UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 08-24
Business Economic Series 03
May 2008

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249607

Nonparametric Estimation of Conditional Beta Pricing Models*

Eva Ferreira,[1] Javier Gil-Bazo[2] and Susan Orbe[3]

Abstract

We propose a new procedure to estimate and test conditional beta pricing models which allows for flexibility in the dynamics of assets' covariances with risk factors and market prices of risk (MPR). The method can be seen as a nonparametric version of the two-pass approach commonly employed in the context of unconditional models. In the first stage, conditional covariances are estimated nonparametrically for each asset and period using the time-series of previous data. In the second stage, time-varying MPR are estimated from the cross-section of returns and covariances, using the entire sample and allowing for heteroscedastic and cross-sectionally correlated errors. We prove the desirable properties of consistency and asymptotic normality of the estimators. Finally, an empirical application to the term structure of interest rates illustrates the method and highlights several drawbacks of existing parametric models.

Keywords: Kernel estimation; Locally stationary processes; Time-varying coefficients; Conditional asset pricing models

JEL Classification: C14;G12;C32

* The authors would like to thank Alfredo Ibáñez for his comments and suggestions. This work has been supported by the MEC grant SEJ2005-05549 and SEJ2007-67448, BBVA grant 1/BBVA 00038.16421/2004, and Grupo MACLAB (IT-241-07) Basque Government.

[1,3] Department of Econometrics and Statistics, University of the Basque Country
[2] Department of Business Administration, University Carlos III of Madrid

1 Introduction

Beta pricing models, such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) or the Arbitrage Pricing Theory (APT) of Ross (1976), have been extensively used in portfolio management, risk management, and capital budgeting applications. In these models, a risky asset's expected return in excess of the risk-free interest rate is linearly related to the covariance of the asset's return with one or more factors capturing market-wide sources of risk. Re-scaled by the variance of each risk factor, covariances are referred to as *betas*, and are interpreted as the asset's exposure to risks that cannot be eliminated through diversification. The slopes of the linear relation, which must be equal for all assets, are interpreted as the rewards per unit of covariance risk or market prices of risk (MPR) associated with each factor.

Implementation of these models has traditionally relied on the assumption of constant MPR and stationarity, which contradicts the mounting empirical evidence that betas and risk premiums vary over the business cycle (see, for instance, Keim and Stambaugh (1986), Fama and French. (1989), Ferson (1989), or Ferson and Harvey (1991)). This has led researchers to propose conditional beta-pricing models in which the linear relation holds period by period, with changing factor sensitivities and MPR. Examples of conditional beta pricing models include those proposed by Harvey (1989), Jagannathan and Wang (1996) and Lettau and Ludvigson (2001). The conditional asset pricing relation is also obtained in arbitrage-free models for option and bond returns, such as Black and Scholes (1973), Cox, Ingersoll and Ross (1985) and their extensions, with discrete returns replaced by instantaneous returns.

In this paper, we propose a new nonparametric procedure to estimate and test conditional beta pricing models, which allows for flexibility in the dynamics of covariances and MPR. The method can be seen as a nonparametric version of the popular two-pass approach developed by Black, Jensen and Scholes (1972), Fama and MacBeth (1973), Shanken (1985) and Shanken (1992) in the context of unconditional models.[1] In the first stage, conditional covariances are estimated nonparametrically for each asset and period, using previous information. In the second stage, time-varying MPR are estimated from the cross-section of returns and estimated covariances (the regressors), using the entire sample and allowing for heteroscedastic and cross-sectionally correlated errors. In particular, a Seemingly Unrelated Regression Equations (SURE) model, introduced by Zellner (1962), is specified in the second pass, with each equation in the system corresponding to one asset. Allowance is made for time-varying slope coefficients (MPR), and they are estimated nonparametrically subject to the constraint of equality of slopes across assets. The method provides a generalization of previous estimation methods of conditional asset pricing models in several aspects. First, conditional covariances are considered under no specified parametric structure. Second, the MPR may be time-varying. Third, locally stationary variables are assumed, as defined in Dalhaus (1997), which permit time-varying mean and, therefore, the usual strong hypothesis

[1] See Shanken and Zhou (2007) for a recent study of the small-sample properties of the two-pass approach and a comparison with alternative procedures.

of stationarity is not needed.

The method proposed for this generalized two-stage regression model is also related to some previous econometric literature in the context of nonparametric time-varying regression models, that extends the original work by Robinson (1989). Orbe, Ferreira and Rodriguez-Póo (2005) analyze a single equation regression model under the assumptions of time-varying coefficients with seasonal pattern and locally stationary variables, although neither the two-step procedure nor a multi-equation model are considered. In Orbe, Ferreira and Rodriguez-Póo (2006) a local constrained least squares estimation method is studied for a single equation regression under the usual assumption of ergodicity. Cai (2007) proposes to estimate a model with time-varying coefficients using local polynomial regression under stationarity of the state variables. Kapetanios (2007) also uses the properties of locally stationary variables to estimate deterministically time-varying variances for the error term in the regression model. As mentioned above, in this paper, a SURE model is first estimated with time-varying coefficients subject to constraints across coefficients corresponding to different equations for each time period. Further, the highest difficulty is related to the fact that, in practice, the explanatory variables (the conditional covariances) are not observed and must be estimated in advance. Hence, we deal with generated regressors that have been widely studied by Zellner (1970) or Pagan (1984), among others, for the classical parametric regression model. In order to avoid the inconsistency problems for the coefficient's estimator derived from the potential correlation between the estimated regressor and the error term, conditional covariances are estimated at each date using only past information.

Most previous tests of conditional beta pricing models impose strong parametric restrictions on the dynamics of covariances and/or MPR. For instance, Bollerslev, Engle and Wooldridge (1988) model conditional covariances as an ARCH process. Harvey (1989) assumes that conditional asset expected returns are a fixed linear function of a vector of lagged instrumental variables capturing conditioning information. Similarly, Jagannathan and Wang (1996) assume that the conditional market premium is linear in one instrument. The most common approach in the literature is, however, to assume that betas are a fixed linear function of the instruments (e.g., Ferson and Schadt (1996), Ferson and Harvey (1999), Lettau and Ludvigson (2001)). Nonetheless, Ghysels (1998) finds that the pricing errors of conditional models may be larger than those of the unconditional model when time variation in betas is incorrectly specified. Brandt and Chapman (2006) report similar consequences when market risk premiums are incorrectly assumed to be linear functions of the state variables. Our method circumvents this problem, since it imposes much weaker assumptions on both factor sensitivities and MPR. In particular, conditional covariances are assumed to vary smoothly, but possibly nonlinearly, in the instruments, while MPR are treated as free parameters that vary smoothly through time. Consistently with Ghysels (1998), our empirical application shows that the linearity assumption can be highly inadequate in some settings.

Our work is closely related to that of Stanton (1997), Wang (2003), and Jones (2006). These authors also estimate flexible conditional beta pricing models of bond, equity and option returns, respectively. In particular, Stanton (1997) first estimates conditional covariances and conditional expected returns nonparametrically, and then obtains MPR by solving directly the system of equations imposed by the conditional asset pricing model for two assets at each point in time. One problem with this approach is that it can generate highly unstable estimates of the MPR. Furthermore, the method does not enable formal inference to be conducted on MPR. Wang (2003) proposes a test statistic for the null hypothesis that conditional expected pricing errors from the conditional CAPM are zero. The test is based on the idea that a simple regression of pricing errors on a vector of instruments should yield zero coefficients. Estimation of pricing errors in the conditional CAPM is possible because the risk factor is the return on the market portfolio, so the conditional market price of risk can be estimated nonparametrically from the time series of excess returns on the market portfolio. The present work departs from that of Wang (2003) in important ways. First, the method proposed does not require that risk factors be portfolio returns, so it can be applied to models where factors are identified with any aggregate variable. Also, although we assume that conditional covariances are functions of the instruments, this assumption is irrelevant for the estimation of MPR. All that is required is that conditional covariances must be consistently estimated. This is a desirable property, since the method of Wang (2003) crucially depends on the choice of instruments. For instance, if a relevant instrument is omitted from the regression of pricing errors on instruments, then Wang (2003) could fail to reject the model even if pricing errors are systematically related to the missing instrument. Moreover, our method allows for multiple risk factors and is, therefore, more general. Finally, Jones (2006) uses Legendre polynomials to approximate conditional expected returns and betas, which are estimated in a Bayesian framework. He then solves for the parameters of the polynomial for the price of risk that minimize mean squared pricing errors for the whole panel of returns. An advantage of our method is that inference can be conducted on the basis of the closed-form asymptotic distribution of the estimators instead of the numerically obtained posterior distribution of the model parameters.

An empirical application to a dynamic model of the term structure of interest rates is presented. In particular, we consider a two-factor model and estimate the MPR of risk associated with each factor. This context is a particularly interesting application for two reasons. On the one hand, there is no consensus in the literature regarding the dynamics driving bond returns and market prices of risk. For instance, Duffee (2002) studies the affine class of term structure models of Duffee and Kan. (1996) and shows evidence that the specification of market prices of risk assumed by completely affine models is empirically implausible. Flexible estimation of market prices of risk may therefore shed further light on the reasons for this failure. On the second hand, it provides an alternative way to test for the number of priced risk factors. For this case, the results provide evidence that there is substantial time variation in MPR and further, that only the first risk factor (which we identify with changes in the short term interest rate) appears to be priced by investors. The latter is consistent with the results reported by Ferreira and Gil-Bazo (2004), which cannot reject the null hypothesis that U.S. bond expected returns are driven by a single priced risk factor.

To summarize, a nonparametric estimator of time-varying market prices of risk in the context of conditional beta pricing model is proposed. The estimation procedure takes into account the defining characteristics of these models: the common market prices of risk imposed by absence of arbitrage or equilibrium; the structure of the error's covariance that follows from the multi-equation model and the fact that the regressors—the covariances between returns and changes in the risk factors—are not observed.

We also contribute to the asset pricing literature by generalizing the traditional estimation procedure proposed in Fama and MacBeth (1973) and the GLS version studied in Shanken (1992). In fact, the previous estimators are particular cases of our proposal. From a technical point of view, the proposed estimator is a type of three-stage estimator because, first of all, the explanatory variables are estimated, secondly the covariance matrix of the error term is estimated and finally the prices of risk are estimated. It should be noted that the usual hypotheses of stationarity are relaxed and locally stationary variables are permitted, which increases the flexibility of the model and gives a more realistic perspective in many empirical situations due to fact that variability of first moments is allowed.

The rest of the paper is organized as follows. Section 2 presents conditional beta pricing models; Section 3 describes the estimation method and presents the main asymptotic results; Section 4 deals with the implementation of the method; Section 5 presents the empirical application; and, finally, Section 6 concludes. The appendix contains the main proofs, tables and figures.

2 The Model

This section presents the general class of conditional beta pricing models followed by the specific model which provides the basis for the main results and the empirical application.

2.1 Conditional beta pricing models

In unconditional beta pricing models, asset returns are assumed to be driven by a set of common risk factors

$$R_{it} = \alpha_i + \beta_{i1}F_{1t} + \ldots + \beta_{ip}F_{pt} + \epsilon_{it}, \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{1}$$

where R_{it} denotes the return on asset i in excess of the risk-free interest rate in period t and $F_{\ell t}$ denotes the realization of the ℓth risk factor in period t, for $\ell = 1, \ldots, p$. Risk factors are assumed to be orthogonal to each other. The error term ϵ_{it} is serially independent with zero mean and nonsingular covariance matrix, conditional on factor realizations. The sample size of the time series is T, and N is the sample size of the cross section.

The standard asset pricing relation is then

$$E(R_{it}) = \gamma_1\beta_{i1} + \ldots + \gamma_p\beta_{ip} \tag{2}$$

where $E(R_{it})$ is the expected return on the ith asset and $\beta_{i1}, ..., \beta_{ip}$ are the coefficients from equation (1). Under the orthogonality condition of the factors, the *betas* represent the sensitivities of the asset's return to changes in the risk factors and are equal to the covariances between the factor and the asset return re-scaled by the variance of the risk factor. Hence, the coefficient γ_ℓ is interpreted as the reward per unit of beta risk associated with factor ℓ.

The first stage of the two-pass estimation procedure consists of estimating betas in equation (1) for each asset from a time-series regression. In the second stage, γ's are estimated as the slope coefficients of a cross-sectional regression of returns on estimated betas. See Shanken (1992) for an analysis of different aspects of the two-pass procedure and a derivation of the asymptotic distribution of the second-pass estimators, and Shanken and Zhou (2007) for a study of the small-sample properties of the methods and a comparison with alternative approaches.

Conditional beta pricing models, such as Harvey (1989), Jagannathan and Wang (1996) or Lettau and Ludvigson (2001), assume that (2) holds period by period so unconditional moments are replaced by conditional moments and the rewards per unit of beta risk are allowed to change over time. Noting that (2) can be rewritten as

$$E(R_{it}) = \gamma_1 \frac{Cov\,(R_{it}, F_{1t})}{Var(F_{1t})} + \ldots + \gamma_p \frac{Cov\,(R_{it}, F_{pt})}{Var(F_{pt})}, \tag{3}$$

then the conditional beta pricing model is

$$E(R_{it}|I_t) = \gamma_{1t} \frac{Cov\,(R_{it}, F_{1t}|I_t)}{Var(F_{1t}|I_t)} + \ldots + \gamma_{pt} \frac{Cov\,(R_{it}, F_{pt}|I_t)}{Var(F_{pt}|I_t)}, \tag{4}$$

where I_t represents investors' information set at the beginning of period t.

In empirical applications, the conditioning information set is replaced by an $m-$dimensional vector of observable variables $X_t = (X_{1t} \ldots X_{mt})^T$. Following Harvey (1989), we are also interested in estimating the reward per unit of covariance risk or market price of risk associated with the ℓth factor. Denoting by $\sigma_\ell^2(X_t)$ the conditional variance of the ℓth factor, and by $c_{i\ell}(X_t)$ the conditional covariance between the asset return and the risk factor, the market price of risk may be defined defined as $\lambda_{\ell t} \equiv \gamma_{\ell t}/\sigma_\ell^2(X_t)$ and the conditional beta pricing model can be rewritten as

$$E(R_{it}|X_t) = \lambda_{1t} c_{i1}(X_t) + \ldots + \lambda_{pt} c_{ip}(X_t) \qquad i = 1, 2, ..., N \quad t = 1, 2, ..., T, \tag{5}$$

which can be estimated from the following set of regression equations

$$R_{it} = \lambda_{1t} c_{i1t}(X_t) + \ldots + \lambda_{pt} c_{ipt}(X_t) + \varepsilon_{it} \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T, \tag{6}$$

where the market prices of risk are restricted to be equal across assets and ε_{it} denotes the error term.

