong, Tat Chee Avenue, Kow Loon, Hong Kong

Received 16 October 1995; accepted 14 November 1996

Abstract

We estimate daily Vasicek, CIR, and spline models on Belgian data and compare the trading profits that can be made on the basis of their residuals. Abnormal returns, measured using three different benchmarks, are negatively related to once- and twice-lagged mispricing. Buying underpriced bonds and (especially) selling overpriced bonds yields significant abnormal returns even when the trade is delayed by up to five days after observing the mispricing. The traditional spline model overfits the data and is least able to detect mispricing. Large model residuals are more likely to be the result of model misspecification or -estimation than are small or medium-sized residuals. © 1997 Elsevier Science B.V. All rights reserved.

JEL classification: G12; G14

Keywords: Bonds; One-factor model; Spline; Duration; Market efficiency

1. Introduction

In this paper, we compare the ability of the Vasicek (1977), Cox–Ingersoll–Ross (CIR) (Cox et al., 1985b), and curve-fitting models to identify mispriced bonds

* Corresponding author. e-mail: piet.sercu@econ.kuleuven.ac.be
[1] efxpwu@cityu.edu.hk

PII S0378-4266(96)00061-1

and generate trading profits. We chose the (one-factor) models by Vasicek (1977) and Cox–Ingersoll–Ross (CIR) (Cox et al., 1985b) because they combine a sound economic basis with tractable closed-form solutions. While the CIR model is already more difficult to estimate than the Vasicek model, this is even more the case for the two-factor models of, for example, Longstaff and Schwartz (1992) and Platten (1994); and empirical work by Grégoire and Platten (1995) on similar data as ours gives no indication that these two-factor models actually do a better job than their one-factor counterparts.

Our work differs from earlier empirical work on the Vasicek, CIR, and spline models in three respects. First, we select the ''best'' model on the basis of an information content criterion (as in Fama, 1990), rather than on purely statistical grounds. That is, unlike other authors we do not want to find out whether or not the Vasicek and CIR models are cross-sectionally as well as longitudinally compatible with the actual data [2]; rather, we want to know whether term structure models, estimated from a single-day cross-section, contain information about future bond returns, and which model seems to be best at identifying mispriced bonds. We also want to find out whether the Vasicek and CIR do any better, in this respect, than the simple five- or four-parameter splines that are still used in the financial community [3]. Our work further differs from earlier work on bond pricing models in that we work with BEF data. Lastly, we have tried to improve the robustness and validity of our findings by introducing some methodological refinements relative to standard market-efficiency tests. This methodology can be summarized as follows.

For every day in the sample, from 1991 through 1992, we first estimate the four competing models from the prices — not the yields — of a particular class of government coupon bonds, and from short-term bond prices constructed from money market interest rates. We estimate the Vasicek and CIR models without any pooling over time or without any inter-temporal constraints on the parameters that were assumed to be constant over time in the derivation of the equilibrium pricing model. In this sense, our approach is similar to standard practice among option traders, who re-estimate volatilities every day or use implicit standard deviations as a basis for trading although their pricing model assumes constant volatilities.

[2] For example, Brown and Dybvig (1986) estimate the CIR model on monthly price quotes for U.S. Treasury issues from 1952 through 1983, and De Munnik and Schotman (1994) test both the Vasicek model and the CIR model with daily data of Dutch Treasury bonds from 1990 through 1991. Related tests on real return data are provided by Brown and Schaefer (1994) and Pearson and Sun (1994). Grégoire and Platten (1995) have tested the CIR, and Longstaff (1992), Longstaff and Schwartz (1992) and Platten (1994) have tested models in the Belgian market.

[3] In view of this, our tests tend to be more like efficiency tests than proper tests of the models. However, the distinction is never very clear-cut: one cannot measure efficiency without a model, nor can one verify/falisfy a model without assuming efficiency. As we use different models to select bonds, our tests also tell us what model works best as a basis for bond trading — an issue that goes beyond the pure efficiency question.

Our day-to-day approach also has the merit that it does not load the dice in favor of the pure curve-fitting techniques, where intertemporal constraints are never imposed.

Having estimated the competing models, we then test whether one can realize abnormal returns by buying (shortselling) bonds that, on that day, were classified as undervalued (overvalued) relative to a particular estimated term structure model. In each of the tests described below, abnormal returns from bond trading are measured relative to three alternative benchmarks. One benchmark is the return on the bond that would have been observed if prices would, at all times, perfectly fit the term structure model that was used to identify the mispricing. Our second benchmark is the contemporaneous realized return on a well-diversified portfolio with the same value and duration as the bond(s) selected by the trading rule, while the third benchmark also matches the traded bonds in terms of convexity. Each of these first-pass estimated abnormal returns is then corrected for the average first-pass abnormal return on a portfolio of all bonds; this correction ensures that, in any given daily cross-section of bonds, the average corrected abnormal return is again exactly zero even after the money market instruments have been left out.

To verify the information content of deviations between observed and fitted prices, we follow two approaches. First, we regress abnormal holding period returns from an individual bond on past term structure model residuals (that is, actual price minus model price). Second, we compute abnormal returns from various trading rules based on differences between observed and model prices — first using a contrarian weighting scheme (with larger positions in bonds that are more mispriced), and then by forming separate portfolios for bonds with different degrees of mispricing. In both the regression tests and the trading rule tests we also introduce various lags between the moment of detecting the mispricing and the implementation of the trade, so as to eliminate biases stemming from bid-ask noise.

Both the regression tests and the results from the trading rule reveal that model residuals are economically useful. In addition, we find that the trading results based on the two economic models, and especially the Vasicek model, are superior to the results obtained when the decisions to buy or sell are based on the standard cubic spline. Lastly, the performance of the cubic spline model improves substantially when one cuts down the number of free parameters from five (as commonly used) to four; we interpret these findings as implying that the standard five-parameter spline model over-fits the data and, therefore, tends to overlook part of the mispricing.

The structure of the paper is as follows. Section 2 deals with the estimation of term structure models. We start with a brief review of the basics of term structure models in general and the Vasicek and CIR models in particular, and then present and discuss the estimates obtained from our sample. Section 3 tests whether the residuals from the estimated term structure model contain any information that would be useful for a trader. Section 4 concludes.

2. Estimation of the bond pricing models

Section 2.1 briefly presents the Vasicek, CIR and spline models. Section 2.2 describes the data and presents the estimation method for our cross-sectional estimation on coupon bond prices. The empirical results are discussed in Section 2.3.

2.1. Three bond pricing models

Let $P(r,t)$ denote the price of a zero-coupon bond or pure discount bond at t and assume that the underlying variable, the short-term interest rate $r(t)$, follows a diffusion process which is continuous over time and exhibits no jumps:

$$\mathrm{d}r = \gamma(r,t)\mathrm{d}t + \sigma(r,t)\mathrm{d}z, \tag{1}$$

where $\mathrm{d}r$ is the change in the short-term interest rate $r(t)$; $\gamma(r,t)$ is the drift rate of $r(t)$ [γ may depend both on $r(t)$ and t]; $\sigma(r,t)$ is the standard deviation of changes in $r(t)$ [σ may depend both on $r(t)$ and t]; and $\mathrm{d}z$ is the standard Wiener process with zero mean and unit per annum variance.

The familiar Black–Scholes (Black and Scholes, 1973), Merton (1973) no-arbitrage pricing equation for any asset that has the short-term interest rate (Eq. (1)) as the underlying factor is:

$$\frac{\partial P}{\partial t} + \frac{\partial P}{\partial r}[\gamma(r,t) - \lambda(r,t)\sigma(r,t)] + \frac{1}{2}\frac{\partial^2 P}{\partial r^2}\sigma^2(r,t) - r(t)P = 0. \tag{2}$$

In this expression, $\lambda(r,t)$ is the price of interest risk at time t, and the factor $[\gamma(r,t) - \lambda(r,t)\sigma(r,t)]$ is the risk-adjusted drift rate of the underlying state variable, in casu the short-term interest rate in Eq. (2). As is well known, the Vasicek and CIR models differ in the way they specify the terms $\gamma(r,t)$, $\sigma(r,t)$ and $\lambda(r,t)$ in Eqs. (1) and (2). We briefly review each model in turn.

In Vasicek (1977), the instant interest rate follows a mean-reverting normal (Ornstein–Uhlenbeck) process,

$$\mathrm{d}r = \kappa(m - r)\mathrm{d}t + \sigma\mathrm{d}z, \tag{3}$$

where κ, m and σ are constants; and $\mathrm{d}z$ is a Wiener process. With Eq. (3), the fundamental differential equation in Eq. (2) becomes:

$$\frac{\partial P}{\partial t} + \frac{\partial P}{\partial r}[\kappa(m - r) - \lambda(r,t)\sigma] + \frac{1}{2}\frac{\partial^2 P}{\partial r^2}\sigma^2 - r(t)P = 0. \tag{4}$$

Recall that $P_T(r,t)$ is the price, at t, of a zero-coupon bond or discount bond maturing at T and contingent on the short-term interest rate $r(t)$. By assuming a constant market price of risk λ over time and using the boundary condition that, at maturity, $P_T(r,T)$ equals unity, the following closed-form pricing model is obtained:

$$P_T(r,t) = \exp\Big[-\phi_0(t)\{1 - e^{-\kappa(T-t)}\} + \phi_1\{1 - \kappa(T-t) - e^{-\kappa(T-t)}\}$$
$$-\phi_2\{1 - e^{-\kappa(T-t)}\}^2\Big], \tag{5}$$

where

$$\phi_0(t) = \frac{r(t)}{\kappa}, \tag{6}$$

$$\phi_1 = \frac{\kappa m - \lambda\sigma}{\kappa^2} - \frac{1}{2}\frac{\sigma^2}{\kappa^3}, \tag{7}$$

$$\phi_2 = \frac{1}{4}\frac{\sigma^2}{\kappa^3}. \tag{8}$$

If the short-term rate $r(t)$ is taken to be unobservable, there are four coefficients to be estimated: κ, ϕ_0, ϕ_1 and ϕ_2. From these estimated coefficients we can derive the implied parameters,

$$\text{implied short-term rate:} \quad r(t) = \kappa\phi_0(t), \tag{9}$$

$$\text{yield on a bond with } T \to \infty: \quad R_L = \kappa\phi_1, \tag{10}$$

$$\text{implied variance of } dr: \quad \sigma^2 = 4\kappa^3\phi_2, \tag{11}$$

$$\text{risk-adjusted drift rate of } r(t): \quad \mu \equiv \kappa(m - r) - \lambda\sigma$$
$$= (\phi_1 + 2\phi_2)\kappa^2 - \kappa r. \tag{12}$$

In contrast, Cox et al. (1985b) adopt a specific general-equilibrium approach that allows them to derive both the interest rate dynamics and the corresponding price of risk:

$$\mathrm{d}r = \kappa(m - r)\mathrm{d}t + \sigma\sqrt{r}\,\mathrm{d}z, \tag{13}$$

$$\lambda(r,t) = \frac{q}{\sigma}\sqrt{r(t)}, \tag{14}$$

where q is a constant. As a result, the general differential Eq. (2) can be specified as

$$\frac{\partial P}{\partial t} + \frac{\partial P}{\partial r}[\kappa(m - r) - qr(t)] + \frac{1}{2}\frac{\partial^2 P}{\partial r^2}\sigma^2 - r(t)P = 0. \tag{15}$$

With the boundary condition $P_T(r,T) = 1$ for a maturing discount bond, the solution to Eq. (15) takes the following specific form:

$$P_T(r,t) = \left\{ \frac{\theta_1 e^{\theta_2(T-t)}}{\theta_2[e^{\theta_1(T-T)} - 1] + \theta_1} \right\}^{\theta_3} \exp\left\{ \frac{-r[e^{\theta_1(T-t)} - 1]}{\theta_2[e^{\theta_1(T-t)} - 1] + 1} \right\}, \tag{16}$$

where

$$\theta_1 = \sqrt{(\kappa + q)^2 + 2\sigma^2}, \tag{17}$$

$$\theta_2 = (\kappa + q + \theta_1)/2, \tag{18}$$

$$\theta_3 = 2\kappa m/\sigma^2. \tag{19}$$

Also in this model there are four coefficients to be estimated: $r(t)$, θ_1, θ_2 and θ_3. From these estimated coefficients we can derive the implied parameters,

$$\text{yield on a bond with } T \to \infty: \quad R_L = \theta_3(\theta_1 - \theta_2), \tag{20}$$

$$\text{implied variance of } dr: \quad \sigma^2 r(t) = 2\theta_2(\theta_1 - \theta_2) r(t), \tag{21}$$

$$\begin{aligned} \text{risk} - \text{adjusted drift rate of } r(t): \quad \mu &\equiv \kappa(m - r) - qr(t) \\ &= \theta_3\sigma^2/2 - (2\theta_2 - \theta_1) r(t). \end{aligned} \tag{22}$$

The cubic spline model [4], finally, is a purely descriptive model without economic foundations. The term structure function consists of a concatenation of a number of third-degree polynomials, spliced together at n "knot points", s_i, $i = 1, \ldots, n$, in a way that ensures continuity in the levels as well as the first and second derivatives:

$$\begin{aligned} P_T(r,t) &= 1 + a_t(T-t) + b_t(T-t)^2 + c_t(T-t)^3 \\ &\quad + \sum_{i=1}^{n} d_{t,i}[\mathrm{Max}\{T - (t + s_i), 0\}]^3. \end{aligned} \tag{23}$$

Usually one selects two knot points — in this study, $s_1 = 2$ years and $s_2 = 4$ years — which implies there are five free parameters in the spline model. As will be illustrated below, in the daily cross-sectional estimations the five-parameter spline tends to produce a better fit than the Vasicek or CIR models. This better fit may stem from two sources: first, this spline has one more free parameter, and second, it imposes less restrictions on the shape of the discount function than the other two models. To be able to sort out the relative importance of each explanation, we have repeated all tests using a four-parameter spline, that is, a spline with just one knot point (set at $s_1 = 2$ years). For the sake of brevity, we will report the results

[4] An alternative to the polynomial spline is the exponential spline (Vasicek and Fong, 1982). However, Shea (1985) finds that the exponential spline is not superior to the latter.

from the four-parameter spline only when they differ markedly from the results obtained with five parameters.

