A non-parametric investigation of risk premia

Peroni, Chiara

UEA, RCEA

11 February 2008

A Non-parametric Investigation of Risk Premia

Chiara Peroni*

Abstract

This paper studies determinants of risk premia using a non-parametric term-structure model of the corporate spread. The model, which measures the extra return of defaultable corporate bonds on their government counterparts, involves the rate of inflation, a key macroeconomic variable that is found to explain the spread non-linearly. This study shows that non-linear methods are useful to investigate features of credit risk and that they give better results than their linear counterparts, enabling testing of affine term-structure specifications. The paper also shows how the non-linear model can be used to forecast the future course of the spread.

JEL CODE: G12, C14.

KEY WORDS: risk premium, corporate spread, default, additive models, non-parametric estimation.

Credit risk, usually defined in terms of *default risk*, measures the possibility of borrowers not being able to pay neither contractual interest nor the principal on their debt obligations. When buying securities, investors try to assess the quality of the borrower in order to reduce the probability of incurring in financial losses, and the higher this risk of credit the higher the required promised payment to compensate for its bearing. Usually, returns on risky securities are higher than returns on securities regarded as "safe" (e.g., yields on corporate bonds are higher than yields on government bonds).

This extra return is known as *risk premium*, a key financial variable which conveys information on the market perception of credit risk and of economic conditions. Its determinants, however, are not very well understood.

To study risk premia, empirical research has often focused on *corporate spreads*. Because these are differences between yields on corporate debt subject to default risk and comparable government bonds free of such risk, they are easily interpreted as direct measures of the *risk premium*. Historical default probabilities, however, are too low to account for the size of observed spreads. Studies report a large non-default component in corporate spreads, which is often left unexplained. Capturing large observed spreads and explaining the link between spreads and default risk are key empirical challenges in this area of finance. These challenges motivate the work presented here.

*Correspondence address: School of Economics, University of East Anglia, NR4 7TJ, Norwich, UK. E-mail address: c.peroni@uea.ac.uk.

This paper investigates the appropriateness of a popular class of model in the analysis of risky assets, Reduced Form Models (RFMs), challenging their underlying linearity assumption. RFMs constitute pricing frameworks, which model the term-structure of activities subject to default risk as a direct extension of risk-free yield curves, and are state of the art in this field. Reduced Form Models derive analytical and simple expressions for risky yields, which are linear in a set of state variables (factors), but impose a strong parametric assumption of linearity on the state variables' time series processes and cross section of yield-prices alike.

To challenge the linearity assumption, this article investigates the use of non-parametric statistical techniques. To better explain the formation of large risk premia, it proposes a non-parametric multi- factor term structure model of the spread. In this model, corporate yields are functions of two sets of factors related to the risk of credit, one based on key macroeconomic variables, the other related to the risk-free term-structure. Functions describing how factors determine spreads are non-linear. It is by using non-parametric techniques, which allow flexible estimation of the model by relying only on the data available, that this study tries to overcome limitations of standard linear approaches. The analysis focuses on the cross-section relation of yields to factors.

Non-parametric statistical techniques are explored in the analysis of financial data. Rather than proposing new methods, it investigates the applicability of existing techniques. It shows that it is possible to divide the non-parametric analysis into the traditional steps involving model specification, estimation and checking. It also investigates post-sample evaluation and forecasting whose inherent difficulty in this non-parametric setting may be obviated with conditional forecasting.

This paper is structured as follows. Section 1 examines the theoretical determinants of risk premia in a Reduced Form framework. Section 2 gives an overview of the empirical methodology used in the paper. After discussing the data (Section 3), Section 4 test for the linearity of the spread in factors. Section 5 specifies and estimates a non-parametric model of the spread. The forecasting performance of this model is evaluated in Section 6. Section 7 specifies and estimates a non-linear but parametric model of the spread, after which Section 8 summarises results and gives concluding remarks.

1 Background

Reduced-Form Models (RFMs), introduced by Duffie and Singleton (1997), are frameworks for the measurement and pricing of credit risk. They are widely popular with academics and practitioners because they offer easily interpretable intuition, but they rely on strong parametric assumptions. RFMs belong to the *affine* class of *term-structure models*, which capture movements of interest rates (yields) and establish determinants of their evolution over time (Dai and Singleton, 2000). In these models, the dynamics of interest rates depends on the evolution of a set of observed, or unobserved, variables (or factors). These variables can be identified with nodes of the term structure itself, or may be macroeconomic variables. Their underlying dynamics is described by affine (linear) *Itô* diffusion processes.

The main advantage of affine models is their tractability: the linearity assumption in the factors' dynamics yields an analytical representation of the term structure and bond price formulas which are easy to interpret and well suited to empirical testing.

RFMs describe the term-structure of yields on defaultable activities as a direct extension of government yield curves. This is done by replacing the risk-free instantaneous rate of interest r, used in conventional term-structure modelling, with a default-adjusted (or risk-adjusted) discount rate R. The price in t of a defaultable zero-coupon zero-recovery bond of maturity T can be written as if the promised payoff were default-free:

$$P_t = E_t^Q[\exp\{-\int_t^T R_s ds\}], \tag{1}$$

where Q indicates assessment under risk-neutral probabilities (Duffie and Singleton, 1997). The risk-adjusted rate R is given by the sum of the risk-free rate r plus a term, λ, which captures default risk and depends on the probability of default (hereafter referred to as the *default rate*):

$$R_s = r_s + \lambda_s; \tag{2}$$

Key assumption here is that default is a "surprise" event, which occurs unexpectedly, and is exogenous to the model. The dynamics of the risk-adjusted rate is modelled as an affine multi-factors diffusion process, examples of which can be found in Duffee (1999), Duffie and Singleton (1999), and Driessen (2005). This gives representation of yields as linear functions of factors. As a result, credit spreads, which are differences between yields on risky assets and yields on risk-free assets, are also linear functions of factors. This is shown as follows, under the simplifying assumption of risk-neutral independence.

Substituting equation 2 in 1 the price of the zero-coupon risky bond can be decomposed into the product of a risk-free price component (Δ) and a price component that depends on the default rate (D):

$$P_t^{(\tau)} = D^{(\tau)}(\lambda)\Delta^{(\tau)}(r), \tag{3}$$

where τ is the time to maturity ($\tau \equiv T - t$). These zero-coupon bond prices are as follows:

$$D^{(\tau)}(\lambda) = A_D(\tau)\exp\{-B_D(\tau)\lambda\}, \tag{4}$$

$$\Delta^{(\tau)}(r) = A_\Delta(\tau)\exp\{-B_\Delta(\tau)r\}; \tag{5}$$

(The coefficients A_Δ, B_Δ, A_D, B_D are deterministic functions of underlying diffusion parameters and time to maturity. The subscripts D and Δ indicates that coefficients refer to, respectively, the default and non-default price component.)[1] The credit spread implied by RFMs is computed as the difference between the yield on the risky bond and the yield on the risk-free bond, using equations 4 and 5. The resulting (observable) spread is linear in the default rate:

$$s(\tau) = -\frac{\log A_D(\tau)}{\tau} + \frac{B_D(\tau)\lambda}{\tau}; \tag{6}$$

[1] Bolder (2001) offers a clear and detailed derivation of pricing formulas for zero-coupon bonds, and gives expressions for deterministic coefficients such as A_Δ, B_Δ, A_D, B_D; one can also see Duffie and Singleton (2003, Chapter 5).

Following Duffee (1999), I now allow the default rate λ to depend (linearly) on the risk-free rate r. This yields a simple and tractable representation of the spread, which is now *linear* in the interest rate AND default rate:

$$s(\tau) = -\frac{\log A(\tau)}{\tau} + \frac{B_D(\tau)}{\tau}\lambda + \frac{(\tilde{B}_D(\tau) - \tilde{B}_\Delta(\tau))r}{\tau}. \tag{7}$$

Here, as above, the coefficients A_Δ, B_D, $\tilde{B}_D$, $\tilde{B}_D$ are deterministic functions of underlying diffusion parameters and time to maturity, and subscripts D and Δ denotes default and non-default components of the risky bond's pricing equation. (The derivation of this formula, and of formula 6, are detailed in appendix A.)

One of the best features of RFMs is this parallel between default-free and defaultable bond price formulas, which gives simple linear expression for yields and spreads. This linearity assumption, however, has been questioned for government yield curves. Substantial evidence, based primarily on the estimation of underlying diffusion processes, suggests that it is too restrictive (Aït-Sahalia, 1996; Gil-Bazo and Rubio, 2004; Arapis and Gao, 2006). The possible presence of non-linearities in data has important implications, suggesting that affine models are misspecified and bond prices formulas are not correct.

There is another well-known shortcoming of these models. While affine models represent yields as linear functions of factors, they fail to provide a clear economic interpretation of those factors. To address this problem, models of risk-free yields have recently been extended to include macroeconomic variables as factors (Ang and Piazzesi, 2003), but this idea has not been applied to credit risk.

Empirical evidence suggests, however, that risk premia are affected by general economic conditions. The observed negative correlations between Treasury rates and credit spreads, documented in Duffee (1998), is often interpreted as evidence that risk premia are correlated with the business cycle. (During economic downturns — associated to low interest rate levels — the risk of bankruptcy increases, driving up yields on risky securities.)

An early empirical model of spreads which explicitly considers the effect of macroeconomic indicators has been proposed by Wadhwani (1986). Wadhwani (1986) explains why the inflation rate affects default risk and spreads. It argues that inflation has an adverse effect on firms' interest payments, creates cash-flow problems and, consequently, increases the number of bankruptcies.

More recently, in the context of duration models, a stream of literature studied the impact of macroeconomic factors on the risk of default. In Figlewski et al. (2008), data on individual firms is used to study the impact of macroeconomic factors such as unemployment, inflation, production and indices of macro performance on default probabilities. The effects of these macro variables are estimated by the coefficients of the parametric component of a Cox regression model. Such effects are measured in isolation and in conjunction with other variables, such as bond, equity markets and firm-specific variables; the latter are related to the credit-rating history of individual firms. Figlewski et al. (2008) find that macroeconomic indicators are capable of explaining the default of individual corporate issuers, although size and sign of the variables' coefficients vary over different model specifications. They also note that the information contained in macro variables appears to be "incremental" to that captured by firm-specific factors. For example, the effect of inflation, insignificant when considered on its own, becomes positive and significant when considered along ratings-related variables. Furthermore, the effects of variables such as GDP and industrial production growth, and general macro indicators, suggests anti-cyclicality of default. The significance and size of macro effects, however, is substantially reduced when they are studied together with other macroeconomic variables, possibly due to the correlation existing among variables.

The analysis of Figlewski et al. (2008) was extended by Couderc et al. (2008). These authors considered the dynamic effect of a similar set of macroeconomic and financial factors on probabilities of default, confirming the explanatory power of business cycle variables, and highlighting the high degree of persistence of economic shocks on the likelihood of default. GDP, industrial production, and personal income growth have negative and persistent effect on default intensities. Unlike the study of Figlewski et al. (2008), inflation has a negative and persistent effect on default probabilities.

The duration studies of default of Figlewski et al. (2008) and Couderc et al. (2008) confirm the negative relation between risk-free interest rates and probabilities of default found in previous literature. They also play down the role of equity market variables, indirectly supporting RFMs on structural models.

Unlike the above authors, Duffie et al. (2007), failed to find a significant effect of macroeconomic factors on default, and consequently focused their attention on a smaller set of variables capturing equity and financial markets conditions, such as the US interest rate and individual firm and market-wide returns. (Their model combines traditional duration analysis with time varying parameters features.)