2.2 A dynamic term-structure model

We next present a dynamic model of the term structure of interest rates that motivates the use of conditional beta pricing models and is the basis of our empirical application. This context is particularly interesting for our purposes for two main reasons. First, there is no consensus in the literature regarding the dynamics driving bond returns and market prices of risk (Duffee (2002) and Cheridito, Filipović and Kimmel (2007)). Flexible estimation of the market prices of risk may therefore shed further light on the reasons for this failure. Second, Ferreira and Gil-Bazo (2004) cannot reject the null hypothesis that U.S. bond expected returns are to be driven by a single priced risk factor. Since the method we propose yields estimates of the market price of risk associated with each factor, it also provides an alternative way to test for the number of priced risk factors.

Consider p state variables $\{X_{\ell t}\}_{\ell=1}^{p}$ capturing the state of the system. Each state variable is assumed to follow a general diffusion process

$$dX_{\ell t} = \mu_{\ell t}(t, X_t)dt + \sigma_{\ell t}(X_t)dW_{\ell t} \tag{7}$$

where the mean function $\mu_{\ell t}(t, X_t)$ is possibly time-varying and depending on the value of the $p-$dimensional vector $X_t = (X_{1t} \ldots X_{pt})^T$, at time t. For the sake of simplicity and without loss of generality, the diffusion function $\sigma_{\ell t}(X_t)$ is assumed to only depend on the state variables X_t. The term $dW_{\ell t}$ denotes the standard derivative of a Wiener processes assumed to be orthogonal to dW_{kt}, for $\ell \neq k$.

Within this context and, considering that N default-free bonds are available in the market, the system of equations for the instantaneous bond returns can be written as

$$\frac{dB_{it}}{B_{it}} = m_{it}(t, X_t)dt + dZ_{it} \quad i = 1, \ldots, N, \tag{8}$$

where the drift and the diffusion can be obtained using the multivariate version of Ito's lemma. In particular, the diffusion is given by

$$dZ_{it} = \sum_{\ell=1}^{p} s_{i\ell t} dW_{\ell t} \qquad \text{with} \qquad s_{i\ell t} B_{it} = \sigma_{\ell t} \frac{\partial B_{it}}{\partial X_{\ell t}}$$

where the arguments have been dropped out.

Assuming a market free of arbitrage opportunities, it can be shown that a vector $\gamma_t = (\gamma_{1t} \ldots \gamma_{pt})^T$ must exist so that

$$m_{it} = r_t + \gamma_{1t} \frac{\partial B_{it}/\partial X_{1t}}{B_{it}} + \ldots + \gamma_{pt} \frac{\partial B_{it}/\partial X_{pt}}{B_{it}}, \quad i = 1, \ldots, N \tag{9}$$

where r_t is the instantaneous risk-free interest rate. Thus, (8) becomes

$$\frac{dB_{it}}{B_{it}} = r_t dt + \sum_{\ell=1}^{p} \gamma_{\ell t} \frac{\partial B_{it}/\partial X_{\ell t}}{B_{it}} dt + \sum_{\ell=1}^{p} s_{i\ell t} dW_{\ell t}, \tag{10}$$

and, defining the conditional covariances, $c_{i\ell t} = Cov\left(\frac{dB_{it}}{B_{it}}, dX_{\ell t} | X_t\right)$,

$$\frac{dB_{it}}{B_{it}} = r_t dt + \sum_{\ell=1}^{p} \gamma_{\ell t} \frac{c_{i\ell t}}{\sigma_{\ell t}^2} + \sum_{\ell=1}^{p} s_{i\ell t} dW_{\ell t}. \tag{11}$$

If a time series of returns is discretely observed for N different bonds, for a sufficiently $(T+1)$ size dense sample of returns, a discretized version of (11) is:

$$R_{it} = \lambda_{1t} c_{i1t} + \lambda_{2t} c_{i2t} + \ldots + \lambda_{pt} c_{ipt} + \varepsilon_{it} \quad i = 1, 2, ..., N \quad t = 1, 2, ..., T, \tag{12}$$

where $R_{it} = \left\{ \left(B_{i(t+1)} - B_{it} \right) / B_{it} \right\} - r_t \Delta t$; Δt is the time step between observations (e.g. if r_t is the annualized interest rate and returns are measured over daily intervals, then $\Delta t = 1/250$), $c_{i\ell t}$ is the conditional covariance between R_{it} and $\Delta X_{\ell t}$, $\lambda_{\ell t}$ is defined as $\gamma_{\ell t}/\sigma^2_{\ell t}$, and the error terms ε_{it} follow a normal distribution. Note that defining the risk factors, F_ℓ, as $X_{\ell t+1} - X_{\ell t}$, the interpretations of $c_{i\ell t}$ and $\lambda_{\ell t}$ are the same as in (6).

It is convenient to remark that, straightforward from expression (11), it follows that the errors ε_{it} are heteroscedastic, serially independent, independent of the explanatory variables ($c_{i\ell t}$) and cross-sectionally related; that is, $E(\varepsilon_{it}\varepsilon_{jt}|X_t) \neq 0$ for $i \neq j$ and $E(\varepsilon_{it}\varepsilon_{js}) = 0$, for all i, j and $t \neq s$.

To estimate λ's in (12), the time series of returns for a single asset and the time-varying covariances between the asset's returns and the risk factors could be used. However, the error term in the set of equations are possibly cross-sectionally related and the absence of arbitrage imposes common market prices of risk. Therefore, it is more efficient to estimate all equations jointly as a Seemingly Unrelated Regression Equations (SURE) model, subject to the restriction that all equations have the same vector of coefficients; that is, for each $t = 1, \ldots, T$, $\lambda_{i\ell t} = \lambda_{j\ell t} = \lambda_{\ell t}$ for all $i, j = 1, \ldots, N$. Thus the model to be estimated is

$$\begin{aligned} R_{1t} &= \lambda_{1t}\ c_{11t} + \lambda_{2t} c_{12t} + \ldots + \lambda_{pt} c_{1pt} + \varepsilon_{1t} \\ R_{2t} &= \lambda_{1t}\ c_{21t} + \lambda_{2t} c_{22t} + \ldots + \lambda_{pt} c_{2pt} + \varepsilon_{2t} \\ &\vdots \\ R_{Nt} &= \lambda_{1t}\ c_{N1t} + \lambda_{2t} c_{N2t} + \ldots + \lambda_{pt} c_{Npt} + \varepsilon_{Nt} \end{aligned} \tag{13}$$

where $\{\lambda_{\ell t}\}_{\ell=1}^{p}$ are the market prices of risk to be estimated. The error term of the system, $\varepsilon_t = [\varepsilon_{1t}\ \varepsilon_{2t} \ldots \varepsilon_{Nt}]^T$ has zero mean and covariance matrix given by

$$E(\varepsilon_t \varepsilon_t^T | X_t) = \Omega_t = \begin{bmatrix} \sigma_{11t} & \sigma_{12t} & \cdots & \sigma_{1Nt} \\ \sigma_{21t} & \sigma_{22t} & \cdots & \sigma_{2Nt} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1t} & \sigma_{N2t} & \cdots & \sigma_{NNt}, \end{bmatrix}$$

where $\sigma_{ijt} = E(\varepsilon_{it}\varepsilon_{jt}|X_t)$ denotes the covariance, conditional on the value of the risk factors, between the error terms corresponding to different equations i and j at time t. Note that this context allows for heteroscedasticity ($\sigma_{iit} = E(\varepsilon^2_{it}|X_t)$) in each equation and for contemporaneous correlations ($\sigma_{ijt} = E(\varepsilon_{it}\varepsilon_{jt}|X_t)$). As mentioned above, all other correlations are zero.

3 Estimation procedure and main results

This section describes the proposed estimator for the coefficients $\lambda_{\ell t}$ in (13). For a better description of the procedure, consider some extra notation: $C_s = (C_{1s} \quad C_{2s} \dots C_{Ns})^T$ is a $N \times p$ order matrix where each term C_{is}^T denotes the row vector $(c_{i1s} \quad c_{i2s} \dots c_{ips})$, for $i = 1, \dots, N$. R_s is a N order column vector $(R_{1s}\ R_{2s} \dots R_{Ns})^T$, $\lambda_t = (\lambda_{1t} \dots \lambda_{pt})^T$ is the p order vector of unknown prices of risk. Finally the realized state-variable $p-$ order column vector is described as $X_t = (X_{1t} \dots X_{pt})^T$. Note that in our example the state variables and the risk factors have the same dimension (p). According to this notation model (13) can be compactly written as

$$R_t = C_t \lambda_t + \varepsilon_t \qquad t = 1, \dots, T. \tag{14}$$

Within this framework, the proposal consists of estimating the time-varying vector of market prices of risk at each time t, λ_t, taking into account the structure of the error covariance matrix, the equality constraints on the coefficients across assets and the assumed smoothness of the coefficients. In order to achieve this goal, we propose to estimate the market prices of risk by minimizing the weighted sum of squared residuals using all available observations:

$$\min_{\lambda_t} \sum_{s=1}^{T} K_{h,ts} (R_s - C_s \lambda_t)^T \Omega_s^{-1} (R_s - C_s \lambda_t), \tag{15}$$

where $K_{h,ts} = (Th)^{-1} K((t-s)/(Th))$, $K(\cdot)$ denotes the kernel weight used to introduce smoothness in the path of coefficients and $h > 0$ is the bandwidth that regulates the degree of smoothness. Solving the normal equations, the resulting estimator has the following closed form

$$\widehat{\lambda}_t = \left(\sum_{s=1}^{T} K_{h,ts} C_s^T \Omega_s^{-1} C_s \right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^T \Omega_s^{-1} R_s, \tag{16}$$

and it can be interpreted as a type of Smoothed Generalized Least Squares estimator. Note that considering the usual standardization $R_s^* = V_s^{-1} R_s$ and $C_s^* = V_s^{-1} C_s$, being V_s the matrix such that $V_s V_s^T = \Omega_s$, the optimization problem (15) can be written as

$$\min_{\lambda_t} \sum_{s=1}^{T} K_{h,ts} (R_s^* - C_s^* \lambda_t)^T (R_s^* - C_s^* \lambda_t), \tag{17}$$

and, therefore, the market prices of risk estimator (16) can be expressed in a more compact form

$$\widehat{\lambda}_t = \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^* \right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} R_s^*. \tag{18}$$

Remark 1 The role of the smoothing parameter h for time-varying coefficients has differences from that of the usual nonparametric regression. In our context, when h is large enough, no time variation is allowed and the resulting estimator leads to the same estimates as in a classical SURE model estimation with constant coefficients, subject to the equality constraints. That is,

$$\widehat{\lambda} = \left(\sum_{s=1}^{T} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} C_s^{*T} R_s^* = \left(\sum_{s=1}^{T} C_s^T \Omega_s^{-1} C_s\right)^{-1} \sum_{s=1}^{T} C_s^T \Omega_s^{-1} R_s. \tag{19}$$

On the contrary, when h is small enough, no smoothness is imposed, so the estimation of each $\widehat{\lambda}_t$ only takes into account the N observations corresponding to the same time period ($s = t$), and it will be estimated independently of the rest of observations ($s \neq t$). That is,

$$\widehat{\lambda}_t = (C_t^{*T} C_t^*)^{-1} C_t^{*T} R_t^* = (C_t^T \Omega_t^{-1} C_t)^{-1} C_t^T \Omega_t^{-1} R_t \tag{20}$$

which is equivalent to estimate independently the cross regressions for each time period.

Remark 2 There is a close relation between the estimator in (18) and the estimator proposed in Shanken (1985), and asymptotically studied in Shanken (1992). Considering constant coefficients ($h \to \infty$) as in (19) and assuming that the covariances between returns (R_{it}) and the risk factors ($X_{\ell(t+1)} - X_{\ell t}$) are time invariant, ($C_s = C \ \ \forall s$), the resulting estimator is

$$\widehat{\lambda} = \left(\sum_{s=1}^{T} C^{*T} C^*\right)^{-1} \sum_{s=1}^{T} C^{*T} R_s^* = (C^T \Omega^{-1} C)^{-1} \ C^T \Omega^{-1} \overline{R}. \tag{21}$$

If we substitute C and Ω by their estimators, respectively, $\widehat{\lambda}$ coincides with the GLS estimator proposed by Shanken (1985).

To study the properties of consistency and asymptotic normality of the general estimator (18), the $MASE$ (Mean Average Squared Error)

$$\begin{aligned} MASE(\hat{\lambda}_t) &= \sum_{\ell=1}^{p} MASE(\hat{\lambda}_{\ell t}) = \sum_{\ell=1}^{p} \left(Bias^2(\hat{\lambda}_{\ell t}) + Var(\hat{\lambda}_{\ell t})\right) \\ &\equiv S^2(\hat{\lambda}_t) + V(\hat{\lambda}_t) \end{aligned}$$

is studied and the following assumptions are considered.

***Assumption* (A1)** The market prices of risk are smooth functions of the time index; that is, $\lambda_{\ell t} = \lambda_\ell(t/T)$ where each λ_ℓ is a smooth function in $C^2[0, 1]$.

***Assumption* (A2)** The weight function $K(u)$ is a symmetric second order kernel with compact support $[-1, 1]$, Lipschitz continuous, and its Fourier transform is absolutely integrable, such that $\int u^2 K^2(u)du$ and $\int K^4(u)du$ are bounded.

***Assumption* (A3)** The conditional covariance can only vary with time through the state vector at time t, X_t. That is, $c_{i\ell t} = c_{i\ell}(X_t)$, where it is assumed that $c_{i\ell}$ is at least twice differentiable for all partial derivatives.