2.2. Data and estimation procedure

The competing models in Eqs. (5), (16) and (23) were estimated from data on BEF interbank deposits and BEF "linear" bonds (*Obligations Linéaires / Lineaire Obligaties*, or OLOs). In this section we describe the data and the estimation procedure.

Like France's *Obligations Assimilables*, OLO bonds are floated in consecutive tranches rather than in one single issue. Each new tranche of a given "line" has identical terms and conditions and is fully fungible (*assimilable*) with earlier tranche issues of the same line. The number of outstanding OLOs is much smaller than the number of ordinary government bonds traded during the same period. However, for the purpose of testing bond pricing models, OLOs have many advantages relative to ordinary bonds. First, OLOs are registered bonds. In contrast, the ordinary government bonds are bearer securities, which are more expensive to trade. Second, because OLOs are registered, they are mainly held by corporations. Because of this, tax clientèle effects are less likely to be a problem for OLOs than for ordinary bonds, which can be held by individuals as well [5]. Third, the coupons from OLOs are not subject to any withholding tax. This makes OLOs more convenient to corporations than ordinary bonds. Fourth, OLOs are more actively traded than ordinary bonds, partly because the primary dealers make a market. In contrast, ordinary bonds are traded either during a (low-volume) daily call auction on the Brussels Exchange, or off the exchange. Finally, OLOs are straight bonds with maturities of up to twenty years, while ordinary bonds are more short-lived and tend to have put or call option features.

Daily OLO price data and BEF Brussels interbank offer rates (BIBOR), from March 27, 1991 through December 30, 1992, were obtained from the *Financieel Economische Tijd* (FET) data service. After deleting non-trading days and some thin-trading days, 421 daily cross-section samples are available. At the beginning of our sample period we have six outstanding OLOs, with times to maturity ranging from about three to twelve years, while at the end we have twelve OLOs with times to maturity ranging from about one to twenty years (Table 1). We report results for the first 351 days only because, as of September 26, 1992 — a few days after the start of heavy tensions in, and a near-collapse of, the European Exchange Rate Mechanism — the term structure became characterized by a

[5] Under personal taxation, interest income on ordinary bonds is subject to a withholding tax of 10% plus, possibly, a (widely evaded) progressive additional tax if worldwide interest income exceeds certain thresholds. Capital gains go untaxed. Corporations, in contrast, all pay the same tax on interest income and capital gains.

Table 1
Belgian government linear bonds (OLOs) [a]

Code	March 27, 1991–December 30, 1992.				
	linear bonds	Issued (year)	Expires year)	Coupon rate (%)	Coupon due date
239.45	OLO01	1989	1999	8.25	Jun. 1
245.51	OLO02	1990	1996	10.00	Apr. 5
247.53	OLO03	1990	2000	10.00	Aug. 1
248.54	OLO04	1991	1998	9.25	Jan. 1
249.55	OLO05	1991	1994	9.50	Feb. 28
251.57	OLO06	1991	2003	9.00	Mar. 1
252.58	OLO07	1991	2001	9.00	Jun. 27
254.60	OLO08	1991	1997	9.25	Aug. 29
257.63	OLO09	1992	2007	8.50	Oct. 1
259.65	OLO10	1992	2002	8.75	Jun. 25
260.66	OLO11	1992	1998	9.00	Jul. 30
262.68	OLO12	1992	2012	8.00	Dec. 24

[a] OLOs are the Belgian government non-callable straight bonds. At the beginning, there are only 6 OLOs available and the number increases to 12 near the end.

trough. As a result, the CIR model estimations no longer converged while the Vasicek model could only be fitted at the cost of implying a negative value for σ^2. Results for the last period do not lead to different conclusions, and are available on request.

The OLO price data reported by the FET are last-trade transaction prices, which implies that they contain bid-ask noise. The maximum allowed bid-ask spread is 25 basis points. Bond price quotes have to be grossed up with accrued interest to obtain the effective invoice price. In addition, bond prices have to be corrected for the one-week settlement effect. That is, the invoice price is actually a one-week forward price. Thus, the bond prices we use for estimation are obtained from the invoice price as follows:

$$P_T = \frac{\text{quote} + \text{accrued interest}}{1 + (7/365)\ \text{BIBOR 1 month}}. \tag{24}$$

We use the 1-month BIBOR because the one-week interest rate is not available to us. Note that while accrued interest on bonds is based on a 360-day year, the Brussels interbank market uses a 365-day year to calculate interest on deposits and loans; this explains the factor (7/365) in the denominator.

To represent the short end of the maturity spectrum we have preferred interbank deposits over treasury bills. It is true that there has been an organized secondary market for treasury bills as of the spring of 1991, which is also the beginning of our sample; however, the T-bill data for the first trading year are rather suspect because, in that period, T-bill yields often exceed the BIBOR rate, by up to 10

Table 2
Brussels interbank offer rates on Belgian Franc (BIBORs)

	Interbank (BIBOR) rates, 27/03/1991D16/09/1992 (351 days)			
	high	low	mean	st. dev.
1-Month	10.250	8.875	9.421	0.299
2-Month	10.125	9.837	9.474	0.259
3-Month	10.063	9.000	9.506	0.228
6-Month	10.030	9.063	9.534	0.189
12-Month	10.000	9.125	9.527	0.173

basis points. This counter-intuitive premium relative to BIBOR reflected the extreme thinness of the T-bill market in the first year of trading. In contrast, the interbank money market is very deep, and has bid-ask spreads of 12.5 basis points per annum except during periods of EMS tensions.

Interbank interest rate data from the *Financieel Economische Tijd* bear on maturities of 1, 2, 3, 6, or 12 months (Table 2). To obtain midpoint prices for short-term discount bond from the BIBOR data, we converted offer rates into mean interbank rates by subtracting half the bid-ask spread and then discounting:

$$P_T = \frac{100}{1 + (T-T)/365 \times [\text{BIBOR}(t,T) - 6.25 \text{ points}]}, \tag{25}$$

where, following the convention in the BEF interbank market, $T - t$ is computed using the actual number of days and a 365-day year. With six to twelve OLOs and five interbank deposits, each cross-section contains eleven to seventeen assets [6].

The pricing equations (5), (16) and (23) refer to zero-coupon bonds, but OLOs are coupon bonds, that is, portfolios of different default-free discount bonds. Thus, the valuation formula for a coupon bond takes the following form:

$$P_T(r,t;c,N(t)) = \sum_{j=1}^{N} \text{CF}_j P_{T_j}(r,t), \tag{26}$$

where $P_T(r,t;c,N)$ is the effective price (quoted price plus accrued interest, and corrected for 7-day settlement effects) of a coupon bond with N annual coupons c and time to maturity $T - t$; $N(t)$ is the number of times cash flows are paid out during the remaining life of the coupon bond; CF_j is the cash flow (c or $100 + c$) received at times T_j, $j = 1, \ldots, N$; $P_{T_j}(r,t)$ is the price of a discount bond with time to maturity $T_j - t$ as given by Eq. (5) (Vasicek), Eq. (16) (CIR), or Eq. (23) (spline).

[6] When a bond was not traded on a particular day, we dropped the bond from the sample, so that the actual number of observations is sometimes smaller than the number of outstanding bonds.

For the economic models — Vasicek and CIR — we use non-linear least-squares to estimate Eq. (26), assuming, like Brown and Dybvig (1986) and De Munnik and Schotman (1994), that empirical bond prices have homoskedastic

Table 3
Cross-sectional estimation of term structure models [a]

(A) The Vasicek model

	Φ_1	Φ_2	Φ_3	κ	r [b] (%)	R_L [c] (%)	μ [d] (%)	σ^2 [e] (%)	SE [f] (%)
Max.	0.0644	0.0557	0.0159	0.0241	10.23	9.05	0.183	0.498	0.324
Min.	0.0098	0.0101	0.0003	0.0041	6.29	8.10	−0.016	0.001	0.037
Mean	0.0248	0.0240	0.0048	0.0101	8.76	8.54	0.035	0.077	0.135
St.D.	0.0071	0.0056	0.0025	0.0018	0.55	0.26	0.025	0.056	0.047
$t > 2.5$ [g]	22.5%	32.2%	42.7%	32.2%					
$t > 2$ [g]	36.2%	52.1%	61.8%	51.9%					
$t > 1.5$ [g]	56.1%	72.4%	77.8%	72.1%					

(B) The CIR model

	θ_1	θ_2	θ_3	r	R_L (%)	μ (%)	$\sigma^2 r$ (%)	SE (%)
Max.	0.0262	0.0220	5.1803	9.76	9.05	0.090	0.917	0.299
Min.	0.0015	0.0015	0.0260	7.63	8.07	−0.005	0.001	0.011
Mean	0.0103	0.0079	0.2061	8.90	8.52	0.019	0.144	0.124
St.D.	0.0052	0.0040	0.4300	0.41	0.26	0.016	0.147	0.054
$t > 2$	59.0%	80.6%	3.7%	94.6%				
$t > 1.5$	78.3%	86.3%	12.0%	96.3%				
$t > 1$	89.2%	2.6%	46.2%	97.4%				

(C) Cubic Spline models

	Five-parameter Cubic Spline model (knot points at 2 and 4 years)						Four-parameter (knot point at 2 years)	
	a_1 ($\times 10^3$)	a_2 ($\times 10^7$)	a_3 ($\times 10^{10}$)	d_1 ($\times 10^{10}$)	d_2 ($\times 10^{10}$)	SE	d_1 ($\times 10^3$)	SE
Max.	−0.2372	1.2159	0.1017	0.6072	0.1008	0.172	0.3636	0.174
Min.	−0.2819	0.1430	−0.5338	−0.1907	−0.0865	0.003	−0.0320	0.019
Mean	−0.2583	0.6472	−0.1954	0.1794	0.0069	0.080	0.1968	0.084
St.D.	0.0098	0.2596	0.1441	0.1722	0.0370	0.027	0.0979	0.027
$t > 2.5$	100.0%	98.6%	79.2%	61.5%	25.4%	84.3%		
$t > 2$	100.0%	99.7%	84.3%	71.8%	31.9%	89.7%		
$t > 1.5$	100.0%	99.7%	86.9%	78.6%	48.4%	94.9%		

errors across maturities [7]. Because our daily cross-sectional samples have at most seventeen data points and we did not want to pool over time (for reasons discussed in Section 1), GMM was deemed unsuitable. For the two spline models we used OLS.

2.3. Discussion of the empirical results

As shown in Fig. 1, during most of the sample period the term structure was characterized by either a steep decline or a positive hump situated around four months to maturity [8]. Table 3 presents mean values, maxima, and minima of the estimated and implied parameters for the Vasicek model (Panel A) and the CIR model (Panel B). As the pricing errors will provide the raw material for the analysis in Section 3, we here only discuss some unconditional moments of these errors, grouped either by model or by bond.

During the sample period, the CIR model marginally outperforms the Vasicek model in terms of goodness-of-fit: the average root mean square error (RMSE) of the regression is somewhat smaller for the CIR model (12.4 basis points for a bond with par value 100) than the Vasicek model (13.5 basis points). This RMSE is roughly equal to the maximum one would expect from a purely random bid-ask bounce: with a maximum bid-ask spread of 25 basis points and equal marginal probabilities that the price is a bid or ask price, the bid-ask bounce generates a RMSE of, at most, $\sqrt{(0.5)^2 \times (0.0025)^2} = 12.5$ basis points. We will provide

[7] The alternative is first to estimate zero-coupon yields, and then to fit the yield-versions of the models, assuming that errors in yields, not errors in prices, are homoskedastic. We prefer to work with prices because the estimation of yields introduces errors, and because transaction costs and bid-ask bounce are proportional to prices, not to yields.

[8] Although it is known that the CIR model can produce a humped term structure, such a shape has not been observed by Brown and Dybvig (1986) or De Munnik and Schotman (1994).

Notes to Table 3:

[a] The Vasicek model and the CIR model are estimated using non-linear least squares, and the Cubic Spline using OLS. Bond invoice prices consist of the daily cross-sectional data of OLOs and short-lived discount bonds converted from BIBORs (par 100) for the period March 27, 1991–September 16, 1992 (351 trading days). Simple annualization is used: where appropriate, daily results are multiplied by 365. The results regarding the first three parameters of the four-parameter spline are qualitatively similar to their five-parameter counterparts, and are not reported.

[b] r is the annualized implied short-term interest rate (i.e., daily rates $\times$ 365).

