Direct Estimation of the risk neutral factor dynamics of affine term structure models

Citation for published version (APA):

Document status and date:
Published: 01/01/2003

DOI:

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Direct estimation of the risk neutral factor dynamics of Gaussian term structure models

Dennis Bams[a,b], Peter C. Schotman[a,c,*]

[a]*Limburg Institute of Financial Economics, University of Maastricht, P.O. Box 616, Maastricht 6200 MD, Netherlands*
[b]*ING-Corporate Reinsurance, P.O. Box 810, 1000 AV Amsterdam, Netherlands*
[c]*Center for Economic Policy Research (CEPR), London, UK*

Abstract

This paper proposes panel data tests of Gaussian affine term structure models. Yield curve data for different moments in time are pooled with the factors treated as fixed effects. With fixed effects the time series properties of the price of risk can be ignored. Results of tests with US interest rate data show that the Gaussian model is able to capture the cross sectional structure of yields as well as unrestricted factor loadings from principal components analysis. However, estimates of the mean reversion parameters in a 3-factor model differ significantly when the model is estimated from yield levels or forward differences, which is inconsistent with the Gaussian model.
© 2003 Elsevier B.V. All rights reserved.

1. Introduction

Models of the term structure of interest rates derive the shape and movements of the yield curve from assumptions about the dynamics of some underlying factors and the price of the risk associated with each factor. Together these assumptions define the risk adjusted (or risk neutral) dynamics of the factors, which is the central object of

---

* Corresponding author. LIFE, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, Netherlands. Fax: +31-43-388-4875.

E-mail address: p.schotman@berfin.unimaas.nl (P.C. Schotman).

interest in applications to interest rate derivatives. A limited number of underlying state variables are assumed to account for the behavior of interest rates of many different maturities. The purpose of this paper is to estimate the dynamic specification of the risk neutral factor process associated with a Gaussian term structure model from a panel of observed term structures without using the actual time series dynamics of the factors, and to test the restrictions of the Gaussian model.

The Gaussian model is a member of the affine class. The affine class has become popular because it is a flexible multifactor model that offers tractability both in derivative pricing and for parameter estimation.[1] The actual factor dynamics of the Gaussian model is a multivariate diffusion with affine drift and constant variance. If in addition the price of risk is constant, it turns out that yields on discount bonds of all maturities are affine functions of the factors, with time invariant intercepts and factor loadings. The single factor version of the Gaussian model is the well known Vasicek (1977) model.

Both assumptions of the Gaussian model are at odds with the data. First, the literature on interest rate dynamics has found stylized facts like nonlinear drift, regime switching and complicated volatility processes.[2] Second, term premia show substantial time variation.[3] These findings lend little empirical support to the Gaussian and other affine models. Dai and Singleton (2000, DS) reject three-factor specifications within a general framework. Other empirical tests are usually concerned with special cases of the models of DS, and therefore subject to the same empirical problems.[4] Ghysels and Ng (1998) reject the model in a semiparametric analysis with observed factors.

These tests are joint tests of the actual time series process and the model for risk prices. The empirical evidence only rejects the subclass of what Duffee (2002) has labelled *completely* affine models. Duffee (2002) and Duarte (2000) propose different specifications for the price of risk. Duffee (2002) shows that much variation in risk prices is possible without creating arbitrage opportunities. Dai and Singleton (2002) and Duffee (2002) show that Gaussian models with this type of time varying risk premia have much better empirical properties. In Duarte (2000) the drift function of the factors under the actual probability measure is no longer required to be affine. But factor processes under the risk neutral probability measure remain affine, so that tractability is maintained. With time varying risk prices the relation between the time series properties of interest rates and the various possible shapes of the yield curve becomes much looser.

Misspecification of the time series part of the model could lead to rejections, even though the cross sectional implications are valid. Our approach focuses on the cross sectional implications of the model. To avoid making any assumptions on the actual time series properties of the factors, we treat them as fixed effects, i.e. as parameters to

---

[1] See Duffie and Kan (1996) and Frachot and Lesne (1993) for a detailed theoretical analysis.

[2] See for example Chan et al. (1992), Andersen and Lund (1997), and Koedijk et al. (1997) for volatility dynamics. Aït-Sahalia (1996), Conley et al. (1997) and Stanton (1997) provide evidence for nonlinear dynamics. See for example Hamilton (1988) and Ang and Bekaert (2002) for regime switching.

[3] See Fama (1984a,b) and Fama and Bliss (1987) for the early empirical evidence on time varying risk premia. See Campbell et al. (1997, chapter 10) for a survey.

[4] See the references in DS, e.g., Chen (1996) and Balduzzi et al. (1996).

be estimated. To do so, we use a complete panel of interest rates with many different maturities.

Our estimation approach differs from earlier studies that have tested term structure models using cross sectional data.[5] These papers consider single factor models, for which the structural parameters are re-estimated every time period using the bond prices at that moment. In this paper we pool yield curves in a panel, keeping the structural parameters constant over time.

Our fixed effects model also contrasts with other panel data studies of affine term structure models.[6] In these studies the cross sectional yield curve model is estimated jointly with a time series process for the factors. Since we do not make any assumption on the time series properties of the risk prices, any misspecification in the fixed effects model will be present a fortiori in models that combine time series and cross sectional information.

In the remainder of this introduction we provide an outline of the paper. Section 2 reviews the Gaussian term structure model. The risk neutral dynamics are most relevant in applications to derivative pricing. Rather than modeling the levels of the yield curve, these applications emphasize models of the yield curve relative to last period's term structure. Let $Y_t(\tau)$ be the yield on a zero-coupon bond with maturity τ, and $\tilde{Y}_t(\tau) = Y_t(\tau) - F_{t-1}(\tau)$ the yield in deviation from last period's forward rate. The Gaussian model implies that the k-vectors of factor loadings $(b(\tau), \tilde{b}(\tau))$ in

$$Y_t(\tau) = a(\tau) + b(\tau)'z_t + v_t(\tau), \tag{1}$$

$$\tilde{Y}_t(\tau) = \tilde{a}(\tau) + \tilde{b}(\tau)'\tilde{z}_t + \tilde{v}_t(\tau), \tag{2}$$

are identical, although the factors $(z_t, \tilde{z}_t)$, intercepts $(a(\tau), \tilde{a}(\tau))$ and measurement errors $(v_t(\tau), \tilde{v}_t(\tau))$ are different for the two specifications. Since levels and deviations from forward rates emphasize very different moments of the data, the implied equality of the factor loadings suggests a powerful test of the Gaussian model.

The test is performed in two stages. First, we consider (1) on its own. Using N maturities we estimate the Gaussian function $b(\tau)$ together with an unrestricted alternative **B**, treating the factors z_t as fixed effects. The first stage test examines how well $b(\tau)$ fits **B**. We repeat this procedure for (2) to obtain a Gaussian function $\tilde{b}(\tau)$ and unrestricted **B̃**. In the second stage we test the equality $b(\tau) = \tilde{b}(\tau)$.

Section 3 discusses the econometric specification. Since our tests are about the cross sectional structure of the model, we prefer a fairly large N. An important part of the specification are the properties of the error terms in a no-arbitrage model that is supposed to be exact. We assume, as in Jacquier and Jarrow (2000), that the errors are small and uncorrelated with the factors. We leave the errors unspecified otherwise. We prefer this option over the alternatives of a tight parametric structure or the introduction of a large number of nuisance parameters.

---

[5] See, among others, Brown and Dybvig (1986), Brown and Schaefer (1994), DeMunnik and Schotman (1994), and Sercu and Wu (1997).

[6] See for example Chen and Scott (1993), Gibbons and Ramaswamy (1993), DeJong (2000), Geyer and Pichler (1999), Babbs and Nowman (1999), Duan and Simonato (1999), Gong and Remolona (1996).

To perform the second stage test, Section 4 follows ideas proposed by Driscoll and Kraay (1998) in the spatial econometrics literature. Their method delivers a $\sqrt{T}$ consistent estimate of the asymptotic covariance matrix of the structural parameters in the function $b(\tau)$ that is robust against both time series and cross sectional correlation and heteroskedasticity, without requiring an estimate of the complete $(N \times N)$ cross sectional covariance matrix.

Section 5 considers the unrestricted alternatives **B** and **B̃**. We explain that in the fixed effects model these are nothing but the factor loadings obtained from a principal components analysis.

The model is tested with data on zero discount yields constructed from US government bonds. Section 6 describes the data and the construction of yield curves. As a preview of the empirical results in Section 7, we find that the estimated functions $(b(\tau), \tilde{b}(\tau))$ are very similar to the unrestricted principal component factor loadings. On the other hand, the formal test rejects the equality of $b(\tau) = \tilde{b}(\tau)$. Section 8 concludes.

2. Gaussian term structure models

Gaussian term structure models are a subclass of the affine models analysed in detail by Duffie and Kan (1996). Yield curves in an affine model have the general structure

$$Y_t(\tau) = a(\tau) + b(\tau)'z_t, \tag{3}$$

where $Y_t(\tau)$ is the yield at time t on a discount bond maturing at time $t + \tau$; z_t is a k-vector of factors; $a(\tau)$ is a scalar and $b(\tau)$ a k-vector of functions of τ. Both $a(\tau)$ and $b(\tau)$ are determined by the parameters that describe the risk neutral time series process of the factors in continuous time. For the Gaussian case the risk neutral dynamics can be written as

$$\mathrm{d}z = \mathscr{K}(\Theta - z)\,\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W, \tag{4}$$

where Θ is k-vector of constants, W is a k-dimensional Brownian motion under the risk-adjusted probability measure, Σ is the $(k \times k)$ instantaneous covariance matrix of the factors and $\mathscr{K}$ is a $(k \times k)$ diagonal matrix with elements κ_i. Despite the diagonal structure for $\mathscr{K}$, the above formulation is the most general Gaussian model that can be identified. In Appendix A it is shown that the model is equivalent to the maximally identified canonical $\mathbb{A}_0(k)$ model in DS.

For pricing of interest rate derivatives the function $b(\tau)$ in (3) is the primary object of interest. Interest rate derivatives can be priced with as inputs the "volatility" function $b(\tau)$, the factor covariance matrix Σ, and the initial term structure.[7] The form of the functions $b(\tau)$ in a Gaussian term structure model is

$$b_i(\tau) = \frac{1 - \mathrm{e}^{-\kappa_i \tau}}{\kappa_i \tau}, \quad i = 1, \ldots, k. \tag{5}$$

The loadings $b_i(\tau)$ for factor i depend solely on the mean reversion of factor i. The factor loadings are decreasing in maturity and satisfy $0 < b_i(\tau) \leqslant 1$ with $b_i(0) = 1$

---

[7] See for example Appendix 22A in Hull (2000) for the detailed valuation of an option on a discount yield in the two-factor Gaussian model.

and $b_i(\infty) = 0$. The factor loadings determine how yields of different maturities move relative to each other over time. Together with Σ they fully describe the covariance structure of the entire yield curve.