***Assumption* (A4)** Both C_{it} and X_{it} are statistically independent of ε_{is}, for all $s \geqslant t$. Moreover, we assume the process (14) with finite distributions such that the sequence $\{X_{it}, C_{it}, \varepsilon_{it}\}$ is strong α-mixing with coefficients $\alpha(k)$ of order 6/5; that is $\alpha(k) = O(k^{-\delta})$, with $\delta > 6/5$. All moments up to order $12+\theta$ exist and they are uniformly bounded, for some positive θ.
***Assumption* (A5)** At each time t, the unconditional expectation $E(C_t^{*T}C_t^*) = G_t$ is symmetric and strictly positive definite, and it can be decomposed as a smooth function of t/T, at least twice differentiable and uniformly bounded, plus a term of order $O(T^{-1})$.
***Assumption* (A6)** The error term ε_t has zero mean conditional on X_t, conditional covariance matrix $\Omega_t = E(\varepsilon_t \varepsilon_t^T | X_t)$ symmetric and positive definite.
***Assumption* (A7)** Let σ^{ijt} be a generic term in Ω_t^{-1}. The p-order matrix

$$\left(\sum_{s=1}^{T} \sum_{i,j=1}^{N} K_{h,ts}\ \sigma^{ijt} C_{i,s} C_{j,s}^T \right), \tag{22}$$

is positive definite and uniformly bounded from above and below.
***Assumption* (A8)** The smoothing parameter h goes to zero and Th goes to infinity, as the sample size T goes to infinity.

Assumption (A1) imposes smoothness on the market prices of risk. (A2) holds for technical reasons in kernel estimation. (A3) imposes smoothness on the explanatory variables. (A4) and (A5) ensure that the generating distribution process for the data is locally stationary, which allows for time-varying means, variances and also serial correlations. These types of processes are very useful and realistic since they can help to model nonstationary variables with a nonexplosive behavior (see Dalhaus (1997), Dalhaus (2000)). Smoothness over error's covariances is also assumed. (A6) excludes equations with exploiting variances and or with lineary dependent error terms and (A7) ensures that the estimator is identified. (A8) is standard in nonparametric estimation.

Theorem 1 *Under the set of assumptions (A1) to (A8), the MASE for the estimator defined in (18), has bias and variance,*

$$S^2(\widehat{\lambda}_t) = \frac{h^4 d_K^2}{4} \mid\mid \lambda_t'' + 2G_t^{-1} G_t' \lambda_t') \mid\mid_2^2 + o(h^4) \tag{23}$$

and

$$V(\widehat{\lambda}_t) = \frac{c_K}{Th}\ tr(G_t^{-1}) + o((Th)^{-1}) \tag{24}$$

where G_t' denotes the matrix with the derivatives of $G_t = E(C_t^{*T}C_t^*)$. The vectors λ_t' and λ_t'' contain the first and second derivatives of λ_t. The constants related with the kernel d_K and c_K are defined respectively as $d_K = \int u^2 K(u)du$ and $c_K = \int K^2(u)du$.

Remark 3 It is important to observe that under assumptions (A1) to (A8), the asymptotic order and the leading terms are the same considering either stationary or locally stationary variables.

Corollary 1 *Consider model (14) and a consistent estimator* $\widehat{\Omega}_s = \hat{V}_s \hat{V}_s^T$ *of* $\Omega_s = V_s V_s^T$*. Then, under the same assumptions in Theorem 1, and if either*

(i) $\widehat{\Omega}_s - \Omega_s = o(MASE(\widehat{\lambda}_t))$*, or*

(ii) The entries in C_s *are bounded*

the estimator

$$\widehat{\lambda}_t^F = \left(\sum_{s=1}^{T} K_{h,ts} C_s^T \widehat{\Omega}_s^{-1} C_s \right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^T \widehat{\Omega}_s^{-1} R_s$$

has the same asymptotic properties than the estimator (17).

All previous asymptotic results have been obtained under the assumption that the explanatory variables are observable and, therefore, they can be directly used in the estimation. However, this is not the case in the context of beta pricing models, in which explanatory variables are not directly observable and must be replaced by proxies. Moreover, the procedure to obtain them should ensure that the properties as the real unobserved variables are preserved.

Taking into account that each element $c_{i\ell t}$ of C_t measures the covariance between returns (R_{it}) and the risk factor ($X_{\ell(t+1)} - X_{\ell t}$) we propose to estimate $c_{i\ell t}$ as a rolling smoothed sample covariance

$$\widehat{c}_{i\ell t} = \widehat{c}_{i\ell}(X_t) = \left(\sum_{s=t-r}^{t-1} K_B(X_s - X_t) \right)^{-1} \sum_{s=t-r}^{t-1} K_B(X_s - X_t) P_{i\ell s}, \tag{25}$$

where we recall that $X_s = (X_{1s} \dots X_{ps})^T$ are the state variables. We define $P_{i\ell s} = R_{is}\Delta X_{\ell s} - \mu_{R_i}(X_s)\mu_{\Delta X_\ell}(X_s)$ being $\mu_{R_i}(X_s)$ and $\mu_{\Delta X_\ell}(X_s)$ the estimated means of R_i and ΔX_ℓ conditional on X_s. K_B is a p-variate kernel $K_B(u) = |B|^{-1/2} K(B^{-1/2}u)$, with smoothing matrix B. Since $E(R_{is}\Delta X_{\ell s}|X_s) - E(R_{is}|X_s)E\left(\Delta X_{\ell s}|X_s\right) = Cov(R_{is}, \Delta X_{\ell s}|X_s) = c_{i\ell}(X_s)$, it can be written that $R_{is}\Delta X_{\ell s} - \mu_{R_i}(X_s)\mu_{\Delta X_\ell}(X_s) = c_{i\ell}(X_s) + u_s$ where $E(u_s|X_s) = 0$. That is, (25) can be read as a one-sided conditional nonparametric estimator in a time series model. It therefore becomes clear that to keep the results, it is crucial to employ a truncated estimator that only uses past information.

Thus, the resulting estimator for the market prices of risk (λ_t) is

$$\widehat{\lambda}_t^{SGLS} = \left(\sum_{s=1}^{T} K_{h,ts} \widehat{C}_s^{*T} \widehat{C}_s^{*} \right)^{-1} \sum_{s=1}^{T} K_{h,ts} \widehat{C}_s^{*T} R_s^{*}, \tag{26}$$

similar to (18), where C is replaced by $\widehat{C}$ and $\widehat{C}_s^* = V_s^{-1}\widehat{C}_s$. Some additional assumptions are required in order to reach the desirable asymptotic results:

Assumption **(C1)** The p-variate kernel K is compactly supported such that $\int K(\mathrm{u})d\mathrm{u} = 1$ and $\int \mathrm{u}\mathrm{u}^T K(\mathrm{u})d\mathrm{u} = \mu_K I_p$, being μ_K a nonnegative scalar where $\int$ is the shorthand for $\int\int \ldots \int_{\mathbb{R}^p}$ and $d\mathrm{u}$ for $du_1 \ldots du_p$.

Assumption **(C2)** Consider a sequence of positively definite diagonal bandwidth matrices $B = diag(b_{i1}^2 \quad b_{i2}^2 \ldots b_{ip}^2)$ for $i = 1, \ldots, N$, such that $|B|^{1/2}$ and $r|B|^{1/2}$ go to zero, $T|B|^{1/2}$ and r go to infinity and r/T goes to zero as the sample size, T, goes to infinity. Note that the bandwidth matrices are considered to be equal for all i, to simplify notation and without loss of generality.

Assumption **(C3)** The distribution of X_t has a Lipschitz of order one time-varying density, $f_t(x) = f(\tau, x)$, where $\tau = t/T$.

Next proposition states the properties for estimator (25).

Proposition 1 *Consider the set of assumptions (A3) to (A6), and (C1) to (C3) then, the estimator defined by (25) is a consistent estimator of* $c_{i\ell}(X_t)$*, with asymptotic bias and variance:*

$$\begin{aligned} Bias(\hat{c}_{i\ell}(X_t|X_t = x_t) &= O(trace(B)) \\ Var(\hat{c}_{i\ell}(X_t|X_t = x_t) &= O\left(\frac{1}{r|B|^{1/2}}\right). \end{aligned}$$

We are now in a position to derive the asymptotic results for the estimator of the market prices of risk when the previous proxies are employed.

Theorem 2 *Under the set of assumptions (A1) to (A8) and (C1) to (C3), using the covariance estimator* $(\widehat{C}_t)$ *defined in (25), the estimator for the market prices of risk* $(\hat{\lambda}_t^{SGLS})$ *defined in (26) is consistent, with the same asymptotic results for the two components of the MASE as in Theorem 1.*

Corollary 2 *Consider model (14) with the consistent estimator of* C_t *defined in (25) and a consistent estimator* $\widehat{\Omega} = \hat{V}_s\hat{V}_s^T$ *for* $\Omega_s = V_s V_s^T$*. Then, under the assumptions in Theorem 2, and if either*

(i) $\widehat{\Omega}_s - \Omega_s = o(MASE(\widehat{\lambda}_t^{SGLS}))$*, or*

(ii) the entries in C *are bounded*

the SFGLS (Smoothed Feasible Generalized Least Squares) estimator

$$\widehat{\lambda}_t^{SFGLS} = \left(\sum_{s=1}^{T} K_{h,ts}\widehat{C}_s^T\widehat{\Omega}_s^{-1}\widehat{C}_s\right)^{-1} \sum_{s=1}^{T} K_{h,ts}\widehat{C}_s^T\widehat{\Omega}_s^{-1}R_s \tag{27}$$

has the same asymptotic properties as in the previous theorems.

The following proposition provides a consistent estimator for the error covariance matrix that must be estimated in advance in order to compute the estimated market prices of risk defined in (27).

Proposition 2 *Consider the estimator for a generic element of the covariance matrix,*

$$\widehat{\sigma}_{ijt} = \left(\sum_{s=1}^{T} K_G(X_s - X_t)\right)^{-1} \sum_{s=1}^{T} K_G(X_s - X_t)(R_{it} - \widehat{\lambda}_t \widehat{C}_{it})^T (R_{jt} - \widehat{\lambda}_t \widehat{C}_{jt}) \tag{28}$$

with $K_G(u) = |G|^{-1/2} K(G^{-1/2}u)$*, being* G *the* p*-order smoothing matrix and* K *a* p*-variate second order kernel. Under assumptions (A1)-(A8), (C3) and the kernel* K_G *satisfying (C1) and (C2) (although no assumption for* r *is needed here), (28) provides a consistent estimator for a generic term of* Ω_t*, for each* t*.*

The next asymptotic distribution (pointwise) for the estimator of λ_t, allows us to test for invariance of the prices of the risk factors through time or to test whether or not the risk premium can be considered significantly non-zero.

Theorem 3 *Assume (A1)-(A8) and (C1)-(C3), consider* $h = o(T^{-1/5})$*, such that the bias tends to zero faster than the variance, and that either (i) or (ii) in Corollary 2 holds. Then, the estimator of* λ_t *at* k *different locations* $t_1, \ldots, t_k$ *converges in distribution to the multivariate normal as,*

$$\left((Th)^{1/2}(\hat{\lambda}_{t_j}^{SGLS} - \lambda_{t_j})\right)_{j=1}^{k} \xrightarrow{p} N(0, c_K G_{t_j}^{-1}) \tag{29}$$

Finally, using the consistent estimator for G_{t_j} *defined in Lemma 1, we can obtain confidence intervals for the* k *selected* λ*'s.*

4 Implementation

The proposed estimator for the seemingly unrelated regression equations model with unknown explanatory variables requires the selection of several smoothing parameters: the matrix of bandwidths B related with the estimation of the proxies; the smoothing parameters related with the time-varying market prices of risk and the smoothing parameter to estimate the covariance matrix.

In general situations, the bandwidths are selected using data driven methods like cross-validation, penalized sum of squared residuals or plug-in methods. For a detailed discussion of each see Härdle (1990), Wand and Jones (1995) or Fan and Gijbels (1996) among others. For multivariate cases, the penalty methods as Rice or Generalized Cross-Validation are appropriate, easy to interpret and faster to compute than the others.

All above data driven methods are based on the fact that minimizing the sum of squared residuals is not adequate for selecting a smoothing parameter. It is well known that a sum of squared residuals equal to zero is easily obtained for bandwidths very close to zero. Nevertheless, in this context, minimizing the sum of squared residuals does not have the same meaning as usual. First, with regards to the smoothness over time imposed on the coefficients (h), it is true that as the value of h increases, the larger is the imposed degree of smoothness, so coefficients become constant eventually. But a bandwidth that tends to zero does not correspond to a zero value for the sum of squared residuals due to the restriction of equality imposed on the coefficients across the N equations ($\lambda_{i\ell t} = \lambda_{j\ell t}$). In this panel setting, a bandwidth close to zero implies that for each time t, coefficients are estimated with a sample of N cross sectional observations. Second, regarding the smoothing parameters involved in the estimation of the proxies, the proposed estimator, by definition, does not include the observation corresponding to each time t when estimating at that point.

To solve the selection problem in practice, we proceed as follows. Since the objective is the estimation of λ's, and in order to increase the dispersion of the proxies (leading to a greater explanatory power) the selection of h and B is addressed jointly. After, the parameter selection for the covariance matrix is addressed.

For the first step—the joint selection of the smoothness parameters for lambdas and proxies—we propose to minimize a penalized sum of squared residuals

$$(NT)^{-1}\sum_{t=1}^{T}(R_t - \widehat{C}_t(B)\widehat{\lambda}_t(h))^T(R_t - \widehat{C}_t(B)\widehat{\lambda}_t(h))\ \mathcal{G}(h,B) \tag{30}$$

where the notation makes the dependence of the smoothing parameters of the estimated $\widehat{C}_t$'s and $\widehat{\lambda}_t$'s explicit, and $\mathcal{G}(h,B)$ denotes the penalizing function. Since the estimator of the covariances defined in (25) does not consider the observation at time t for the estimation, there is no need to penalize the selection of B. Thus, we will use $\mathcal{G}(h,B) = \mathcal{G}(h)$ that only accounts for the h parameter.

If we consider Generalized Cross Validation (GCV) method, then the penalty is

$$\mathcal{G}(h) \approx \left(1 - (NT)^{-1} trace P(h)\right)^{-2} \tag{31}$$

where $P(h)$ is the projection matrix

$$\frac{K(0)}{Th}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[C_{is}^T\left(X_i'\mathcal{K}_{h,t}X_i\right)^{-1}X_i^T\right]Z_t, \tag{32}$$

and $C_{is} = (c_{i1t}\ c_{i2t}\dots c_{ipt})^T$ has already been defined, the $T\times p$ order matrix $X_i = (C_{i1}\ C_{i2}\dots C_{iT})^T$ is the data matrix corresponding to the ith equation, $\mathcal{K}_{h,t} = diag\{K_{h,ts}\}_{s=1}^T$ is a T order diagonal matrix with kernel weights, and Z_t is a T order column vector with tth element equal to one and rest of elements equal to zero.