[c] R_L is the annualized yield on a very long-term zero coupon bond ($T \to \infty$).

[d] μ is the annualized risk adjusted drift rate of the short-term interest rate.

[e] The annualized implied variance of changes in r is σ^2 in the Vasicek model but $\sigma^2 r$ in the CIR model.

[f] SE (RMSE) stands for standard error (root mean squared error) of regression [e.g., 0.10 means 10 basis points (for a par value of 100)].

[g] Percentages of parameter estimates (per parameter) with t-ratio (in absolute values) greater than, say, 2.5.

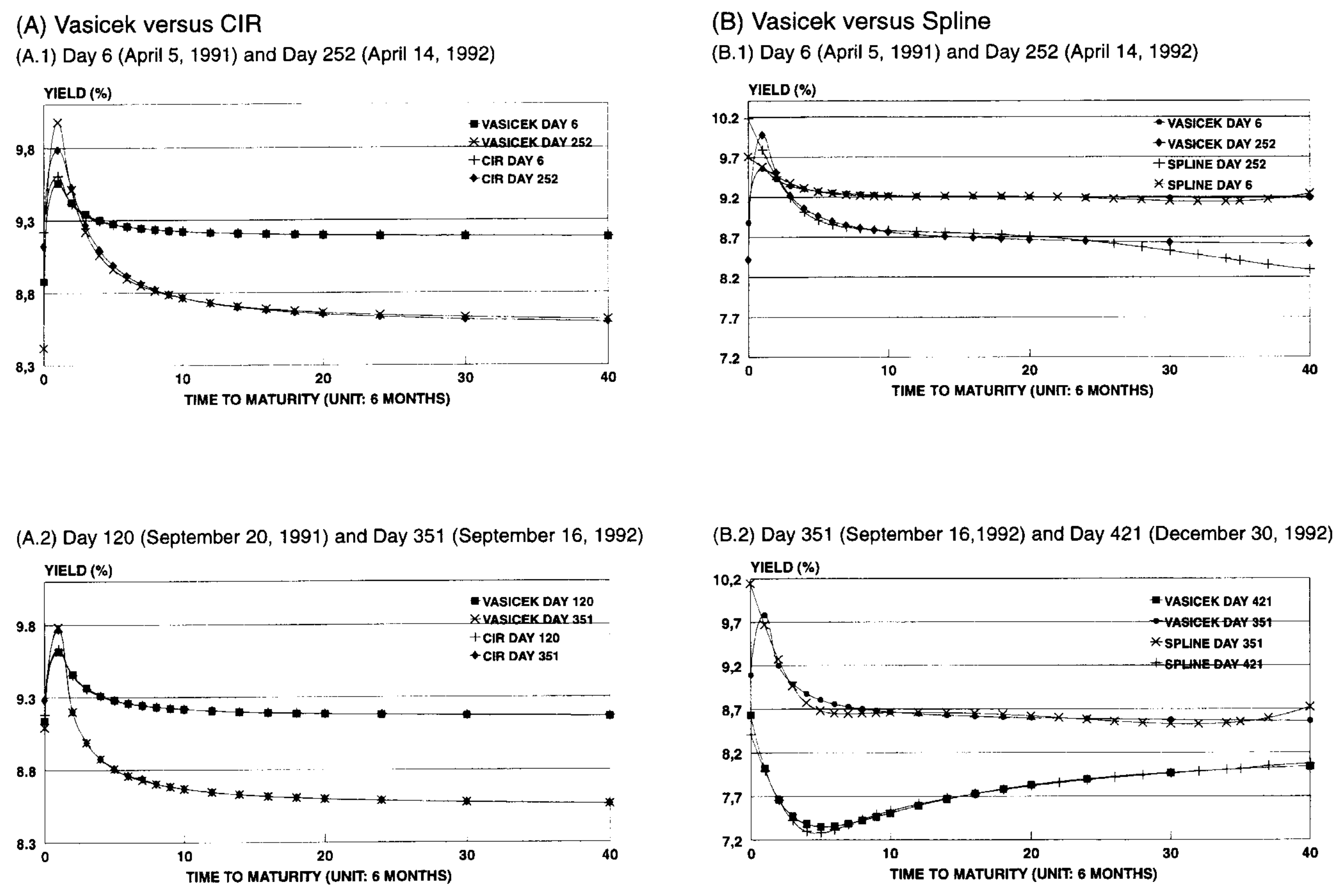

Fig. 1. Yield curve comparison. The Vasicek, the CIR, and the Cubic Spline Yield curves: Vasicek versus CIR (Panel A) and Vasicek versus Cubic Spline (Panel B) are plotted in pair on a few representative dates. The two knots for the Cubic Spline are set at the maturities of 2 and 4 years, respectively. For some dates, the maximum maturity of 20 years shown in the graphs requires extrapolation of the estimated term structure beyond actually available maturities. Time to maturity is measured in days.

evidence, below, that the residual RMSE is not just random bid-ask bounce, though; that is, the actual spreads must, on average, have been below the legal maximum of 25 basis points. While the residual RMSE produced by the two economic models is already low relative to the maximal bid-ask bounce and relative to the results obtained by De Munnik and Schotman (1994) [9], both cubic spline models easily beat the other two models in this respect: the mean RMSE is for the five- (four-) parameter spline is a mere 8.0 (8.4) basis points. This lower RMSE suggests that either the actual bid-ask spread probably was below the legal maximum 25 basis points — otherwise it would be hard to explain RMSEs below 12.5 basis points — , or that the spline model has a tendency to over-fit.

Deviations between actual prices and model prices can also be analyzed longitudinally, i.e. per asset rather than per cross-section, so as to verify whether or not the model consistently misprices some individual bonds. Table 4 and Fig. 2 report the mean error and the mean absolute error (MAE) per asset. Mean errors exhibit no clear pattern across assets, but the mean absolute errors (MAE) are more revealing. In both economic models (Vasicek and CIR) the MAEs tends to be smaller for interbank deposits than for bonds, with figures well below ten basis points and increasing with time to maturity of the deposit. The MAEs of OLO lines 02, 03, 07, and 09 exceed ten basis points; in addition, for OLO03 and 09 the size of the MAE is also close to the size of the mean error, which means that virtually all of the errors have the same sign-negative for OLO03, and positive for OLO09 [10]. The pricing errors obtained for OLOs 03 and 09 from the five-parameter spline functions are less consistently of the same sign, and much smaller in absolute value, than the errors obtained from the two economic models [11]. Similar results (not shown) were obtained with a four-parameter spline.

Like the average cross-sectional, the low MAEs for most bonds (with the exception for OLO03 and 09) seem to suggest that the MAEs may merely reflect

[9] De Munnik and Schotman (1994) found an average standard error of 18 basis points for the Dutch market. The difference between their and our results is unlikely to be explained by a higher turbulence during the Dutch sample period: while the yield curves obtained by De Munnik and Schotman are almost flat, we have steeply declining and humped curves. The higher standard deviations in De Munnik and Schotman are more likely to be the result of pooling data over one week, something we did not do. During the last 70 days in our sample, however, the residual standard deviations seem to have been substantially higher in both the spline and Vasicek models.

[10] None of the traders we talked to has provided any reason why these lines would behave abnormally. Five primary dealers have created a market in stripped bonds based on OLO09, but this occurred only after the (351-day) sample period. Thus, the stripping of OLO09 cannot affect the sample results.

[11] Thus, in the trading rule tests reported in Section 3, the economic models will systematically generate purchase signals for OLO03, and sale signals for OLO09, that are not followed, on average, by price corrections. We assume that the trader does not learn from these systematic errors and continues to buy (sell) OLO03 (OLO09) whenever the regressions suggest mispricing. In light of our finding that the Vasicek and CIR models still outperform the spline model in terms of trading profits, this assumption of "no learning" is conservative.

Table 4
Cross-sectional model residuals [a]

		(A) Vasicek model residuals					(B) CIR model residuals					(C) Cubic Spline residuals				
		max.	min.	mean	MAE	AC	max.	min.	mean	MAE	AC	max.	min.	mean	MAE	AC
1-Month	351	0.088	−0.136	−0.015	0.022	0.48	0.018	−0.078	−0.015	0.017	0.80	0.116	−0.046	0.006	0.011	0.65
2-Month	351	0.170	−0.150	−0.003	0.024	0.12	0.035	−0.086	−0.017	0.019	0.79	0.114	−0.042	0.004	0.017	0.80
3-Month	351	0.245	−0.149	0.013	0.031	0.24	0.060	−0.118	−0.013	0.020	0.73	0.067	−0.059	−0.002	0.023	0.87
6-Month	351	0.281	−0.150	0.034	0.052	0.55	0.220	−0.142	0.000	0.022	0.43	0.086	−0.069	−0.011	0.027	0.84
12-Month	351	0.308	−0.209	0.019	0.074	0.68	0.271	−0.143	0.010	0.054	0.71	0.105	−0.151	0.015	0.033	0.85
OLO01	314	0.253	−0.227	0.011	0.082	0.78	0.253	−0.305	−0.013	0.090	0.81	0.301	−0.157	0.068	0.097	0.78
OLO02	322	0.335	−0.402	0.023	0.119	0.90	0.287	−0.391	0.016	0.105	0.88	0.117	−0.216	−0.016	0.048	0.78
OLO03	324	0.133	−0.496	−0.195	0.198	0.96	0.045	−0.531	−0.216	0.217	0.93	0.066	−0.289	−0.116	0.118	0.64
OLO04	315	0.170	−0.267	−0.000	0.063	0.78	0.185	−0.263	−0.016	0.071	0.75	0.148	−0.149	0.018	0.045	0.70
OLO05	300	0.302	−0.257	−0.013	0.074	0.72	0.275	−0.176	−0.011	0.049	0.55	0.129	−0.155	−0.006	0.026	0.65
OLO06	318	0.392	−0.257	0.055	0.103	0.84	0.314	−0.415	0.004	0.082	0.77	0.279	−0.063	0.024	0.041	0.88
OLO07	282	0.253	−0.332	−0.036	0.136	0.93	0.207	−0.425	−0.082	0.130	0.86	0.217	−0.211	0.031	0.086	0.86
OLO08	219	0.189	−0.231	0.017	0.057	0.70	0.174	−0.264	0.010	0.063	0.69	0.151	−0.124	0.018	0.048	0.72
OLO09	118	0.700	−0.007	0.252	0.253	0.98	0.643	−0.132	0.198	0.202	0.92	0.003	−0.020	−0.007	0.007	0.47
OLO10	51	0.115	−0.293	−0.067	0.093	0.88	0.096	−0.293	−0.070	0.091	0.86	0.128	−0.143	−0.003	0.062	0.72
OLO11	31	0.130	−0.101	−0.031	0.056	0.74	0.130	−0.101	−0.031	0.056	0.74	0.129	−0.083	−0.001	0.044	0.69

[a] Cross-sectional model residuals (pricing errors) are defined as actual bond trade prices minus model prices for each individual bonds (par 100). Maximum, minimum, mean, absolute mean and autocorrelation are reported for the period: March 27, 1991–September 16, 1992. BIBORs are converted into discount bonds (par 100). MAE stands for mean absolute pricing error or model residual. AC stands for autocorrelation in individual model residuals and all results are significant at 1% level. Results for OLO12 (6 obervations) are not reported.

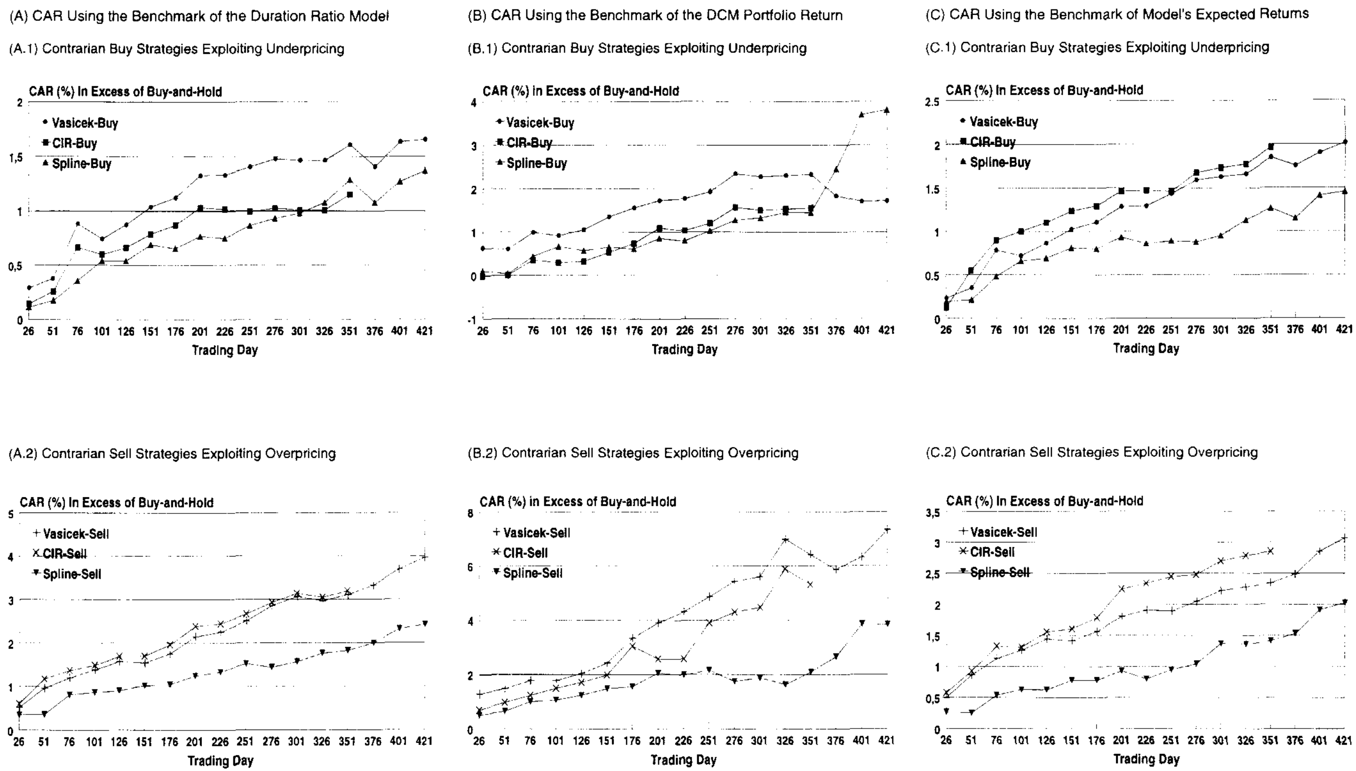

Fig. 2. CAR on the buy-and-hold portfolio. The buy-and-hold portfolio consists of OLOs only. This chart shows the CARs for this buy-and-hold portfolio computed for each of the benchmarks — the duration ratio model, the duration-and-convexity-matched (DCM) portfolio return, and the three model specific benchmarks (Vasicek model's expected return, the CIR model's expected return, and the Cubic Spline model's expected return). The CIR result is only available before September 17, 1992 (or 351 trading days).