The intercept $a(\tau)$ is constant over time and affects the shape of the average yield curve. In this paper we are concerned with the covariance structure of yield curve movements and will not test theoretical restrictions on $a(\tau)$. Neither do we explore the actual time series properties of the factors. When risk prices are time varying, the time series mean reversion could be different from $\mathcal{K}$.

Model (3) relates the factors to the level of discount yields. In much of the options literature one is more interested in movements of the yield curve relative to the current yield curve.[8] To relate to that literature we define the forward differences $\tilde{Y}_t(\tau) = Y_t(\tau) - F_{t-1}(\tau)$, where

$$F_{t-1}(\tau) = \frac{\tau+1}{\tau} Y_{t-1}(\tau+1) - \frac{1}{\tau} Y_{t-1}(1) \tag{6}$$

is the forward rate at time $t-1$ for the yield on a discount bond starting at time t and maturing at time $t+\tau$. In Appendix A it is shown that the forward differences have the same factor structure

$$\tilde{Y}_t(\tau) = \tilde{a}(\tau) + b(\tau)'\tilde{z}_t, \tag{7}$$

but with an intercept $\tilde{a}(\tau)$ and factors $\tilde{z}_t$ that differ from (3). Estimation of (3) and (7) emphasize different moments of the data.[9] By requiring that $b(\tau)$ must be the same in both models we obtain overidentifying moment conditions for a test of the Gaussian model.

The factor structure in (7) is the same as in the extended Vasicek model in Hull and White (1993), often used for pricing term structure derivatives. Their derivation of the extended Gaussian model requires less restrictive assumptions about the form of the initial forward rate curve. In deriving (7) we assumed that $F_{t-1}(\tau)$ is itself determined by the Gaussian model, whereas in Hull and White (1993) it is given exogenously.

The Gaussian model differs from the other members of the affine class by having a constant factor covariance matrix Σ. In other affine models the covariance matrix Σ depends on one or more of the factors. DS show that this form of stochastic volatility leads to admissibility restrictions in order to keep the covariance matrix positive definite for all possible realizations of z_t. The most important of these restrictions is that non-Gaussian factors must be positive and positively correlated among themselves. DS and Campbell (2000) find that much of the empirical improvement in their models comes from allowing negative correlations among at least two of the factors.

In our parameterization of the Gaussian model there are no restrictions on z_t or its covariance structure. The factors can therefore be treated as unrestricted parameters. Treating z_t as parameters in other, non-Gaussian, affine models leads to many inequality

---

[8] See, e.g., Hull (2000) for a textbook treatment and further references.

[9] Many empirical studies work with excess holding period returns. This leads to an equivalent expression, since the excess holding period return $R_t(\tau) = -\tau\tilde{Y}_t(\tau)$ is a scaled version of the forward difference. In the econometric analysis the scaling implies different cross sectional heteroskedasticity.

restrictions that greatly complicate estimation and testing. This paper therefore focuses exclusively on the Gaussian model.

Although the Gaussian model seems the most flexible for fitting yield curves, rejection of the Gaussian model does not imply that we can reject all affine models. Different submodels of the affine class are not nested. Candidate non-Gaussian affine models can have different number of factors and have a functional form for the factor loadings that differs from (5).[10]

3. Specification

For time t we define y_t as an N-vector of discount yields $y_{it} = Y_t(\tau_i)$ $(i = 1, \ldots, N)$. We further define the $(N \times k)$ matrix B containing the factor loadings $b_{ij} = (1 - e^{-\tau_i \kappa_j})/\tau_i \kappa_j$, and an N-vector a with unrestricted intercepts a_i. Adding an N-vector of error terms v_t our econometric specification for testing the Gaussian model is the panel,

$$y_t = a + Bz_t + v_t, \quad t = 1, \ldots, T, \tag{8}$$

where z_t is the k-vector of factors. We assume that the errors have zero mean and cross sectional covariance matrix Ψ, with elements $\psi_{ij} = \mathsf{E}[v_{it}v_{jt}]$. The model for forward differences is analogous: replace y_t, a, z_t, v_t and Ψ by $\tilde{y}_t$, $\tilde{a}$, $\tilde{z}_t$, $\tilde{v}_t$ and $\tilde{\Psi}$. Only B should be the same in both formulations. Our primary interest is in the functional form of the factor loadings, which depend on the parameter vector $\kappa = (\kappa_1, \ldots, \kappa_k)'$.

For identification we assume that z_t has mean zero. We also assume that the unconditional factor covariance matrix $\Omega = \mathsf{E}[z_t z_t']$ exists and has rank k. The main difference with most other term structure panel studies is that we treat the factors z_t as fixed effects in order to avoid assumptions about the actual time series properties.

The remainder of this section discusses two important aspects in the specification of (8): data and error terms. Below we discuss the levels model in detail. We comment on the differences with the model for forward differences when necessary.

3.1. Data

The panel model (8) is specified with discount yields as the dependent variable. However, discount yields for longer maturities are not directly observable, but must be estimated from raw data on coupon bonds or swap rates. Our main motivation for applying a first stage estimator that converts the data to discount yields is computational. Coupon bond prices and swap rates are nonlinear functions of the discount yields and therefore nonlinearly related to the factors z_t. When the time varying factors z_t cannot be concentrated out analytically, the nonlinear least squares algorithm needs to optimize

---

[10] Dai and Singleton (2002, p. 423) discuss the same issue: "Since (...) the family of Gaussian models (family $\mathbb{A}_0(3)$) gives the most flexibility to the structure of factor correlations and conditional means, one might conjecture a priori that these models would perform at least as well as other affine models." However, "(...) the dimensionality of our models and the richness of the moments used in ML estimation seem to preclude more precise theoretical statements (...)".

over Tk parameters simultaneously, which is a formidable numerical task in a panel with moderately large T.[11]

We construct the yield curve non-parametrically from coupon bond prices using cubic splines as in McCulloch (1975) and McCulloch and Kwon (1993). The result is a smooth function $Y_t(\tau)$ of τ. The yield curve is evaluated at maturities $\tau_i = i\Delta_N$ ($i = 1, \ldots, N$) with interval length $\Delta_N = \tau_{max}/N$. Increasing the number of observations means that we take an ever finer grid of maturities τ_i without changing the endpoint τ_{max}. The cross sectional dimension N is therefore arbitrarily large, but the data information is limited by the raw data from which the yield curve is constructed and the flexibility of the functional forms used to estimate the yield curve. For inference we therefore rely on asymptotics that are valid for fixed N and large T. Section 6 contains further details about the data and the construction of the yield curve.

3.2. Error terms

The introduction of error terms is always slightly uneasy in asset pricing models that have been derived with no-arbitrage arguments, since they are inconsistent with the assumptions of the no-arbitrage model. DS, and many others in the empirical literature, avoid error terms by setting the number of factors equal to the number of maturities ($N = k$). With this assumption the factors can be recovered exactly from the data by inverting (3). The parameters are identified by the time series properties of the factors and tests of the model rely on strong additional assumptions on the price of risk. Since our aim is to test the cross sectional specification of the model, without relying on auxiliary assumptions on the price of risk, we need to include more maturities than factors. This necessarily leads to a specification that includes error terms.

Jacquier and Jarrow (2000) discuss the properties of error terms in no-arbitrage models. They distinguish between measurement errors and model errors. As discussed above, measurement error arises from the construction of discount yields from coupon bond prices. Other sources of error in the yield curve consist of bond price errors (bid/ask effects), tax effects, liquidity and transactions costs, none of which are considered in developing the theoretical term structure model. According to Jacquier and Jarrow (2000) the combined effect of the two types of errors should (i) be small, (ii) be unrelated to observable model inputs or variables, and (iii) have similar properties in different samples. Condition (i) is obvious for the model to have any credibility as a theoretically error-free no-arbitrage model. The size of the errors, measured as root mean-squared error for each maturity τ_i, is an important criterion in assessing the empirical results. Regarding (ii), it will be reasonable to assume that the errors are independent of the factors, i.e. $\mathsf{E}[v_t z'_{t-s}] = 0$ for all s. As for property (iii) we assume that the time series v_t is stationary. Empirically, we compare results for different subperiods.

---

[11] For single factor models pooled over short time periods it is numerically feasible to estimate the Gaussian (and also the CIR) model directly with coupon bond prices. See for example DeMunnik and Schotman (1994) for weekly pooled estimates, and Brown and Dybvig (1986) and Brown and Schaefer (1994) for period by period nonlinear least squares.

Since the error term deals with unmodeled phenomena, there are no other apparent economic restrictions to guide the specification of an econometric model for the error term. Yields of all maturities are subject to measurement error. Since both the yield curve and the fitted values $b(\tau)'z_t$ are smooth functions of τ, we expect that the errors v_{it} will also be a smooth function of τ, leading to strong positive cross sectional correlation between neighboring error terms v_{it} and $v_{i+1,t}$. This is confirmed by the results of De Jong (2000) and Lamoureux and Witte (2002). [12]

From the results of, for example, Chen and Scott (1993) and De Jong (2000), we also expect strong positive time series autocorrelation of the errors in the model for yield levels. Forward differences data do not exhibit much autocorrelation. Some negative autocorrelation could arise, since at longer maturities the measurement error in $\tilde{y}_{it}$ is the difference of measurement error at times t and $t - 1$.

In preliminary work (Bams and Schotman, 1998) we experimented with simple parametric structures assuming first order autocorrelation in both τ_i and t, and cross sectional heteroskedasticity depending on τ_i, but had to conclude that the true structure is much more complex. [13] In further preliminary analysis of a 3-factor model we also learned that with yield levels (instead of log prices) as the dependent variable, the residual variances were of the same order of magnitude for all maturities. Similarly, by working with forward differences (instead of excess returns), we also obtained residuals that are of the same order of magnitude over the entire maturity range. Therefore v_{it} are cross sectionally nearly homoskedastic. Lacking plausible assumptions, we decided to leave the error process unspecified. We give up some efficiency, but we avoid problems related to poorly estimated or misspecified error structures. Since the errors are small (recall condition (i) of Jacquier and Jarrow, 2000), even an inefficient estimator is expected to be precise enough to allow meaningful inference on κ. Empirically, this appears the case (see Section 7).