Once the first parameters h and B have been selected, the second step is to select the smoothing parameter matrix G for the error's covariance matrix. For fixed h and B, we propose to select G minimizing the weighted sum of squared residuals

$$\sum_{t=1}^{T}(R_t - \widehat{C}_t\widehat{\lambda}_t)^T\widehat{\Omega}_t^{-1}(R_t - \widehat{C}_t\widehat{\lambda}_t) \tag{33}$$

where the estimator of any generic term of $\widehat{\Omega}$, $\widehat{\sigma}_{ijt}$, is given by (28).

Taking into account the estimated $\widehat{\Omega}$ with the selected G, once the error covariance matrix is plugged in the final estimator, the new market prices of risk are reestimated by (27). For this reason, it is possible that the smoothing parameters selected in the first step are not optimal. To avoid this, it is convenient to perform some iterations in order to refine the method. Hence, we suggest fixing h and B but reestimate the Ω matrix using the residuals coming from the new estimated λ's. As in parametric context, this procedure should converge to a final covariance matrix and, consequently, to final λ's. However, changing h and B in this iterative procedure does not provide a convergent method. We do not ensure an adequate identification of the systematic part of the model.

5 Empirical application

In this section, the method is applied to estimate the term structure model presented in Subsection 2.2. The data set consists of 9,485 daily observations of the term structure of U.S. Treasury interest rates covering the period from January 1969 to December 2006. More specifically, *Treasury Constant Maturity Rates* are used as proxies for default-free interest rates of different maturities. The data are available from the Federal Reserve Bank of St. Louis' FRED online database. These rates are interpolated by the Treasury from the yield curve, which is estimated on a daily basis using a cubic spline model from closing market bid yields on actively traded Treasury securities in the over-the-counter market. Yields for different maturities are displayed on Figure 1 and descriptive statistics are shown in Table 1.

To construct the series of excess bond returns, we first recover bond prices from yields, and then compute one-period returns from buying one-year, three year, five-year and ten-year zero-coupon bonds and selling them one day later. We then subtract the (continuously compounded) three-month interest rate, which we take as a proxy for the short term interest rate. Figure 2 displays the four series of daily excess returns. These graphs are suggestive of the presence of conditional heteroscedasticity, which justifies the use of a time-varying covariance matrix. Table 2 contains the descriptive statistics for excess bond returns.

In order to estimate conditional covariances, a set of state variables needs to be selected, that are observable at the beginning of each period. In our model, all available information will ultimately be reflected in the term structure of interest rates. Observation of the yield curve is, therefore, sufficient to know the state of the system. We further assume that the yield curve can be summarized by two variables: the *level* and the *slope* of the yield curve, as proxied by the continuously compounded three-month rate, and the spread between the 10-year rate and the 3-month rate, respectively, both continuously compounded. The two series are displayed for our sample period in Figure 3. Risk factors are defined accordingly as changes in the level and changes in the slope of the yield curve. In order to guarantee factor orthogonality, we regress by OLS changes in the spread on changes in the short rate and take the residuals from the regression as the second risk factor.

In the first pass, we estimate the covariances between each bond's returns and the risk factors conditional on the state variables following the procedure defined in Section 3, equation (25). The results are shown in Figures 4 and 5. For the sake of brevity, only conditional covariances with changes in the short rate are shown. In particular, Figure 4 plots estimated conditional covariances against the short rate holding the term spread constant and equal to its sample median value and Figure 5 plots estimated conditional covariances against the spread holding the short rate constant and equal to its sample median value.

The graphs in Figures 4 and 5 suggest that conditional covariances are nonlinear functions of the state variables. This intermediate result is important *per se* since it contradicts the dynamic assumptions of some popular interest rate models. For instance, the one-factor model of Vasicek (1977) implies constant covariances, while Cox et al. (1985) implies that covariances are linear in the single risk factor.

Once conditional covariances have been estimated, we may estimate the market prices of risk using (27). Figure 6 displays the evolution of market prices of covariance risk associated with each risk factor in our sample period. A number of conclusions can be drawn from Figure 6. First, changes in the level of the yield curve appear to play a bigger role in determining differences in risk premia across bonds. To see this, note from Figures 4 and 5 that covariance risk is about ten times larger for the first factor than for the second factor. Figure 6 further shows that the absolute value of the market price of risk is not smaller on average for the first factor throughout the sample period and is higher when specific intervals within the sample period are considered. This is confirmed by Table 3, which shows market prices of risk on the first day of each year and the corresponding asymptotic t-statistics. The market price of risk associated with the first factor is statistically significant (at least at the 5 percent significance level) for 16 percent of the dates.

Second, although the market price of risk associated with changes in the short rate is generally negative–which implies a positive risk premium since conditional covariances are negative–it changes signs and becomes positive, particularly in the period 2004–2006, a period of low (and very stable) interest rates. Table 3 confirms that the market price of risk at the beginning of 2005 was positive and statistically significant. This can be taken as further evidence against the completely affine class of term structure models which do not allow for the market price of covariance risk to be positive (Duffee (2002)). The estimates of λ_{1t} take the most negative values in 1970, 1991, and 2000. Interestingly, all three peaks in the market price of interest rate risk correspond to a level of the short term interest rate of 6 percent, which is roughly the sample average. Finally, the market price of interest rate risk appears to have been stable and closer to zero in the so-called Volcker years (1979–1987), characterized by high (and volatile) interest rates. Both the high positive value of lambda in the period 2004–2006 and the low absolute value in the period 1979–1987 are consistent with a relatively stable risk premium (expected return in excess of the risk free rate) associated with the risk of changes in the short rate. This would imply a higher absolute value of the market price of risk in those periods in which interest rate risk is lower and a lower absolute value of market price of risk when interest rate risk is higher.

Third, the market price of risk associated with changes in the spread appears to play no role in the determination of bond expected returns. Table 3 confirms this observation: the market price of risk associated with the second factor is never statistically significantly different from zero. This finding is consistent with Ferreira and Gil-Bazo (2004) who report evidence that a single priced risk factor (not necessarily associated with changes in the short term interest rate risk) is sufficient to explain daily bond premia.

Finally, to assess the economic significance of compensation for risk associated with each risk factor, in Figures 7 and 8, we plot the time series of estimated risk premia for changes in the short rate and changes in the spread, computed as the product of each market price of risk and the corresponding covariance. Results confirm that the risk premium for bearing interest rate risk is not only more statistically significant than the risk premium associated with changes in the spread, but also that it accounts for a larger fraction of total risk premium.

6 Summary and conclusions

In this paper, we estimate consistently the time-varying parameters of a very general conditional beta pricing model. The proposed nonparametric estimation procedure for a SURE model makes it possible to estimate market-prices of risk from observed asset returns without imposing any parametric structure on the asset return dynamics or the dependence of the market price of risk function on time or the state of the system. The method can be seen as a nonparametric analogue of the two-pass approach developed by Black et al. (1972), Fama and MacBeth (1973), Shanken (1985) and Shanken (1992) to estimate and test unconditional beta pricing models.

Similarly to the nonparametric method proposed by Wang (2003), the estimation method proposed in this paper is not subject to Ghysels' critique (Ghysels (1998)) who states that misspecification of time-varying conditional moments and market prices of risk may induce larger pricing errors than those obtained by unconditional beta pricing models. The method can be applied to a much more general family of models than the one considered in Wang (2003). At the same time, the procedure retains the simplicity and intuitive approach of the two-pass estimator, commonly used to estimate and test unconditional models.

The application of the method to the U.S. Treasury bond data, yields a number of interesting insights. First, conditional covariances appear to be highly non-linear in the yield curve, which casts doubt on the empirical plausibility of previous attempts to model conditional covariances as linear functions of the state variables. Second, only one risk factor, related to changes in the short term interest rate, appears to be priced by the markets, consistently with the evidence in Ferreira and Gil-Bazo (2004). Further, the results provide evidence of changes of sign in the market price of risk associated with changes in the short rate.

Acknowledgements

The authors would like to thank Alfredo Ibañez for his comments and suggestions. This work has been supported by the MEC grant SEJ2005-05549 and SEJ2007-67448, BBVA grant 1/BBVA 00038.16421/2004, and Grupo MACLAB (IT-241-07) Basque Government.

Appendix

In order to prove Theorem 1 the following lemma are first needed.

Lemma 1 *Under Assumptions (A2) to (A5), and (A8), it holds that*

$$\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^* \xrightarrow{a.s.} G_t$$

$$(Th)\sum_{s=1}^{T} K_{h,ts}^2 C_s^{*T} C_s^* \xrightarrow{p} c_K G_t$$

Proof of Lemma 1

For the ease of notation consider any generic scalar term of $TK_{h,ts}C_s^{*T}C_s^*$ as $Z_s = TK_{h,ts}c_s^*c_s^*$. Z_s is a $\alpha-$mixing sequence of size 6/5 with the proper bounded moments, and $E(Z_s) = \frac{1}{h}K\left(\frac{t-s}{Th}\right)g_s$, that tends to g_t. Therefore, applying the SLLN in White (1984), Corollary 3.48 for dependent variables under mixing conditions the first result follows. For the second use similar steps. □

Proof of Theorem 1

First we write the Mean Average Squared Error

$$\begin{aligned} MASE(\widehat{\lambda}_t) &= trE[(\widehat{\lambda}_t - \lambda_t)(\widehat{\lambda}_t - \lambda_t)^T] \\ &= \|\, Bias(\widehat{\lambda}_t)\,\|_2^2 \; + trVar(\widehat{\lambda}_t) = S^2(\widehat{\lambda}_t) + V(\widehat{\lambda}_t). \end{aligned}$$

Then, note that the estimator of λ_t

$$\begin{aligned} \widehat{\lambda}_t - \lambda_t &= \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} R_s^* - \lambda_t \\ &= \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^* (\lambda_s - \lambda_t) \\ &\quad + \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s \end{aligned}$$

has a random denominator. We overcome this problem working with a redefined bias and variance term, using the weight $W_t^* = G_t^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*$. Hence, the redefined bias is $Bias^*(\widehat{\lambda}_t) = Bias(W_t^* \widehat{\lambda}_t)$. For technical reasons, we use different bandwidths for W_t^* and for $\widehat{\lambda}_t$, say h^* and h respectively, such that condition next holds:

$$\frac{E \parallel W_t^* - I \parallel^2}{E \parallel \widehat{\lambda}_t - \lambda_t \parallel^2} = o(1) \tag{34}$$

as T goes to infinity. This condition establishes that W_t^* goes to the identity at a faster rate than the mean square error goes to zero, and this implies that the rate of convergence for the mean square error must be suboptimal; for this case it means slower than $T^{-4/5}$.

Considering the term defined by $Bias^*$

$$\begin{aligned} Bias^*(\widehat{\lambda}_t) &= G_t^{-1} \sum_{s=1}^{T} K_{h,ts} E(C_s^{*T} C_s^*)(\lambda_s - \lambda_t) + G_t^{-1} \sum_{s=1}^{T} K_{h,ts} E(C_s^{*T} \varepsilon_s^*) \\ &= G_t^{-1} \sum_{s=1}^{T} K_{h,ts} E(C_s^{*T} C_s^*)(\lambda_s - \lambda_t) \end{aligned}$$

since $E(C_s^{*T} \varepsilon_s^*) = E(C_s^{*T} E(\varepsilon_s^* | C_s^{*T})) = 0$. Using the Taylor expansion with $t - s = Thu$,

$$\begin{aligned} Bias^*(\widehat{\lambda}_t) &= G_t^{-1} \sum_{s=1}^{T} K_{h,ts} G_s(\lambda_s - \lambda_t) \\ &= G_t^{-1} \int K(u)[G_t - huG_t' + o(h^2)][-\lambda_t' hu + \frac{1}{2}\lambda_t''(hu)^2 + o(h^2)] \\ &= \frac{1}{2} d_k h^2 (\lambda_t'' + 2G_t^{-1} G_t' \lambda_t') + o(h^2) \end{aligned}$$

where G_t' denotes the matrix with the derivatives of G_t and λ_t' and λ_t'' denote the vectors for the first and second derivatives of λ respectively. Thus

$$S^2(\widehat{\lambda}_t) = \frac{1}{4} d_k^2 h^4 \mid\mid \lambda_t'' + 2G_t^{-1} G_t' \lambda_t' \mid\mid_2^2 + o(h^4).$$

The variance term

$$\begin{aligned} Var(\widehat{\lambda}_t) &= Var(\widehat{\lambda}_t - \lambda_t) = Var\left[\left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} R_s^*\right] \\ &= Var\left[\left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*(\lambda_s - \lambda_t)\right. \\ &\quad \left. + \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*\right)^{-1} \sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s\right] \end{aligned}$$

and using the redefined variance term, $Var^*(\widehat{\lambda}_t) = Var(W_t^* \widehat{\lambda}_t)$, we have that

$$Var^*(\widehat{\lambda}_t) = Var\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*(\lambda_s - \lambda_t)\right] + Var\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s\right] \tag{35}$$

and since the cross term cancels due to $E(\varepsilon_s|C_s^*) = 0$, the sum of variances can be split into two terms:

$$\begin{aligned} V(\widehat{\lambda}_t) &= trVar^*(\widehat{\lambda}_t) = trVar\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*(\lambda_s - \lambda_t)\right] \\ &+ trVar\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s\right] = V_1 + V_2. \end{aligned} \tag{36}$$