purely random bid-ask bounce (which would generate a MAE of, at most, $(|12.5| + |-12.5|)/2 = 12.5$ basis points). However, such a conclusion would be unwarranted. First, there is a substantial MAE for many deposits, too; and as market values for interbank deposits are based on the mean interest rate, bid-ask noise is absent from these data. Second, the last columns of each panel of Table 4 reveal that, for all assets, the first-order autocorrelations in the model residuals are significantly positive. Purely random bid-ask bounce cannot be a source of autocorrelation in pricing errors (as opposed to returns, or residual returns, where bid-ask bounce causes negative autocorrelation). It follows that the major sources of apparent mispricing must be either highly autocorrelated errors in the specification or estimation of the model, or highly autocorrelated true mispricing, or both, rather than purely random bid-ask bounce. In Section 3 we will have a closer look at the model residuals for the OLOs, and verify whether they allow any profitable trading strategies or successful forecasts about holding period returns.

3. The information content in the model residuals

One conceptual weakness of models that, like the Vasicek or CIR model, postulate an interest rate or another non-price process as the driving state variable, is that such a model does not take the current term structure as given and is, therefore, likely to deem all outstanding bonds to be mispriced. Clearly, at least some of this apparent mispricing is likely to be due to model misspecification. On the other hand, in the presence of noise trading by uninformed or time-pressed investors it is quite likely that bonds are, to some extent, effectively mispriced relative to the (unidentified) "true" model. In this section, we verify whether the apparent mispricing in the Vasicek and CIR models is entirely due to model misspecification and mis-estimation or whether such a model is also able to detect some genuine mispricing due to noise trades. If there is genuine mispricing, trading on the basis of model residuals should be profitable. In short, in this part of the paper we view the CIR and Vasicek estimated term structure models as (somewhat complicated) curve-fitting techniques, and we do not worry about non-constancy of those parameter estimates that, in the logic of the model, should be constant [12]. The focus is on how useful the model residuals are to a bond trader, and whether the economic models outperform simple spline models for the purpose of identifying mispriced bonds.

This part of the paper is structured as follows. Section 3.1 defines the holding period returns and the equilibrium expected holding period returns that serve as benchmarks in our subsequent regression and trading rule tests. We use three alternative benchmark returns. One is the duration ratio model — a single index

[12] Grégoire and Platten (1995) do test for the statistical acceptability of the cross-temporal constraints in the Belgian market, and find that all models fail in this respect.

model that, like the market model for stocks, compares the realized return on the trading portfolio to the return on a diversified portfolio with the same value and risk (duration). The second benchmark is the return on a portfolio that matches the bond in terms of both duration and convexity. Both these benchmark are rather ad hoc, but they have the advantage of being independent of the details of the term structure model upon which the trading rule is based. The last benchmark is the conditional expected bond return implied by the change in the fitted model prices. A first test of the potential usefulness of the term structure residuals is conducted in Section 3.2, where we regress each of these measures of abnormal bond returns on the previous trading day's percentage mispricing. The second test is a trading rule test, described in Sections 3.3 and 3.4. We compute CARs in calendar time for three trading strategies: (1) buy underpriced bonds; (2) shortsell overpriced bonds; and (iii) combine both. In Section 3.3 the weights within each portfolio are proportional to the degree of initial mispricing relative to the model that is being used (Vasicek, CIR, or spline), while in Section 3.4 the weights are equal but the mispricing has to exceed a give filter size.

3.1. Bond holding period returns and expected returns

From each day's estimated Vasicek term structure, we compute the day's Vasicek residual for each bond, i.e. the actual bond price minus the model price or fitted value. The procedure is repeated for the CIR and spline models. If a given bond pricing model is correct and reliably estimated, then a positive residual implies that the corresponding bond is overvalued, while a negative model residual implies that the bond is undervalued. Subsequent holding period returns can then be analyzed to verify or falsify that model's diagnosis.

In this section we describe the three benchmarks that are used to eliminate the "normal" component in these holding period returns. Event studies or trading rule tests in the stock market frequently use benchmarks like the market model or the ex post CAPM, a procedure which filters out price changes due to general market movements while simultaneously taking into account differences in market sensitivity (β). The advantage of doing so is that, when holding period returns are corrected for market movements, the standard error of the abnormal return becomes smaller and the tests more powerful. Below, we propose three alternative benchmark returns that likewise intend to filter out general market movement from the raw bond returns defined in Eqs. (27) and (29).

The first way to eliminate the normal return uses the normative prices, at times $t-1$ and t, implied by the model that is being considered. Define Φ_t as the set of model parameter estimates obtained for day t. From the estimated model for day $t-1$, we can compute the model's equilibrium price for any bond i, which we denote by $\hat{P}_{i,t-1}$. We can also compute the fitted next-day equilibrium price using the time-t estimated parameters, denoted by $\hat{P}_{i,t}$. These two equilibrium prices imply an equilibrium holding period return $E_t(\mathrm{HP}_{it}|\Phi_{t-1},\ \Phi_t)$ and a corresponding abnormal return (AR), as follows:

$$E_t(\mathrm{HP}_{i,t}|\Phi_{t-1},\Phi_t) = \frac{\hat{P}_{i,t} - \hat{P}_{i,t-1} + \text{coupon payment}}{\hat{P}_{i,t-1}}, \tag{27}$$

and

$$\mathrm{AR}_{i,t} = \mathrm{HP}_{i,t} - E_t(\mathrm{HP}_{i,t}|\Phi_{t-1},\Phi_t). \tag{28}$$

While this expected return captures movements of the market as a whole between days t and $t-1$, and implicitly takes into account the sensitivity of the bond to shifts in the term structure, the procedure has the drawback that it assumes the validity of the very model whose forecasting performance is being tested. This may introduce some degree of circularity into the tests. We therefore employ two additional benchmarks for expected returns.

The stock market model defines the abnormal return on a share as the estimated residual $\epsilon_{i,t}$ from the regression $\mathrm{HP}_{i,t} = \alpha_i + \beta_i \mathrm{HP}_{m,t} + \epsilon_{i,t}$, where $\mathrm{HP}_{i,t}$ is the return on the stock between t and $t-1$, and $\mathrm{HP}_{m,t}$ is the contemporaneous realized return on the market portfolio m. For bonds, estimating β from a times series regression creates conceptual problems, since the β-coefficient of a bond is changing with its time to maturity. To avoid time series estimation of β we adopt a duration model similar to the one in Reilly and Sidhu (1980) and Elton and Gruber (1991), who suggest to use the ratio of duration of the individual bond over duration of the (equally weighted) market portfolio as an approximation for β. The one-factor duration model is:

$$\mathrm{HP}_{i,t} - \alpha_{i,t-1} = \beta_{i,t}\left[\overline{\mathrm{HP}}_t - \bar{\alpha}_{t-1}\right], \tag{29}$$

where $\alpha_{i,t-1}$ = the per annum continuously compounded yield on bond i, times Δt $(= 1/365)$; and $\beta_{i,t} = D_{i,t}/D_{m,t}$, the relative duration beta.

In the presence of noise, we can append an error term to Eq. (29) which, in an otherwise efficient market has a zero expectation [13]. Given the change in the term structure as summarized by $\overline{\mathrm{HP}}_t$, we can compute the abnormal return (AR) as follows:

$$\mathrm{AR}_{i,t} = \mathrm{HP}_{i,t} - \left[\alpha_{i,t-1} + \beta_{i,t}\left[\overline{\mathrm{HP}}_t - \bar{\alpha}_{t-1}\right]\right]. \tag{30}$$

The duration model (29) has the advantage that it does not assume the validity of the model that is being tested. This advantage comes at a cost: as is well known, the duration model underlying Eqs. (29) and (30) assumes that the consecutive term structures are parallel to each other, and that changes are minute

[13] If the market portfolio contains a sufficiently large number of assets, such noise will not materially affect the market return $\overline{\mathrm{HP}}_t$. In our case the market portfolio contains just the eleven to seventeen assets. With such a small bond portfolio, an abnormally high (low) return in one of the OLO's will also affect the market return upwards (downwards), which then implies that the excess return as computed from Eq. (29) is biased towards zero. Thus, the benchmark is overly conservative. Since we do find abnormal returns, the existence of a small-sample bias actually reinforces our conclusions.

and non-stochastic. It is true that, when the intervals are very short (one day) and only medium- to long-term bonds are considered, these assumptions are less likely to cause major problems. However, at a small cost we can also use a second-degree approximation that better accommodates changes over finite intervals and linear twists of the term structure.

Thus, we compute as our third benchmark the return on a portfolio that matches the trading portfolio as far as present value, duration $(-\partial P_i/\partial R \cdot P^{-1})$, and convexity $(0.5 \cdot \partial^2 P_i/\partial R^2 \cdot P^{-1})$ are concerned. This value-, duration- and convexity-matched (DCM) portfolio uses three equally-weighted portfolios. Our first portfolio contains the one-, two-, and three-month interbank deposits, the second portfolio the six- and twelve-month deposits, and the last portfolio all OLOs except the OLO that is being matched.

3.2. Regression test

The question to be answered in the remainder of this paper is whether the amount of mispricing, as identified from the cross-sectional term structure estimates, carries any information for the subsequent holding period. The logic is as follows. The deviation between the observed price and the model price consists potentially of: (1) a purely apparent (spurious) mispricing that is due to model misspecification or mis-estimation; and (2) genuine mispricing relative to the (unidentified) ''true'' valuation model. If all of the observed deviations between model prices and actual quotes stem from model mis-specification or -estimation [component (1)], then there is no reason why this deviation should be informative about subsequent returns. If, on the other hand, a non-trivial part of the deviation corresponds to genuine mispricing, then this mispricing should, on average, disappear over time. That is, truly undervalued (overvalued) bonds should provide above-normal (below-normal) holding period returns later on. To sort out this issue, the holding period returns in excess of the benchmark returns, as defined in Section 2, are analyzed in two ways. In this section we discuss the results from regression tests where the initial mispricing is related to subsequent abnormal returns. In later sections we test a trading rule.