4. Estimation and testing

We estimate κ by ordinary least squares. To eliminate the fixed effects z_t and a we take all yields in deviation of their time series average $\bar{y} = (1/T)\sum_t y_t$ and premultiply (8) by the matrix $M = I - B(B'B)^{-1}B'$,

$$M(y_t - \bar{y}) = M(v_t - \bar{v}) = \hat{v}_t. \tag{9}$$

The least squares criterion for estimating κ from (9) is

$$s^2(\kappa) = \frac{1}{2NT} \sum_t \hat{v}_t'\hat{v}_t = \frac{1}{2N} \operatorname{tr}(MV), \tag{10}$$

---

[12] This contrasts with many studies, for example Jegadeesh and Pennacchi (1996), Babbs and Nowman (1999) and Brandt and He (2002), that assume that measurement error is uncorrelated, both in t and τ. That assumption is often made for computational reasons, since they jointly estimate the factor dynamics and the measurement equations.

[13] In many cases we also had great difficulty with the numerical optimization of the quasi-likelihood function.

where V is the $(N \times N)$ sample covariance matrix of y_t. Differentiating (10) with respect to κ gives the first order conditions. Using the calculations in Appendix B the first order conditions are written explicitly as the orthogonality conditions

$$\frac{1}{T} \sum_t \hat{z}_t \odot \hat{u}_t = 0, \tag{11}$$

where the dot-product operator $\odot$ denotes elementwise multiplication of two matrices, $\hat{z}_t = (B'B)^{-1}B'(y_t - \bar{y})$ are estimates of the factors, and $\hat{u}_t = (1/N)C'_\perp(y_t - \bar{y})$ is a vector of k linear combinations of the data with $C_\perp = MC$ and

$$C = \left(\frac{\partial b_1}{\partial \kappa_1}, \ldots, \frac{\partial b_k}{\partial \kappa_k} \right),$$

an $(N \times k)$ matrix that contains the derivatives of the columns in B with respect to the mean reversion parameter of that column. The system (11) contains k equations that must be solved numerically to obtain the least squares estimator $\hat{\kappa}$.

Let κ_0 be the true value of κ. Evaluating the matrices B and C at $\kappa = \kappa_0$, and letting $T \to \infty$ for fixed N, the orthogonality condition (11) converges to (see Appendix B)

$$\text{plim} \frac{1}{T} \sum_t \hat{u}_t \odot \hat{z}_t = \frac{1}{N} \, \text{dg}((B'B)^{-1}B'\Psi C_\perp), \tag{12}$$

where $\text{dg}(A)$ is a vector containing the main diagonal of A. Since for general error covariance matrix Ψ the result is not zero, the least squares criterion (10) is not minimized at $\kappa = \kappa_0$. Hence the OLS estimator $\hat{\kappa}$ is not consistent for large T and fixed N. Without assumptions on Ψ the estimator is still not consistent when $N \to \infty$, since $B'\Psi C_\perp$ is of order $\text{O}(N^2)$ for general Ψ. The nature of the inconsistency is an identification problem. A general full rank cross sectional error covariance matrix Ψ is not identified due to the loss of k degrees of freedom from the fixed effects z_t. [14] For a consistent estimator we must assume that Ψ is such that (12) is zero. Condition (12) states that estimation error in z_t, equal to $(B'B)^{-1}B'v_t$, must be uncorrelated with $\frac{1}{N}C'_\perp v_t$. Many studies, for example Chen and Scott (1993) and Duffee (2002), make the stronger assumption that k out of N yields are observed without error, so that the factors can be estimated without error. [15]

For the computation of standard errors we follow a suggestion of Driscoll and Kraay (1998) for spatial data with cross sectional dependence. Their method does not require an estimate of the $(N \times N)$ cross sectional error covariance matrix Ψ, but just the k-dimensional scores $h_t(\kappa) = \hat{z}_t \odot \hat{u}_t$, which are cross sectional averages with only a time subscript. It follows from standard $\sqrt{T}$ asymptotics that

$$\sqrt{T}(\hat{\kappa} - \kappa_0) \sim N(0, H^{-1}SH^{-1}), \tag{13}$$

---

[14] Identification of an unrestricted Ψ, as in DeJong (2000) or Lamoureux and Witte (2002), requires a model for the dynamic process of the factors. That is exactly what we avoid in this paper.

[15] Even without this assumption the asymptotic bias is likely to be negligible when Ψ is small relative to Ω, as it should under property (i) of Jacquier and Jarrow (2000). We conclude this from a linear approximation of the first order conditions (11) around $\kappa = \kappa_0$.

where H is the $(k \times k)$ matrix of second order derivatives of $s^2(\kappa)$, and

$$S = \mathsf{E}\left[\left(\frac{\sum_t h_t(\kappa_0)}{\sqrt{T}}\right)\left(\frac{\sum_t h_t(\kappa_0)}{\sqrt{T}}\right)'\right]. \tag{14}$$

The covariance matrix S is estimated using the Newey-West procedure as

$$\hat{S}_T = W_0 + \sum_{\ell=1}^{L}\left(1 - \frac{\ell}{L+1}\right)(W_\ell + W_\ell'), \tag{15}$$

$$W_\ell = \frac{1}{T}\sum_{t=\ell+1}^{T}(\hat{z}_t\hat{z}_{t-\ell}') \odot (\hat{u}_t\hat{u}_{t-\ell}'). \tag{16}$$

The estimator and standard errors were derived for the model on yield levels. The same methods are also directly applicable for the forward differences with parameter vector $\tilde{\kappa}$ and scores $\tilde{h}_t$.

For a Wald test of the equality of $b(\tau)$ in the levels and forward differences model, we need the joint covariance matrix of $\hat{\kappa}$ and $\hat{\tilde{\kappa}}$. Let $\mathbf{h}_t' = (h_t(\hat{\kappa})'\tilde{h}_t(\hat{\tilde{\kappa}})')$ be the combined score vector, and let $\hat{\mathbf{S}}_T$ be an estimate of the $(2k \times 2k)$ covariance matrix of the scores, obtained analogously to (15) above. Let $\mathbf{H} = \text{diag}(H, \tilde{H})$ be the $(2k \times 2k)$ Hessian, and $\mathbf{R} = (I - I)$ the $(k \times 2k)$ matrix of restrictions. The Wald statistic is computed as

$$W = T(\hat{\kappa} - \hat{\tilde{\kappa}})'(\mathbf{R}\mathbf{H}^{-1}\hat{\mathbf{S}}_T\mathbf{H}^{-1}\mathbf{R}')^{-1}(\hat{\kappa} - \hat{\tilde{\kappa}}). \tag{17}$$

Under the null the test will be asymptotically chi-squared with k degrees of freedom.

5. Principal components

The least squares function (10) can be expanded as

$$s^2(\kappa) = \frac{1}{N}\,\text{tr}(MV) = \frac{1}{N}(\text{tr}(V) - \text{tr}((B'B)^{-1}B'VB)). \tag{18}$$

Minimizing (18) with respect to κ is equivalent to maximizing the second term. Let λ_i $(i = 1, \ldots, N)$ be the eigenvalues of V in descending order. When all elements of B could be chosen freely, the second term has the upper bound,

$$\max \text{tr}((B'B)^{-1}B'VB) = \sum_{i=1}^{k}\lambda_i. \tag{19}$$

The upper bound will be attained if the factor loadings B span the space of the eigenvectors corresponding to the largest k eigenvalues. The same eigenvalues also define the principal components of y_t. The fixed effects model (8) can thus be viewed a restricted version of the principal components model

$$y_t = a + \mathbf{B}z_t + v_t, \tag{20}$$

where $\mathbf{B}$ is an $(N \times k)$ matrix of free parameters. The Gaussian model implies that $\mathbf{B}$ is a tightly parameterized function of the k-vector κ of mean reversion parameters.

Without further restrictions $\mathbf{B}$ is not identified, since for every $(k \times k)$ nonsingular matrix Q the matrix $\mathbf{G} = \mathbf{B}Q^{-1}$ spans the same space of factor loadings. After characterizing one particular solution for $\mathbf{G}$, all solutions $\mathbf{B} = \mathbf{G}Q$ are equivalent. Starting with an initial $\mathbf{G}$ we wish to find a rotation matrix Q, such that $\hat{\mathbf{B}} = \mathbf{G}Q$ matches the theoretical factor loadings of the Gaussian model as close as possible. The least squares solution is

$$\hat{\mathbf{B}} = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'B, \tag{21}$$

where B contains the elements $(1 - e^{-\kappa_j \tau_i})/\kappa_j \tau_i$ for a vector of mean reversion parameters κ.[16] The projection (21) is very useful for a graphical assessment of the correspondence of the restricted B and the unrestricted $\mathbf{B}$. It does not easily lend itself to a statistical test, since the number of elements in $\mathbf{B}$ increases with N so that we cannot apply the Driscoll and Kraay (1998) method.

Eqs. (18) and (19) imply that

$$s_{PC}^2 = \frac{1}{N} \sum_{j=k+1}^{N} \lambda_j. \tag{22}$$

is a lower bound for the residual variance from the OLS estimates of the Gaussian model. We therefore use the percentage reduction in residual variance as a goodness of fit statistic for a k-factor Gaussian model.

Time series data on yield levels could well be integrated or nearly integrated. The first principal components then extract the common trend components from interest rates (see Stock and Watson, 1988). The leading example is when interest rates are cointegrated and spreads are all stationary as in Campbell and Shiller (1987) and Hall et al. (1992). In that case the first column of $\mathbf{B}$ is proportional to a vector of ones. In the Gaussian model this corresponds to $\kappa_1 = 0$, so that mean reversion is absent both in the actual and risk neutral dynamics. This is a boundary case, because in the Gaussian term structure models all factors must be stationary.

Like the estimation of the structural parameters, the principal components analysis on forward differences $\tilde{Y}_t(\tau)$ is completely analogous to the levels case. Again the Gaussian model implies that after rotation the factor loadings corresponding to the largest k eigenvalues must be identical to the Gaussian function $b(\tau)$. In addition, after rotation the factor loadings from principal components analysis should be the same for levels and forward differences. Of course, the eigenvalues and the components themselves will be different.

Principal components analysis has been an important tool in the interest rate derivatives literature for some time.[17] Most of the literature has extracted the principal

---

[16] Brown and Schaefer (2000) reverse the use of (21). They start from the first k principal components of first differences to obtain the normalised $(N \times k)$ factor loadings matrix $\mathbf{G}$. They estimate the mean reversion parameters κ in the Gaussian model by minimizing the squared elements of $B - \mathbf{G}Q$.