For the first term and taking into account that $G_s = E(C_s^{*T} C_s^*)$ we have that

$$\begin{aligned} V_1 &= trVar\left(G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*(\lambda_s - \lambda_t)\right) = trG_t^{-1} Var\left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} C_s^*(\lambda_s - \lambda_t)\right) G_t^{-1} \\ &= trG_t^{-1} E\left[\sum_{s=1}^{T} K_{h,ts}\left(C_s^{*T} C_s^* - G_s\right)(\lambda_s - \lambda_t)\left(\sum_{s=1}^{T} K_{h,ts}\left(C_s^{*T} C_s^* - G_s\right)(\lambda_s - \lambda_t)\right)^T\right] G_t^{-1} \\ &= trG_t^{-1} E\left(\sum_{s=1}^{T}\sum_{s'=1}^{T} K_{h,ts} K_{h,ts'}\left(C_s^{*T} C_s^* - G_s\right)(\lambda_s - \lambda_t)(\lambda_{s'} - \lambda_t)^T\left(C_{s'}^{*T} C_{s'}^* - G_{s'}\right)\right) G_t^{-1} \\ &= tr\sum_{s=1}^{T}\sum_{s'=1}^{T} K_{h,ts} K_{h,ts'}(\lambda_s - \lambda_t)(\lambda_{s'} - \lambda_t)^T E\left[\left(C_{s'}^{*T} C_{s'}^* - G_{s'}\right) G_t^{-1} G_t^{-1}\left(C_s^{*T} C_s^* - G_s\right)\right] \\ &= tr\sum_{s=1}^{T}\sum_{s'=1}^{T} K_{h,ts} K_{h,ts'}(\lambda_s - \lambda_t)(\lambda_{s'} - \lambda_t)^T Q_{s,s'} \end{aligned} \tag{37}$$

where $Q_{s,s'} = E\left[\left(C_{s'}^{*T} C_{s'}^* - G_{s'}\right) G_t^{-1} G_t^{-1}\left(C_s^{*T} C_s^* - G_s\right)\right]$ is a bounded p order square matrix. Expression (37) can be divided in two, those corresponding to same terms and the cross terms. When $s = s'$ we have that

$$tr\sum_{s=1}^{T} K_{h,ts}^2(\lambda_s - \lambda_t)(\lambda_s - \lambda_t)^T Q_{s,s}$$

where Q_{ss} is bounded and has same order than

$$\begin{aligned} \sum_{s=1}^{T} K_{h,ts}^2(\lambda_s - \lambda_t)(\lambda_s - \lambda_t)^T &= (Th)^{-1}\int K^2(u)(-hu\lambda_t' + o(h))(-hu\lambda_t' + o(h))^T du = \\ &= (Th^{-1})h^2\lambda_t'(\lambda_t')^T\left(\int u^2 K^2(u)du\right) + o(h^2) = O\left(\frac{h}{T}\right). \end{aligned}$$

For the cross terms, $s \neq s'$

$$tr \sum_{\substack{s,s'=1 \\ s\neq s'}}^{T} K_{h,ts} K_{h,ts'} (\lambda_s - \lambda_t)(\lambda_{s'} - \lambda_t)^T Q_{s,s'}$$

has same order than

$$\sum_{\substack{s,s'=1 \\ s\neq s'}}^{T} K_{h,ts} K_{h,ts'} (\lambda_s - \lambda_t)(\lambda_{s'} - \lambda_t)^T = O\left(\frac{h}{T} + \frac{1}{T^2}\right).$$

Thus $V_1 = O\left(\frac{h}{T} + \frac{1}{T^2}\right).$

For the second term in (36) and taking into account that $E(\varepsilon_s^*|C_s^*) = 0$:

$$\begin{aligned}
V_2 &= trVar\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^*\right] = trG_t^{-1} Var\left[\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^*\right] G_t^{-1} \\
&= trG_t^{-1} E\left[\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^* \left(\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^*\right)^T\right] G_t^{-1} \\
&= trG_t^{-1} E\left[\sum_{s=1}^{T}\sum_{s'=1}^{T} K_{h,ts} K_{h,ts'} C_s^{*T} \varepsilon_s^* \varepsilon_{s'}^* C_{s'}^*\right] G_t^{-1}.
\end{aligned}$$

Now, since $E(\varepsilon_{is}\varepsilon_{js'}) = 0$ for all $s \neq s'$, $E(\varepsilon_{is}^*\varepsilon_{js'}^*) = 0$ and

$$\begin{aligned}
V_2 &= trVar\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^*\right] = trG_t^{-1} E\left[\sum_{s=1}^{T} K_{h,ts}^2 C_s^{*T} \varepsilon_s^* \varepsilon_s^* C_s^*\right] G_t^{-1} \\
&= trG_t^{-1} E\left[\sum_{s=1}^{T} K_{h,ts}^2 C_s^{*T} E\left(\varepsilon_s^* \varepsilon_s^* |\, C_s^*\right) C_s^*\right] G_t^{-1}.
\end{aligned}$$

Then, as $E\left(\varepsilon_s^* \varepsilon_s^* |\, C_s^*\right) = I$ and using the result (34) of Lemma 1 we have that

$$V_2 = trVar\left[G_t^{-1}\sum_{s=1}^{T} K_{h,ts} C_s^{*T} \varepsilon_s^*\right] = \frac{c_k}{Th}\, trG_t^{-1} G_t G_t^{-1} + o((Th)^{-1}) = \frac{c_k}{Th}\, trG_t^{-1} + o((Th)^{-1}).$$

Finally, since the order of V_1 is negligible with respect to V_2 and (34) holds, we have

$$V(\widehat{\lambda}) = \frac{c_k}{Th}\, trG_t^{-1} + o((Th)^{-1}),$$

which provides that the order of the leading term in the variance coincides with the order for the variance term in standard results. □

Proof of Corollary 1

Either condition (i) or (ii) provides, together with the rest of assumptions, the sufficient conditions of regularity to check that the convergence of $\widehat{\Omega}_s$ to Ω_s implies the equivalence between the asymptotic properties of $\widehat{\lambda}_t^F$ and $\widehat{\lambda}_t$ □

Proof of Proposition 1

In order to deal with the random denominator, a modified bias is defined,

$$Bias^{\star}(\widehat{c}_{i\ell}(X_t)|X_t = x_t) = Bias\left[\frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1} K_B(X_s - x_t)\widehat{c}_{i\ell}(x_t)\right]$$

with $\tau = t/T$. Then

$$Bias^{\star}(\widehat{c}_{i\ell}(X_t)|X_t = x_t) \quad = \quad \frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\Big(E\left[K_B\left(X_s - x_t\right)P_{i\ell s}\right] - f(\tau, x_t^T)c_{i\ell}(x_t)\Big)$$

and since $P_{i\ell s} = c_{i\ell}(X_s) + u_s$ being $E(u_s|X_s) = 0$:

$$\begin{aligned} Bias^{\star}(\widehat{c}_{i\ell}(X_t)|X_t = x_t) &= \frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\Big(E\left[K_B(X_s - x_t)c_{i\ell}(X_s)\right] - f(\tau, x_t^T)c_{i\ell}(x_t)\Big) \\ &= \frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\left[\int K_B\left(\omega - x_t\right)c_{i\ell}(\omega)f\left(s/T, \omega^T\right)d\omega - f(t/T, x_t^T)c_{i\ell}(x_t)\right]. \end{aligned}$$

Now, define $f_s(x) = f(s/T, x)$, $\mathcal{D}_c(x_t) = \left(\frac{\partial c_{i\ell}(x_t)}{\partial x_{1t}}\dots\frac{\partial c_{i\ell}(x_t)}{\partial x_{pt}}\right)^T$, $\mathcal{D}_{f_s}(x_t^T) = \left(\frac{\partial f_s(x_t^T)}{\partial x_{1t}}\dots\frac{\partial f_s(x_t^T)}{\partial x_{pt}}\right)$ and the $(p\times p)$ order matrices $\mathcal{H}_c(x_t)$ and $\mathcal{H}_{f_s}(x_t^T)$ having as generic terms, (j, j'), $\frac{\partial^2 c_{i\ell}(x_t)}{\partial x_{jt}\partial x_{j't}}$ and $\frac{\partial^2 \partial f_s(x_t^T)}{\partial x_{jt}\partial x_{j't}}$ respectively. Using a standard multivariate kernel of order two and the Lipschitz condition for the density f; we have that

$$\begin{aligned} Bias^{\star}(\widehat{c}_{i\ell}(X_t)|X_t = x_t) &= \frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\left[\int K(z)c_{i\ell}(x_t + B^{1/2}z)f_s\left(x_t^T + B^{1/2}z^T\right)dz - f_t(x_t^T)c_{i\ell}(x_t)\right] \\ &= \frac{1}{rf(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\left[\int K(z)\left(c_{i\ell}(x_t) + (B^{1/2}z)^T\mathcal{D}_c(x_t) + \frac{1}{2}(B^{1/2}z)^T\mathcal{H}_c(x_t)(B^{1/2}z) + o(trace(B))\right)\times\right. \\ &\left.\left(f_s(x_t^T) + (B^{1/2}z^T)^T\mathcal{D}_{f_s}(x_t^T) + \frac{1}{2}(B^{1/2}z^T)^T\mathcal{H}_{f_s}(x_t^T)(B^{1/2}z^T) + o(trace(B))\right)dz - f_t(x_t^T)c_{i\ell}(x_t)\right] \end{aligned}$$

$$= \frac{1}{rf(\tau, x_t^T)} \sum_{s=t-r}^{t-1} \frac{1}{2} tr\left(BH_c(x_t) \int K(z) z z^T dz \right) f_s(x_t^T)$$

$$+ \frac{1}{rf(\tau, x_t^T)} \sum_{s=t-r}^{t-1} c_{i\ell}(x_t) \left(f_s(x_t^T) - f_t(x_t^T) \right) + O(trace(B))$$

$$= \frac{1}{rf(\tau, x_t^T)} \frac{\mu_K}{2} tr\left(BH_c(x_t) \right) \sum_{s=t-r}^{t-1} f_s(x_t^T) + \frac{1}{rf(\tau, x_t^T)} \sum_{s=t-r}^{t-1} c_{i\ell}(x_t) \left(f_s(x_t^T) - f_t(x_t^T) \right) + O(trace(B))$$

$$= \frac{\mu_K}{2} tr\left(BH_c(x_t) \right) + \frac{1}{rf(\tau, x_t^T)} \sum_{s=t-r}^{t-1} c_{i\ell}(x_t) O\left(\frac{r}{T} \right) + O(trace(B))$$

$$= O(trace(B)) + O\left(\frac{r}{T} \right)$$

Next, we obtain the redefined variance for a generic term $\widehat{c}_{i\ell}(X_t)$:

$$Var^\star(\widehat{c}_{i\ell}(X_t) | X_t = x_t) = \frac{1}{r^2 f^2(\tau, x_t^T)} Var\left[\sum_{s=t-r}^{t-1} K_B(X_s - x_t) p_{i\ell s} \right]$$

$$= \frac{1}{r^2 f^2(\tau, x_t^T)} \left[\sum_{s=t-r}^{t-1} Var\left(K_B(X_s - x_t) p_{i\ell s} | x_t \right) \right.$$

$$\left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov\left(K_B(X_s - x_t) p_{i\ell s}, K_B(X_{s'} - x_t) p_{i\ell s'} | x_t \right) \right]$$

$$= \frac{1}{r^2 f^2(\tau, x_t^T)} \left[\sum_{s=t-r}^{t-1} Var\left(K_B(X_s - x_t)(c_{i\ell}(x_s) + u_{i\ell s}) | x_t \right) \right.$$

$$\left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov\left(K_B(X_s - x_t)(c_{i\ell}(x_s) + u_{i\ell s}), K_B(X_{s'} - x_t)(c_{i\ell}(x_{s'}) + u_{i\ell s'}) | x_t \right) \right]$$

$$= \frac{1}{r^2 f^2(\tau, x_t^T)} \left[\sum_{s=t-r}^{t-1} Var\left(K_B(X_s - x_t)\ u_{i\ell s} | x_t \right) + \sum_{s=t-r}^{t-1} Var\left(K_B(X_s - x_t)\ c_{i\ell}(x_s) | x_t \right) \right.$$

$$+ \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov\left(K_B(X_s - x_t) c_{i\ell}(X_s), K_B(X_{s'} - x_t) c_{i\ell}(X_{s'}) | x_t \right)$$

$$\left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov(K_B(X_s - x_t)\ u_{i\ell s}, K_B(X_{s'} - x_t)\ u_{i\ell s'} | x_t) \right]$$

$$= \frac{1}{r^2 f^2(\tau, x_t^T)} \left[\sum_{s=t-r}^{t-1} Var(K_B(X_s - x_t)\ u_{i\ell s} | x_t) + \sum_{s=t-r}^{t-1} Var\left(K_B(X_s - x_t)\ c_{i\ell}(x_s) | x_t \right) \right.$$

$$+\sum_{\substack{s,s'=t-r\\ s\neq s'}}^{t-1} Cov\left(K_B(X_s-x_t)c_{i\ell}(X_s), K_B(X_{s'}-x_t)c_{i\ell}(X_{s'})|x_t\right)\Bigg] = T_1+T_2+T_3$$

since for $s \neq s'$ the conditional expectation $E(u_{i\ell s}u_{i\ell s'})$ cancels and, therefore, only the diagonal terms remain. For T_1

$$\begin{aligned}
T_1 &= \frac{1}{r^2 f^2(\tau, x_t^T)}\left[\sum_{s=t-r}^{t-1} E(K_B^2(X_s-x_t)E(u_{i\ell s}^2|X_s)|x_t)\right]\\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x_t^T)}\left[\sum_{s=t-r}^{t-1} E(K_B^2(X_s-x_t)|x_t)\right]\\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x_t^T)}\left[\sum_{s=t-r}^{t-1} \int K_B^2(z-x_t)f\left(s/T, z^T\right)dz\right]\\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x_t^T)|B|^{1/2}}\sum_{s=t-r}^{t-1}\int K^2(u)\left(f(\tau,x_t^T)+O(traceB^{1/2})+O\left(\frac{r}{T}\right)\right)du\\
&= \frac{\sigma_{u_{i\ell}}^2}{rf(\tau, x_t^T)|B|^{1/2}}\int K^2(u)du + h.o.t. \ = O\left(\frac{1}{r|B|^{1/2}}\right)+h.o.t.
\end{aligned}$$

For T_2

$$\begin{aligned}
T_2 &= \frac{1}{r^2 f^2(\tau, x_t^T)}\sum_{s=t-r}^{t-1} Var\left(K_B(X_s-x_t)\ c_{i\ell}(x_s)|x_t\right)\\
&= \frac{1}{r^2 f^2(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\left[\int K_B^2(w-x_t)c_{i\ell}^2(w)f_s(w^T)dw - \left(\int K_B(w-x_t)c_{i\ell}(w)f_s(w^T)dw\right)^2\right]\\
&= \frac{1}{r^2 f^2(\tau, x_t^T)}\sum_{s=t-r}^{t-1}\Bigg[|B|^{-1/2}\int K^2(z)c_{i\ell}^2(x_t+B^{1/2}z)f_s(x_t^T+B^{1/2}z^T)dz\\
&\qquad - \left(\int K(z)c_{i\ell}(x_t+B^{1/2}z)f_s(x_t^T+B^{1/2}z^T)dz\right)^2\Bigg]\\
&= \frac{1}{r^2 f^2(\tau, x_t^T)|B|^{1/2}}\sum_{s=t-r}^{t-1}\Bigg[\left(c_{i\ell}^2(x_t)f_s(x_t^T)\int K^2(u)du + O(traceB^{1/2})\right)\\
&\qquad -|B|^{1/2}\left(c_{i\ell}(x_t)f_s(x_t^T)+O(traceB^{1/2})\right)^2\Bigg]\\
&= \frac{c_{i\ell}^2(x_t)}{r^2 f^2(\tau, x_t^T)|B|^{1/2}}\int K^2(u)du\sum_{s=t-r}^{t-1}\left(f(\tau,x_t^T)+O(traceB^{1/2})+O\left(\frac{r}{T}\right)\right)+\text{h.o.t}\\
&= \frac{c_{i\ell}^2(x_t)}{rf(\tau, x_t^T)|B|^{1/2}}\int K^2(u)du + h.o.t.\ =\ O\left(\frac{1}{r|B|^{1/2}}\right)+h.o.t.
\end{aligned}$$

And finally for the third term, T_3,

$$T_3 = \frac{1}{r^2 f^2(\tau, x_t^T)} \sum_{\substack{s,s'=t-r \\ s\neq s'}}^{t-1} Cov\left[K_B(X_s - x_t)c_{i\ell}(X_s), K_B(X_{s'} - x_t)c_{i\ell}(X_{s'})|x_t\right]$$

Using (A4)

$$\sum_{k=1}^{r} Cov\left[K_B(X_s - x_t)c_{i\ell}(X_s), K_B(X_{s+k} - x_t)c_{i\ell}(X_{s+k})|x_t\right]$$

is uniformly bounded and, hence, the order of T_3 is $O(r^{-1})$, negligible with respect to T_1 and T_2.