To test whether there is a genuine mispricing component in the term structure model residuals, we first focus on the very short run: we regress abnormal rates of returns of a bond between days $t-1$ and t on the bond's percentage residual observed at $t-1$. Thus, the first regression is:

$$\mathrm{AR}_{i,t} = a + b\frac{\mathrm{RES}_{i,t-1}}{P_{i,t-1}} + e_t, \tag{31}$$

where $\mathrm{AR}_{i,t}$, the abnormal return on bond i, defined as the return in excess of either the model-implied normal return, the DM portfolio return, or the DCM portfolio return; and $\mathrm{RES}_{i,t-1} = P_{i,t-1} - \hat{P}_{i,t-1}$ where $P_{i,t-1}$ is the actual bond price at $t-1$ and $\hat{P}_{i,t-1}$ is the fitted value of the price at $t-1$ computed from the time $t-1$ Vasicek, CIR, or spline model.The competing hypotheses are:

Table 5
Regression tests [a]

(A) Vasicek residuals

	Obs.	(A.1) Benchmark: DM portfolio return				(A.2) Benchmark: DCM portfolio return				(A.3) Benchmark: Return from fitted prices			
		lag = 0		lag = 1		lag = 0		lag = 1		lag = 0		lag = 1	
		a	b	a	b	a	b	a	b	a	b	a	b
		$(\times 10^{-4})$		$(\times 10^{-4})$		$(\times 10^{-4})$		$(\times 10^{-4})$		$(\times 10^{-4})$		$(\times 10^{-4})$	
OLO01	376	0.454	−0.175	0.398	−0.151	0.367	−0.189	0.268	−0.157	0.277	−0.195	0.169	−0.146
		(1.38)	(−5.24) **	(1.17)	(−4.73) **	(0.91)	(−4.98) **	(0.65)	(−4.04) **	(0.79)	(−5.90) **	(0.46)	(−4.49) **
OLO02	383	0.624	−0.074	0.542	−0.043	1.304	−0.116	1.217	−0.055	0.088	−0.129	−0.046	−0.059
		(2.36) **	(−3.03) **	(1.99) *	(−2.00) *	(2.73) **	(−3.01) **	(2.47) **	(−1.38)	(0.28)	(−5.02) **	(−0.14)	(−2.44) **
OLO03	388	−2.113	−0.120	0.829	−0.047	−3.660	−0.182	−2.215	−0.099	−2.825	−0.160	−0.844	−0.050
		(−2.75) **	(−3.11) **	(−1.51) **	(−1.69) *	(−3.72) **	(−3.78) **	(−3.04) **	(−2.73) **	(−3.85) **	(−4.34) **	(−1.45)	(−1.67) *
OLO04	379	0.520	−0.134	0.520	−0.102	0.820	−0.114	0.804	−0.086	0.268	−0.251	0.127	−0.131
		(1.68)	(−3.93) *	(1.71) *	(−2.71) **	(2.06) *	(−2.16) *	(1.94) *	(−1.59)	(0.97)	(−5.94) **	(0.43)	(−3.43) **
OLO05	354	0.609	−0.088	0.0.0611	−0.085	1.127	−0.082	1.185	−0.015	−0.534	−0.247	−0.235	−0.078
		(2.52) **	(−2.91) **	(2.35) **	(−2.86) **	(2.85) **	(−1.60)	(2.88) *	(−0.29)	(−1.57)	(−5.81) **	(−0.64)	(−1.75) *
OLO06	381	0.120	−0.117	−0.388	−0.034	−0.643	−0.321	−1.878	−0.119	0.765	−0.191	0.065	−0.053
		(0.26)	(−2.24) *	(−0.77)	(−0.72)	(−0.76)	(−3.55) **	(−2.02) *	(−1.26)	(1.99) *	(−4.58) **	(0.15)	(−1.29)
OLO07	345	−0.593	−0.057	−0.474	−0.030	−1.764	−0.098	−1.596	−0.077	−0.545	−0.077	−0.260	−0.031
		(−1.47)	(−2.84) **	(−1.21)	(−1.53)	(−3.35) **	(−3.77) **	(−3.03) **	(−2.88) **	(−1.46)	(−3.87) **	(−0.70)	(−1.51)
OLO08	283	1.458	−0.115	1.182	−0.059	1.698	−0.140	1.525	−0.117	1.164	−0.178	0.701	−0.082
		(3.45) **	(−3.74) **	(3.29) **	(−1.24)	(2.97) **	(−2.51) **	(2.64) **	(−1.87) *	(2.90) **	(−5.12) **	(1.82) *	(−1.66) *
OLO09	186	−2.566	0.046	−2.495	0.034	2.759	−0.223	−0.222	−0.102	2.676	−0.115	1.139	−0.051
		(−2.00) *	(1.03)	(−2.05) *	(0.81)	(0.52)	(−1.09)	(−0.04)	(−0.49)	(2.93) **	(−2.96) **	(1.37)	(−1.27)
OLO10	119	−2.259	−0.256	−2.578	−0.318	−2.170	−0.175	−2.072	−0.187	−1.042	−0.176	−0.639	−0.144
		(1.84) *	(−2.57) **	(−2.11) *	(−3.52) **	(−1.80) *	(−1.78) *	(−1.69) *	(−1.94) *	(−1.50)	(−3.01) **	(−0.90)	(−2.55) **
OLO11	95	1.686	−0.170	1.697	−0.175	0.210	0.006	0.315	−0.031	0.928	−0.174	1.006	−0.187
		(1.86) *	(−2.47) **	(1.88) *	(−2.66) **	(0.14)	(0.05)	(0.19)	(−0.26)	(1.44)	(−2.91) **	(1.49)	(−3.11) **

(B) CIR residuals

	Obs.	(B.1) Benchmark: DM portfolio return				(B.2) Benchmark: DCM portfolio return				(B.3) Benchmark: Return from fitted prices			
		lag = 0		lag = 1		lag = 0		lag = 1		lag = 0		lag = 1	
		a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b
OLO01	313	0.367	−0.153	−0.148	−0.074	−0.377	−0.158	−0.143	−0.073	−0.566	−0.200	−0.288	−0.075
		(−1.09)	(−5.11) **	(−0.42)	(−2.36) **	(−0.92)	(−4.61) **	(−0.33)	(−2.05) *	(−1.49)	(−6.00) **	(−0.69)	(−2.20) *
OLO02	321	0.521	−0.070	0.383	−0.040	1.269	−0.102	1.125	−0.056	0.254	−0.137	−0.020	−0.050
		(1.84) *	(−2.50) **	(1.35)	(−1.84) *	(2.49) **	(−2.49) **	(2.19) *	(−1.37)	(0.73)	(−4.05) **	(−0.06)	(−1.57)
OLO03	323	1.432	−0.072	−0.715	−0.037	−2.722	−0.114	−2.025	−0.078	−6.617	−0.343	−0.196	−0.012
		(−1.50)	(−1.76) *	(−0.75)	(−0.88)	(−2.23) *	(−2.20) *	(−1.64)	(−1.48)	(−6.55) **	(−7.02) **	(−0.15)	(−0.20)
OLO04	314	0.045	−0.079	−0.085	0.029	0.521	−0.086	0.485	−0.076	−0.354	−0.255	−0.163	−0.116
		(0.20)	(−2.62) **	(0)	(−2.85) **	(1.32)	(−1.90) *	(1.19)	(−1.52)	(−1.11)	(−5.72) **	(−0.49)	(−2.35) **
OLO05	299	0.471	−0.131	0.465	−0.092	1.115	−0.134	1.066	−0.064	−0.629	−0.456	−0.285	−0.128
		(1.91) *	(−2.66) **	(1.60)	(−1.34)	(2.56) **	(−1.67) *	(2.33) **	(−0.81)	(−1.97) *	(−5.90) **	(−0.75)	(−1.47)
OLO06	318	0.164	−0.099	−0.273	−0.025	−2.031	−0.204	−2.514	−0.008	0.049	−0.234	−0.121	−0.091
		(−0.43)	(−1.94) *	(−0.68)	(−0.65)	(−2.03) *	(−2.33) **	(−2.40) **	(−0.10)	(0.13)	(−4.93) **	(−0.30)	(−2.33) **
OLO07	281	0.387	0.050	−0.169	−0.015	−1.702	−0.089	−1.548	−0.069	−1.449	−0.195	−0.512	−0.069
		(−1.01)	(−1.98) *	(−0.46)	(−0.61)	(−3.01) **	(−2.62) **	(−2.68) **	(−2.08) *	(−2.89) **	(−5.18) **	(−0.89)	(−1.93) *
OLO08	218	0.557	−0.139	0.536	−0.122	1.259	−0.206	1.097	−0.177	0.387	−0.303	0.165	−0.106
		(2.00) *	(−3.60) **	(1.87) *	(−3.45) **	(2.52) **	(−3.32) **	(2.13) *	(−2.81) **	(1.04)	(−5.58) **	(0.40)	(−1.98) *
OLO09	117	1.751	−0.107	0.981	−0.078	8.672	−0.601	4.903	−0.419	5.331	−0.274	2.042	−0.106
		(1.72) *	(−2.55) *	(0.92)	(−1.64)	(1.75) *	(−2.87) **	(0.86)	(−1.55)	(2.87) **	(−3.11) **	(1.30)	(−1.34)
OLO10	50	1.234	−0.164	−0.694	−0.116	−0.970	−0.057	0.751	0.113	−1.918	−0.248	−0.054	−0.117
		(−1.34)	(−1.63)	(−0.62)	(−0.92)	(−0.65)	(−0.53)	(0.43)	(0.74)	(−2.12) *	(−3.05) **	(−0.43)	(−1.05)
OLO11	30	0.435	−0.194	−1.498	−0.464	−0.735	−0.247	−0.159	−0.158	−0.469	−0.235	−1.197	−0.419
		(−0.39)	(−1.17)	(−1.96) *	(−3.67) **	(−0.45)	(−1.09)	(−0.10)	(−0.68)	(−0.42)	(−1.46)	(−1.44)	(−2.85) **

Table 5 (continued)

(C) Spline residuals, five parameters

	Obs.	(C.1) Benchmark: DM portfolio return				(C.2) Benchmark: DCM portfolio return				(C.3) Benchmark: Return from fitted prices			
		lag = 0		lag = 1		lag = 0		lag = 1		lag = 0		lag = 1	
		a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b	a ($\times 10^{-4}$)	b
OLO01	376	1.588	−0.255	1.375	−0.198	1.511	−0.231	1.222	−0.196	1.485	−0.218	0.830	−0.138
		(4.17) **	(−6.21) **	(3.52) **	(−5.40) **	(3.14) **	(−5.60) **	(2.62) **	(−4.53) **	(3.90) **	(−6.41) **	(2.45) **	(−3.75) **
OLO02	383	0.178	−0.114	0.265	−0.074	0.593	−0.184	0.778	−0.121	−0.594	−0.163	−0.170	−0.030
		(0.65)	(−2.80) **	(0.94)	(−2.08) *	(1.20)	(−2.59) **	(1.50)	(−1.77) *	(−2.91) **	(−4.80) **	(−0.82)	(−0.87)
OLO03	388	−1.096	−0.097	−0.628	−0.055	−2.343	−0.168	−1.510	−0.091	−1.900	−0.166	−0.663	−0.056
		(−1.79) *	(−2.05) *	(−1.35)	(−1.51)	(−2.87) **	(−2.74) **	(−2.48) **	(−0.98) *	(−3.99) **	(−4.03) **	(−1.71) **	(−1.54)
OLO04	379	0.820	−0.240	0.601	−0.102	0.900	−0.109	0.751	−0.023	0.629	−0.338	0.263	−0.127
		(2.17) *	(−3.57) **	(1.82) *	(−1.87) *	(2.05) *	(−1.38)	(1.63)	(−0.28)	(2.82) **	(−7.62) **	(1.10)	(−3.05) **
OLO05	354	0.649	−0.197	0.630	−0.176	1.124	−0.251	1.108	−0.163	−0.185	−0.293	−0.073	−0.098
		(2.48) **	(−2.36) **	(2.31) *	(−2.04) *	(2.87) **	(−2.29) *	(2.72) **	(−1.41)	(−1.29)	(−5.11) **	(−0.46)	(−1.21)
OLO06	381	−0.427	0.034	−0.654	0.052	−1.867	−0.003	−2.436	0.036	0.473	−0.153	0.329	−0.078
		(−0.95)	(0.64)	(−1.46)	(1.00)	(−1.81) *	(−0.03)	(−2.34) **	(0.40)	(2.10) *	(−4.31) **	(1.18)	(−2.10) *
OLO07	345	−0.223	−0.083	−0.280	−0.037	−1.129	−0.129	−1.110	−0.078	−0.061	−0.098	−0.091	−0.026
		(−0.60)	(−3.13) *	(−0.74)	(−1.32)	(−2.47) **	(−3.89) **	(−2.35) **	(−2.32) *	(−0.22)	(−4.06) **	(−0.31)	(−0.94)
OLO8	283	1.376	−0.121	1.126	−0.058	1.207	−0.058	1.403	−0.113	1.013	−0.179	0.565	−0.055
		(3.00) **	(−2.66) **	(3.22) **	(−0.92)	(1.96) *	(−0.80)	(2.47) **	(−1.39)	(3.16) **	(−4.95) **	(2.08) *	(−1.18)
OLO09	186	−1.855	−0.531	−2.300	−0.797	−0.971	2.203	−2.443	−0.175	−0.074	−0.356	0.330	0.006
		(−2.43) **	(−1.51)	(−2.68) **	(−1.32)	(−0.31)	(1.66) *	(−0.74)	(−0.07)	(0.38)	(−1.93) *	(1.68) *	(0.04)
OLO10	119	0.069	−0.230	0.273	−0.266	−0.348	−0.219	−0.481	−0.135	0.786	−0.227	0.205	−0.056
		(0.01)	(−1.73) *	(0.37)	(−1.97) *	(0.47)	(−1.88) *	(−0.60)	(−1.06)	(1.26)	(−3.44) **	(0.29)	(−0.74)
OLO11	95	1.587	−0.205	1.496	−0.193	−0.130	0.115	−0.145	0.099	0.699	−0.242	0.647	−0.221
		(1.68) *	(−1.70) *	(1.67) *	(−1.48)	(−0.08)	(0.56)	(−0.09)	(0.50)	(1.21)	(−2.82) **	(1.09)	(−2.52) **

H_1: $b = 0$ and $a = 0$: In setting the next day's price, the market ignores the estimated mispricing, either because the deemed mispricing is irrelevant or because the market does not react within one day;

H_2: $b = -1$: All of the estimated mispricing is corrected within one day;

H_3: $-1 < b < 0$: Some of the estimated mispricing is only apparent, and/or the market needs more than one day to fully correct the error.