Using a different methodology, Huang and Kong (2003) also investigated the effects of macro factors on credit risk. These authors regress credit spread changes on indices of macroeconomic and equity market performance. They also consider the effect of risk free interest rates, as measured by changes in treasury yield indices and the yield curve slope. Despite some indication that macro factors have some explanatory power for credit spread changes, macro indices seem to have a significant effect on speculative-grade rather than investment-grade yields, and the highest t-ratios in the regressions are associated with equity market factors. Overall, credit spread changes seem more closely related to equity market factors and volatility. In later work (Huang and Kong, 2005), the same authors considered the effects of (scheduled) macroeconomic news announcements on spreads and their volatility patterns. Once again, they find that such announcements, and in particular those regarding leading economic indicators and employment reports, affect mainly speculative-grade bonds.

In summary, studies on determinants of default risk suggests that macroeconomic factors play some role in determining default probabilities. The studies reviewed above, however, interpret macroeconomic variables as business cycle indicators, or indicators of global economic health. They do not suggest an economic mechanism for the determination of credit risk, and do not explain the reason why specific variables would increase — or decrease — default risk. Existing evidence is inconclusive and, at times, contradictory (in terms of size, sign, and significance of estimated effects); nonetheless, it is useful because it provides clues that help to specify observable and estimable models of the spread.

The above studies suggest ways to overcome the major difficulty in the estimation of model (7), which is the inclusion of the crucial but unobservable variable λ (the default rate). In theoretical models, λ depends on the probabilities of default, but historical default rates are too low to account for observed spreads, even for very short-term securities (this fact is often referred to as the *credit spread puzzle*). It has been argued that this puzzle emerges because credit spreads are determined by more factors than the risk of default. Empirical studies have found evidence of non-default components in risk premia, often linked to liquidity and taxation, and reported a range of estimates on its size (Elton et al., 2001). Recently, however, Longstaff et al. (2004) concluded that default risk accounts for a large part of the corporate spread (this uses credit-default swaps). This suggests that the problem lies in the *measurement* of default risk, in particular in the measurement of the market's *perception* of risk, rather than in its relative explanatory ability.

This paper measures the link between default risk and spreads by choosing appropriate factors which are observable. The idea is to use a macroeconomic variable as the observable. Thus, the spread model estimated here includes a key macro variable related to default risk: *the inflation rate*. This extends RFMs to include macroeconomic variables, enabling the investigation relationship between macroeconomic conditions and the risk premium.

2 Methodology

As said above, this paper uses non-linear non-parametric methods to analyse corporate spread indices. This is done to validate RFMs. To apply non-linear methods, we first need to test for the linearity of the spread in factors. The methodology adopted here uses a general non-parametric model to specify the alternative to the null linear model. A non-parametric multiple regression model is also used to specify a non-linear regression model of the spread, which is used to check the significance of the individual non-linear effects.

The following gives a brief overview of the main techniques used here. First, additive modelling is presented. This simplifies the estimation of non-parametric regression models with multiple regressors, while retaining some structure and facilitates the interpretation of results. Then, a general framework for testing in a non-parametric setting is presented.

2.1 Additive modelling

To study the relation between a set of explanatory variables (or predictors) $X_1, \ldots, X_D$ and a dependent variable (or response) Y using a sample of n observations, a non-parametric regression model describes the dependence of Y on the Xs as a smooth function plus an additive error term. In the case of a single predictor, this is as follows:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, .., n; \tag{8}$$

(Standard assumptions apply to the error term ϵ.) Estimation of the regression function f is performed using *local smoother*. Smoothers are statistical tools based on the local averaging of data: the regression estimator at a point x_i, $\hat{f}(x_i)$, is the weighted average of those observations y_js with predictors in a neighborhood of x_i (weights depends on the predictors' distance from x_i). There are several smoothers available (see Härdle, 1989). This article uses the *local-linear method* (Fan, 1992), which proceeds by dividing the sample into intervals of equal length (along the X direction) and runs linear regressions on each interval.[2] The local-linear estimator is also used in the estimation of additive models (see below).

With more than one predictor, the model of equation 8 generalises to the regression surface

$$y_i = m(x_{1i}, \ldots, x_{Di}) + \epsilon'_i, \qquad i = 1, .., n; \tag{9}$$

In this case, the model's estimation is difficult due to a problem known as the *curse of dimensionality*. This indicates the worsening performance, in the statistical sense, of local smoothers as the number of variables increases.[3] A related problem is data sparsity. As noted above, surface estimation is based on the principle of local averaging; when data are smoothed over multiple dimensions, however, to find non-empty neighborhoods one should increase their size, and locality would be lost (Härdle, 1989, Chapter 10). Surface smoothers are also difficult to interpret, as the regression output cannot be displayed when the number of predictors is greater than 2, thus loosing an important advantage of non-parametric regression, ie the visualisation of regression lines. These shortcomings motivate the use of dimension-reduction models, such as *Generalised Additive Models* (GAMs) (Hastie and Tibshirani, 1990), that only involve one dimensional functions.[4]

In GAMs, the dependent variable is modelled as the sum of smooth functions of the explanatory variables:

$$y_i = \alpha + m_1(x_{1,i}) + \ldots\ldots + m_D(x_{D,i}) + \epsilon''_i, \qquad i = 1, .., n; \tag{10}$$

Here, the m_js are univariate smooth functions of the explanatory variables $X_1, \ldots, X_D$.

The model of equation 10 is clearly a non-parametric version of the multiple linear regression model. In this model, the contribution of each explanatory variables to the

[2]The local-linear estimator $\hat{m}_h(x)$ is given by the a^* which minimises the weighted least squares problem:

$$\min_{a,b} \sum_i \{y_i - a - b(x_i - x)\}^2 K(x_i - x, h);$$

where $a - b(x_i - x)$ is a polynomial of order 1, which approximates the regression function around x; $K(x - .)$ is a Gaussian *kernel*, which weights the observations falling in the interval around x; h (the *bandwidth*) controls the width of the kernel, ie the size of the local neighborhood. The latter is a crucial element of the estimation as it controls the degree of smoothing performed in the estimation. On the various available methods for choosing the bandwidth one can see Härdle (1989), chapter 5.

[3]The convergence rate of a non-parametric estimator depends on the number of regressors. Consequently, the "distance" between the non-parametric estimator and the true value collapses at a much slower speed as the number of predictors increases (Härdle, 1989, Section 4.1).

[4]The dimension-reduction principle, which refers to convergence rate properties of statistical models, is illustrated in Stone (1985).

response is additive, which may be regarded as a strong assumption. However, equation 10 provides a very convenient way to represent non-parametric multiple regressions, for the following reasons: (1) it is flexible enough to allow departures from linearity; (2) it is possible to obtain a graphical output, so that the estimated effects are far easier to interpret than a d-dimensional surface; (3) the additive feature provides a basis for inference in the non-parametric context. Residual sum of squares (RSS) and approximate degrees-of-freedom (df) can be calculated, and these quantities can be used to assess the significance of each smooth term in modelling the response. Examples of additive modelling, its properties and inferences are presented in Hastie and Tibshirani (1990), which remains the most comprehensive source for GAMs to date. The use of GAM to model non-linear time series is discussed in Fan and Yao (2003, Chapter 8.).

The effects of the explanatory variables on the response are estimated using an iterative procedure, known as *back-fitting algorithm*. (The algorithm is described in Appendix B.) A detailed analysis, and a rigorous justification, of the algorithm can be found in Buja et al. (1989). Opsomer and Ruppert (1997) examined theoretical properties of the back-fitting estimator for an additive model with two explanatory variables when the local-linear smoother is used. They provide expressions for smoother matrices, asymptotic bias and variance, and show that the model has the same convergence rate of a univariate local linear regression. Opsomer and Ruppert (2000) extended this analysis to the general case of more than two explanatory variables. The implementation of the algorithm in the programming language S is presented in Chambers and Hastie (1992, Chapter 7).

Additive models have been widely used in nonparametric modelling, due to their intuitive appeal and the availability of software, but they have been rarely applied to study of financial and economic data. This article aims to show their usefulness also in this context.

2.2 Testing linearity

This article uses a Generalised Likelihood Ratio (GLR) approach (Fan and Jiang, 2007) to test for the linearity of spread in factors. This is a framework for testing hypothesis in a non-parametric setting, based on a general version of the likelihood-ratio (LR) statistic.

This GLR approach is favoured here on other linearity tests proposed in the statistical literature for two reasons: 1) unlike other tests for linearity, the null distribution of the GLR test statistic is known and it is independent of nuisance (unknown) parameters, which renders the test applicable to practical problems;[5] 2) the GLR test is derived against a general alternative hypothesis, specified using non-parametric modelling, instead of being restricted to a specific non-linear alternative. (Examples of linearity tests with parametric alternatives, and their derivation, can be found in Granger and Teräsvirta, 1993, Chapter 3.)

Fan et al. (2001) demonstrated that a GLR test statistic, based on appropriate non-parametric estimators, follows asymptotically a χ^2 distribution under the null hypothesis. The authors applied this result to testing a simple linear regression model against a non-parametric alternative, as follows problem is as follows:

$$H_0 : y = \alpha + \beta x \quad vs. \quad H_1 : y = f(x); \tag{11}$$

(Here, α and β denote the unknown parameters of the linear regression; f is a smooth function.) The appropriate GLR statistic for this testing problem is given by:

$$\lambda_n(h) = \log(H_1) - \log(H_0) = \frac{n}{2} \log \frac{RSS_0}{RSS_1}; \tag{12}$$

where $log(H_1)$ and $log(H_0)$ are the log-likelihood functions for the alternative and the null model, and RSS_0 and RSS_1 the residual sum of squares (RSS) for the null and alternative model. The main difficulty here is that LR tests require the model be estimated under the alternative. In practice, the test works by substituting the maximum likelihood estimator of the alternative model by a reasonable non-parametric regression estimator, namely the local linear method. The null hypothesis is rejected when $\lambda_n(h)$ is too large.

For the testing problem 11, Fan et al. (2001) established the following result:

Under the null hypothesis and certain conditions, if $nh^{3/2} \to \infty$ the following result holds:

$$r_K \lambda_n(h) \overset{a}{\sim} \chi^2_{d_n(h)}, \tag{13}$$

where

$$d_n(h) = r_K c_K \left|\Omega\right| h^{-1};$$

[5] Several tests for linearity in the non-parametric setting have been proposed. An example is the test analysed by Härdle and Mammen (1993), and based on measures of distance between parametric and non-parametric fits. Azzalini et al. (1989) studied a test based on the F statistic, which is based on the restrictive assumption that linear and non-parametric models are nested. All these tests have a major problem: the null distribution of the test statistic is, in general, unknown and/or depends on nuisance parameters/functions.

Here, h denotes the bandwidth, $d_n(h)$ the degrees of freedom of the χ^2 statistic, Ω the length of support of the covariate x; r_K and c_K are constants. (Values of r_K and c_K for different choices of the Kernel used in the local-linear estimation are tabulated in Fan et al., 2001, table 2, p.170). Intuitively, degrees of freedom depend on the amount of smoothing performed, through the term $|\Omega| h^{-1}$. Once the degrees of freedom are calculated, critical values can easily be found based on the known null distribution. The validity of this result is discussed in (Fan et al., 2001, section 4.1); the conditions under which the result holds are stated in (Fan et al., 2001, section 3.1).

The GLR approach outlined above can be extended to a variety of models/testing problems, as shown in Fan and Jiang (2007). Fan and Jiang (2005) extended the GLR approach to testing in the multiple regression context. Consider the following problem, in which the linearity of a multiple regression model with d explanatory variables is tested against a general non-linear alternative:

$$H_0 : y = \alpha + \beta_1 x_1 + ... + \beta_D x_D \quad vs. \quad H_1 : y \neq \alpha + \beta_1 x_1 + ... + \beta_D x_D; \tag{14}$$

The testing procedure proposed by Fan and Jiang (2005) comprise the following steps:

1. Write the alternative non-parametric model using a GAM model structure, so that H_1 is given by:

$$y = \alpha + m_1(x_1) + m_2(x_2) + ... + m_D(x_D) = \alpha + \sum_d m(x_d), \tag{15}$$

   (Recall that ms are univariate non-parametric functions of each explanatory variable);

2. Estimate the alternative model, using the back-fitting algorithm, and compute the GLR statistics as:

$$\lambda_n(h) = \frac{n}{2} \log \frac{RSS_0}{RSS_1}, \tag{16}$$

   where n is the number of observations, and RSS_0 and RSS_1 residuals sum of squares from the null and alternative models;

3. Use bootstrap methods to simulate the null distribution of the GLR statistics, and compute p-values.

Fan and Jiang (2005) showed that

$$r_K \lambda_n(h) \overset{a}{\sim} \chi^2_{d_n(h)}, \tag{17}$$

where

$$d_n(h) = r_K c_K \sum_d \frac{|\Omega_d|}{h_d};$$

(For a discussion of this result, see Fan and Jiang (2005), Section 3.2. There expressions for the constant r_K and c_K are also given.) One should note that, in this setting, the computation of degrees of freedom is more complicated than for result 13. This is why a bootstrap method is required to simulate the null distribution of $\lambda_n(h)$.