Therefore, the final expression for each (i, ℓ) variance term is

$$Var^{\star}(\widehat{c}_{i\ell}(X_t)|X_t = x_t) = \frac{c_{i\ell}^2(x_t) + \sigma_{u_{i\ell}}^2}{rf(\tau, x_t^T)|B|^{1/2}} \int K^2(u)du + h.o.t.$$

and the proof is complete. □

Proof of Theorem 2

It is sufficient to check that the proof of Theorem 1 follows considering the estimated covariances instead of the real ones. First, note that (A4) holds for the estimated covariances ($\widehat{C}$) and that (A5) holds up to order $o(1)$; that is, $E(\widehat{C}_t^{*T}\widehat{C}_t^*) = E(C_t^{*T}C_t^*) + o(1) = G_t + o(1)$.

Now, the steps of the proof of Theorem 1 follows straightforward using $\widehat{C}$ instead of C. Only the second term for the variance (36) need an extra step.

The second term for the variance can be written as,

$$\begin{aligned}
& Var\left(G_t^{-1}\sum_{s=1}^{T} K_{h,ts}\widehat{C}_s^{*T}\varepsilon_s^*\right) = \\
= \; & G_t^{-1}E\left[E\left(\sum_{s} K_{h,ts}^2\widehat{C}_s^{*T}\varepsilon_s^*\varepsilon_s^{*T}\widehat{C}_s^* + \sum_{s\neq s'} K_{h,ts}K_{h,ts'}\widehat{C}_s^{*T}\varepsilon_s^*\varepsilon_{s'}^{*T}\widehat{C}_{s'}^*|\widehat{C}_s^*\right)\right]G_t^{-1} \\
= \; & G_t^{-1}E\left[\sum_{s} K_{h,ts}^2\widehat{C}_s^{*T}\varepsilon_s^*\varepsilon_s^{*T}\widehat{C}_s^* + \sum_{s<s'} K_{h,ts}K_{h,ts'}\widehat{C}_s^{*T}\varepsilon_s^*E(\varepsilon_{s'}^{*T}|\widehat{C}_s^*, \widehat{C}_{s'}^*, \varepsilon_s)\widehat{C}_{s'}^*\right. \\
& \left. + \sum_{s>s'} K_{h,ts}K_{h,ts'}C_s^{*T}E(\varepsilon_s^*|\widehat{C}_s^*, \widehat{C}_{s'}^*, \varepsilon_{s'})\varepsilon_{s'}^{*T}\widehat{C}_{s'}^{*T}\right]G_t^{-1} \\
= \; & G_t^{-1}E\left[\sum_{s} K_{h,ts}^2\widehat{C}_s^{*T}\varepsilon_s^*\varepsilon_{s'}^{*T}\widehat{C}_s^*\right]G_t^{-1}
\end{aligned}$$

since ε_s is independent of the past information. Using the fact that $E(\widehat{C}_t^{*T}\widehat{C}_t^*) = E(C_t^{*T}C_t^*) = G_t + o(1)$, it finally holds

$$Var\left(G_t^{-1}\sum_{s=1}^{T} K_{h,ts}\widehat{C}_s^{*T}\varepsilon_s^*\right) = \frac{c_k}{Th}G_t^{-1} + o((Th)^{-1})$$

and this step completes the proof. $\square$

Proof of Corollary 2

Apply the same arguments than in Corollary 1. $\square$

Lemma 2 *Under Assumptions (A3) to (A5) and (C1) to (C3); it holds that*

$$\frac{1}{r}\sum_{s=t-r}^{t-1} K_G(X_s - x_t) \xrightarrow{a.s.} f(\tau, x_t) \tag{38}$$

where $\tau = t/T$.

Proof of Lemma 2

Following similar steps than in Lemma 1, define $Z_s = K_G(X_s - x_t)$. The sequence Z_s has mean $f(s/T, x_s)$ and therefore $E((1/r)\sum_{s=t-r}^{t-1} Z_s = f(\tau, x_t) + o(1)$. A direct application of White (1984), (see Corollary 3.48) drives to the result. $\square$

Proof of Proposition 2

It holds following the proof of Proposition 1. $\square$

Proof of Theorem 3

Consider the sequence of variables Z_t defined as

$$Z_t = \sum_{s=1}^{T} K_{h,ts}\widehat{C}_s^{*T}\varepsilon_s^*. \tag{39}$$

Using White and Domowitz (1984), it is sufficient to verify that, since their Assumption A holds, the result in their Theorem 2.4 applies. Since the bias is negligible with respect to the variance term, the result follows straightforward by applying Crammer.

$\square$

References

Black, F., M. Jensen, and M. Scholes, 1972, The capital asset pricing model: some empirical findings, in M. Jensen, ed., Studies in the Theory of Capital Markets. Praeger, New York.

Black, F. and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities. The Journal of Political Economy 81, 637–654.

Bollerslev, T., R. Engle, and J. Wooldridge, 1988, A capital asset pricing model with time-varying covariances. Journal of Political Economy 96, 116—-131.

Brandt, M. and D. Chapman, 2006, Linear Approximations and Tests of Conditional Pricing Models, National Bureau of Economic Research Cambridge, Mass., USA.

Cai, Z., 2007, Trending time-varying coefficient time series models with serial correlated errors. Journal of Econometrics 136, 163–188.

Cheridito, P., D. Filipović, and R. Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence. Journal of Financial Economics 83, 123–170.

Cox, J., J. Ingersoll, and S. Ross, 1985, An Intertemporal General Equilibrium Model of Asset Prices. Econometrica 53, 363–384.

Dalhaus, R., 1997, Fitting time series models to nonstationary processes. The Annals of Statistics 25, 1–37.

Dalhaus, R., 2000, A likelihood approximation for locally stationary processes. The Annals of Statistics 28, 1–37.

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

Duffee, G. and R. Kan, 1996, A yield-factor model of interest rates Mathematical Finance 6, 379–406.

Fama, E. and K. French, 1989, Business conditions and the expected returns on bonds and stocks. Journal of Financial Economics 25, 23–50.

Fama, E. and J. MacBeth, 1973, Risk, return, and equilibrium: Empirical test. Journal of Political Economy 81, 607–636.

Fan, J. and I. Gijbels, 1996, Local Polynomial Modelling and its Applications, Chapman and Hall, London.

Ferreira, E. and J. Gil-Bazo, 2004, Beyond single-factor affine term structure models. Journal of Financial Econometrics 2, 565–591.

Ferson, W., 1989, Changes in Expected Security Returns, Risk, and the Level of Interest Rates. The Journal of Finance 44, 1191–1217.

Ferson, W. and C. Harvey, 1991, The variation of economic risk premiums. Journal of Political Economy 99, 385–415.

Ferson, W. and C. Harvey, 1999, Conditioning Variables and the Cross Section of Stock Returns. The Journal of Finance 54, 1325–1360.

Ferson, W. and R. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions. The Journal of Finance 51, 425–461.

Ghysels, E., 1998, On stable factor structures in the pricing of risk: Do time varying betas help or hurt. The Journal of Finance 53, 549–574.

Härdle, W., 1990, Applied Nonparametric Regression, Cambridge University Press, Cambridge.

Harvey, C., 1989, Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289–317.

Jagannathan, R. and Z. Wang, 1996, The conditional capm and the cross-section of expected returns. The Journal of Finance 51, 3–54.

Jones, C., 2006, A nonlinear factor analysis of SandP 500 index option returns. The Journal of Finance 41, 2325–2363.

Kapetanios, G., 2007, Estimating deterministically time-varying variances in regression models. Economics Letters 97, 97–104.

Keim, D. and R. Stambaugh, 1986, Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357–390.

Lettau, M. and S. Ludvigson, 2001, Resurrecting the c)capm: A cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–27.

Orbe, S., E. Ferreira, and J. Rodriguez-Póo, 2005, Nonparametric estimation of time varying parameters under shape restrictions. Journal of Econometrics 126, 53–77.

Orbe, S., E. Ferreira, and J. Rodriguez-Póo, 2006, On the estimation and testing of time varying constraints in econometric models. Statistica Sinica 16, 1313–1333.

Pagan, A., 1984, Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25, 221–247.

Robinson, P., 1989, Nonparametric estimation of time varying parameters, in P. Hackl, ed., Statistical Analysis and Forecasting Economic Structural Change. Springer-Verlag, 253–264.

Ross, S., 1976, The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Shanken, J., 1985, Multivariate tests of the zero-beta capm. Journal of Financial Economics 14, 327–348.

Shanken, J., 1992, On the estimation of beta-pricing models. Review of Financial Studies 5, 1–33.

Shanken, J. and G. Zhou, 2007, Estimating and testing beta pricing models: Alternative methods and their performance in simulations. Journal of Financial Economics 84, 40–86.

Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk. The Journal of Finance 33, 885–901.

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

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

Wand, M. and M. Jones, 1995, Kernel Smoothing, Chapman and Hall, London.

Wang, K., 2003, Asset pricing with conditioning information: A new test. The Journal of Finance 58, 161–196.

White, H., 1984, Asymptotic Theory for Econometricians, Academic Press, London.

White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations. Econometrica 52, 143–161.

Zellner, A., 1962, An efficient method for estimating seemingly unrelated regressions and tests for aggregate bias. Journal of the American Statistical Statistical Association 57, 348–368.

Zellner, A., 1970, Estimation pf regression relationships containing unobserved variables. International Economic Review 11, 441–454.

Table 1
Descriptive Statistics of Interest Rates, in %

	Maturity				
Sample	**3 month**	**1 year**	**3 years**	**5 years**	**10 years**
Full					
Mean	5.98	6.61	7.07	7.31	7.56
S.D.	2.85	3.00	2.77	2.63	2.47
Min	0.80	0.88	1.34	2.08	3.13
Max	17.14	17.31	16.59	16.27	15.84
1969-1978					
Mean	5.95	6.64	6.97	7.12	7.22
S.D.	1.39	1.35	0.96	0.85	0.78
Min	2.99	3.48	4.27	4.74	5.38
Max	9.74	10.52	9.59	9.35	9.16
1979-1988					
Mean	9.01	9.95	10.36	10.51	10.69
S.D.	2.86	2.86	2.49	2.36	2.15
Min	5.03	5.52	6.13	6.40	6.95
Max	17.14	17.31	16.59	16.27	15.84
1989-1998					
Mean	5.20	5.71	6.31	6.62	6.95
S.D.	1.57	1.57	1.35	1.22	1.13
Min	2.61	2.96	4.01	3.95	4.16
Max	9.10	9.85	9.88	9.75	9.53
1999-2006					
Mean	3.22	3.55	4.06	4.41	4.84
S.D.	1.73	1.70	1.38	1.09	0.75
Min	0.80	0.88	1.34	2.08	3.13
Max	6.24	6.44	6.88	6.83	6.79

Table 2
Descriptive Statistics of Daily Excess Bond Returns, in %

	Maturity			
Sample	**1 year**	**3 years**	**5 years**	**10 years**
Full				
Mean	0.0025	0.0049	0.0065	0.0095
S.D.	0.0879	0.2354	0.3697	0.6660
Min	-0.9379	-2.4147	-3.0837	-5.6154
Max	0.9380	2.2084	3.5954	7.0847
1969-1978				
Mean	0.0011	0.0004	-0.0009	-0.0061
S.D.	0.0697	0.1764	0.2563	0.3826
Min	-0.4618	-0.9371	-1.1938	-1.9173
Max	0.3948	1.3099	1.8515	2.7961
1979-1988				
Mean	0.0041	0.0060	0.0071	0.0110
S.D.	0.1423	0.3487	0.5419	0.9991
Min	-0.9379	-2.4147	-3.0837	-5.6154
Max	0.9380	2.2084	3.5954	7.0847
1989-1998				
Mean	0.0036	0.0096	0.0146	0.0251
S.D.	0.0535	0.1785	0.2961	0.5434
Min	-0.3318	-1.1176	-1.8961	-3.5674
Max	0.3748	0.8830	1.3882	2.6997
1999-2006				
Mean	0.0011	0.0032	0.0048	0.0077
S.D.	0.0424	0.1811	0.3016	0.5550
Min	-0.2127	-0.8618	-1.3445	-2.3493
Max	0.4849	1.4595	1.8433	2.0508

Table 3
Estimated market prices of risk of changes in the level ($\widehat{\lambda}_{1t}$) and the slope ($\widehat{\lambda}_{2t}$) of the yield curve