Bid-ask noise may bias these tests in favor of the information-content hypotheses, H_2 and H_3. Specifically, assume that midpoint prices fully correspond to the predictions of the model that is being tested. As our data are transaction prices rather than midpoints, bid-ask noise would nevertheless induce spurious under- or overpricing; and this measured initial mispricing would, on average, disappear the next day because the next price is equally likely to be a bid price or an ask. This apparent error correction in the prices would then result in a spuriously negative estimate of b. To avoid this bid-ask induced bias in the slope coefficient of Eq. (31), we therefore add a new regression test, which differs from Eq. (31) in that that the regressor is taken from the last trading but one:

$$\mathrm{AR}_{i,t} = a + b\frac{\mathrm{RES}_{i,t-2}}{P_{i,t-2}} + e'_t. \tag{32}$$

The disadvantage of introducing the lag is that, if at time $t-2$ there is genuine mispricing (rather than just bid-ask noise), this genuine mispricing may be partly or entirely gone by time $t-1$, when the holding period starts. Thus, for the purpose of detecting genuine mispricing, the regression coefficients in Eq. (32) are biased against H_2 and H_3 rather than in favor of H_2 and H_3 (as is the case with Eq. (31)).

The empirical results for Eq. (31) ("lag = 0") and Eq. (32) ("lag = 1") are presented in Table 5; in Panel A, the Vasicek percentage residual is used as the regressor, while in Panels B and C the regressor is the percentage residual from the CIR and five-parameter spline model, respectively. Each of these panels has three subparts depending on the benchmark used in computing the abnormal part

Notes to Table 5:

[a] $\mathrm{AR}_t = a + b(\mathrm{RES}_{t-1-\mathrm{lag}})/(P_{t-1-\mathrm{lag}}) + e_t$ (Lag = 0 or 1; t-ratios in parentheses)

OLO data are from March 27, 1991 (or from the first trade) through December 30, 1992 (September 16, 1992, for the CIR model). The regressand is the percentage deviation between the observed price and the fitted price obtained from either the Vasicek (Panel A), CIR (Panel B), or five-parameter spline model (Panel C). The regressor is the deviation between the observed return on a benchmark portfolio which is either duration-matched (DM-Panels A1, B1, C1) or duration-and-convexity-matched (DCM-Panels A2, B2, C2), or the deviation between the observed return and the return on the fitted prices implied by the time structure model (Panels A3, B3, C3). The regressor is either the one observed at the beginning of the one-day holding period ("lag = 0"), or one trading day before ("lag = 1"); the former probably is likely to bias the slope coefficients towards more negative values, while the latter biases against detection of genuine pricing errors. In all regressions, t-statistics use standard errors which adjust for heteroscedasticity. One asterisk denotes significance at the 0.10 level and two asterisks denote significance at the 0.05 level for a two-tailed test. Adjusted R^2's (not reported) are 7% or less. Results for OLO12 (6 observations) are omitted.

of the return — the duration-matched (DM) portfolio, the duration-and-convexity-matched (DCM) portfolio, or the own-model implied return. All but three b-estimates at lag zero are negative, and the three exceptions are insignificant. For lag zero, most of these are also significantly below zero. As expected, the number of significantly negative coefficients drops after introducing a one-day lag, but about half of the t-statistics remain below -2 [14]. All this clearly rejects H_1: $b=0$. Also the hypothesis H_2: $b=-1$ is rejected resoundingly (t-statistics not shown). This leaves us with H_3: there is some information content in the estimated pricing errors, but either part of the so-called error is spurious or the market reacts slowly to such errors.

3.3. Trading rule tests

To obtain an impression of the economic relevance of the predictability of returns on the basis of deviations between observed and model prices, we test a contrarian trading rule: we buy (sell) assets that are deemed to be undervalued (overvalued) [15], and the positions we take become larger the more important the degree of mispricing. The trading rule is tested in calendar time rather than in event time, to detect possible subperiods where the rule worked better than average and to avoid problems with event-time tests when there are long runs of under- or overpricing. (See Bjerring et al. (1983) for a discussion of calendar-time versus event-time tests.)

3.3.1. Design of the test

We only consider OLOs. On any day, we form a portfolio of underpriced bonds (subscript p, short for purchase), a portfolio of overpriced bonds (subscript s, short for sale), weighted by the size of the mispricing ($\mathrm{RES}_{i,t-1-L}$, where L is the implementation delay). For example, if the number of underpriced bonds on day t is N_{pt}, then the mean abnormal return for day t on the purchase portfolio is:

$$\overline{\mathrm{AR}}_{p,t} = \sum_{i=1}^{N_{p,t}} \frac{\mathrm{RES}_{i,t-1-L}}{\sum_{i=1}^{N_{p,t}} \mathrm{RES}_{i,t-1-L}} \mathrm{AR}_{i,t}, \tag{33}$$

[14] The results obtained when the own-model implied return is taken as the benchmark are related to the autocorrelation tests in Table 4. This is because, with the own-model benchmark, the regressand is approximately equal to the change in the regressor. That is, the regression is, approximately, $[\mathrm{RES}_t - \mathrm{RES}_{t-1}]/P_{t-1} = a + b\mathrm{RES}_{t-1}/P_{t-1} + e_t$, so that b is, approximately, unity minus the autocorrelation coefficient of the cross-sectional model residual. Thus, these regression tests confirm the mean reversion (or gradual correction) that was already indicated by the autocorrelations in Table 4.

[15] Thus, trading is based purely on the residuals. We have also implemented a test that incorporates the information in the intercepts of regressions (31) and (32). These intercepts estimate expected abnormal returns assuming perfect pricing the next day. The conclusions of this test are similar to the conclusions reported here.

where $\overline{\mathrm{AR}}_{x,t}$, $x = \{p,s\}$,is the abnormal return on the purchase (sale) portfolio; $N_{x,t}$, $x = \{p,s\}$, is the number of bonds in the purchase (sale) portfolio on day t; $\mathrm{RES}_{i,t-1-L} = P_{i,t-1-L} - \hat{P}_{i,t-1-L}$, the residual for bond i in the day $t-1-L$ cross-sectional term structure model; and $\mathrm{AR}_{i,t}$ = the abnormal return realized between $t-1$ and t, defined relative to the DM portfolio, the DCM portfolio, or the own-model implied return.

The parameter L is varied from 0 to 5 — that is, the delay in trading is varied from zero to five trading days. For $L \geq 1$, there is a delay of at least one day between the decision to trade and the actual implementation, which should eliminate the bid-ask bounce bias that arises for $L = 0$. Similarly, the abnormal return from shortselling the portfolio of overpriced bonds is:

$$\overline{\mathrm{AR}}_{s,t} = -\sum_{i=1}^{N_{s,t}} \frac{\mathrm{RES}_{i,t-1-L}}{\sum_{i=1}^{N_{s,t}} \mathrm{RES}_{i,t-1-L}} \mathrm{AR}_{i,t}. \tag{34}$$

Before implementing the rule, we first verified the validity of the three benchmarks. The duration benchmark is designed so as to yield a zero cross-sectional average abnormal return across all assets — OLOs as well as bank deposits. In this respect, when applied for the duration benchmark, Eq. (34) is similar to the (equally weighted) market model, where by construction the cross-sectional sum of all residuals $\epsilon_{i,t}$ from $\mathrm{HP}_{i,t} = \alpha_i + \beta_i \mathrm{HP}_{m,t} + \epsilon_{i,t}$ is zero every period. However, there is no reason why stock market residuals, when averaged over a non-random subset of assets — say, low-β stocks —, should be zero. In fact, the

Table 6
Abnormal returns on a buy-and-hold portfolio [a]

Benchmarks	351 Trading days		421 Trading days	
	CAR (%)	t	CAR (%)	t
(I) Duration-matched portfolio return	0.50	2.76 **	0.46	1.97 *
(II) DCM portfolio return	−0.45	−2.27 *	0.17	0.26
(III) Vasicek model's expected return	−0.08	−0.21	−0.16	−0.37
(IV) CIR model's expected return	−0.14	−0.35		
(V) Cubic Spline model's expected return	0.07	0.39	0.14	0.71

[a] CARs of the buy-and-hold portfolio of all OLOs are reported for the first period: March 27, 1991–September 16, 1992 (351 trading days) and the whole period: March 27, 1991–December 30, 1992 (421 trading days). Abnormal returns of individual bonds are measured using five alternative benchmarks: (I) the return on a portfolio of OLOs and deposits with the same value and duration; (II) the return on a portfolio of OLOs and deposits with the same value, duration, and convexity; and the return implied by the fitted prices from (III) the Vasicek model, (IV) the CIR model, and (V) the five-parameter Cubic Spline model. The table shows the CARs for an equally weighted portfolio of all OLOs, that is, without the deposits. For the t-ratios, standard errors use the Newey–West correction with 4 lags. One asterisk denotes significance at 0.05 level and two asterisks at 0.01 level for a one-tailed test.

size effect familiar from CAPM tests suggests that an average return computed over a subset of low-β stocks would systematically deviate from zero. Likewise, the cross-sectional average abnormal return computed over OLOs only — the high-duration assets — may deviate systematically from zero. Analogously, for the own-model return benchmark the average pricing error, across deposits and OLOs, is zero at each date, but this does not guarantee that the average return across all bonds is zero. To check for a possible non-zero average "abnormal" return in the benchmark, we computed abnormal returns averaged over all OLOs for each day t, and cumulated them over all days. The results are shown in Table 6, and depicted in Fig. 2. For the three own-model implied return benchmarks, the cumulative abnormal return on the buy-and-hold-all-OLOs portfolio is consistently small, both statistically and algebraically. For the duration benchmark, however, the cumulative abnormal return on a portfolio of all OLOs gradually increases to reach a grand total of 0.50% over 351 days — not enormous in the economic sense, but nevertheless significant from a statistical point of view. For the DCM benchmark, finally, the cumulative abnormal return on the buy-and-hold portfolio of all OLOs after 351 days is significantly negative (at −0.45%). To remove possible bias, we work with a corrected average abnormal return, $\Delta\overline{\mathrm{AR}}$, defined as follows [16]:

$$\Delta\overline{\mathrm{AR}}_{x,t} = \sum_{i=1}^{N_{x,t}} \frac{\mathrm{RES}_{i,t-1-L}}{\sum_{i=1}^{N_{x,t}} \mathrm{RES}_{i,t-1-L}} H_{i,t-1-L}\left(\mathrm{AR}_{i,t} - \sum_{k=1}^{O_t} \frac{\mathrm{AR}_{k,t}}{O_t}\right), \quad x=p,\ s. \tag{35}$$

where O_t = the number of outstanding OLOs at time t; and $H_{i,t-1-L} = +1\ (-1)$ of bond i is underpriced (overpriced) on day $t-1-L$. That is, from the abnormal returns on individual bonds we subtract the corresponding abnormal return from holding an equally weighted portfolio containing all OLOs. This ensures that the modified abnormal returns, when averaged across all OLOs, are now exactly equal to zero on any given day t. Lastly, the average return from the combined trading portfolio (subscript c) is

$$\Delta\overline{\mathrm{AR}}_{c,t} = \frac{\Delta\overline{\mathrm{AR}}_{p,t} + \Delta\overline{\mathrm{AR}}_{s,t}}{2}. \tag{36}$$

If a trading strategy can outperform the naive buy-and-hold portfolio, $\Delta\overline{\mathrm{AR}}_t$ should be positive, on average. To test this, we compute the cumulative average abnormal return, starting from day 1 until day τ:

$$\mathrm{CAR}_{x,\tau} = \sum_{t=1}^{\tau} \Delta\overline{\mathrm{AR}}_{x,t}, \quad x=p,\ s,\ c. \tag{37}$$

[16] $\Delta\overline{\mathrm{AR}}_t$ is set equal to zero if the day-t trading portfolio contains no assets.

where τ is the calendar time, measured in trading days. The t-test is based on the Newey–West standard deviation of $\Delta\overline{\mathrm{AR}}$ corrected for 4th degree autocorrelation.

3.3.2. Validity issues

Conrad and Kaul (1993) discuss three potential pitfalls in tests of contrarian trading rules: compounding of upward bias in asset returns over long holding periods, transaction costs, and bias stemming from bid-ask bounce in the data. In this section, we describe how these three issues are dealt with in our tests.

(1) Upward drift. As we have seen, the returns we use are corrected for the return on a benchmark portfolio — the model's implied normal return, the duration-matched (DM) return, or the Duration and Convexity Matched (DCM) return. Each such benchmark controls for market-wide movements while taking into account also the bond's own characteristics. In addition, we eliminate the remaining average bias that shows up in the subsample of OLOs. This procedure should eliminate most of the potential bias stemming from the compounding of upward drift in asset returns over long holding periods: on any given day, the average cross-sectional abnormal return is exactly equal to zero.

(2) Transaction costs. In this paper we only present gross returns from trading, that is, abnormal returns before transaction costs, for the following reasons. First, although transaction costs are relevant for arbitrage-motivated trades, the level of these costs very much depends on the size of the trade and the capacity of the trader. Accordingly, we follow Fama (1991)'s suggestion and let the arbitrageur decide whether or not the gross returns from arbitrage are larger than the transaction costs. Second, transaction costs are irrelevant if the trade is inspired by exogenous in- or outflows of cash into a bond portfolio; thus, the gross returns will tell us whether it is worthwhile to select bonds on the basis of fitted bond prices (rather than just picking an issue at random) before such a liquidity-inspired trade is made.