The method above is also applicable to test non-parametric null hypothesis versus non-parametric alternatives, for example to test the significance of one — or more — smooth term in the additive model of equation Fan and Jiang (2007, 2005). Jiang et al. (2007) suggest to extend the GLR method to a semi-parametric setting, to test whether one or more components of the additive model of equation should be specified as non-linear. The authors look at testing problem of the type:

$$H_0 : y = \alpha + \beta_1 x_1 + ... + \beta_D x_D \quad vs. \quad H_1 : y = \alpha + \beta_1 x_1 + ... + m_{D-1}(x_{D-1}) + m_D(x_D). \quad (18)$$

(Note that the alternative here is a semi-parametric model.) The authors give a null distribution for the GLR test statistic and establish its properties.

The following starts by presenting the dataset. Then, the testing procedure presented above is applied to a models that represents spreads as a linear function of several factor, such as the model of equation 6.

3 Data

In this analysis I use a data set constructed from government yield curve data compiled by the Federal Reserve and corporate yields indices from Moody's database, which were obtained from the FRED database.[6] The data set is made up of monthly time series ranging from April 1953 to June 2006. Spreads are computed as differences between yields on investment-grade corporate bond (rated by Moody's as Aaa and Baa) and constant-maturity-Treasury yields of comparable maturity. Table 1 reports summary statistics for the spread series.[7]

Figure 1 presents time series of Aaa and Baa spreads (right panel) and of the *relative spread* (right panel), which measures the difference between Baa and Aaa yields. The spread series look very correlated. They never intersect and the Baa spread is always higher, and more variable, than the Aaa spread. Aaa spreads have been well above zero over most of the sample period, and have been ever increasing since the early 80s. Baa spreads have been high and volatile during the decades 1970-2000 and featured an increasing trends in recent years. Most high observations for Aaa spreads were recorded in the years following the Asian crisis, whereas high values of Baa spreads occur during the 70s and 80s. The relative spread has peaked in the first half of the 80s, after the 1979 turning point in U.S. monetary policy. Following a period of decrease, it has increased again in recent years, following the outbreak of the Asian crisis and Russia's default (1998). One can also see the large movement which occurred in the aftermath of the 9/11/2001 terrorist attacks.

[6] http://research.stlouisfed.org/fred2/

[7] Both Aaa and Baa spreads exhibit moderate skewness and kurtosis. The Jarque-Bera test rejects normality in both series. The unit-root hypothesis is also rejected at 5% confidence level by a Dickey-Fuller (DF) test. (These results, however, should be interpreted carefully. In general, DF regressions are based on linear AR models, so their results could be misleading if the true dynamics is non-linear.)

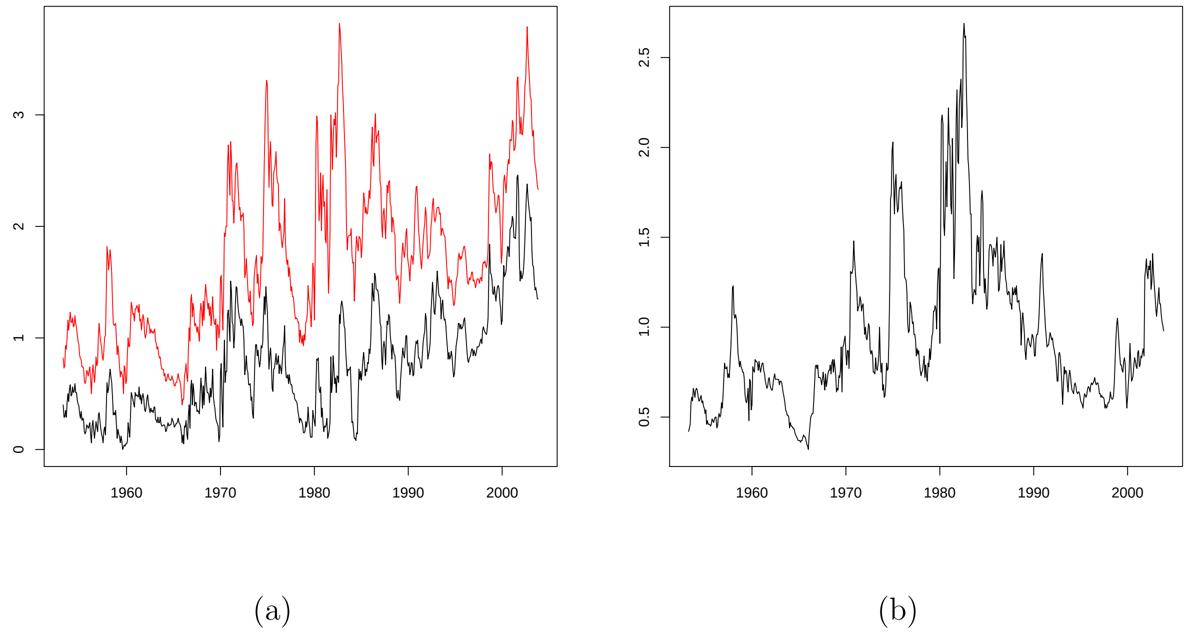


Figure 1: Time series of **Aaa** (in black) and **Baa** (in red) spreads (a) and relative **Baa-Aaa** spread (b).

Statistics	S^{Aaa}	S^{Baa}
Mean	0.75	1.69
Standard dev.	0.50	0.72
Minimum	0.00	0.40
Maximum	2.46	3.82
Kurtosis	3.33	2.57
Skewness	0.82	0.44
JB stat	71.92	24.23
ADF stat	-4.48	-3.86

Table 1: *Summary Statistics for Spread Data*
Legend: S^{Aaa}, S^{Baa} denote spread on Aaa-rated bonds, and on Baa-rated bonds; JB is the Jarque-Bera test for normality, and ADF is the Augmented Dickey-Fuller test for unit root.

Kernel density estimates of the spreads, in figure 2, evidence non-normal features in the data, such as heavy tails and skewness. These features are not unusual in financial data and imply high probabilities of observing extreme values.

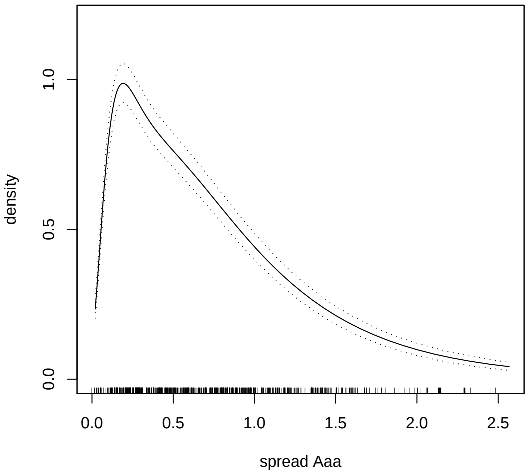
Although kernel density plots are useful summaries of the data, they ignore the time series nature of the data. When data are time series, conditional densities give a better description of the data generating process than marginal densities. Conditional densities provide insights into the dependence structure in the data. They also informs on features such as non-normality and non-constancy of mean and variance. Figure 3 shows that densities of spreads conditional on their past values shares features with estimated marginal densities, such as skewness and heavy-tails, but do vary over the conditioning variable. (Here, for reasons of space, I only report conditional densities for a lag of 6-months.) There is also a suggestion of bimodality in the upper boundary of the data. (One should be aware, however, that densities in this region is noisy, due to the sparsity of observations).[8]

So, spread densities have non-normal features, and do not have constant features over time.

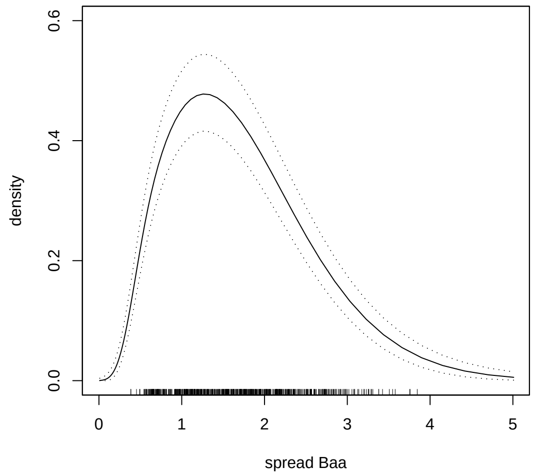
[8]The conditional densities of the spread are estimated using a direct Kernel-based approach, which uses the definition of conditional distribution as the ratio of joint and marginal densities. The density of the spread s, conditional on its past values s_{t-k}, where k denotes the time lag, is given by:

$$\hat{f}(s|s_{t-k}) = \frac{\hat{f}(s, s_{t-k})}{\hat{f}(s_{t-k})} = \frac{\sum_{i=1}^{n} W_h(s_{t-k,i} - s_{t-k}) K_b(s_i - s)}{\sum_{j=1}^{n} W_h(s_{t-k,j} - s_{t-k})}; \tag{19}$$

Here, W and K are kernel functions, and b and h are bandwidths for smoothing, respectively, over the variable s and its lagged value. In principle, one should estimate the above density on *each* value assumed by the conditioning variable, which makes conditional Kernel density estimation computationally expensive. To overcome this problem, I use "stacked conditional density plots" (Hyndman et al., 1996), which allow plotting the density for selected values of the conditioning variable, stacked by each other in a perspective plot. Conditional density plots have been produced using the package `hdrcde`, by R. Hyndman.

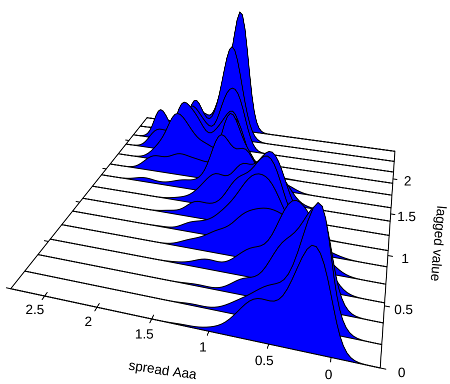


(a)

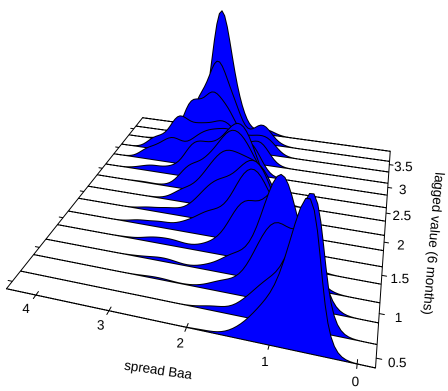


(b)

Figure 2: Kernel density estimates of Aaa (a) and Baa spreads (b).
(The histogram at the basis of each plot represents the frequency of the observations; dotted lines are variability bands.)



(a)



(b)

Figure 3: Stacked conditional densities of Aaa (a) and Baa spreads (b) on past values (6 months).

4 Testing linearity in Reduced Form Models

This section tests a fundamental assumption of affine models of credit risk: the linearity of yields-spreads. This is done folloing the procedure outlined in Section 2.2. That is, the linear specification of the spread, obtained by solving RFMs for the price of bonds, is tested against a model in which spreads are non-linear functions of factors using the GLR principle. This section assumes that corporate spreads are explained by factors capturing the risk-free term-structure and macroeconomic conditions.