Time period (t)	$\widehat{\lambda}_{1t}$	t-stat	$\widehat{\lambda}_{2t}$	t-stat
08/01/1970	-754.36	-1.90	211.47	0.75
04/01/1971	-470.03	-2.85	174.30	1.14
03/01/1972	-117.02	-0.87	19.00	0.14
02/01/1973	198.36	1.24	-110.40	-0.68
02/01/1974	-34.14	-0.33	50.43	0.32
02/01/1975	-251.23	-2.22	207.74	1.32
02/01/1976	-183.97	-1.55	27.98	0.20
03/01/1977	-39.47	-0.33	-38.25	-0.28
03/01/1978	87.83	0.99	40.55	0.32
02/01/1979	37.86	0.62	46.93	0.36
02/01/1980	-23.07	-0.60	39.65	0.68
02/01/1981	19.64	0.84	24.26	0.60
04/01/1982	-9.42	-0.47	-7.14	-0.18
03/01/1983	-37.56	-1.51	-58.50	-1.12
03/01/1984	-13.64	-0.30	42.87	0.52
02/01/1985	-45.20	-0.82	-68.97	-0.84
02/01/1986	-90.28	-1.35	-77.74	-0.92
02/01/1987	-166.65	-1.51	124.16	1.41
04/01/1988	29.02	0.24	16.82	0.18
03/01/1989	47.15	0.28	-62.19	-0.47
02/01/1990	-229.20	-1.18	122.11	0.79
02/01/1991	-454.80	-2.42	144.14	1.11
02/01/1992	-336.75	-2.16	-2.07	-0.02
04/01/1993	49.42	0.29	-173.71	-1.46
03/01/1994	160.35	0.63	34.04	0.24
03/01/1995	-8.00	-0.04	-29.47	-0.22
02/01/1996	-205.59	-0.96	41.90	0.32
02/01/1997	-185.22	-0.61	36.99	0.26
02/01/1998	54.24	0.13	-176.65	-1.08
04/01/1999	-37.29	-0.14	43.78	0.34
03/01/2000	156.17	0.30	-11.17	-0.06
02/01/2001	-748.92	-3.91	70.79	0.63
02/01/2002	-354.20	-2.50	18.84	0.29
02/01/2003	-80.02	-0.43	-70.62	-1.00
02/01/2004	223.12	0.51	-27.66	-0.31
03/01/2005	523.60	2.44	-82.99	-1.01
03/01/2006	236.84	1.51	-37.81	-0.34

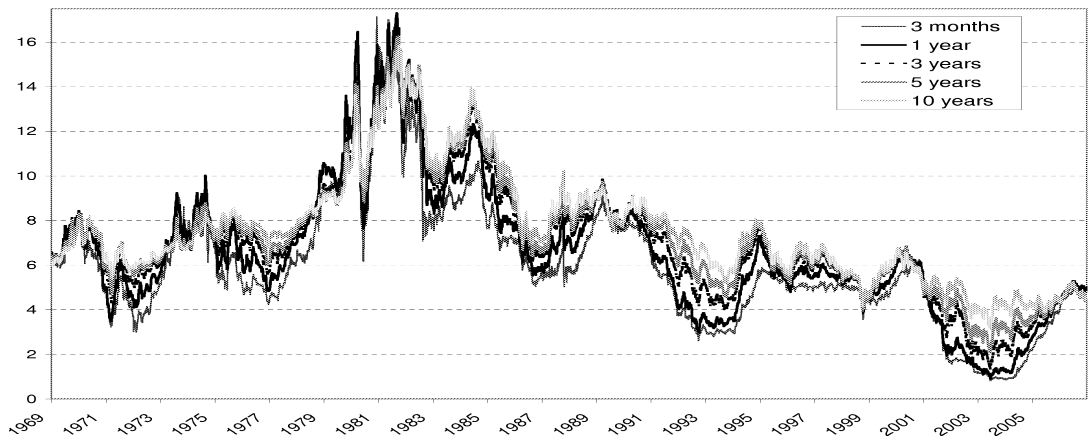

Fig. 1. Time series of default-free interest rates of different maturities in the period 1969-2006.

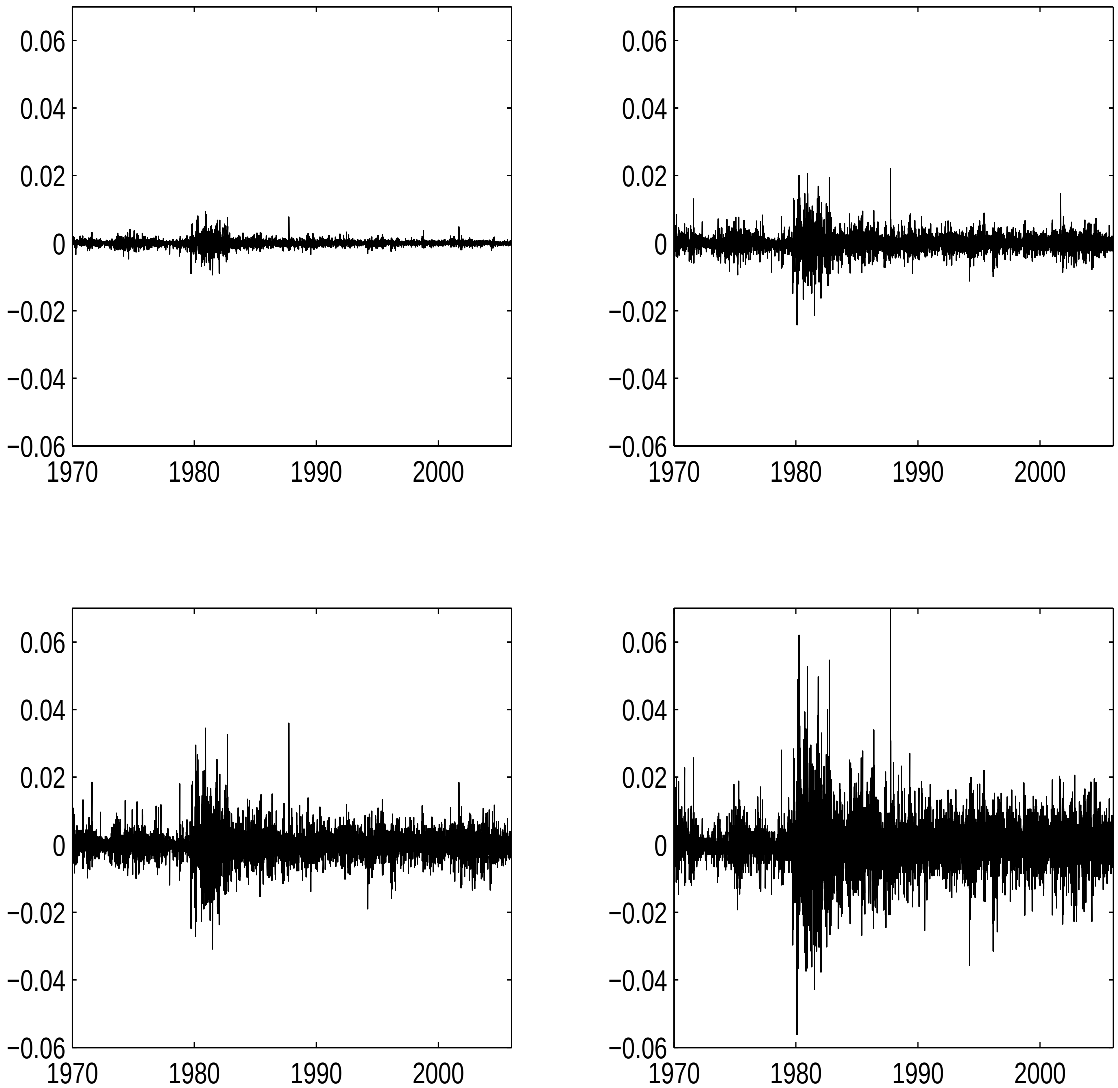

Fig. 2. Daily excess returns on 1-year (upper left), 3-year (upper right), 5-year (bottom left) and 10-year (bottom right) default-free bonds.

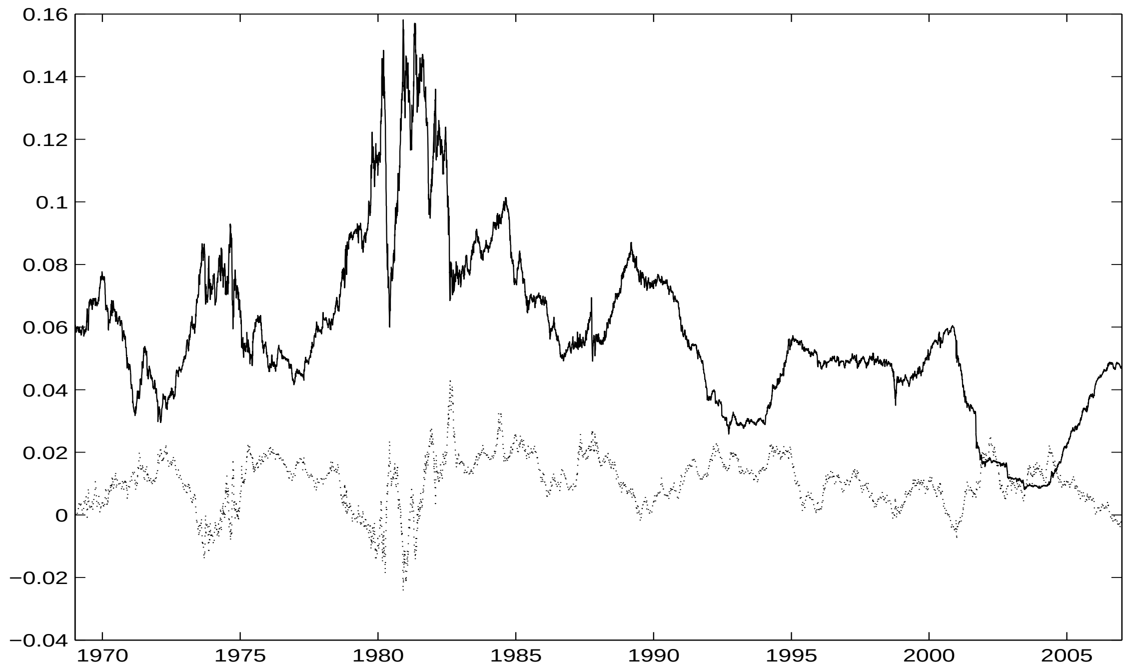

Fig. 3. Time series of the short term interest rate (solid line) and the spread between the 10-year and 3-month rate (dotted line) in the period 1969-2006.

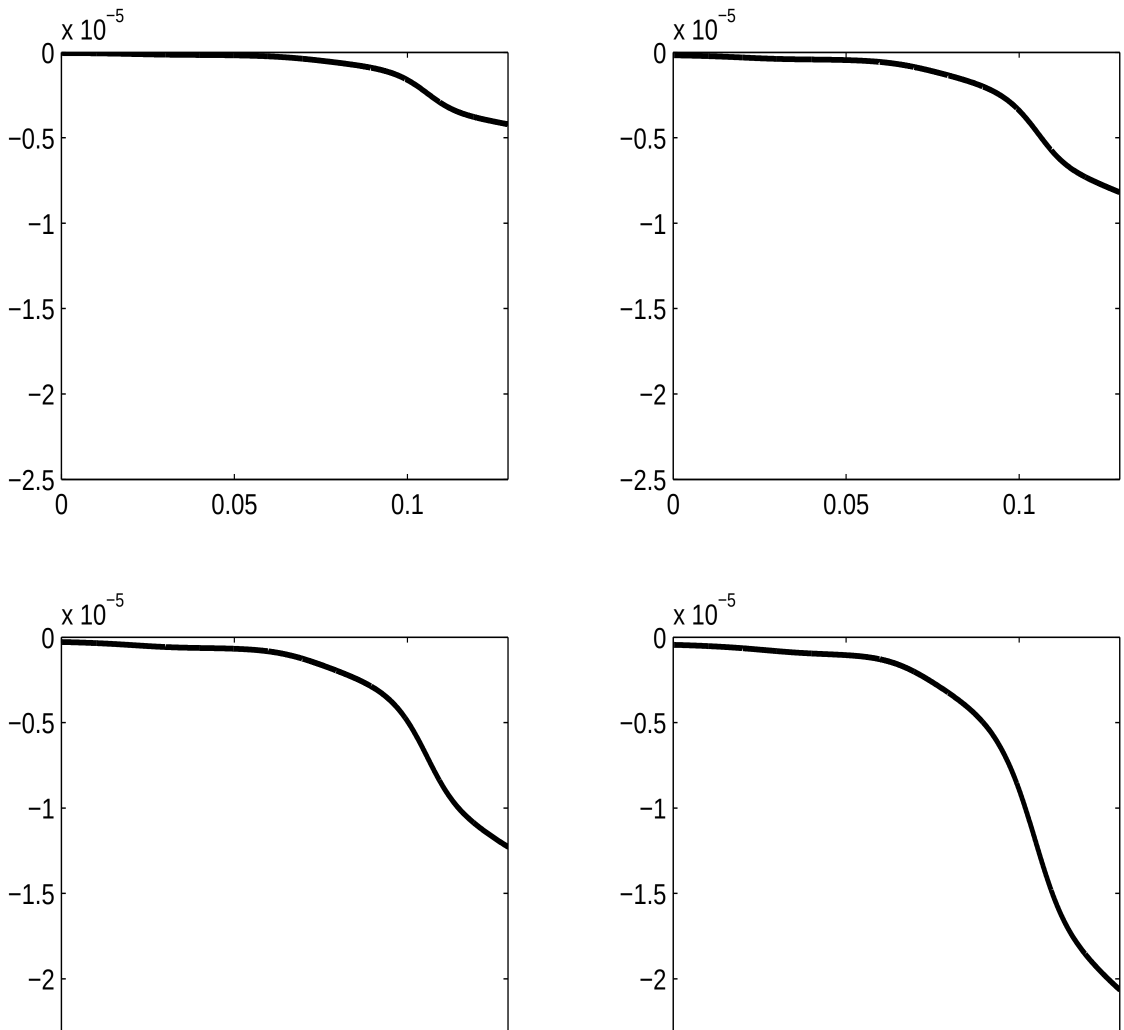

Fig. 4. Estimated conditional covariances between daily bond returns and changes in the short rate as a function of the short rate. The term spread is held constant and equal to its sample median value. Graphs correspond to 1-year (upper left), 3-year (upper right), 5-year (bottom left) and 10-year (bottom right) default-free bonds.