(2) Bid-ask bounce. If a last-trade price is a bid (ask) price, the bond is more likely to be classified as being underpriced (overpriced). But the trader has to buy an ''underpriced'' bond at the ask rather than the bid, and the seller likewise trades at the bid rather than the ask. Thus, if it is assumed that the contrarian trader can immediately deal at the last observed price, the computed return will tend to overstate the true return before transaction costs. To deal with this, we introduce lags of one to five days between the decision to trade and the actual implementation of the trade. For example, in the case of a one-day lag, the trader buys at the close of the trading day following the identification of an underpriced bond. The introduction of such a lag will, on average, eliminate the bias stemming from bid-ask bounce under the assumption that the probability that today's last trade is a purchase is independent of whether the previous day's last trade was a purchase or not. There is no a priori reason to doubt this assumption; and direct tests in the US stock market have not rejected this hypothesis (Lehmann, 1990; Ball et al., 1995).

The introduction of lags between the decision to trade and the actual implementation of the transaction is conservative for three reasons. First, although bid-ask

Table 7
CARs from contrarian trading strategies

Lag	Benchmark: DM			Benchmark: DCM			Benchmark: Model prices		
	buy	sell	both	buy	sell	both	buy	sell	both
(A) CARs from trading on the basis of Vasicek model residuals:									
0	**3.74**	**4.16**	**3.95**	**5.29**	**9.63**	**7.46**	**4.69** **	4.88	**4.78**
	(4.52) **	(4.15) **	(4.70) **	(4.68) **	(5.19) **	(5.62) **	(5.67) **	(5.71) **	(6.10) **
1	1.61	3.11	**2.36**	2.33	**6.42**	**4.37**	1.83	2.34	2.09
	(2.80) **	(3.70) **	(3.92) **	(2.49) **	(3.53) **	(3.66) **	(3.54) **	(3.94) *	(4.42) **
2	1.79	2.37	2.08	**2.74**	4.71	**3.72**	**2.09**	**1.97**	**2.03**
	(3.00) **	(2.65) **	(3.23) **	(2.80) **	(2.38) **	(2.79) **	(4.08) **	(3.33) **	(4.21) **
3	1.34	**2.98**	**2.16**	**1.95**	**3.78**	2.87	**1.69**	**1.96**	**1.83**
	(2.26) *	(3.19) **	(3.52) **	(2.12) *	(1.75) *	(2.19) *	(3.10) **	(2.99) **	(3.69) **
4	1.17	**2.26**	**1.72**	**1.65**	5.74	3.70	1.03	1.35	1.19
	(2.29)	(2.41) **	(2.82) **	(2.32) *	(2.60) **	(2.99) **	(2.10) *	(2.18) *	(2.59) **
5	**1.09**	1.96	**1.52**	**1.26**	**5.06**	**3.16**	**1.35**	**1.79**	**1.57**
	(2.01) *	(2.13) **	(2.37) **	(1.57)	(2.53) **	(2.53) **	(2.94) **	(3.27) **	(3.54) **
(B) CARs from trading on the basis of CIR model residuals:									
0	3.19	4.03	3.61	3.93	8.61	6.27	3.65	**5.53**	4.59
	(4.51) **	(4.87) **	(5.35) **	(3.98) **	(4.76) **	(5.17) **	(5.49) **	(6.43) **	(6.77) **
1	1.15	**3.19**	2.17	1.56	5.31	3.43	**1.96**	**2.85**	**2.41**
	(2.42) **	(4.61) **	(4.48) **	(2.20) *	(3.18) **	(3.44) **	(3.87) **	(4.40) **	(4.88) **
2	1.59	**2.67**	**2.13**	2.02	**4.95**	3.48	1.55	1.49	1.52
	(3.11) **	(3.41) **	(3.97) **	(2.52) **	(2.57) **	(2.90) **	(3.20) **	(1.95) *	(2.91) **
3	0.58	2.96	1.77	0.81	3.24	2.03	0.76	1.29	1.03
	(1.02)	(3.35) **	(3.04) **	(0.89)	1.49)	(1.57)	1.47)	(1.86) *	(1.98) *
4	1.01	1.82	1.42	1.24	**6.78**	**4.01**	0.92	**1.55**	**1.24**
	(2.09) *	(2.09) *	(2.68) **	(1.76) *	(3.03) **	(3.34) **	(2.07) *	(2.57) **	(2.97) **
5	0.56	**1.96**	1.26	0.69	4.06	2.37	0.88	1.23	1.05
	(1.29)	(2.33) **	(2.26) **	(1.05)	(2.08) *	(2.05) *	(1.85) *	(1.96) *	(2.22) *

(C) CARs from trading on the basis of five-paremeter spline model residuals:

0	2.91	3.37	3.14	3.55	3.33	3.44	3.40 **	4.73	4.37
	(5.05) **	(5.45) **	(5.89) **	(4.25) **	(3.84) **	(5.06) **	(6.94) **	(7.70) **	(7.72) **
1	1.29	1.82	1.55	1.45	2.09	1.77	1.27	1.41	1.34
	(2.64) **	(3.56) **	(3.81) **	(1.89) *	(2.64) **	(2.27) **	(3.04) **	(2.76) **	(3.14) **
2	1.46	1.53	1.50	1.74	1.41	**1.58**	1.10	1.74	1.42
	(2.90) **	(2.97) **	(3.77) **	(2.05) *	(1.93) *	(2.80) **	(2.36) **	(3.65) **	(3.31) **
3	1.14	1.47	1.30	1.28	1.64	1.46	1.15	1.30	1.06
	(2.29) *	(3.43) **	(3.70) **	(1.31)	(2.53) **	(2.80) **	(1.93) *	(2.82) **	(2.68) **
4	1.13	1.19	1.16	0.34	1.38	0.86	0.84	0.95	0.89
	(2.23) *	(2.65) **	(3.45) **	(0.45)	(2.09) *	(1.81) *	(1.98) *	(1.97) *	(2.24) **
5	0.51	0.82	0.66	0.38	0.64	0.38	0.45	0.32	0.38
	(1.01)	(1.56)	(1.72) *	(0.72)	(0.91) *	(0.72)	(1.11)	(0.68)	(0.98)

(D) CARs from trading on the basis of four-paremeter spline model residuals:

0	3.38	3.48	3.43	3.83	3.63	3.73	4.07	4.78	4.43
	(5.44) **	(5.50) **	(6.04) **	(4.06) **	(4.07) **	(5.00) **	(6.87) **	(7.23) **	(7.34) **
1	**1.66**	2.07	1.86	**2.40**	1.96	2.18	1.45	1.47	1.46
	(3.45) **	(3.95) **	(4.59) **	(2.60) **	(2.40) **	(3.66) **	(3.26) **	(3.24) **	(3.50) **
2	**1.93**	1.59	1.76	**3.13**	1.18	2.15	1.36	1.95	1.66
	(3.69) **	(2.90) **	(4.05) **	(3.62) **	(1.41)	(3.62) **	(2.65) **	(3.43) **	(3.46) **
3	**1.43**	1.70	1.57	**2.07**	2.02	2.05	1.17	1.45	1.31
	(2.86) **	(3.66) **	(4.00) **	(1.91) *	(2.34) **	(3.13) **	(2.65) **	(3.43) **	(3.46) **
4	**1.44**	1.30	1.37	0.87	1.37	1.12	**1.09**	1.44	**1.26**
	(2.75) **	(2.66) **	(3.68) **	(0.83)	(1.53)	(1.58)	(2.49) **	(3.71) **	(3.51) **
5	0.55	1.02	0.79	0.79	1.37	1.12	0.31	0.50	0.41
	(1.24)	(1.80) *	(3.13) **	(0.99)	(0.93)	(1.35)	(0.71)	(1.08)	(1.05)

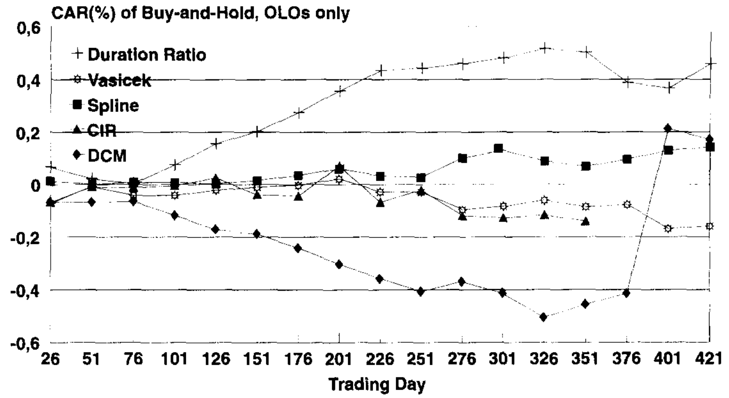

Fig. 3. CARs from contrarian strategies when trading takes place with a one-day lag.

bounce should no longer bias the estimated mean excess return once a delay is introduced, the bounce still boosts the variance of the returns and, therefore, makes it harder to obtain statistically significant results. Second, the longer the delay, the more likely it becomes that the initial mispricing will have partly or wholly disappeared. Thus, our computed results are likely to be inferior to the ones that can be obtained in practice because, in reality, the trader is able to buy or sell at the next opening rather than at the close of the nth next trading day. A last point, related to the second one, is that in our tests the trader is assumed to act upon the initial under- or overpricing signal without considering the current price of the bond that was mispriced n days ago. Thus, with a lag between decision and implementation, our tests will include some trades that would have been deemed unprofitable by a real-world trader because the initial mispricing has already disappeared or has even been reversed.

3.3.3. Results

The results for the Vasicek, CIR, and spline models are reported in Table 7 and shown in Fig. 3. The key findings are as follows:

Notes to Table 7:

[a] In the period March 27, 1991–September 16, 1992 (351 trading days) we trade, in calender time, on the basis of residuals from the daily cross-sectional estimations of four models: Vasicek, CIR, and a cubic spline with five or four parameters. If, in a cross-sectional regression on day $t-1$, the residual is negative (positive) the bond is bought (sold), and at each date the portfolio weights are set proportionally to the size of the mispricing (contrarian weighting scheme). The trade is implemented with a lag that is varied from 0 to 5 trading days. The normal return is either the return on the portfolio matched in terms of value and duration (DM) or duration and convexity (DCM), or the return implied by the model's fitted prices. In addition, the return is corrected for the average bias, across all OLOs, that remains after subtracting each normal return (as described in Table 6). Figures in parentheses are t-ratios, in which standard errors use the Newey–West correction with 4 lags. One (two) asterisk(s) denotes significance at the one-tailed 0.05 (0.01) level. Bold numbers indicate the highest return, across models (that is, for a given strategy (buys, sell, or both, as indicated in the column heading) and benchmark (DM, DCM, model-implied return))

- First, across all four models (Vasicek, CIR, five-parameter spline, and four-parameter spline) and benchmarks (DM, DCM, and own-model implied return), the cumulative abnormal returns in excess of buy-and-hold are positive and significant when there is no delay in trading. The abnormal returns that would be obtained if trading were immediate (at the price that provides the signal) range from 3% to almost 6% over a period of about 351 trading days for the DM and own-model benchmarks, and occasionally up to 10% if convexity is taken into account in the matching portfolios [17].
- Second, about half of this profit disappears if the trade is delayed one working day. It is impossible for us to say to what extent this drop in profits is due to the elimination of the bid-ask bounce bias rather than genuine corrections in the mid-point prices. However, the results for lag = 1 (that is, when trading takes place with a one-day delay) remain significantly positive everywhere. As, in practice, a trader can deal within a shorter delay and with more recent information, we conclude that before-cost profits from bond-picking on the basis of term structure residuals was surely profitable.
- Third, the adjustment in market prices takes time: trading profits remain positive and significant even if the trade is delayed by four or five days after the signal (see lines "lag 2–5" in Table 7). Note also that the trading profits become smaller the longer the delay — that is, market prices and model prices do converge over time. This suggests that all models are to some extent able to detect genuine mispricing.
- Fourth, the abnormal returns that use the own-model implied return as a benchmark are not systematically higher than the abnormal returns computed from the two duration-based models. This suggests that the abnormal returns are not likely to be the result of a circular application of the model.
- Fifth, for any given trading delay and benchmark, the results from trading on the basis of the five-parameter spline model residuals are inferior to the results based on the economic-oriented models: the Vasicek model outperforms its competitors more often than any other model, CIR comes in second, and the five-parameter spline is a distant last. Combined with our earlier finding of a better fit in the cross-sectional estimation, this suggests that the traditional spline model, with its five free parameters and its flexible form, is actually over-fitting the data. [18]

---

[17] We have no clear explanation why the results for the DCM benchmark seem uniformly better. One element may be that, unlike the other benchmarks, the DCM-matched portfolio contains short-term deposits. Also, with three portfolios needed to match a given bond, the matching portfolios contain few assets and are, therefore, more noisy. Lastly, the DCM-matching portfolio does not contain the bond that is being matched; in contrast, duration matching uses the equally weighted market portfolio of all assets (including the mispriced bond), and the own-model benchmark likewise uses all bond prices.

[18] Recall that the two economic models consistently misprice OLOs 03 and 09, and that we assume that the trader never learns from past mistakes. Thus, the results from Vasicek and CIR probably understate the results a real-life trader would have made. This reinforces the conclusion that these models do best.