[17] See Bliss (1997) and James and Webber (2000, chapter 16) for a survey and for details on the relation between principal components analysis and Gaussian term structure models.

components from first differences of yields following Litterman and Scheinkman (1991). In practice the principal components from forward differences $\tilde{Y}_t(\tau)$ or standard first differences $Y_t(\tau) - Y_{t-1}(\tau)$ will be very similar, since for longer maturities the forward rate $F_{t-1}(\tau)$ and the lagged spot rate $Y_{t-1}(\tau)$ are almost the same (see Section 6).[18]

6. Data

We construct yield curves from the CRSP database of monthly US government coupon bond data for the period from January 1970 until December 1995. The most popular methods for constructing the yield curve are the cubic splines introduced by McCulloch (1975) and the parsimonious smooth functional forms of Nelson and Siegel (1987) extended by Bliss (1997).[19] Being based on just a few parameters the parsimonious models impose a lot of a priori structure on the yield curves. Cross sectional tests of the Gaussian model will then be difficult to interpret. A rejection of the Gaussian model could be due to misspecification of the functional form of the parsimonious curves. Alternatively, not rejecting the Gaussian model could be due to the close relation between the Nelson/Siegel specification and the Gaussian model.

We therefore prefer the cubic spline methods. Cubic spline data have been used extensively in empirical work, often based on the well known McCulloch–Kwon (McCulloch and Kwon, 1993) term structure data. A drawback of the cubic splines is that their polynomial form does not work well for the very long term maturities. For this reason, and because long term bonds have not always been available in the last three decades, we restrict our attention to discount yields with maturities up to ten years. The parameters of the spline functions are estimated month by month following the procedures outlined in McCulloch (1975, 1990).[20]

Fig. 1 shows the full data panels. Summary statistics are presented in Fig. 2. Since these interest data are well known, we only discuss the features most relevant to the tests of the Gaussian model. Apart from noisy behavior at the short end, monthly yield changes are almost horizontal. Volatility decreases with maturity, as it should when interest rates are mean reverting. At longer maturities volatility is still far from zero, indicating that implied mean reversion will be small. From a time series perspective mean reversion will also be small. The first order autocorrelation is above 0.95 for all maturities, consistent with (near) unit root behavior.

The term structure of volatilities for the levels is flatter than that for the changes and the forward differences. This is primarily due to the initial steep decrease in volatility for the very short term rates (maturities six months and less). As is well documented, e.g. Duffee (1996), very short rates behave differently from the rest of the term structure. Except at the shorter maturities, the volatility of first differences $Y_t(\tau) - Y_{t-1}(\tau)$ is almost identical to the volatility structure for the forward differences $Y_t(\tau) - F_{t-1}(\tau)$. The same holds for the autocorrelations, except at the shortest maturities.

---

[18] Some studies, e.g., Heath et al. (1990), have applied principal components analysis to changes in forward rates. These can be transformed to the volatility structure for forward differences.

[19] See James and Webber (2000, chapter 15) for an overview and further references.

[20] Details about the construction of the data are available on request.
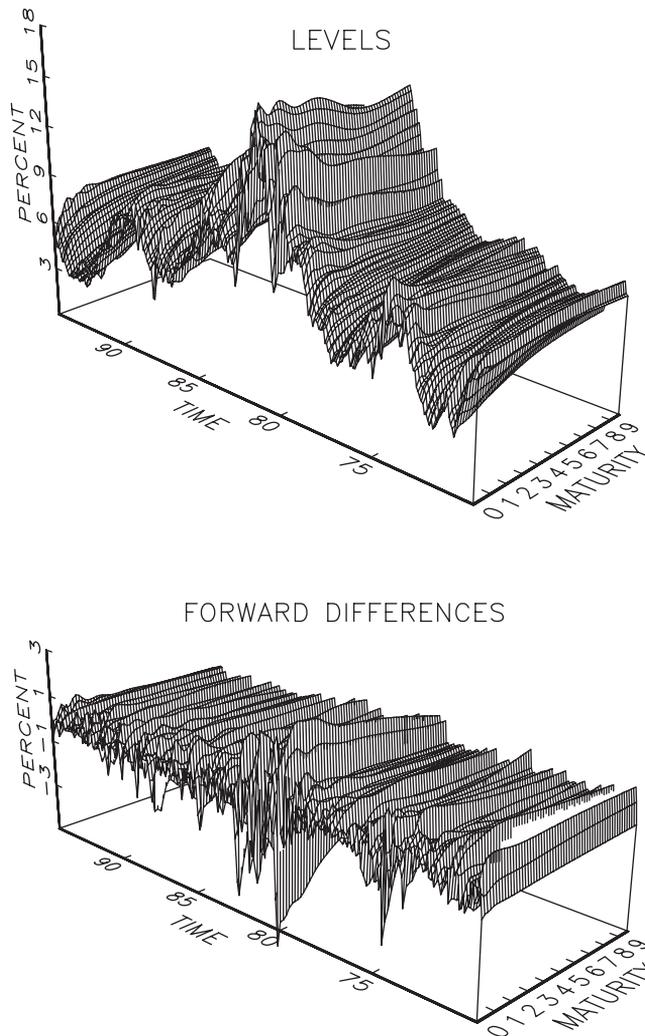
Fig. 1. US Interest Rates. The upper panel shows monthly term structures of US discount yields $Y_t(\tau)$ for the period 1970–1995 and maturities between 0 and 10 years. The lower panel shows monthly forward differences $Y_t(\tau) - F_{t-1}(\tau)$ for the same sample period.

The lower panel of Fig. 2 shows summary statistics for the subsample January 1983–December 1995. For the data after 1982 the autocorrelations of yield levels are lower than for the full sample and the volatility of the yield levels decreases sharply for maturities longer than eight years. Both are signs of increased mean reversion. For the forward differences the autocorrelations are very negative at maturities between six months and one year. In this subsample there seems to be severe measurement error around these maturities. This could also explain the irregular peak in the volatility structure.

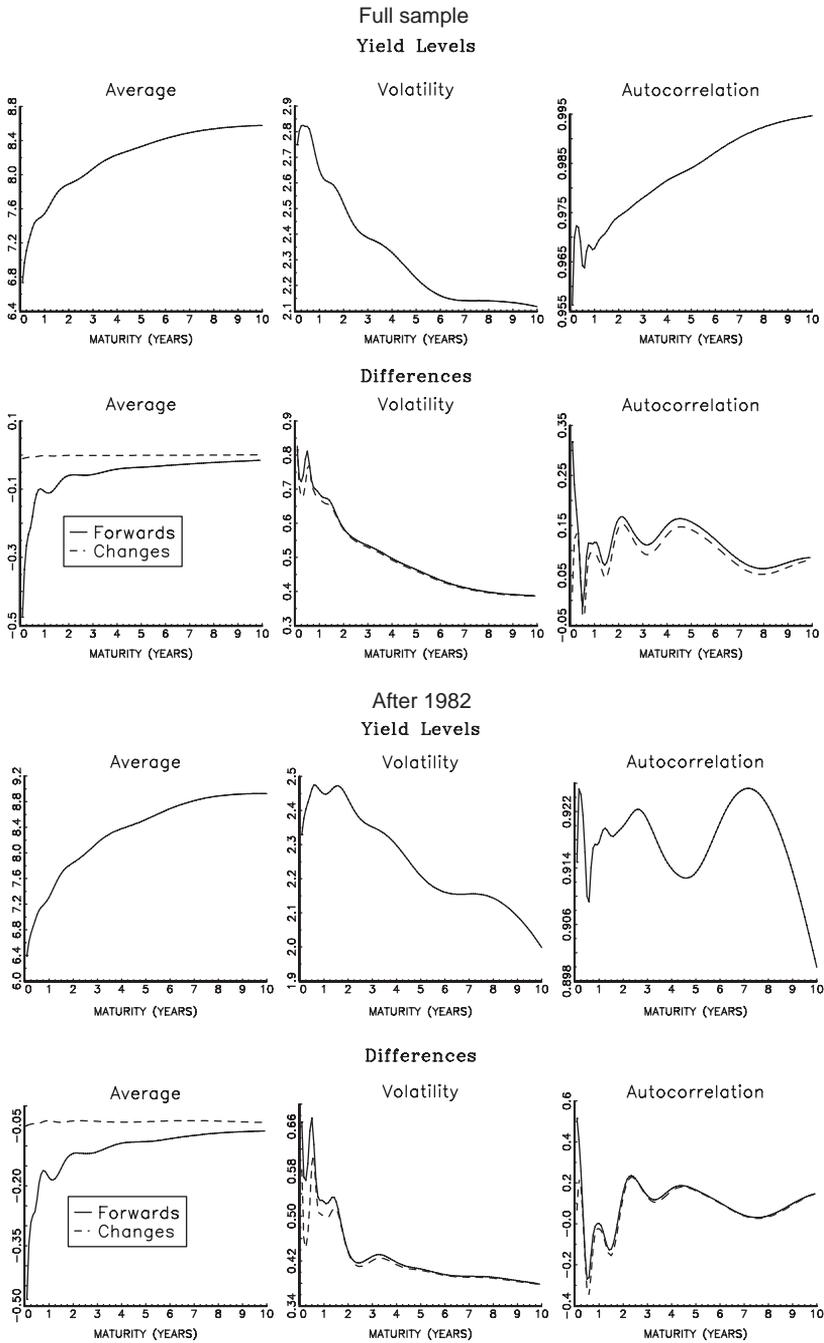Fig. 2. Data summary. The figure shows the time series average, standard deviations and first order autocorrelations for the discount yields $Y_t(\tau)$ levels, forward differences $Y_t(\tau) - F_{t-1}(\tau)$ and ordinary first differences $Y_t(\tau) - Y_{t-1}(\tau)$. Yields are expressed in percent per year. *Full Sample* refers to the sample period January 1970–December 1995; *After 1982* uses data starting January 1983.

Table 1
Eigenvalues

Data	Sample	Number of factors				
		1	2	3	4	
Yield	Full sample	624.28	19.47	1.28	0.51	Eigenvalue
levels		96.6%	99.6%	99.8%	99.9%	Cumulative %
		0.429	*0.148*	*0.105*	*0.082*	s_{PC}
	After 1982	592.13	11.22	1.01	0.41	Eigenvalue
		97.8%	99.7%	99.8%	99.9%	Cumulative %
		0.376	*0.115*	*0.086*	*0.067*	s_{PC}
Forward	Full sample	26.72	2.46	0.64	0.43	Eigenvalue
differences		86.0%	93.9%	96.0%	97.4%	Cumulative %
		0.191	*0.126*	*0.103*	*0.083*	s_{PC}
	After 1982	18.59	2.13	0.63	0.30	Eigenvalue
		83.4%	93.0%	95.8%	97.2%	Cumulative %
		0.164	*0.102*	*0.082*	*0.066*	s_{PC}

Notes: The first row of entries for each sample reports the largest four eigenvalues λ_k of the covariance matrix of either yield levels $Y_t(\tau)$ or forward differences $\tilde{Y}_t(\tau)$. The second row is the cumulative percentage of variance explained by the first k principal components. The third row in *italics* is the residual standard deviation defined in (22). This number serves as the lower bound for the residual variance of the OLS estimates of the Gaussian model. The yield curve is evaluated at 120 maturities from one month to ten years at monthly intervals. The covariance matrices are of order (120×120).