As shown in Section 1, RFMs imply a spread which is linear in factors. The model of equation 7 suggests a simple linear regression model, in which corporate spreads depend on the risk-free rate and default risk:

$$s_t = \alpha + \beta_1 r_t + \beta_2 \lambda_t + \varepsilon_t, \quad t = 1, \ldots, T; \tag{20}$$

Here, s denotes the spread, r the risk-free rate, and λ the premium that captures the risk of default; the βs are unknown parameters, and ε an *iid* error term.

To test the validity of the affine specification above, we first need to identify a suitable alternative model. Following Fan and Jiang (2005), this is specified using a GAM framework, which offers a general non-parametric alternative to model 20 that is computable in an efficient way. The GAM model describes the spread as the sum of non-linear functions of factors, as follows:

$$s_t = \alpha + m_r(r_t) + m_\lambda(\lambda_t) + \varepsilon'_t; \tag{21}$$

Here, the ms are univariate smooth functions of factors. Formally, the testing problem is as follows:

$$H_0: \quad s_t = \alpha + \beta_1 r_t + \beta_2 \lambda_t + \varepsilon_t, \quad vs. \quad H_1: \quad s_t = \alpha + m_r(r_t) + m_\lambda(\lambda_t) + \varepsilon'_t; \tag{22}$$

To estimate the models of equations 20 and 21, it is essential to include observable variables that capture default risk. The following assumes that spreads are explained by factors capturing the risk-free term structure and macroeconomic conditions. This choice is motivated by existing theoretical and empirical literature on default risk, which suggests that risk-free rates and macroeconomic conditions are among determinants of yields on risky bonds (see Section 1). Among macroeconomic variables, inflation is a good candidate to explain the risk of default (see Wadhwani, 1986). The *term spread*, also known as the slope of the yield curve, is included along the short rate to better capture the the risk-free term-structure.[9] This leads to the following empirical model of the spread:

$$s_t = \alpha + \beta_1 \pi_t + \beta_2 r_t + \beta_3 y_t + \varepsilon_t; \tag{23}$$

[9]Single-factor models of the term-structure have been criticised for their inability to explain the observed variability of the yield curve through time and across maturities. Authors have argued that the term-structure dynamics is too complex to be summarised by a single-source of uncertainty. To address this, multi-factor representations, where the yield are explained by several state variables, have been introduced. Short and long-term rates are being used to explain intermediate maturities in the non-defaultable bond market (e.g. Knight et al., 2006).

Here, s denotes the corporate spread, r the risk-free short rate, y the term spread (i.e.) and π the inflation rate. (Here, the short-rate is measured by the 3-months Treasury Bill rate, and the term-spread by the difference between the 10-years Treasury note yield and the 3-month Treasury Bill. Inflation is measured by the 12-months percent change in the Consumer Price Index released by the Bureau of Labor Statistics.)

The next step in the testing procedure consists in computing the value of the GLR statistics for the testing problem above. This compares Residuals Sum of Squares (RSS) from the null and alternative model (the latter being estimated using the back-fitting algorithm), as follows:

$$\lambda_n(h) = n/2 \log(RSS_0/RSS_1); \tag{24}$$

This computation produces high values of the test statistics. This can be seen in table 2, where the second column reports observed values of the GLR statistics. p-values are computed using a version of the bootstrap method detailed in (Fan and Jiang, 2005, section 4.2). (The idea is that, since the asymptotic null distribution is independent on nuisance parameters/functions, for finite samples the null distribution can be approximated by a bootstrap method.) This bootstrap procedure comprises the following steps:

1. Fix the value of the bandwidth at its estimate $\hat{h}$. For the original data, compute the observed value of the test statistics $\lambda_n(\hat{h})$ according to eq. 24;

2. Sample randomly and with replacement from the residuals obtained at step 1. Define the bootstrap responses $s^b = \hat{\alpha} + \hat{\beta}_1 r + \hat{\beta}_2 s + \hat{\epsilon}^b$. This forms a bootstrap sample $\{S^b; R, Y\}$;

3. use the bootstrap sample to obtain the GLR statistics $\lambda_n^b(\hat{h})$;

4. Repeat steps 2 and 3 n times to obtain a sample of GLR statistics. Here the number of bootstrap replications is set at $n = 1000$.

The test rejects the null hypothesis of linearity. This can be seen, again, in table 2. Bootstrap p-values are listed in the third column. (P-values are computed as the proportion of times that the bootstrap statistics $\lambda_n^b(\hat{h})$ exceeds the observed value $\lambda_n(\hat{h})$.)

This procedure presents two difficulties: first, it is well known that p-values depend heavily on the sample size; second, the available data are time-series. Although the model-based bootstrap adopted here can be applied to the analysis of time series data (Davison and Hinkley, 1997, Chapter 8), it is useful to analyse sub-samples of data to assess the robustness of our results. Hence, we take random sub-samples of 200 observations for analysis. We repeat the procedure 100 times, as it is computationally very expensive, and obtain p-values for each sub-sample. The fourth and five columns 2 report average p-values for the sub-samples. This provides evidence that the non-parametric model is appropriate for the sub-samples at the nearly zero significance level.

The test above refutes the linearity assumption that underpins RFMs of yield-spreads. To mitigate this problem, the following uses non- linear models of spread determination.

Model	GLR test	p-value	sub-sample p-value
Aaa spread	73.04	0	0.0
Baa spread	50.60	0	0.0

Table 2: *Linearity test of yield-spread: GLR statistic.*

5 An empirical analysis of spread determination

This section explores the role of risk-free yields and inflation rate in determining spreads in the context of a multi-factor non-linear model of risky yields. It performs a graphical analysis of the spread determination using two important variables: the risk-free short-term interest rate and the inflation rate.

5.1 Spread and factors: a first look

This section estimates a simple regression model of spreads to the risk-free short-term rate, which allows departures from linearity. The model is estimated using a locally-linear regression technique (Fan, 1992), where the bandwidth is selected by cross-validation (Hardle and Marron, 1985).

The scatter-plots in figure 4, which show spreads against the short-rate, evidence several clusters. These correspond to two groups of observations which associate low spread values with low interest rates, and mid-range spread values with mid-range interest rates. In the top-left area of the graphic, there is a group of observations characterised by the association of high spreads with low interest rates; this group consists of data for the period post 9/11. (In figure 4, corresponding observations for this latter group are denoted by triangles.)

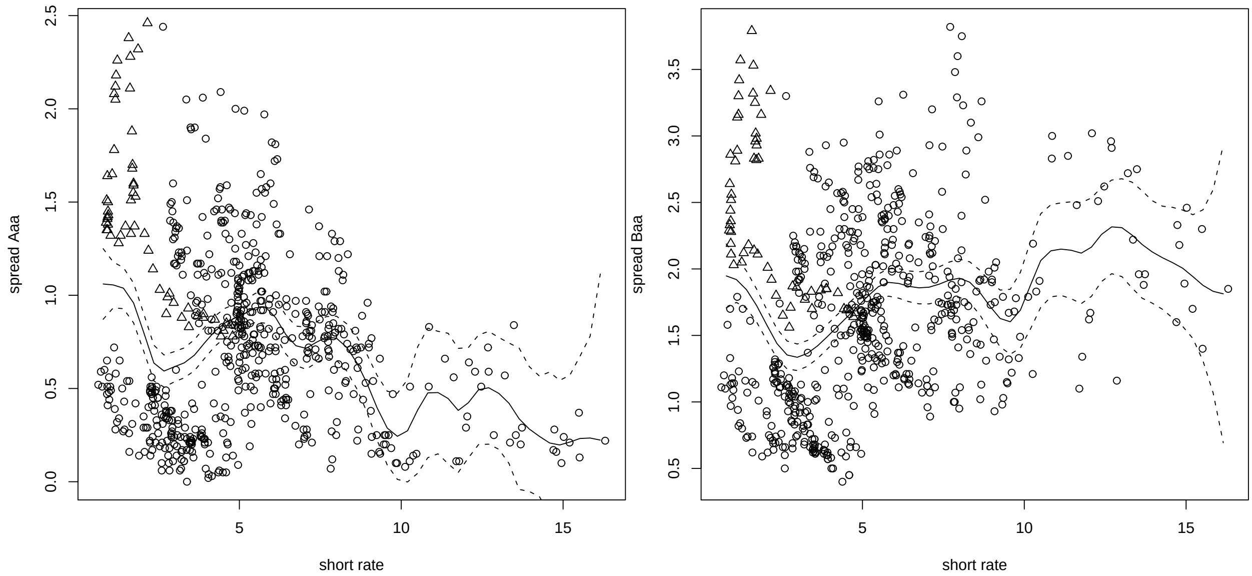
The following non-linear model summarises the relationship of spreads to the risk-free rate:

$$s_t^i = m(r_t) + e_t, \qquad t = 1, \dots, T; \tag{25}$$

here, s is the spread, r the risk-less rate, $i = Aaa, Baa$ (in the following, the index i is omitted); m is a smooth function whose shape is unrestricted, hence allowing departures from linearity, and e is an *iid* error term with zero mean and standard deviation σ. The continuous curves in figure 4 represent the non-parametric estimates of the regression function m, with variability bands.[10] These estimates show a non-linear relation between

[10]In non-parametric analysis, variability bands quantify the variance of the estimate without accounting for the non-zero bias, offering a measure of the degree of variability present in the estimate; so, the term variability band is used in order to distinguish such measures from proper confidence bands. In practice, variability bands are computed using twice the standard error, derived by the asymptotic variance of the non-parametric estimator for the model of equation 25:

$$var(\hat{m}(x)) \approx a(k)\frac{\sigma^2}{Thf(x)}$$



(a) (b)

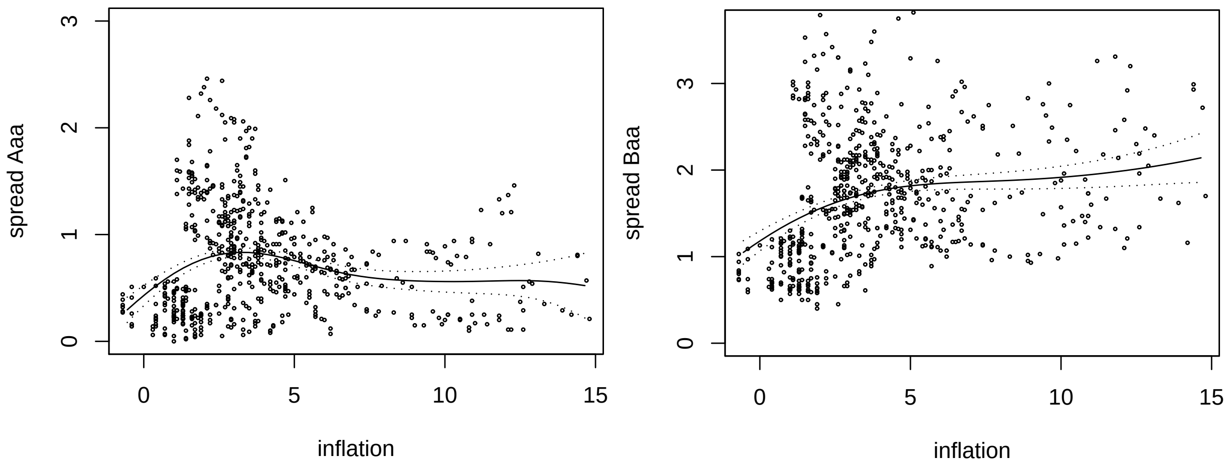
Figure 4: Aaa (a) and Baa spread (b) vs. short-rate: scatter-plot and non-parametric regression curve, with indication of its variability. (Observations denoted by triangles refers to post 9/11 observations.)

short-term yields and spreads. The relation between Aaa spreads and the short-rate highly non-linear (left panel). The relation between Baa spreads and the short-rate, negative for low level of the short rate, becomes positive for higher values of the short-rate.[11]

Figure 5 present scatter-plots and regression estimates of spreads against inflation rates. The relation between Aaa spreads and inflation is increasingly concave for low to mid-range inflation rates, turning negative for higher inflation rates ($> 5\%$). Smooth regressions indicates a positive concave relationship between Baa spreads and inflation rates, implying that yield spreads increases with inflation at decreasing rates. This seems in line with the view that increasing inflation correspond to increasing spreads.

where a is a constant that depends on the kernel choice (k), h the bandwidth, T the sample size and f the density of the observations. (This method follows what proposed in Bowman and Azzalini, 1997, chapter 4.) One can see that the variance depends inversely on the local density of the data, given by $Thf(x)$. This explains why bands depicted in figure 4 tend to be larger at the right boundary of the data, where data are sparse.