- Sixth, a substantial part of the overfitting by the five-parameter spline can be avoided by eliminating one free parameter: when the number of knot points is cut down from two to one, the resulting four-parameter spline does a consistently better job than the five-parameter spline, and occasionally even beats the Vasicek and especially the CIR model.
- Lastly, we note that for virtually all models, benchmarks, and lags, the abnormal returns from selling overpriced bonds tend to be higher than the abnormal returns from buying underpriced issues. This suggests that, at least during the test period, short-selling restrictions may have been important in practice. This is not a foregone conclusion: overpricing should quickly disappear if arbitrageurs have sufficient long positions in the bonds that are overpriced, or if there is a sufficiently large flow of liquidity-motivated sales. An alternative explanation of the persistence of overpricing could be taxes on capital gains; but for Belgian corporations such taxes are waived if the transaction is an ''arbitrage'' transaction, that is, if the realized capital gains are reinvested within a short period.

3.4. Filter rule tests

The contrarian weighting scheme assumes that it is optimal to buy (or shortsell) more of a bond the larger the estimated initial mispricing. In this section we verify this assumption empirically, by having the trade decision depend on the size of the initial mispricing. The results will also shed some light on our conjecture that the spline model's better cross-sectional fit is, actually, the result of overfitting.

The test works as follows. We start on day 25 [19]. If, on a given day, an OLO is deemed to be sufficiently overvalued in the sense that its time $t-1$ estimated pricing error is positive and larger than a certain number of basis points (the filter), we short-sell the overvalued bonds. Similarly, if the residual for an OLO is negative and below (minus) the filter size, we say that the bond is sufficiently undervalued, and we buy and add it to the portfolio. For every given filter size, we again report the results for the purchase-rule and shortselling-rule separately as well as pooled. In the pooled results, the filter is symmetric; that is, the percentage overpricing that triggers the sale is the same as the percentage underpricing that triggers a purchase. The amounts invested in each mispriced bond are assumed to be equal, with day-to-day portfolio rebalancing, such that the abnormal return from the portfolio is given by the equally-weighted average abnormal return adjusted for bias on day t, $\Delta\overline{\mathrm{AR}}_t$, over the N_t bonds in the portfolio:

$$\Delta\overline{\mathrm{AR}}_t = \sum_{i=1}^{N_t} \frac{\Delta \mathrm{AR}_{i,t} H_{i,t-1}}{N_t} \tag{38}$$

[19] We lose 24 days at the beginning of the period to compute standard deviations for the average abnormal returns.

where

$\overline{\mathrm{AR}}_t$ = the average abnormal return on day t;

$N_t = \sum_{i=1}^{N_t} |H_{i,t-1}|$ is the number of bonds in the portfolio on day t;

$\mathrm{AR}_{i,t}$ = the abnormal return realized between $t-1$ and t, defined as in either Eq. 28 or Eq. 32;

$$H_{i,t-1} = \begin{cases} +1 \text{ if the bond is underpriced and if the trading rule allows buying;} \\ -1 \text{ if the bond is overpriced and if the trading rule allows shortselling;} \\ 0, \text{ otherwise} \end{cases}$$

As before, the abnormal returns for all benchmarks were corrected for the corresponding abnormal return on the buy-and-hold portfolio of all OLOs. Abnormal returns are then cumulated over time, and t-tests are computed as in Bjerring et al. (1983) [20].

To avoid repetition, Table 8 reports only the results for the best- and worst-performing models (Vasicek and the five-parameter spline), using as benchmarks the DM and own-model implied return. These results can be summarized as follows. First, the underperfermance of trading on the basis of spline model residuals, relative to trading on the basis of the Vasicek model, seems to hold for any given filter size. Thus, the spline model again appears too flexible and, therefore, less able to distinguish mispricing or bid-ask noise from true equilibrium values. A second conclusion from Table 8 is that, when increasing the size of the filter,

[20] If at least one bond is included in the day-t trading portfolio, we trace back the history of the portfolio's average abnormal return (adjusted for bias, as in Eq. (35)) over days $t-24$, $t-23$, ..., $t-5$, and calculate the Newey–West 4th-order autocorrelation adjusted standard deviation, σ_t. $\Delta\overline{\mathrm{AR}}_t$ is then standardized into a Student's variable $Z_t = \Delta\overline{\mathrm{AR}}_t/\sigma_t$ with, under the null hypothesis that the trading rule yields no systematically positive returns, mean zero and standard deviation $\sqrt{20/(20-2)} = 1.0541$. Still under the same null, the statistic $1/\sqrt{T-26}\,\mathrm{SUM}_{t=25}^{T} Z_t/1.0541)$ converges to a unit normal if T is sufficiently large. In this test, $T < 420$ because in some days the trading portfolio is empty.

Notes to Table 8:

[a] If, in a cross-sectional regression on day $t-1$, the residual exceeds the size of the pre-set filter (varied between 0 and 30 basis points), the bond is added to an equally weighted portfolio. The return is corrected by either (1) the return on the duration-matched portfolio of OLOs and deposits or (2) the return implied by the model's fitted prices. In addition, the return is corrected for the average bias, across all OLOs, that remains after subtracting each normal return (as in Table 6). CARs are then cumulated in calendar time. t-statistics are as in Bjerring et al. (1983). * (**) denotes significance at 0.05 (0.01) level, one-tailed. The best result across filters, per column and benchmark, is printed in bold.

Table 8
Profits of filter rules [a]

Fltr.	Buy strategies			Sell strategies			Combined strategies		
(bp)	obs.	CAR	t	obs.	CAR	t	obs.	CAR	t
(A.1) Vasicek residuals — Benchmark: return on DM portfolio:									
0	396	3.23	3.04 **	396	3.22	2.89 **	396	2.90	3.01 **
5	396	3.61	3.36 **	396	4.64	2.98 **	396	3.76	3.19 **
10	396	4.03	2.89 **	396	5.70	3.09 **	396	4.91	3.15 **
15	396	**4.70**	2.47 **	393	**5.91**	4.96 **	396	**6.02**	3.13 **
20	379	2.96	2.47 **	323	3.88	2.90 *	382	4.43	3.58 **
25	271	2.35	2.28 *	205	2.10	1.42	313	3.04	2.70 **
30	183	1.22	0.55	108	1.38	−0.70	226	1.92	0.23
(A.2) Vasicek residuals — Benchmark: return from fitted prices:									
0	396	4.12	3.86 **	396	3.65	3.16 **	396	3.54	3.29 **
5	396	4.63	4.07 **	396	4.92	3.05 **	396	4.51	4.00 **
10	396	5.60	4.67 **	396	5.73	2.75 **	396	5.99	3.30 **
15	396	**6.26**	3.19 **	393	**6.48**	4.97 **	396	**7.30**	4.28 **
20	379	3.96	5.06 **	323	4.84	7.70 **	382	6.17	6.65 **
25	271	2.48	2.78 **	205	2.38	5.81 **	313	3.51	3.26 **
30	183	1.36	1.07	108	1.60	2.17 *	226	2.27	1.32
(B.1) Spline residuals — Benchmark: return on DM portfolio:									
0	396	2.36	2.35 **	396	2.50	2.75 **	396	2.29	2.73 **
5	396	**3.38**	2.44 **	396	3.34	1.85 *	396	3.13	2.82 **
10	396	3.10	−6.99 **	396	**4.98**	2.76 **	396	**4.43**	2.33 **
15	312	1.34	1.97 *	334	3.21	3.14 **	352	3.16	1.30
20	121	1.14	0.50	191	1.58	1.18	221	1.65	−0.37
25	42	0.63	2.07 *	36	0.91	1.85 *	45	0.99	1.35
30	26	−0.09	0.22	23	0.45	0.00	26	0.05	0.29
(B.2) Spline residuals — Benchmark: return from fitted prices:									
0	396	3.34	4.48 **	396	3.26	3.45 **	396	3.10	3.29 **
5	396	**4.71**	5.30 **	396	4.95	2.66 **	396	4.54	3.08 **
10	396	3.92	4.02 **	396	**5.36**	3.06 **	396	**5.12**	2.96 **
15	312	1.80	3.20 **	312	3.03	2.39 **	352	3.21	1.96 *
20	121	1.40	1.18	121	1.62	1.43	221	2.12	1.98 *
25	42	0.88	3.21 **	36	0.74	0.87	45	1.15	1.60
30	26	0.05	0.21	23	0.28	0.00	26	0.05	0.26

profits tend to go up first but then tend to go down. Thus, the contrarian weighting scheme — which places greater emphasis on bonds that are deemed to be highly mispriced — is not optimal. The finding that very large residuals lead to lower average profits suggests that, for all models, large residuals are more likely to be the result of model mis-specification or -estimation rather than mispricing. Third, we find that the optimal filters tend to be smaller for the spline model than for the Vasicek model. Conversely, large residuals from the spline model (which, one may recall, are also relatively rare) are even more suspect, on average, than large residuals from the Vasicek models.

4. Conclusions

We estimate daily Vasicek/CIR bond models on BEF government bonds and interbank deposits, 1991/1992. The Vasicek model produces slightly larger MSE's than the CIR model, but the results are otherwise very similar. The five- and four-parameter cubic spline models, on the other hand, easily beat the two economic models in terms of average fit. Regression tests reveal that part of the deviation between observed price and model price are reversed the next day, and also the second day after the observation of the initial mispricing. This suggests that the estimated residuals do reflect genuine pricing errors, not just model mis-specification or mis-estimation and bid-ask bounce bias. After correction for market-wide changes, a strategy of buying underpriced bonds or (especially) selling overpriced bonds turns out to be profitable, yielding a significant 3–9% more, over eighteen months, than a buy-and-hold bond portfolio. The best results are obtained if trading is based on the Vasicek and CIR models. The traditional five-parameter spline model, being more flexible, seems to overfit the data and is, therefore, less able to detect mispricing; but the spline's performance can be improved by cutting the number of knot points down to one. Lastly, large model residuals are more likely to be the result of model misspecification or -estimation than are small or medium-sized residuals. Thus, it is better not to adopt a contrarian strategy of increasing one's stake in a bond the greater its degree of mispricing.

Acknowledgements

We are grateful to Ray Ball, Stan Beckers, Nai-Fu Chen, Jin-Chuan Duan, Leora Klapper, Michelle Lee, Peter Schotman, Raman Uppal, Cynthia Van Hulle, and Lambert Vanthienen, and other participants in workshops at K.U. Leuven, the 1994 EAA Doctoral Tutorial, the 1995 EFA Conference, and the City University of Hong Kong, and especially to two anonymous referees, whose comments helped us to improve both the contents and the presentation. All remaining errors are our sole responsability.

References

Ball, R., S.P. Kothari and C.E. Wasley, 1995, Can we implement research on stock trading rules?, Journal of Portfolio Management 21, 54–63

Bjerring, J.H., J. Lakonishok and T. Vermaelen, 1983, Stock prices and financial analysts' recommendations, Journal of Finance 38, 187–204.

Black and Scholes, 1973, The pricing of options and corporate liabilities, Journal of Political Economy 81, 637–659.

Brown, R.H. and S.M. Schaefer, 1994, The term structure of real interest rates and the Cox, Ingersoll, and Ross model, Journal of Financial Economics 35, 3–42.

Brown, S. and P. Dybvig, 1986, The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates, Journal of Finance 41, 617–630.

Conrad, J. and G. Kaul, 1993, Long-term market overreaction or biases in computed returns?, Journal of Finance 48, 39–63.

Cox, J.C., J. Ingersoll and S. Ross, 1985b, A theory of the term structure of interest rates, Econometrica, 53, 385–407.

De Munnik, J. and P. Schotman, 1994, Cross sectional versus time series estimation of term structure models: Empirical results for the Dutch bond market. Journal of Banking and Finance 18, 997–1025.

Elton, E.J. and M.J. Gruber, 1991, Modern Portfolio Theory and Investment Analysis, 4th ed. (Wiley, New York).

Fama, E.F., 1990, Term-structure forecasts of interest rates, inflation, and real returns, Journal of Monetary Economics 25, 59–76.

Fama, E.F., 1991, Efficient markets: II, Journal of Finance 46, 1575–1617.

Grégoire, Ph. and I. Platten, 1995, An empirical study of general equilibrium models on the Belgian bond market, Working Paper (Facultés Universitaires Notre-Dame de la Paix, Namur).

Lehmann, B.N., 1990, Fads, martingales, and market efficiency. Quarterly Journal of Economics, 1–28.

Longstaff, F., 1992, Multiple equilibria and term structure models, Journal of Financial Economics, 32, 333–344.

Longstaff, F. and E. Schwartz, 1992, Interest-rate volatility and the term structure: A two-factor general-equilibrium model, Journal of Finance 47, 1259–1282.

Merton, R.C., 1973, Theory of rational option pricing, Bell Journal of Economics and Management Science 4, 141–183.

Pearson, N.D. and T.S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, Ross model, Journal of Finance 49, 1279–1304.

Platten, I., 1994, Non-linear general equilibrium models of the term structure: Comments and two-factor generalization, Finance 15, 63–78.

Reilly, F.K. and R.S. Sidhu, 1980, The many uses of bond duration, Financial Analysts Journal 46, 59–72.

Shea, G., 1985, Interest term structure estimation with exponential splines: A note, Journal of Finance 40, 319–325.

Vasicek, O., 1977, An equilibrium characterization of the term structure, Journal of Financial Economics 5, 177–188.

Vasicek, O. and W. Fong, 1982, Term structure modeling using exponential splines, Journal of Finance 37, 339–348.