Table 1 gives an impression about the covariances among yields. The first principal component of yield levels explains 96.6% of the variance. For the forward differences the first factor is a little less dominant, but three factors still account for 96% of the variance. As shown in Section 5 the residual variance from the principal components is a lower bound on what can be obtained from the Gaussian model. With three factors the residual standard deviation for levels is about 10 basis points. This is not far above the average of 3 basis points pure measurement error for cubic splines in McCulloch and Kwon (1993). [21] In the Gaussian model additional factors are merely noise.

7. Results

We have estimated models with one, two and three factors, for different subsamples, both with data on yield levels $Y_t(\tau)$ and data on forward differences $Y_t(\tau) - F_{t-1}(\tau)$. Table 2 reports all parameter estimates.

We start the discussion of the results with the model for yield levels for the full sample period. Fig. 3 shows the residual standard deviation by maturity τ_i, providing more detail on the fit of the models. For the one factor model the residual variance has the same U-shape found in Brandt and He (2002). Both from the table and the figure

---

[21] This number is based on the file zeroerr2.txt on their website (see references).

Table 2
Parameter estimates

	Full Sample				After 1982			
	1 fact.	2 fact.	2 fact. $\tau > 6m$	3 fact.	1 fact.	2 fact.	2 fact. $\tau > 6m$	3 fact.
Yield levels								
κ_1	0.058	0.030	0.019	0.001	0.039	0.076	0.054	0.037
	(0.025)	(0.013)	(0.011)	(0.016)	(0.019)	(0.022)	(0.017)	(0.015)
κ_2	...	0.582	0.607	0.707	...	0.397	0.445	0.563
	...	(0.087)	(0.083)	(0.128)	...	(0.084)	(0.064)	(0.070)
κ_3	...	...	...	2.036	...	...	...	4.877
	...	...	...	(0.726)	...	...	...	(1.062)
$s(\kappa)$	0.431	0.154	0.119	0.111	0.336	0.138	0.100	0.097
MAE	0.296	0.102	0.085	0.076	0.234	0.086	0.068	0.065
ρ^2	1.1%	7.8%	7.6%	10.0%	2.2%	10.6%	9.5%	8.9%
AR_t	0.90	0.63	0.62	0.53	0.88	0.61	0.56	0.48
AR_τ	0.97	0.89	0.95	0.90	0.96	0.86	0.95	0.92
Forward differences								
κ_1	0.159	0.194	0.127	0.092	0.053	0.077	0.042	0.028
	(0.038)	(0.057)	(0.039)	(0.027)	(0.045)	(0.048)	(0.030)	(0.023)
κ_2	...	0.639	0.713	1.172	...	0.732	0.991	1.568
	...	(0.164)	(0.011)	(0.301)	...	(0.232)	(0.188)	(0.396)
κ_3	...	...	...	3.287	...	...	...	5.688
	...	...	...	(1.214)	...	...	...	(1.239)
s	0.193	0.129	0.103	0.105	0.177	0.119	0.091	0.093
MAE	0.124	0.080	0.067	0.066	0.111	0.073	0.060	0.062
ρ^2	1.3%	4.1%	2.3%	3.3%	1.3%	8.1%	6.3%	9.4%
AR_t	−0.13	−0.24	−0.34	−0.34	−0.15	−0.20	−0.33	−0.34
AR_τ	0.93	0.87	0.93	0.89	0.92	0.84	0.93	0.91

Notes: The table reports the estimation results for the models

$$y_{it} = a_i + \sum_{j=1}^{k} b_j(\tau_i) f_{jt} + v_{it}, \quad \text{(Yield levels)}$$

$$\tilde{y}_{it} = \tilde{a}_i + \sum_{j=1}^{k} b_j(\tau_i) \tilde{f}_{jt} + \tilde{v}_{it}, \quad \text{(Forward differences)}$$

with unrestricted maturity specific intercepts a_i and $\tilde{a}_i$. Yields are expressed as percent per year. Mean reversion is measured 'per year', so that $\ln 2/\kappa$ is the halflife in years. Standard errors are in parentheses. *Full sample* refers to all monthly data and all 120 maturities τ_i between one month to 10 years at monthly intervals; *After 1982* only uses data starting January 1983; $\tau > 6m$ excludes maturities less than six months; $s(\kappa)$ is the residual standard deviation; MAE is the mean absolute error; $\rho^2 = (s^2(\kappa) - s_{PC}^2)/s^2(\kappa)$ measures the variance reduction from the Gaussian model to the principal components model. AR_t and AR_τ measure the first order autocorrelation of the residuals v_{it} over time and across maturities, respectively.

it is immediately clear that the one factor model fits very poorly. Since problems with a single factor model have been well documented, we will not further discuss it here.

For 2- and 3-factor models the fit is more uniform across the entire maturity range. The 3-factor model fits the data very well over the entire maturity range. Both the
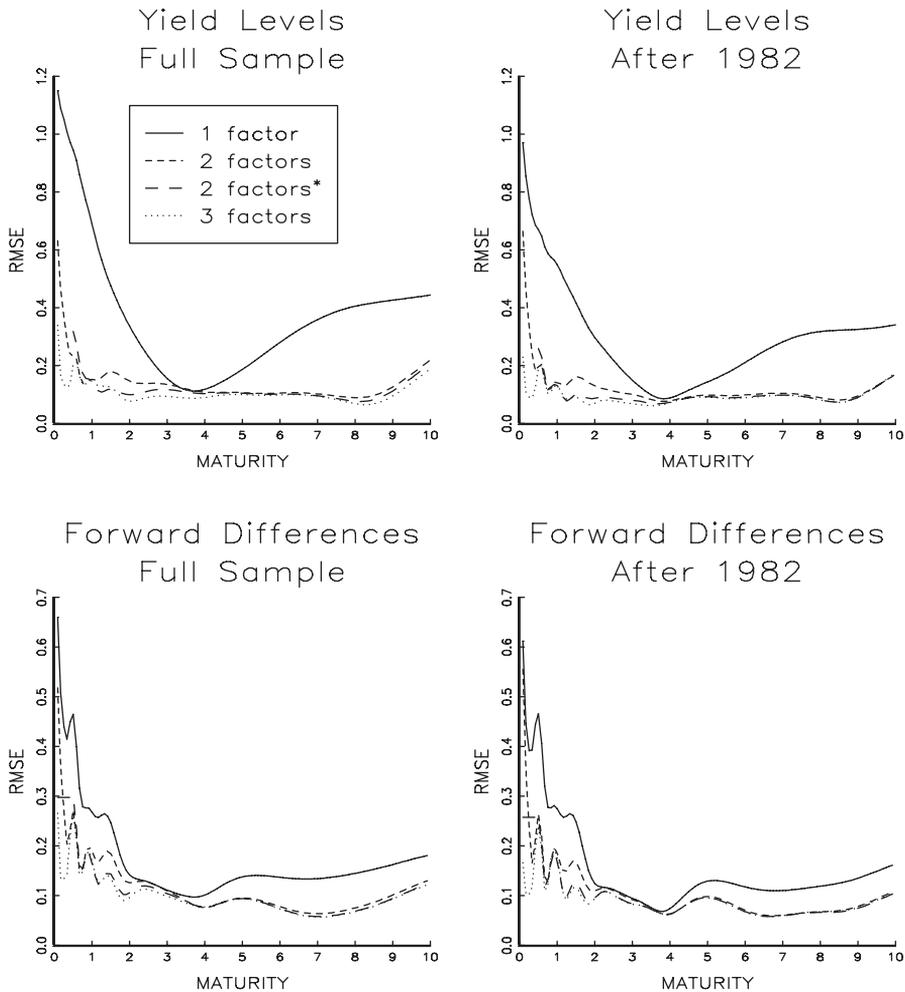
Fig. 3. RMSE by maturity. The figure shows the RMSE of the residuals of the estimated models with various numbers of factors. The asterisk in the legend refers to the model with maturities $\tau > 6m$. The top panels refers to the model for discount yields $Y_t(\tau)$ levels. The lower panels shows the same statistics for the models estimated with the forward differences $Y_t(\tau) - F_{t-1}(\tau)$. The left panels are based on the full sample; the right panels on the sample starting January 1983.

average absolute residual and the standard deviation of the residuals are about 0.1%, which is very close to the principal components lower bound. The better fit of the 3-factor model relative to the 2-factor model is entirely due to the shortest maturities. A 2-factor model estimated without maturities less than six months is almost identical to the 3-factor model. For further reference we label this model the 2*-model.

In all models the mean reversion of the first, dominant, factor is close to zero. Although the standard error of κ_1 is small, the point estimate is not significantly different

from zero in both the 2^*-factor model and the 3-factor model. A unit root in the implied dynamics means that $b_1(\tau)=1$ for all maturities and that the dominant movements are parallel shifts in the yield curve. The mean reversion of the second factor is much larger and significantly different from zero. This indicates that there is at most one nonstationary trend in the risk-neutral dynamics of interest rates.

In the top row of graphs in Fig. 5 the functions $b_j(\tau)$ of the 2^*- and 3-factor models are compared to the principal component factor loadings, i.e. linear combinations of the first two or three eigenvectors of the sample covariance matrix. For each of the factors the principal components factor loadings are almost identical to the Gaussian function $b_j(\tau)$. The structure on the factor loadings imposed by the Gaussian model corresponds almost exactly with the structure in the covariance matrix of yield levels.

The first three principal components are usually interpreted as level, slope and curvature factors. After rotation the same interpretation can be given to the factors in the Gaussian model. The rightmost column of graphs in Fig. 5 shows the original orthonormal factor loadings of the first three principal components and a rotation of the functions $b_j(\tau)$. After rotation they have the familiar shapes of a flat curve, a downward sloping curve and a third curve that goes down steeply and then rebounds.