[11]The estimation of the same model, leaving out the post 9/11 observations, provided a clearer non-linear pattern. There is evidence of two regimes in the Aaa spread to short-rate relationship: this is increasing for low interest rates, and decreasing for high interest rates (that is, greater than 5%). The relationship between Baa spreads and short-rate is increasing and concave.



(a) (b)

Figure 5: Aaa spread and Baa spread vs. inflation rate: scatter-plot and non-parametric regression curve (with variability bands).

Scatter-plots and regression lines give interesting insights into the relationships of interest: correlations between interest rate and spread, and inflation and spread, vary among rating classes, and are non-linear. This contrast with the existing literature, which indicates a negative sign for the correlation between spreads and short-rate (see, for example, Duffee, 1998). This analysis, however, remains purely descriptive. Yields are known as better described by multi-factors models, as opposed to single-factor models. This is addressed in the following section.

5.2 An additive model of the spread

This section discusses specification and estimation of the non-linear relationship of spread to factors uncovered by the linearity test performed in section 4. The actual form of the function that relates factors and spreads is unknown, and economic and financial theory is of little help. As seen above, empirical and theoretical studies of default risk suggest the choice of factors, but postulate simple linear relationships between those factors and risky yields. Because non-parametric models do not place assumptions on the functional form of the relationship of interest, they are the natural choice. Consider the following general non-parametric model:

$$s_t = m(\pi_t, r_t, y_t) + \epsilon_t; \tag{26}$$

Here, r denotes the short rate, y the term-spread, and π the inflation rate, m is a smooth function of factors, and ϵ an *iid* error term. In principle, such model can be estimated using standard surface-smoothers techniques. However, this method suffers the curse of dimensionality and calls for the use of dimension-reduction techniques such as GAMs (see

Section 2.1). An additive framework offers a convenient way of modelling spreads as a function of several variables whose effect is non-linear:

$$s_t = m_\pi(\pi_t) + m_r(r_t) + m_y(y_t) + \epsilon'_t; \tag{27}$$

Here, the response variable is the sum of one-dimensional non-linear functions of factors. Little is assumed about the shapes of these function, apart from smoothness. The non-linear functions are estimated non-parametrically, using the *back-fitting* algorithm (Buja et al., 1989), whose basic building block is the local-linear smoother. The estimation output is essentially graphical, as plots of individual smooth terms allow us to to examine the possible non-linear effect of each explanatory variable.[12]

Table 3 presents results of the estimation for the non-parametric model of equation 28. (Table 4 gives estimates for a linear model of the spread, such as the one of equation 23, for comparison.) Models have been estimated separately for Aaa and Baa spreads. Table 3 reports values of approximate significance tests for each smooth term. All non-parametric terms are statistically significant when compared to critical values of a F distribution. (Note that this amounts to test the validity of a model in which the corresponding component is linear.)[13]

Figure 6 presents estimates of the smooth terms of equation (28). (Those plots on the left column are for the model estimated with Aaa spreads data; those on the right to Baa spreads data.)

The inflation effect is clearly non-linear. Interestingly, the shape of the smooth functions looks very similar for both models. Slopes are positive at very low inflation rates, turn negative at medium-range rates (between 3% and 7%), and again positive at high levels of inflation (above 7%). In the region characterised by high inflation rates, where only few observations are available, confidence bands are large, but not dramatically large, indicating that the estimation is reasonably stable.

High and increasing inflation is often associated with economic growth. One would naturally expect to observe a negative correlation of inflation to spreads (see, for example, Couderc et al., 2008). However, the mechanism through which inflation affects the risk of default is possibly more complicated than wat a business-cycle perspective suggests. There is another effect that should be taken into account: higher inflation leads to an increase in the cost of borrowing. This has been effectively illustrated in Wadhwani (1986): an increase in inflation causes cash-flow problems for firms, due to an unexpected rise in (nominal) interest rate payments. Indeed, the positive relation between spread and inflation rate uncovered here is consistent with the view that rising inflation increases the number of

[12] Estimation and graphics are performed using the package `gam` in `R`.

[13] The F test here is based on (approximate) degrees of freedom and residual sum of squares. Hastie and Tibshirani (1990) discuss the computation of degrees of freedom for additive models and argue that, despite the null distribution of the F statistics is unknown, F tables can be used as guidance (see Hastie and Tibshirani, 1990, Sections 5.4.4 and 5.4.5). Clearly, the F statistics here is also an approximation of the GLR statistics of Fan and Jiang (2005), as $\lambda_n(h) = n/2 \log(RSS_0/RSS_1) \approx n/2(RSS_0 - RSS_1)/RSS_1$ for testing significance of parametric components.

bankruptcies. Thus, bond-holders request compensation for the higher default risk, which leads to an increase in the corporate spread.

This interpretation is complicated by the fact that the response of spreads to inflation varies with different levels of inflation. Indeed, it can be observed that the positive relationship spread-inflation corresponds to high and low levels of inflation. This can be interpreted as follows: (a) low levels of inflation, usually associated to a stagnating economy (deflation), increase the market's perception of risk and boost corporate yields; (b) investors' perceive high rate of inflation as a sign of economic instability, and this, in addition to the cash-flow effect described above, determines an upward pressure on spreads. (Indeed, historically, the US economy was characterised by high levels of inflation in the 70s, the period known as "stagflation", a mixture of high inflation and stagnation.) In contrast, if investors perceive mid-range inflation rates as a symptom of growing economy, corporate spreads narrow. Noticeably, such levels of inflation have been observed during good times for the U.S. economy, such as mid 80s and mid 90s.

The term-spread effect is positive and significant in both models, but looks weaker for the Aaa spread model. The function is S-shaped. This result is consistent with findings in Couderc et al. (2008), who found a large significant effect of the slope of the term structure on default rates.[14] The slope of the risk-free yield curve is often interpreted as reflecting inflation and growth expectations. So, a positive slope indicates expectations of future growth, whereas a negative slope (or curve inversion) indicates worsening economic conditions. Indeed, one can see that negative values of the term-spread are associated with rising spreads in the Baa model. A negative slope, however, do not seem to affect Aaa spreads, perhaps indicating the greater sensitiveness of lower rated securities to expectations of negative growth. (Recall that Baa spreads are the lowest rated of investment-graded bonds.) The association of a positive slope with increasing spreads is more difficult to interpret. There are two possible explanations. An increase in slope due to a decrease in the short rate of interest, usually associated to recessions, could lead to an increase in the risk of default. Otherwise, an increase in slope due to an increase in the long rate, which suggests that inflation is expected to rise, would lead investors to expect a deterioration in credit quality and to require higher yields to hold corporate bonds.

The short rate effect is also highly non-linear. At low and medium-range levels, the non-linear pattern is similar in both models, first decreasing and then increasing. For higher rates (i.e. greater than 7.5%), the smooth function has a decreasing trend in the Aaa spread model, and an increasing trend in the Baa data. (Here, confidence bands are large for values of the interest rate greater than 10%.)

A negative short rate effect is consistent with the view that economic downturns lead both to lowering interest rates and a deterioration in credit quality, thus to widening spreads. The positive slope in the medium-range interest rates region may be linked to a recovery in the economy and to the fear of inflationary pressures, leading to increasing

[14]Unlike Couderc et al. (2008), Duffee (1998) failed to find a significant effect of the slope on corporate spreads. Figlewski et al. (2008) looked at the impact of the long-term rate, and found a significant positive effect of this variable on transition to default. In contrast, Duffie et al. (2007) did not find long rates capable of explaining default probabilities.

spreads. The overall impression, however, is that of a decreasing trend in the Aaa spread-short rate relation.

On the other hand, widening Baa spreads are associated to increasing short rates. In this case, rising short-term interest rates seem to lead investors away from defaultable bonds, lowering bonds prices, and increasing yields. These short-rate effects patterns confirm the explorative results in the previous section, and contrast with negative correlations found in previous studies.

Model	term	Npar *df*	Npar F	$Pr\ F$	RSS (*df*)	AIC
Aaa spread	$m(\pi)$	2.6	18.25	0	94.28 (558)	620.40
	$m(r)$	3.4	24.38	0		
	$m(y)$	3.2	6.82	0		
Baa spread	$m(\pi)$	2.6	9.49	0	159.75 (558)	922.05
	$m(r)$	3.4	20.77	0		
	$m(y)$	3.1	5.10	0.001		

Table 3: **GAM model of the spread: approximate significance of smooth terms.** Legend: Npar *df* denotes degrees of freedom, and Npar-F the approximate F value for each smooth term; *pr* Fs give corresponding p-values. RSS is the sum of squared residuals, with degrees of freedom (*df*) in parentheses, and AIC is the Akaike Information Criterion for the overall model.

Model	term	estimate (se)	t value	$Pr\ (>\lvert t\rvert)$	Adj. R^2	F-stat
Aaa spread	int.	0.619 (0.049)	12.541	0.000	0.169	44.32
	π	0.026 (0.009)	2.911	0.003		
	r	-0.034 (0.009)	-3.539	0.000		
	y	0.157 (0.015)	9.886	0.000		
Baa spread	int.	0.783 (0.06)	12.496	0.000	0.342	111.4
	π	0.069 (0.01)	5.919	0.000		
	r	0.037 (0.01)	3.032	0.002		
	y	0.334 (0.02)	16.507	0.000		