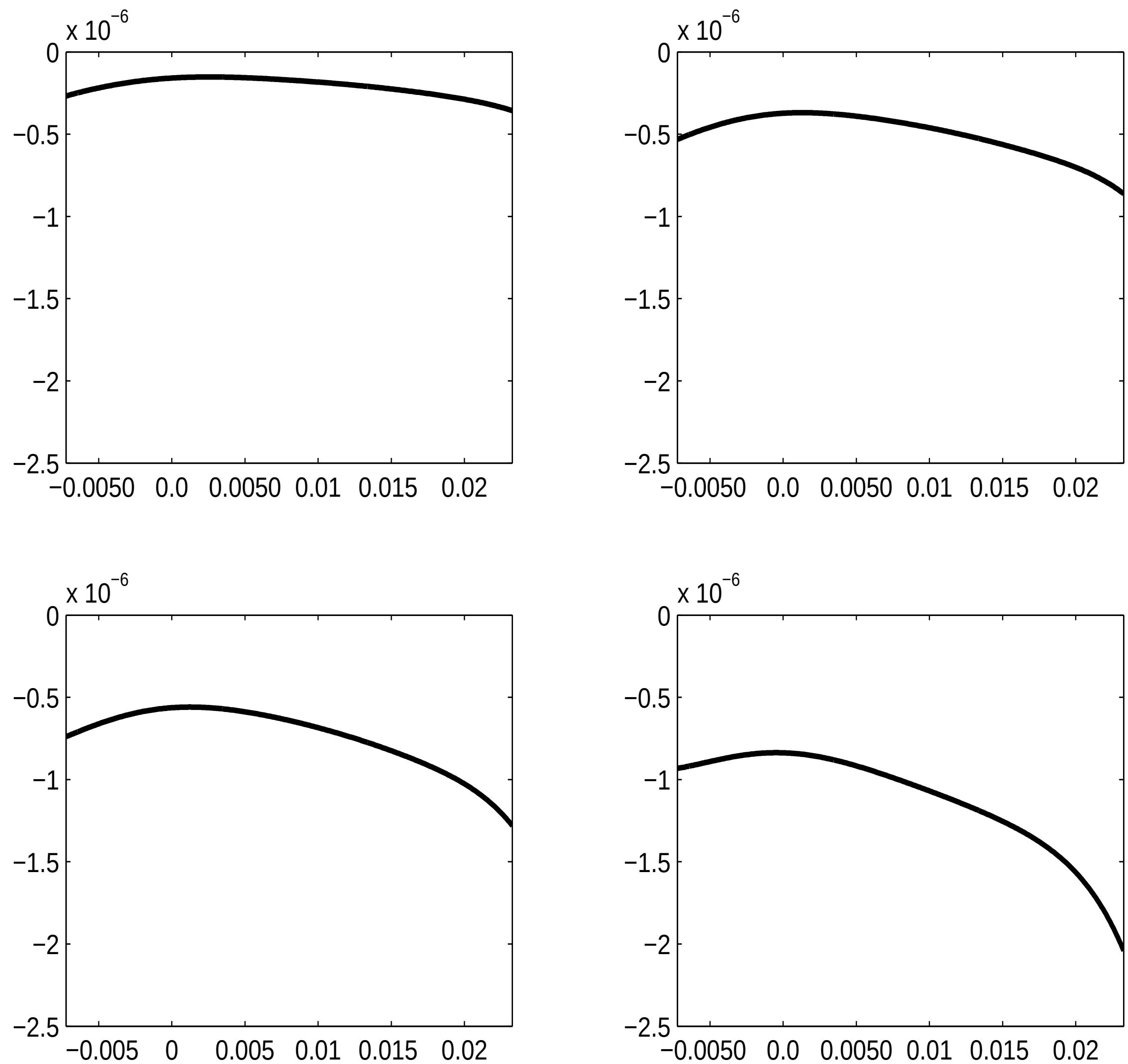

Fig. 5. Estimated conditional covariances between daily bond returns and changes in the short rate as a function of the term spread level. The short rate is held constant and equal to its sample median value. Graphs correspond to 1-year (upper left), 3-year (upper right), 5-year (bottom left) and 10-year (bottom right) default-free bonds.

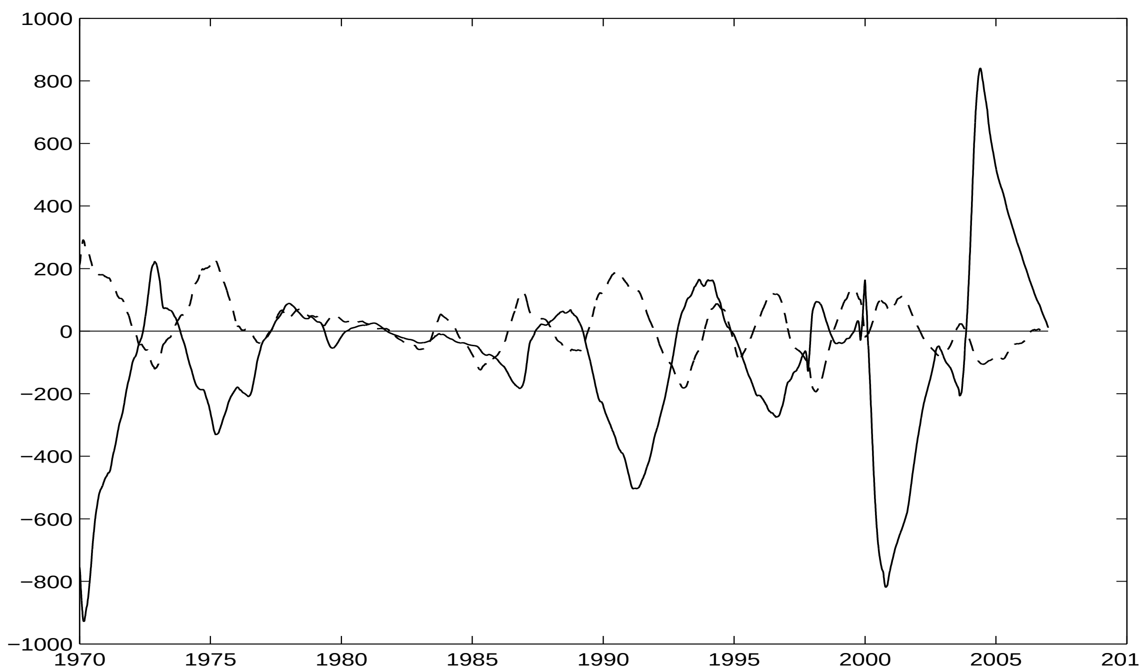

Fig. 6. Time series of estimated market prices of risk associated with changes in the short term interest rate (solid line) and changes in the term spread (dashed line).

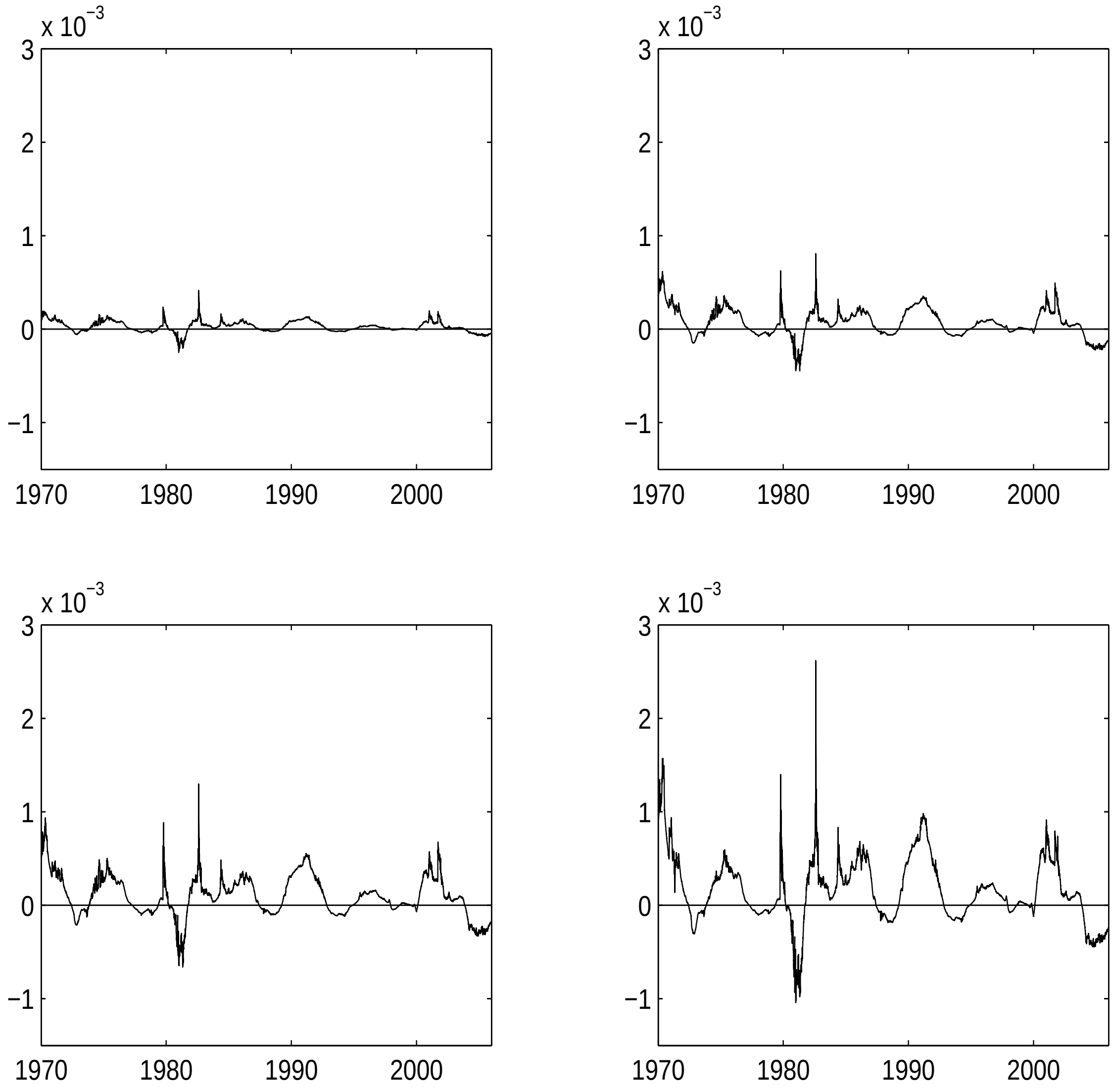

Fig. 7. Time series of estimated daily risk premia associated with changes in the short rate for 1-year (upper left), 3-year (upper right), 5-year (bottom left) and 10-year (bottom right) default-free bonds.

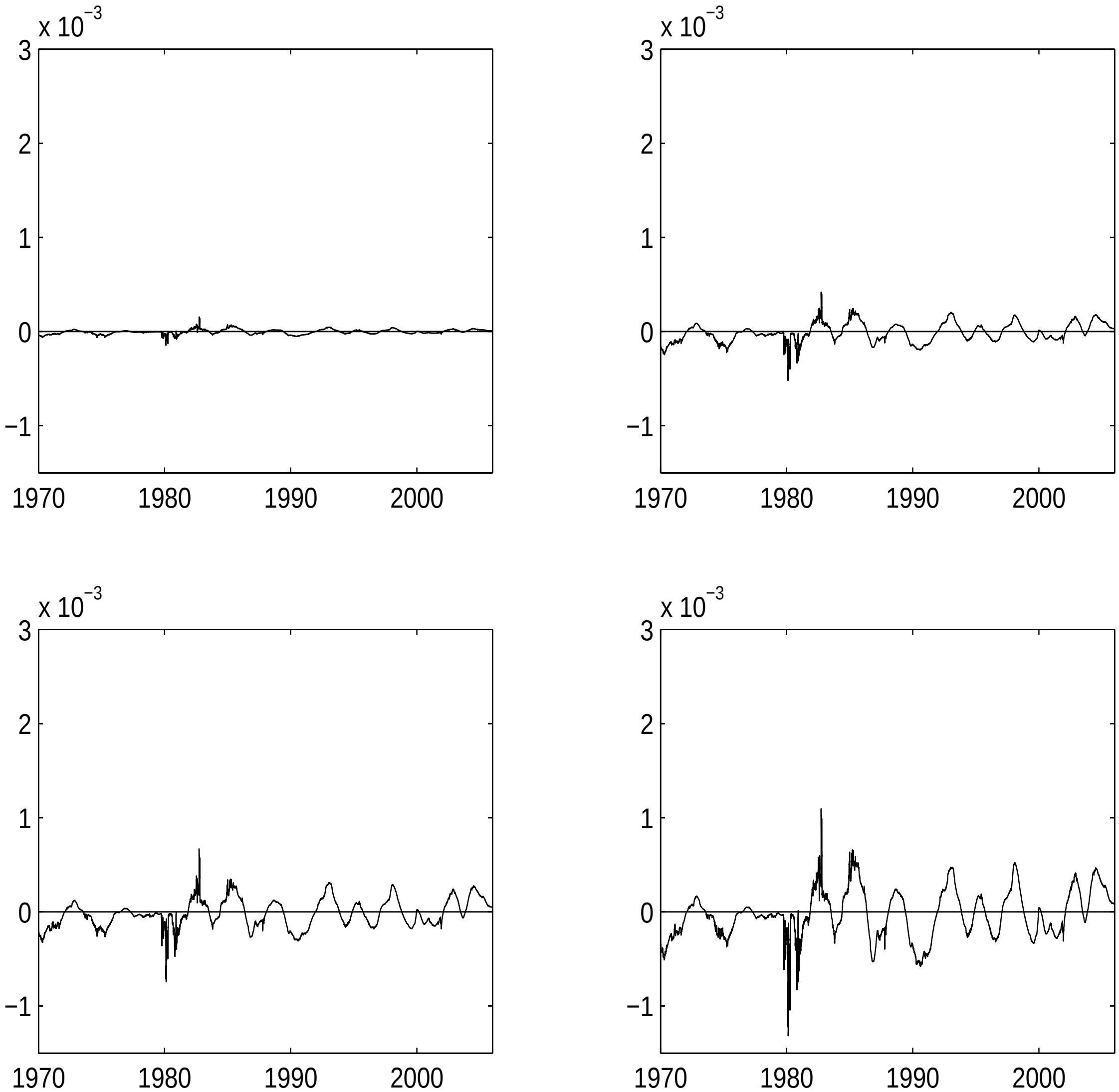

Fig. 8. Time series of estimated daily risk premia associated with changes in the term spread for 1-year (upper left), 3-year (upper right), 5-year (bottom left) and 10-year (bottom right) default-free bonds.