Fig. 4 shows the residual RMSE by month for the 3-factor model. Like any term structure model, the period 1979–1982 is problematical. To investigate whether the increase in RMSE in this period is residual heteroskedasticity or an indication of structural differences in regime, we estimate all models on the subsample after 1982. Detailed estimation results for the post-1982 sample are in the upper right panel of Table 2. For the period after 1982 we find substantial and significant mean reversion in the dominant factor for all models. In the 3-factor model the mean reversion of the third factor implies rapidly declining factor loadings for the third factor. The third factor only accounts for the largely idiosyncratic behaviour of the short term interest rates; it has negligible effect on longer term yields. Most important is that the principal components factor loadings are again almost identical to the Gaussian factor structure, see the third row of graphs in Fig. 5.

Two of the most salient residual diagnostics of the models are the first order autocorrelations, both cross sectionally and over time.[22] Both are positive and large. The cross sectional correlation is very persistent, and does not seem to depend at all on the number of factors that are included. We conjecture that this correlation is due to the measurement error induced by the splines. The first order time series autocorrelation reduces with the number of factors. But even for the 3-factor model the autocorrelation remains substantial at 0.53. Although this is consistent with earlier literature, for example Chen and Scott (1993), the persistent measurement error suggests that some bonds with similar maturity must be mispriced systematically, perhaps because of omitted coupon effects.

The lower panels of Table 2 repeat the empirical analysis using data on forward differences. The remaining rows in Fig. 5 compare the Gaussian factor loadings $b(\tau)$

---

[22] They are calculated as the pooled correlations, $AR_t = \sum_t \sum_i v_{it} v_{i,t-1} / \sum_t \sum_i v_{it}^2$ and $AR_\tau = \sum_t \sum_i v_{it} v_{i-1,t} / \sum_t \sum_i v_{it}^2$, respectively.
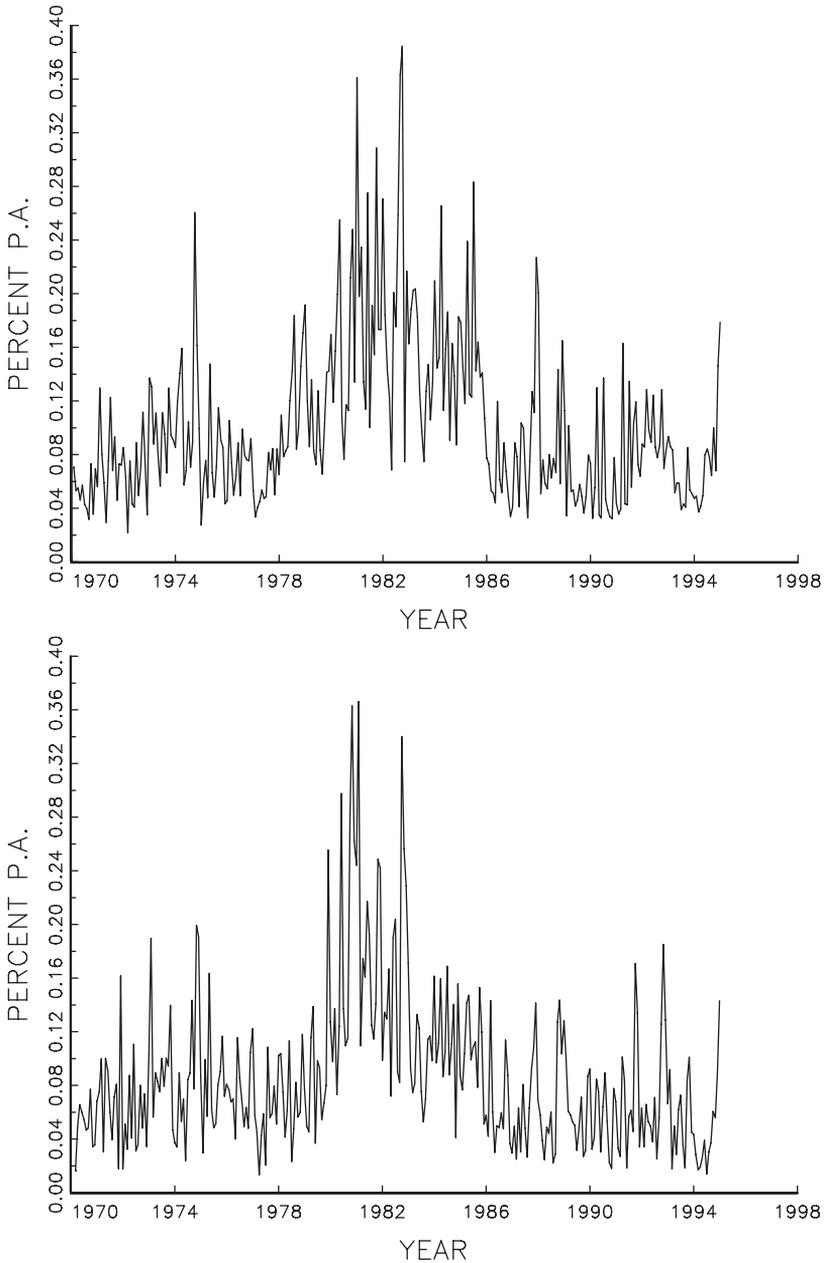
Fig. 4. RMSE by month. The top figure shows the RMSE of the residuals of the 3-factor model for yield levels over all maturities for every month in the sample. The bottom figure shows the same RMSE for the 3-factor model estimated on forward differences. Parameters have been estimated over the full sample period.
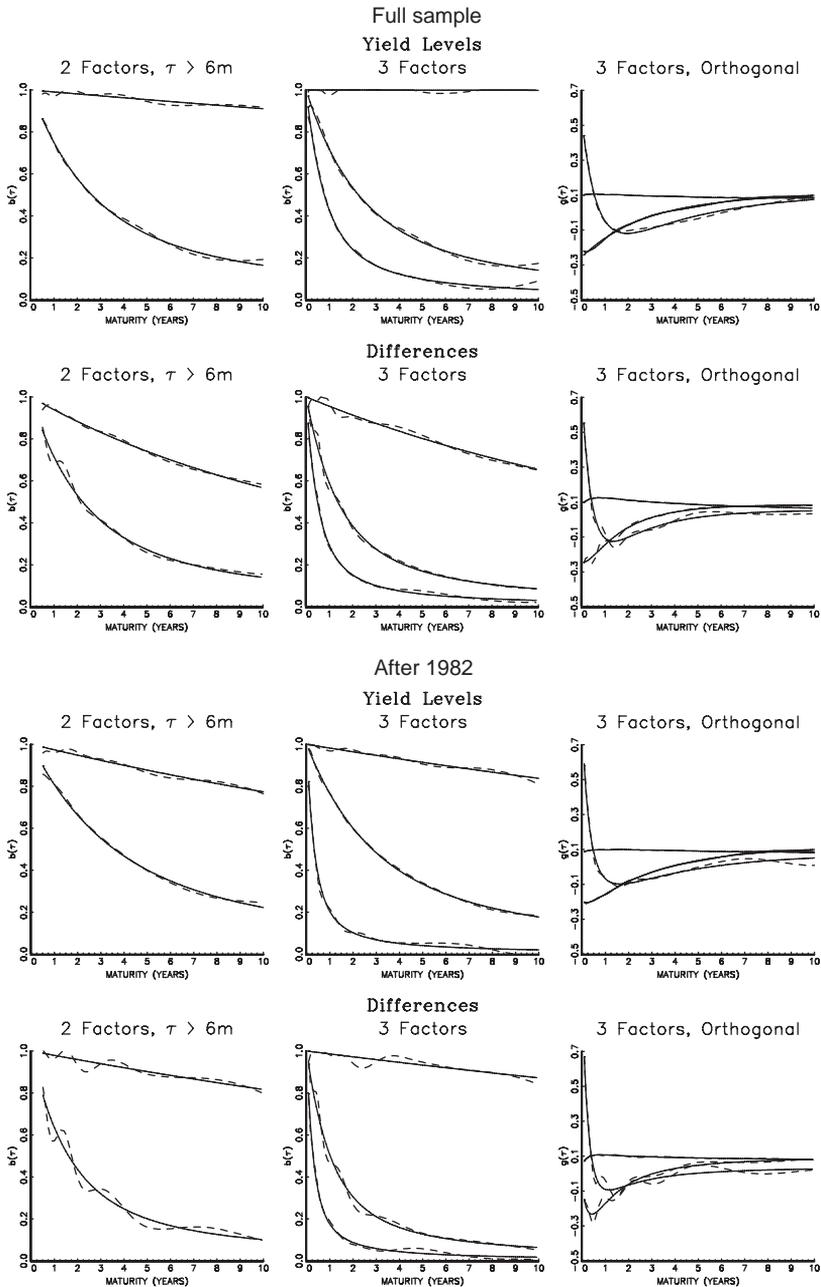
Full sample

After 1982

Fig. 5. Factors loadings. The solid lines are the factor loadings $b(\tau)$ according to the Gaussian affine model. The dashed lines are the (rotated) factor loadings from the principal components analysis. The first column of graphs refers to the 2^*-factor model which excludes maturities less than six months. The second column refers to the 3-factor model. The third column of graphs is a rotation of the second column, such that the factors are orthogonal. Input data for the four rows of graphs are sample covariance matrices: levels or forward differences and full sample or data starting in 1983.

Table 3
Overidentification tests

	Full sample		After 1982	
	2* fact.	3 fact.	2* fact.	3 fact.
$t(\kappa_1 - \tilde{\kappa}_1)$	−2.62	−2.65	0.49	0.39
$t(\kappa_2 - \tilde{\kappa}_2)$	−0.71	−1.26	−3.24	−2.71
$t(\kappa_3 - \tilde{\kappa}_3)$	...	−0.98	...	−0.52
$\chi^2(k)$	12.56	24.37	11.04	37.07

Notes: The table reports the test results for the hypothesis that parameters in the levels model are equal to parameters in the model with forward differences: $\kappa = \tilde{\kappa}$. The $t(\kappa_i - \tilde{\kappa}_i)$ are t-statistics for individual parameters; $\chi^2(k)$ reports the Wald statistics. Parameter estimates correspond to the "2-factor ($\tau > 6m$)" and "3-factor" models in Table 2.

and the principal component factor loadings. Just as for the levels model, the Gaussian model almost perfectly fits the principal components.

The parameter estimates differ in a few respects from the estimates for the yield levels. First of all, for the full sample period the first mean reversion parameter is significantly positive. Table 3 shows that it is also significantly different from the estimate of κ_1 estimated from the yield levels. The difference between the first mean reversion parameter is the main reason why the overall test of the overidentifying restrictions leads to rejecting the null hypothesis.

For the post-1982 subsample the difference between the estimates of κ_1 between levels and forward differences is much smaller. The point estimates from the forward differences are even smaller than those from the levels, and the differences are not significant. Instead, the two sets of estimates now differ with regard to the second factor, both in the 2*- and 3-factor models. Overall the null hypothesis of the overidentifying moment conditions is still rejected.