Table 4: **Linear model of the spread.**

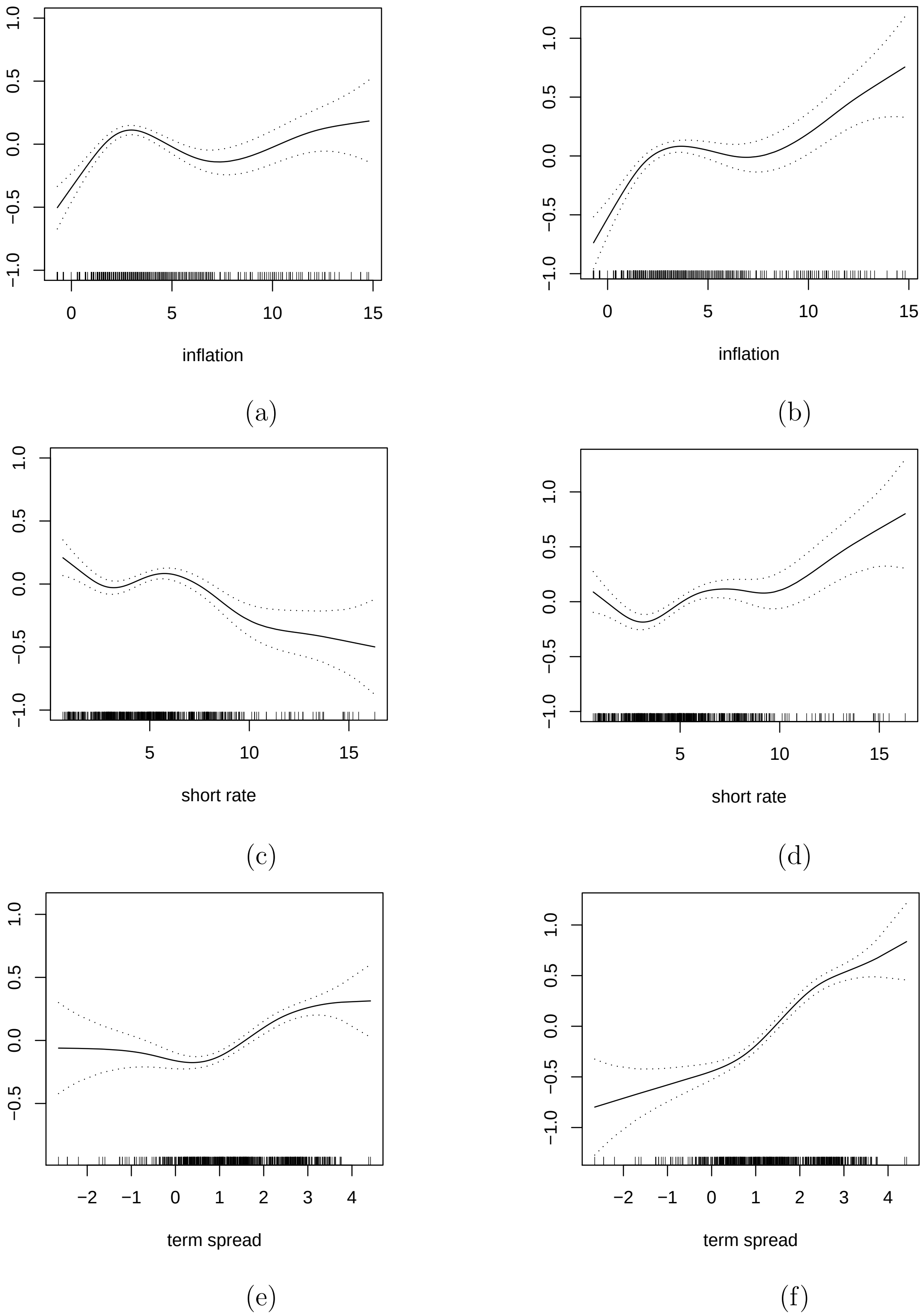
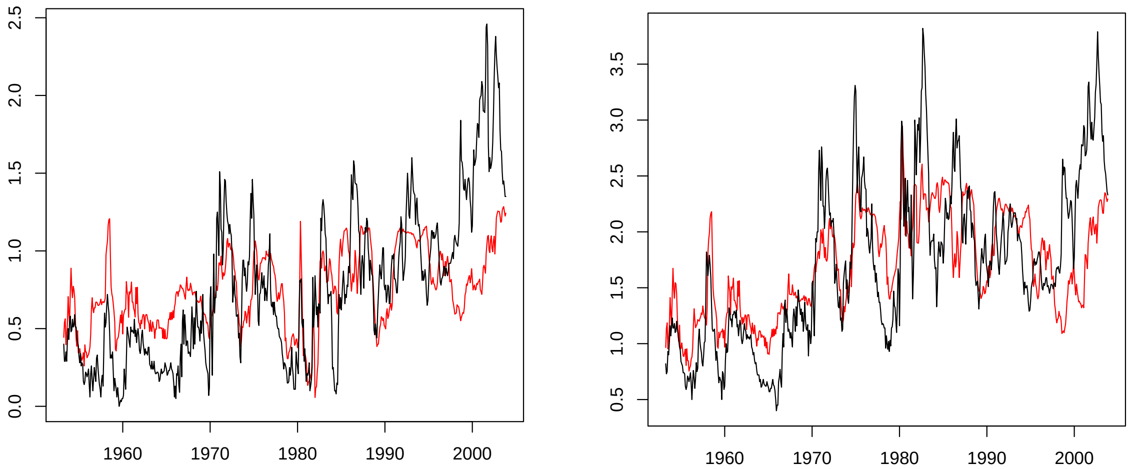


Figure 6: GAM model, estimates of smooth terms: Aaa (left column) and Baa spread (right column). (The "histogram" at the basis of each plot represents the frequency of the observations, dashed lines are variability bands.)

The graphs of figure 7 compare historical Aaa and Baa spread values with spread values predicted from the non-parametric model of this section. The left panel compares fitted and observed Aaa spread. As it can be observed, the model is good at capturing movements in the actual series. It is capable of predicting the increasing trend in the spread which occurred since the mid 90s (although the actual spread is clearly underestimated), and it also follows the actual series closely in the first two decades of the sample. The model's performance, however, weakens in the middle years of the estimation period; in particular, it fails to predict the sudden increases and subsequent contractions that characterised the spread series during the 80s. The right panel, which show observed and fitted values of the Baa spread, presents similar results. The non-parametric model is good at capturing features of the actual series, but its performance worsens in the middle years of the sample (e.g. it cannot detect the drop in Baa spread that follows its historical maximum in 1982).



(a) (b)

Figure 7: GAM models with inflation: comparing observed (in black) and fitted values (in red) of Aaa spread (a) and Baa spread (b).

Overall, the empirical results of this section show that non-parametric models are capable of capturing the behaviour of corporate spreads, and, in doing so, uncovering novel and interesting relations among variables. Inflation is found to determine spread in a non-linear fashion. The relation between spread and risk-free term structure is also non-linear.

6 Forecasting the spread

This section conducts a simple *out-of-sample* forecasting experiment. This helps determining whether the non-linear pattern in historical corporate spread data identified in previous sections is useful to forecast the future course of the spread. This uses the idea of conditional forecasting, which enables making *what-if* questions. For example, if inflation is equal to a certain value, what is the expected value of the spread.

Conditional forecasting is based on the interpretation of a regression model as a predictive model. When a non-parametric regression model is used, conditional forecasts give expected values of the response *conditional* on functions of explanatory variables. Previously, the spread (s) was modelled as a non-parametric function of the term-spread (y), the short rate of interest (r), and the inflation rate (π):

$$s_t = m_\pi(\pi_t) + m_r(r_t) + m_y(y_t) + \epsilon'_t; \tag{28}$$

Based on the model above, forecasts for the out-of-sample period are formed conditionally on the values of the explanatory variables, which are assumed known. Thus, following estimation over $t = 1, ..., T$, 1-step-ahead forecasts are computed as follows:

$$\hat{s}_{T+h} = \hat{m}_\pi(\pi_{T+h}) + \hat{m}_r(r_{T+h}) + \hat{m}_y(y_{T+h}), \qquad h = 1, .., H; \tag{29}$$

where h indices out-of-sample observations. This additive framework provides a simple and parsimonious way of generalising (in the non-parametric, non-linear sense) a predictive linear model. The advantage of this GAM representation is amenability of computation. There are procedures capable of evaluating the smooth fitted functions at new (out-of-sample) values of the covariates, at least if these are in the domains of the original data (Chambers and Hastie, 1992).

In order to perform this experiment, the sample is divided into estimation and forecasting period; the 30 most recent observations are reserved for forecasting purposes (these go from 2004:1 to 2006:6).

The model of equation 29 is estimated using in-sample data; then, the fit is used to produce predicted values of the spread over the out-of-sample period. Graphical analysis enables forecasts to be compared to observed spreads. The left panel in figure 8 presents actual (s_t) and fitted values ($\hat{s}_t$) of the Aaa spread (the latter are reported in red, and refers to the forecasting period only); the right panel does the same for the Baa spread. These results are encouraging: the fitted models capture the declining trend in the spread over recent years. However, an interesting feature of the graphs is the suggested upward trend in the spread forecasts for the last few observations. A closer look at observed spreads shows that the decreasing trend that followed the 2001 peak has recently slowed down. In the last few months of the sample period, spreads have been increasing again, as well as corporate yields. Others prominent features of the data for the forecasting period are the increase in inflation, and the large increase in the short rate of interest, which led to a flattening of the yield curve. Forecasts of the spread model seem capable of capturing these tendencies in the data.

Figure 8 shows that, at the end of the forecasting period, the model underestimates realised spreads. In general, forecasts are almost never precisely accurate. Thus, to evaluate the forecasting performance of a model one should compare its predictions to those of other competing models, on the basis of some appropriate criteria. Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are widely used measures of forecast accuracy (Gooijer and Hyndman, 2006), given by:

$$RMSE = \sqrt{\frac{1}{H}\sum_{t=1}^{H}[y_t - f_t]^2}; \tag{30}$$

$$MAPE = 100/H \sum_{t=1}^{H} \frac{|y_t - f_t|}{y_t}; \tag{31}$$

Here f is the forecast, y the observed value of the series, and H the length of the forecasting period. These measures are based on the forecast error in t, defined as $e_t = y_t - f_t$, and on a loss function $L = l(e)$: the RMSE, for example, is based on a quadratic loss function. The MAPE is based on the average distance between forecasts and actual values, without regard to whether individual forecasts are overestimates or underestimates. The RMSE also shows the size of the error without regard to sign, but it gives greater weight to larger errors.

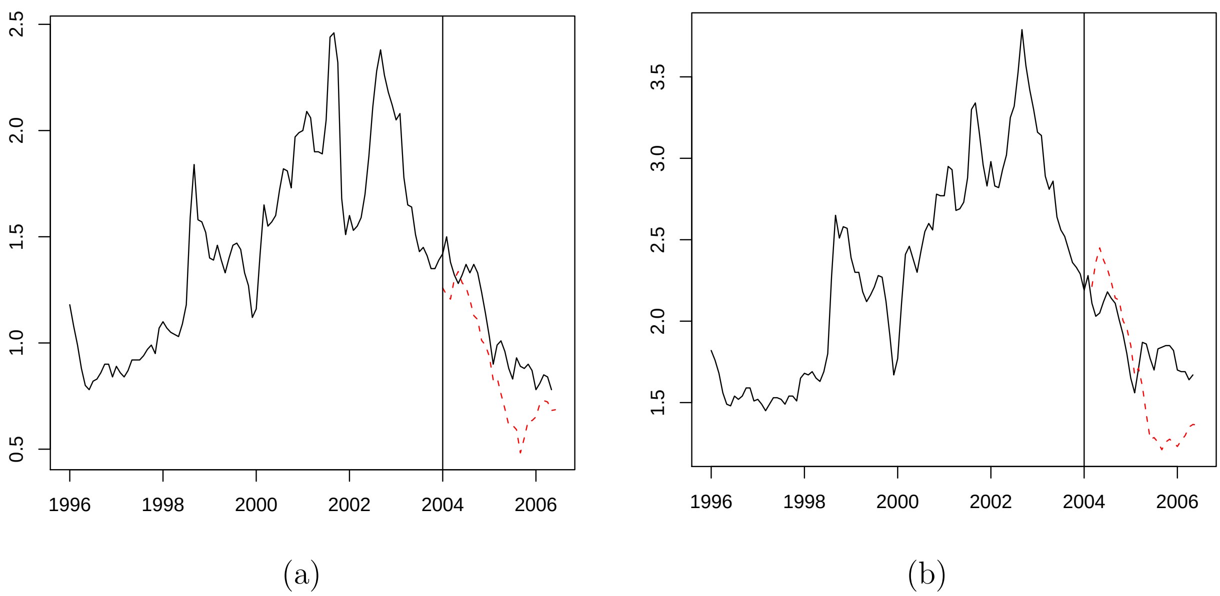


Figure 8: **Aaa spread** (a) and **Baa spread** (b): realised and predicted values (in red).

The following compares the forecasting performance of the non-parametric spread model of equation 28 (the base model) to several alternative specifications: 1) a linear version of the base model; 2) an additive model which includes past values of the spread; 3) a benchmark linear first-order auto-regression, denoted by $AR(1)$. The linear model is as follows:

$$s_t = \alpha + \beta_\pi \pi_t + \beta_r r_t + \beta_y y_t + \varepsilon_t; \tag{32}$$

The additive models with lagged spread is specified as:

$$s_t = m_\pi(\pi_t) + m_r(r_t) + m_y(y_t) + m_l(s_{t-4}) + \epsilon''_t; \tag{33}$$

(Here, the selected past value of the spread is the spread observed at the beginning of the previous quarter.) Models estimates are used to construct forecasts and forecast errors.[15]

Tables 5 and 6 report, along with measures of (in-sample) goodness of fit, the two criteria of forecasting performance, RMSE and MAPE. For both Aaa and Baa spreads, the RMSE and MAPE have their minima for the non-parametric model with past values of the spread. The base model, however, does better than the linear model and the $AR(1)$ model when estimated using Aaa spread data. Its performance, however, worsens for Baa spread data.

Figures 9 – 11 present comparisons of forecasts and observed values for the additive model with lags, the linear model, and the $AR(1)$ dynamics. Looking at the various graphs, one can see that the non-parametric model with lag, and the AR(1) model, tend to overestimate the observed spread. The linear model, instead, tends to underestimate it. These tendencies are not detected by accuracy criteria such as RMSE and MAPE, which do not distinguish between positive and negative forecast error. Furthermore, forecasts based on AR(1) do not capture the increase in spreads at the end of the sample. Thus, we believe that the non-parametric model of equation 28 still provides useful information for forecasting, and is preferable to models based on linear specifications.

The analysis of this section showed that non-linear models lead to forecast improvements on linear specifications. These results are encouraging, however, non-parametric conditional forecasting has several limitations: a) it is not possible to form forecast when new values of explanatory variables are outside the domain of the available observations; b) the computation is restricted to 1-step-ahead forecasts, as it is not possible to find recursive formulas such as those used in time series forecasting. These problems are due to the model-free nature of the estimation. The second issue may be addressed using time series processes of the explanatory variables to form multi-step ahead forecasts. However, explanatory factors follow diffusion processes in term-structure analysis, which should be approximated with discrete auto-regressions. Furthermore, these processes themselves are likely to be non-linear, in which case multi-step forecasts quickly become extremely complex. This would introduce both uncertainty and bias in the forecasts. The first issue could be addressed specifying non-linear but parametric models. However, forecasts based

[15]The $AR(1)$ model has been chosen accordingly to the AIC criterion. The model estimates and forecasts are produced using the package `stats` and `forecast` (the latter by R. Hyndman). The additive model with lags includes the spread observed at the beginning of the previous quarter (i.e. at time $t-4$). We chose to include only one lagged value because of the dimensionality problem that comes with additional explanatory terms in non-parametric estimation.

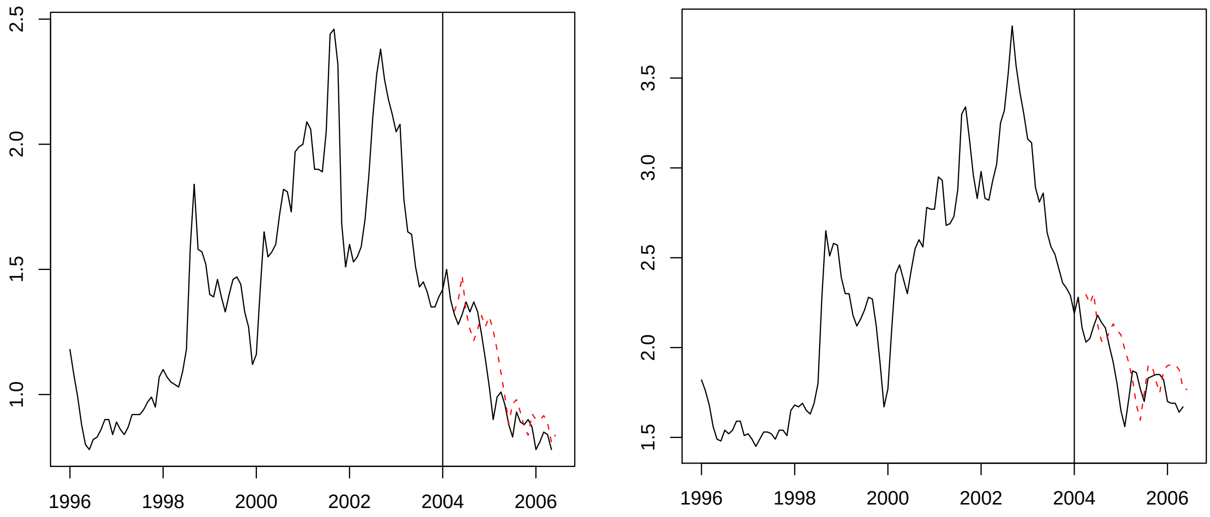
on parametric non-linear models often suffers of the same problems of non-parametric forecasts listed above, unless it is possible to identify a simple relation among variables: the following illustrate an attempt to specify a parametric model of the corporate spread using the information provided by the non-parametric model.

Model	RSS	AIC	RMSE	MAPE
Model 28	98.69	648.10	0.182	15.64%
Model 32	126.71	781.93	0.248	21.60%
Model 33	25.76	-196.30	0.112	8.65%
$AR(1)$	9.60	-788.17	0.184	17.13%

Table 5: *Forecasting models of the Aaa spread*: relative performance. Legend: RSS denotes the residual sum of squares, AIC the Akaike information criterion; RMSE is the root mean square forecast error, and MAPE the mean absolute percentage error.

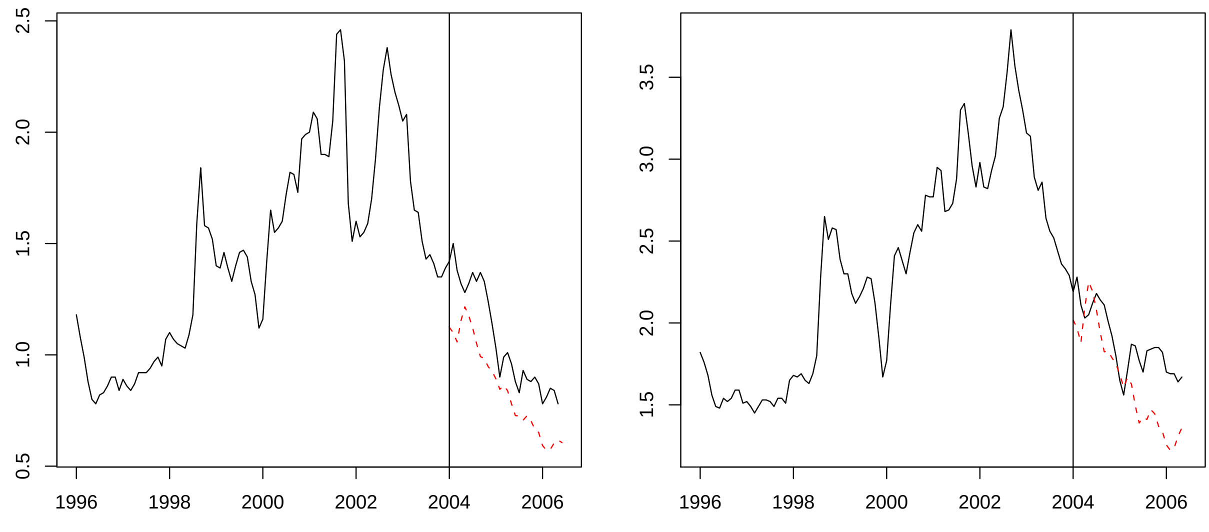
Model	RSS	AIC	RMSE	MAPE
Model 28	171.41	984.85	0.367	16.44 %
Model 32	203.84	1070	0.323	15.39%
Model 33	48.18	220.12	0.177	8.35%
$AR(1)$	17.63	-421.83	0.204	9.80%

Table 6: *Forecasting models of the Baa spread*: relative performance.(Legend: as above.)



(a) (b)

Figure 9: Additive model with lags (eq. 33): observed and forecasted values (in red) for **Aaa spread** (a) and **Baa spread** (b).



(a) (b)

Figure 10: Linear model (eq. 32): observed and forecasted values (in red) for **Aaa spread** (a) and **Baa spread** (b).

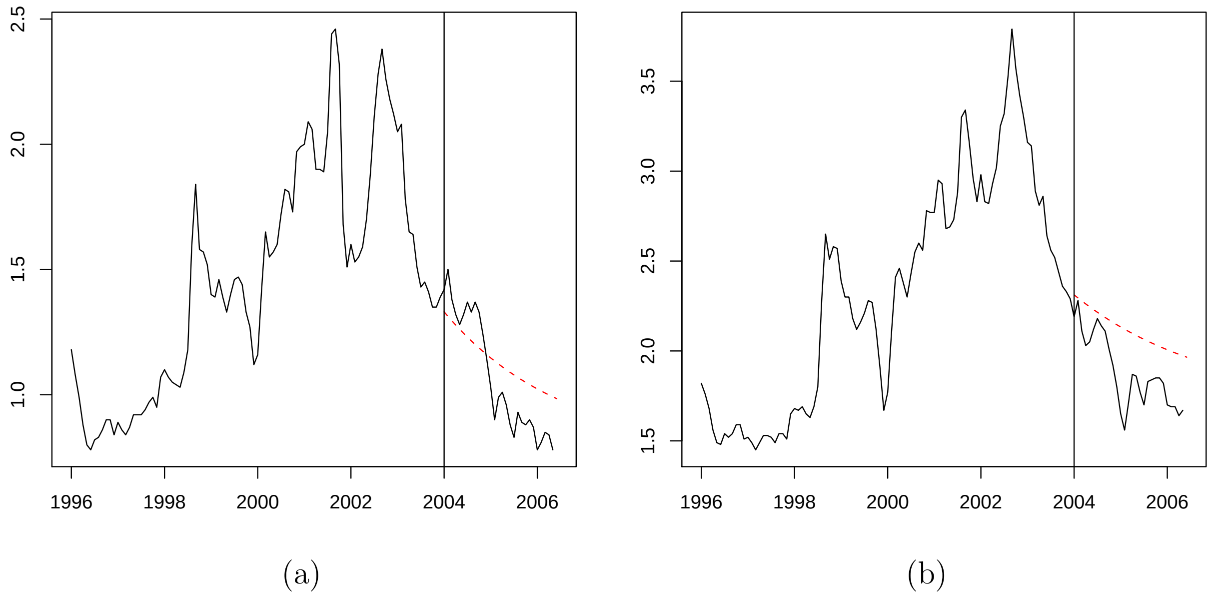


Figure 11: AR(1) model: observed and forecasted values (in red) for **Aaa spread** (a) and **Baa spread** (b).

7 A parametric non-linear model of the spread

This paper has illustrated several uses of non-parametric regression techniques, such as model specification, estimation, testing and forecasting. One of the purposes of non-parametric analysis is to aid parametric modelling, by suggesting specific linear or non-linear models. This section does this focusing on the relation of corporate spreads to inflation. The analysis is conducted using the smooth transition regression approach.

The analysis of an additive spread model in section 5.2 suggested a possible threshold behaviour in the response of spreads to inflation, with the response changing slowly over values of inflation. To study threshold behaviour of economic variables, regime switching models have been proposed. Such models interpret economic relations as characterised by "regimes", each regime being associated to a particular state of the world. (As a result, regression models are piecewise linear specifications.) Among regime switching models, *Smooth Transition Regressions* (STRs) (Teräsvirta, 1994) are preferred here as they model changes in regime as continuous functions of explanatory variables. STRs describe the dependence of the spread (s) on the vector of explanatory variables (X) as follows:

$$s_t = \beta_0' X_t + F(w_t)\beta_1' X_t + u_t, \qquad t = 1, \ldots, T; \tag{34}$$

Here, the β_is are vectors of parameters, w a transition variable, and u an *iid* error term. F, the function that determines the transition between regimes, often takes the logistic form:

$$F(w_t) = \frac{1}{1 + \exp\{-\lambda \Pi_{k=1}^{2}(w_t - c_k)\}}; \tag{35}$$

Regimes are associated with the extreme values of the transition function, $F = 0,\ F = 1$, which changes monotonically as the transition variable w increases. The parameters c_js are interpreted as thresholds between regimes, and λ as the speed of transition. Different values of k are associated to different types of regime-switching behaviour. While a value of $k = 1$ is capable of characterising asymmetric behavior, $k = 2$ is appropriate in a situation in which the behaviour of the relation of interest is similar at low and high value of s and different in the middle. Clearly such structure does not only allow interactions between transition and explanatory variables, but also permits intermediate positions between regimes — hence, the name of *smooth transition* regression. When the transition function is logistic, STRs are called Logistic Smooth Transition Regressions (LSTRs). van Dijk et al. (For details on STR modelling see, for example 2002).