The residuals of the forward differences exhibit strong negative autocorrelation. This is usually a symptom of overdifferencing. Here it is more likely related to the induced measurement error from constructing forward rate curves.

The rejection of the Gaussian model means that the transformation from levels to forward differences, $y_{it} - (i/(i+1))y_{i+1,t-1} + (1/(i+1))y_{1,t-1}$, does not preserve the factor structure. The factor structure has strong implications for the implied variances of interest rate innovations. To explore the differences in implications, we calculate the variance $\sigma^2(\tau) = b(\tau)'\Sigma b(\tau)$ implied by both models. The factor innovation covariance matrix should be the same in the actual and in the risk neutral measure, so that we can estimate Σ from the time series of the estimated factors z_t and $\tilde{z}_t$, respectively. Since Σ is the covariance matrix of the innovations in the factors, the residuals of a first order vector autoregression of the estimated factors are used to estimate Σ. Also, since Σ is calculated from the factor innovations, yield levels and forward differences should lead to comparable estimates. The results are reported in Table 4. They confirm that the factors have strong negative mutual correlations, which was one of the main arguments to work with the Gaussian model instead of a model with square root factors that cannot be negatively correlated. The results are remarkably robust over the different

Table 4
Volatility

	Full sample				After 1982			
	Levels		Forwards		Levels		Forwards	
	2* fact.	3 fact.	2* fact.	3 fact.	2* fact.	3 fact.	2* fact.	3 fact.
σ_1	0.450	0.365	0.833	0.569	0.649	0.505	0.510	0.440
σ_2	0.772	1.417	1.007	1.729	0.758	0.854	0.713	1.361
σ_3	...	1.375	...	1.809	...	0.913	...	1.745
ρ_{12}	−0.23	−0.08	−0.63	−0.20	−0.73	−0.52	−0.48	−0.27
ρ_{13}	...	−0.02	...	0.04	...	0.10	...	0.09
ρ_{23}	...	−0.86	...	−0.89	...	−0.73	...	−0.91

Notes: The table reports estimates of the conditional standard deviations and correlations of the estimated factors f_t and $\tilde{f}_t$. The estimates are derived from the covariances of the residuals of a first order vector autoregression of the factors. The 2-factor model in this table refers to the model estimated with maturities $\tau \geqslant 6m$.

subsamples, numbers of factors, and between levels and forward differences. The first and second factor are always negatively correlated. A third factor is almost uncorrelated with the first factor and strongly negatively correlated with the second factor.

Fig. 6 shows the functions $\sigma(\tau) = \sqrt{b(\tau)' \Sigma b(\tau)}$ for all different models. Mostly the yield levels and forward differences give the same results despite the significant differences in the estimates of κ. For the sample after 1982 the 3-factor model generates the familiar hump-shaped volatility structure, except for the steep decline for the shortest maturities less than 3 months. Again this indicates that the behavior of the very short term maturities does not easily fit into the Gaussian model structure.

8. Summary and conclusions

This paper proposes a cross sectional analysis of multivariate Gaussian term structure models. The main feature is that it does not depend on a specific parameterization of factor risk prices. Under our assumptions the risk neutral dynamics are Gaussian, but the actual time series process for the factors is left unspecified. Tests of the Gaussian model are carried out by (i) comparing the implied factor loadings for different maturities to unrestricted factor loadings obtained from principal components analysis, and (ii) comparing the factor loadings obtained from yield levels to the factor loadings from forward differences. The latter test exploits the overidentifying moment conditions that the factor loadings of the Gaussian model should be the same for levels and forward differences.

In the empirical results we find that the factor loadings implied by the Gaussian model are very close to the principal component factor loadings, both for levels as well as forward differences. On the other hand we reject the overidentifying conditions that the factor loadings for the levels are the same as those of the forward differences.
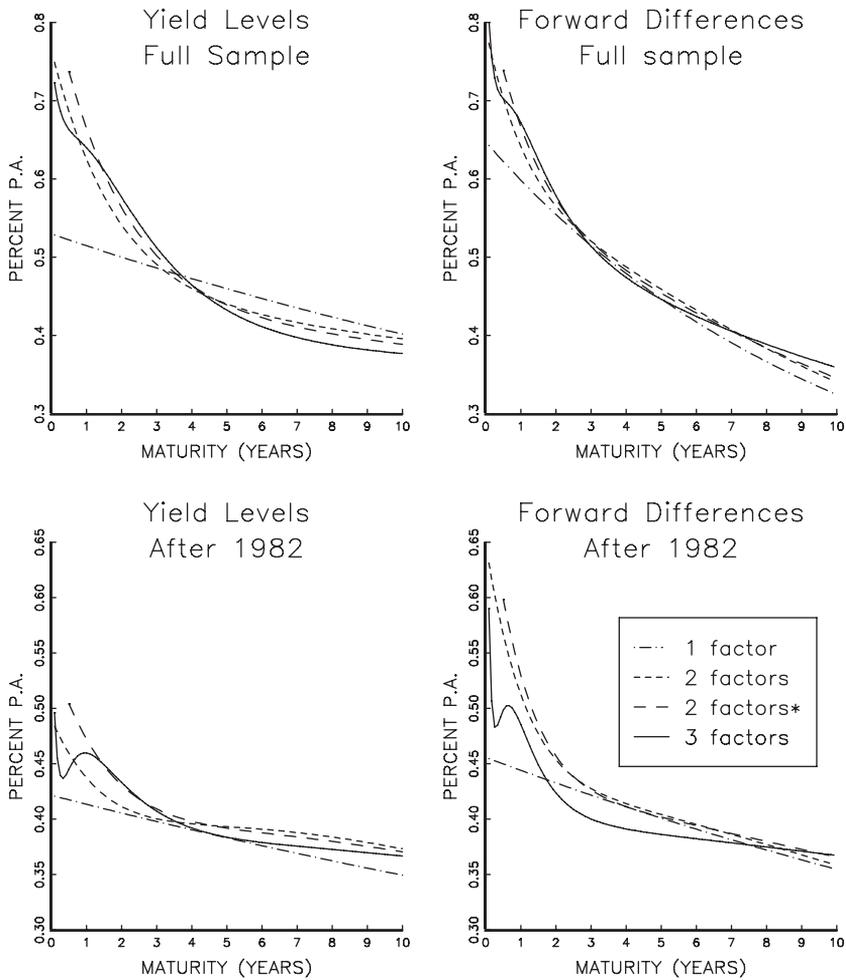
Fig. 6. Volatility. The figures show the implied volatility of yields as a function of maturity, calculated as $\sqrt{b(\tau)'\hat{\Sigma}b(\tau)}$, where $\hat{\Sigma}$ is contains the factor covariance matrix and $b(\tau)$ depends on the mean reversion parameters κ. The factor covariance matrix is calculated as the residual covariance matrix of a first order vector autoregression of the estimated factors z_t and $\tilde{z}_t$, respectively.

Acknowledgements

Bi-annual Panel Data Conference (Amsterdam), York Econometrics Day, Aarhus Business School, CEMAF (Lisbon), Tilburg University, Stockholm School of Economics, and HEC (Paris) for their comments. Of course all errors are our own.

Appendix A. Gaussian model

A.1. Derivation of (3) and (5)

DS consider the following specification for the risk neutral dynamics of the factors z_t in a Gaussian model with constant risk prices:

$$\mathrm{d}z = \mathscr{K}(\Theta - z)\,\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W, \tag{A.1}$$

$$r_t = \delta_0 + \sum_{i=1}^{k} \delta_i z_{it}, \tag{A.2}$$

where, in addition to notation explained in Section 2, r_t is the instantaneous spot rate and δ_i are scalar parameters. The canonical model of DS has the following identification restrictions (see their paragraph I.B.1):

$$\mathscr{K} \quad \text{is a lower triangular matrix,}$$
$$\Sigma = I, \quad \text{the identity matrix.} \tag{A.3}$$

The parameter vector δ is unrestricted. Alternative equivalent parameterizations are obtained through affine transformations of the factor dynamics. For a parameterisation in which the factor covariance matrix Σ is unrestricted, we shift the restrictions on Σ to $\mathscr{K}$ and δ. We start by premultiplying z with a $(k \times k)$ matrix $\mathbf{L}_1$,

$$z^{(1)} = \mathbf{L}_1 z, \tag{A.4}$$

such that $\mathbf{L}_1 \mathscr{K} \mathbf{L}_1^{-1}$ is diagonal, so that $\mathbf{L}_1$ are the eigenvectors of $\mathscr{K}$.

Next we take the spot rate r_t as the sum of the factors as in Duffie and Kan (1996). This is achieved by a transformation with a diagonal matrix $\mathbf{L}_2$ with diagonal elements δ_i. The new factors are $z^{(2)} = \mathbf{L}_2 z^{(1)}$. Since $\mathbf{L}_2 \mathscr{K} \mathbf{L}_2^{-1} = \mathscr{K}$, when $\mathbf{L}_2$ is diagonal, the transformation matrix $\mathbf{L}_2$ leaves the mean reversion matrix unchanged. Compared to the DS parameterization we now have an unrestricted factor covariance matrix $\Sigma = \mathbf{L}_2 \mathbf{L}_1 \mathbf{L}_1' \mathbf{L}_2$, with $\frac{1}{2} k(k+1)$ free parameters. To compensate we place $\frac{1}{2} k(k-1)$ restrictions on the lower triangle of $\mathscr{K}$ and k restrictions on δ.

The functions $a(\tau)$ and $b(\tau)$ in (3) are found by solving a system of Ricatti differential equations, given as equations (6) and (7) in DS. Our interest in this paper is only in $b(\tau)$. In our notation their differential equation for $B(\tau) = \tau b(\tau)$ is

$$\frac{\mathrm{d}B(\tau)}{\mathrm{d}\tau} = -\mathscr{K}B(\tau) + \iota, \tag{A.5}$$

where ι is a vector of ones. Boundary conditions are $B(0) = 0$. The general solution to the set of linear differential equations (A.5) is

$$B_i(\tau) = \frac{(1 - \mathrm{e}^{-\kappa_i \tau})}{\kappa_i}. \tag{A.6}$$

Dividing by τ we obtain (5) in the text.

A.2. Derivation of (7)

Start from the Gaussian model for discount bond prices (3). The forward difference $\tilde{Y}_t(\tau) = Y_t(\tau) - F_{t-h}(\tau)$ follows from the Gaussian model as

$$\tilde{Y}_t(\tau) = Y_t(\tau) - \frac{\tau + h}{\tau} Y_{t-h}(\tau + h) + \frac{h}{\tau} Y_{t-h}(h)$$

$$= \tilde{a}(\tau) + b(\tau)' z_t - \left(\frac{\tau + h}{\tau} b(\tau + h) - \frac{h}{\tau} b(h) \right)' z_{t-h}, \tag{A.7}$$

with new intercept $\tilde{a}(\tau) = a(\tau) - ((\tau + h)/\tau)a(\tau + h) + (h/\tau)a(h)$. From the definition of $b(\tau)$ in (5) we find that its jth component satisfies

$$\frac{\tau + h}{\tau} b_j(\tau + h) - \frac{h}{\tau} b_j(h) = \frac{1 - \mathrm{e}^{-\kappa_j(\tau + h)}}{\kappa_j \tau} - \frac{1 - \mathrm{e}^{-\kappa_j h}}{\kappa_j \tau} = \mathrm{e}^{-\kappa_j h} b_j(\tau). \tag{A.8}$$

Therefore the dependence of $\tilde{Y}_t(\tau)$ on the factors z_t follows as

$$b(\tau)' z_t - \left(\frac{\tau + h}{\tau} b(\tau + h) - \frac{h}{\tau} b(h) \right)' z_{t-h} = \sum_{j=1}^{k} b_j(\tau)(z_{jt} - \mathrm{e}^{-\kappa_j h} z_{j,t-h})$$

$$= b(\tau)' \tilde{z}_t, \tag{A.9}$$

where $\tilde{z}_{jt} = z_{jt} - \mathrm{e}^{-\kappa_j h} z_{j,t-h}$ are the new factors.

Appendix B. Estimation details

B.1. Derivation of (11)

Recall the least squares criterion (10) and rewrite $\hat{v}_t$ as

$$\hat{v}_t = y_t - \bar{y} - B\hat{f}_t, \tag{B.1}$$

$$\hat{f}_t = (B'B)^{-1} B'(y_t - \bar{y}). \tag{B.2}$$

Differentiating the sum of squares (10) with respect to κ gives

$$\frac{\partial s^2}{\partial \kappa_i} = \frac{1}{NT} \sum_t \frac{\partial \hat{v}_t'}{\partial \kappa_i} \hat{v}_t = -\frac{1}{NT} \sum_t \left(\frac{\partial B}{\partial \kappa_i} \hat{z}_t + B \frac{\partial \hat{z}_t}{\partial \kappa_i} \right)' \hat{v}_t. \tag{B.3}$$

The first order derivatives can be simplified considerably by noting that $\partial B/\partial \kappa_i$ is an $(N \times k)$ matrix of zeros, except for the ith column, which is given by the N-vector c_i with elements

$$c_{\ell i} = \frac{\partial b_{\ell i}}{\partial \kappa_i} = \frac{\tau_\ell e^{-\kappa_i \tau_\ell} - b_{\ell i}}{\kappa_i}, \quad \ell = 1, \ldots, N. \tag{B.4}$$

We join the column vectors c_i in the $(N \times k)$ matrix $C = (c_1, \ldots, c_k)$. We also define the k-vector $\hat{u}_t = \frac{1}{N} C' \hat{v}_t$. With this notation and using that $B' \hat{v}_t = B' M(y_t - \bar{y}) = 0$ we can rewrite (B.3) as

$$\frac{\partial s^2}{\partial \kappa_i} = -\frac{1}{T} \sum_t \hat{z}_{it} \hat{u}_{it}. \tag{B.5}$$

Stacking the derivatives in a k-vector we obtain the first order conditions (11) in the text.

B.2. Derivation of (12)

Using the definitions of $\hat{z}_t$ and $\hat{u}_t$ the first order derivatives can be written alternatively as

$$\frac{\partial s^2}{\partial \kappa_i} = -\frac{1}{NT} \sum_t e_i'(B'B)^{-1}B'(y_t - \bar{y})(y_t - \bar{y})'MCe_i, \tag{B.6}$$

where e_i is the ith column of the k-dimensional identity matrix. For the plim at $\kappa = \kappa_0$ as $T \to \infty$ and N fixed we find

$$g_{0i} = -\frac{1}{N} e_i'(B'B)^{-1}B' VMCe_i$$

$$= -\frac{1}{N} e_i'(B'B)^{-1}B' \Psi MCe_i. \tag{B.7}$$

The last line in (B.7) follows from the factor assumption $\mathrm{Var}(y_t) = B\Omega B' + \Psi$ and the orthogonality $B'M = 0$. Stacking this result for all elements in κ gives (12) in the text.

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Andersen, T.G., Lund, J., 1997. Estimating continuous time stochastic volatility models of the short term interest rate. Journal of Econometrics 77, 343–377.

Ang, A., Bekaert, G., 2002. Short rate nonlinearities and regime switches. Journal of Economic Dynamics and Control 26, 1244–1275.

Babbs, S.H., Nowman, K.B., 1999. Kalman filtering of generalized Vasicek term structure models. Journal of Financial and Quantitative Analysis 34, 115–130.

Balduzzi, P., Das, S.R., Foresi, S., Sundaram, R.K., 1996. A simple approach to three factor affine term structure models. Journal of Fixed Income 6, 43–53.

Bams, D., Schotman, P.C., 1998. Direct estimation of the risk neutral factor dynamics of affine term structure models. CEPR DP 2034.

Bliss, R.R., 1997. Movements in the term structure of interest rates. Federal Reserve Bank of Atlanta Economic Review, 4th quarter 16–33.

Brandt, M.W., He, P., 2002. Simulated likelihood estimation of affine term structure models from panel data. University of Pennsylvania, Working paper.

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

Brown, R.H., Schaefer, S.M., 2000. Why long term forward interest rates (almost) always slope downwards. London Business School, Working paper.

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

Campbell, J.Y., 2000. Asset pricing at the millennium. Journal of Finance 55, 1515–1567.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.

Campbell, J.Y., Shiller, R.J., 1987. Cointegration and tests of present value relations. Journal of Political Economy 95, 1062–1088.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.

Chen, L., 1996. Interest Rate Dynamics, Derivatives Pricing, and Risk Management. Springer, Berlin.

Chen, R.R, Scott, L., 1993. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income 3, 14–31.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 5, 531–552.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

Dai, Q., Singleton, K., 2002. Expectation puzzles, time-varying risk premia, and dynamic models of the term structure. Journal of Financial Economics 63, 415–441.

De Jong, F., 2000. Time-series and cross section information in affine term structure models. Journal of Economics and Business Statistics 18, 300–314.

De Munnik, J.F.J., Schotman, P.C., 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.

Driscoll, J.C., Kraay, A.C., 1998. Consistent matrix estimation with spatially dependent panel data. Review of Economics and Statistics 80, 549–580.

Duan, J.C., Simonato, J.G., 1999. Estimating and testing exponential-affine term structure models by Kalman filter. Review of Quantitative Finance and Accounting 13, 111–135.

Duarte, J., 2000. The Relevance of price risk in affine term structure models. University of Chicago, Working paper.

Duffee, G.R., 1996. Idiosyncratic variation in treasury bill yields. Journal of Finance 51, 527–551.

Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Duffie, D., Kan, R., 1996. A yield factor model of interest rates. Mathematical Finance 6, 379–406.

Fama, E.F., 1984a. Term premiums in bond returns. Journal of Financial Economics 13, 529–546.

Fama, E.F., 1984b. The information in the term structure. Journal of Financial Economics 13, 509–528.

Fama, E.F., Bliss, R.R., 1987. The information in long maturity forward rates. American Economic Review 77, 680–692.

Frachot, A., Lesne, J.-P., 1993. Econometrics of linear factor models of interest rates. Banque de France, Working paper.

Geyer, A.L.J., Pichler, S., 1999. A state-space approach to estimate and test multi-factor Cox–Ingersoll–Ross models of the term structure. Journal of Financial Research 22, 107–130.

Ghysels, E., Ng, S., 1998. A semiparametric factor model of interest rates and tests of the affine term structure. Review of Economics and Statistics 80, 535–548.

Gibbons, M.R., Ramaswamy, K., 1993. A test of the Cox, Ingersoll, and Ross model of the term structure. Review of Financial Studies 6, 619–658.

Gong, F., Remolona, E.M., 1996. A three factor econometric model of the US term structure. Federal Reserve Bank New York, Working paper.

Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasury bill yields. Review of Economics and Statistics 74, 116–126.

Hamilton, J.D., 1988. Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. Journal of Economic Dynamics and Control 12, 385–423.

Heath, D., Jarrow, R., Morton, A., 1990. Contingent claim valuation with a random evolution of the term structure. Review of Futures Markets 9, 54–76.

Hull, J.C., 2000. Options, Futures, and Other Derivatives, 4th Edition. Prentice-Hall, Englewood Cliffs, NJ.

Hull, J.C., White, A., 1993. One factor interest-rate models and the valuation of interest-rate derivative securities. Journal of Financial and Quantitative Analysis 28, 235–254.

Jacquier, E., Jarrow, R.A., 2000. Dynamic evaluation of contingent claim model error. Journal of Econometrics 94, 145–180.

James, J., Webber, N., 2000. Interest Rate Modelling. Wiley, New York.

Jegadeesh, N., Pennacchi, G.G., 1996. The behavior of interest rates implied by the term structure of eurodollar futures. Journal of Money Credit and Banking 28, 426–446.

Koedijk, C.G., Nissen, F.G.J.A., Schotman, P.C., Wolff, C.C.P., 1997. The dynamics of short-term interest rate volatility reconsidered. European Finance Review 1, 105–130.

Lamoureux, C.G., Witte, H.D., 2002. Empirical analysis of the yield curve: the information in the data viewed through the window of Cox, Ingersoll, and Ross. Journal of Finance 57, 1479–1520.

Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.

McCulloch, J.H., 1975. The tax-adjusted yield curve. Journal of Finance 30, 811–830.

McCulloch, J.H., 1990. U.S. government term structure data, 1947–1987. In: Friedman, B.M., Hahn, F. (Eds.), Handbook of Monetary Economics. North-Holland, Amsterdam, pp. 672–715.

McCulloch, J.H., Kwon, H.C., 1993. U.S. Term Structure Data, 1947–1991. OSU Working Paper # 93-6, http://economics.sbs.ohio-state.edu/jhm/ts/mcckwon/mccull.htm.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. Journal of Business 60, 473–489.

Sercu, P., Wu, X., 1997. The information content in bond model residuals: an empirical study on the Belgian bond market. Journal of Banking and Finance 21, 85–720.

Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973–2002.

Stock, J.H., Watson, M.W., 1988. Testing for common trends. Journal of the American Statistical Association 83, 1097–1107.

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