We start the analysis by testing the linear specification against the LSTR type of non-linearity. (Inflation is selected as transition variable.) This is done using a small sample version of the test proposed by Luukkonen et al. (1988), which compares the observed test statistic with critical values from a F distribution with $3D$ and $T - 4D - 1$ degrees of freedom, where D is the length of the vector X (see van Dijk et al., 2002, p.13). The LM-type linearity of Luukkonen et al. (1988) is based on a third-order Taylor-series expansion of the transition function of equation 35, which yields the following auxiliary regression:

$$s_t = \beta_0' X_t + \sum_{j=1}^{3} \beta_j' \tilde{X}_t w_t^j + \varepsilon; \tag{36}$$

The null hypothesis is $H_0 :\ \beta_1 = \beta_2 = \beta_3 = 0$. (Under the null, the test statistic follows a χ^2 asymptotic distribution.) Once linearity is rejected in favour of LSTR nonlinearity, we need to select the appropriate form of the transition function. Here, the choice is between a model with a single switching mechanism ($k = 1$ in equation 35) and a model with two switching mechanisms ($k = 2$) (hereafter referred to as, respectively, LSTR(1) and LSTR(2)). This is because, as seen above, the LSTR(2) model is especially useful in the case of reswitching. To choose between LSTR(1) and LSTR(2), Teräsvirta (1994) suggests the following test sequence:

$$\begin{aligned} H_{03} &: \beta_3 = 0 \\ H_{02} &: \beta_2 = 0 | \beta_3 = 0 \\ H_{01} &: \beta_1 = 0 | \beta_3 = \beta_2 = 0 \end{aligned}$$

The LSTR(2) specification should be preferred when the test of H_{02} yields the strongest rejection.

Table 7 gives the results of the linearity test, which support non-linearity of the LSTR type. The data strongly reject the null hypotheses of linearity (H_0) for both Aaa and Baa spreads when inflation is selected as transition variable. The test sequence based on hypotheses H_{03}-H_{02}-H_{01}, however, does not provide a clear answer in favour of either $k = 1$ or $k = 2$. So, I estimate both LSTR(1) and LSTR(2) models and postpone the choice between the two specifications to the stage of model fitting and evaluation.

Hypothesis	F-stat		
Aaa Spread			
H_0	$F(9, 625) =$	23.02	(0.0000)
H_{03}	$F(3, 625) =$	10.27	(0.0000)
H_{02}		24.88	(0.0000)
H_{01}		28.21	(0.0000)
Baa Spread			
H_0	$F(9, 625) =$	14.57	(0.0000)
H_{03}	$F(3, 625) =$	20.28	(0.0000)
H_{02}		14.13	(0.0000)
H_{01}		6.88	(0.0001)

Table 7: F **statistics for the linearity tests of the spread models.** (p-Values in parentheses; transition variable: inflation, π.)

Since linearity is rejected by the F test, we proceed by estimating the LSTR models, which yields results reported in table 8 for the Aaa spread and table 9 for the Baa spread.

A first look at the estimation results shows that all coefficients are significant, with large t-ratios. One interesting feature is that the estimates of γ, the parameter that controls the speed of the transition among regimes, are very large, such that the transition between regimes is rapid.[16] (Transition functions for both models are depicted in figures 12 and **??**, panel (a), in appendix C.) The LSTR(1) model captures two regimes. The estimates of γ and c are such that the change of the logistic function from zero to 1 takes place rapidly for an inflation rate of about 2%.

A comparison of the LSTR(1) and LSTR(2) specifications shows there is little difference between the two models in terms of overall fit and parameters' significance. (For example, in the Aaa spread model, there is only a 0.2% gain in terms of residuals standard deviation when moving from a LSTR(1) to a LSTR(2) specification.) Estimation results casts serious doubts on the ability of the LSTR(2) model to fit the data. A problem is that the LSTR(2) model does not produce plausible estimates of the thresholds. (For example, for the Aaa spread model, c_1 and c_2 are equal to, respectively, -0.722 and 2.268. These values are nowhere near those suggested by the estimates of the non-parametric additive model, which can be seen in figure 6.) This is also evidenced by the plots of the transition functions,

[16]The estimates of γ have t-ratios close to zero. Teräsvirta (1994) points out that, in such case, the large standard deviation and small t-value are consequences of lack of information around the thresholds, rather than of genuine lack of significance of the parameter.

which show a remarkable lack of data to the left (right) of the first (second) threshold in the model for Aaa (Baa) spreads.

To clarify this point, we proceed by imposing restrictions on the parameters of the LSTR models so that risk-free rates are present only in the linear part of the model. This exploits general trends in the relation of risk-free rates to spread indicated by the estimation of the additive model and linear model of the spread. The estimation of the restricted model gives thresholds values closer to those suggested by the non-parametric model. The thresholds estimates, however, have very low t-ratio, which suggests they are not reliable. (Estimates for the restricted models are reported in the last columns of table 8 and 9.)

This analysis showed that LSTR model of the spread seems capable of capturing one of the switching mechanism identified by the non-parametric model in the relation between inflation and spreads. Noticeably, the threshold of a LSTR(1) model of the spread captures the first turning point of the inflation effect highlighted by the non-parametric estimation. The non-parametric model, however, seems to suggest that more than one regime is present in the relation of spread to inflation. The LSTR(2) model, however, does not seem capable of reproducing this feature. One possible explanation for this is that the remaining non-linearity is more complex than what is possible to render with a non-linear but parametric model.

model: coefficient	**LSTR(1)**	**LSTR(2)**	Restricted
linear part			
int	-0.777 (-7.031)	-0.822 (-7.127)	1.066 (14.910)
π	0.132 (2.745)	0 .171 (3.059)	-0.040 (-1.900)
r	0.176 (6.203)	0.175 (6.161)	-0.044 (-4.943)
y	0.593 (13.467)	0.586 (13.261)	0.123 (8.312)
nonlinear part			
int	1.887 (14.947)	1.932 (14.813)	-0.758 (-9.778)
π	-0.137 (-2.804)	-0.176 (-3.110)	0.100 (4.965)
r	-0.235 (-7.853)	-0.233 (-7.815)	
y	-0.53528 (-11.479)	-0.528 (-11.296)	
transition parameters			
γ	166.683 (0.451)	158.791 (0.481)	513.27 (0.001)
c_1	2.266 (28.625)	-0.722 (-14.851)	1.406 (0.197)
c_2		2.268 (31.904)	5.695 (1.131)
goodness-of-fit measures			
se	0.382	0.381	0.404
AIC	-1.912	-1.911	-1.797
$\overline{R}^2$	0.412	0.414	0.340

Table 8: **Aaa Spread: summary estimates of LSTR models** . (Values in parentheses are estimated t-ratios; *se* is the standard deviation of residuals. Full estimation results for these models are shown in Table 10, Appendix C.)

model: coefficient	**LSTR(1)**	**LSTR(2)**	Restricted
linear part			
int	-0.822 (-5.589)	1.195 (14.525)	1.241 (15.654)
π	0.144 (2.235)	0.036 (-1.131)	-0.018 (-1.042)
r	0.284 (7.500)	0.024 (2.005)	0.032 (2.694)
y	0.935 (16.026)	0.227 (10.853)	0.309 (15.261)
nonlinear part			
int	1.989 (11.845)	-1.976 (-11.871)	-0.750 (-8.680)
π	-0.101 (-1.529)	0.038 (1.030)	0.121 (6.998)
r	-0.262 (-6.574)	0.731 (12.352)	
y	-0.707 (-11.430)	0.265 (6.757)	
transition parameters			
γ	113.297 (1.032)	34.288 (0.589)	425.64 (0.001)
c_1	2.232 (52.506)	2.228 (40.370)	1.404 (0.359)
c_2		14.423 (332.546)	9.350 (0.139)
goodness-of-fit measures			
se	0.505	0.504	0.541
AIC	-1.350	-1.351	-1.217
R^2	0.493	0.495	0.418

Table 9: **Baa Spread: summary estimates of LSTR model.** (Full estimation results for these models are shown in Table 11, Appendix C.)

8 Conclusions

This paper tested affine RFMs of credit risk against their non-parametric counterparts to better explain the determinants of risk premia. The analysis, based on corporate spread indices, showed that, despite their common use, there is no empirical evidence to support the restrictions imposed by affine models. On the contrary, evidence suggests that parametric choices in affine RFMs are too restrictive.

This paper demonstrated the non-linearity of functions that describe how the models' factors contribute to the determination of spreads. This was confirmed using a formal test for the linearity of a regression function. The goodness-of-fit of the non-linear models over the estimation period, and the adequacy of the non-parametric approach, was further evaluated by comparing observed spreads and predicted values from non-parametric models. The non-parametric models have proved to be able to fit the data well.

The paper showed that the inflation rate plays an important role in explaining corporate spreads. Increasing inflation is associated to widening spreads, and this relationship is stronger when inflation is low (or high), which is consistent with the view that higher inflation increases the number of bankruptcies, leading to an increase in the spread. This non-linear effect of inflation is also confirmed by the estimation of a non-linear but parametric model of the spread.

This analysis showed that the use of non-parametric techniques together with non-linear but parametric modelling offers interesting insights, revealing relations of interests among economic and financial time series. It also showed the relevance of non-parametric techniques, as non-linear but parametric models are often unable to capture in full non-linear dynamics.

This paper provided evidence in favour of non-parametric methods to improve spread forecasts. Conditional forecasts based on non-parametric models out-perform forecasts based on linear models. Linear time series models are widely used in forecasting. In general, this is motivated by the assumption that data are normally distributed. Yet, as shown in this paper, many financial variables do not have a normal distributions. Non-parametric modelling avoids the problem of structural instability in the parameters, which is well known to cause the break-down in the forecasting performance of linear predictive models, because it naturally accomodates changes in parameters. Regime changes, varying parameters, and more complex non-linearities are captured without the need of pre-specifying models' functional forms. Nonetheless, this work pointed out important limitations of conditional non-parametric forecasting, for which recursive formulas are unavailable. This make non-parametric forecasting difficult and calls for further research in this area. Future work will look at the use the time series process of the variables of interest to form multi-step ahead forecasts.

Computations. The non-parametric analysis was performed in this paper have been carried out using the open-source statistical software `R` (R Development Core Team, 2007). The estimation of the LSTRs was carried out using the freely available software `JMulTi` (see Lütkepohl and Krätzig, 2004).

Acknowledgements. Thanks are due to Peter Spencer, my PhD supervisor, to my advisors Karim Abadir and Gabriel Talmain, and to Peter Moffatt and Marianne Sensier for advise and useful discussion. Thanks are also due to seminars participants at the University of York, Manchester, East Anglia, and at the Rimini Center for Economic Analysis. This paper has been presented at the Symposium on Non-linear Dynamics and Econometrics (San Francisco, April 2008) and at the European Meeting of the Econometric Society (Milan, August 2008).

References

Y. Aït-Sahalia. Non-parametric pricing of interest rate derivative securities. *Econometrica*, 64:527–560, 1996.

A. Ang and M. Piazzesi. A no-arbitrage vector autoregression of term structure dynamics with macroeconomics and latent variables. *Journal of Monetary Economics*, 50:745–787, 2003.

M. Arapis and J. Gao. Empirical comparisons in short-term interest rate models using nonparametric methods. *Journal of Financial Econometrics*, 4:310–45, 2006.

A. Azzalini, A. Bowman, and W. Härdle. On the use of nonparametric regression for model checking. *Biometrika*, 76:1–11, 1989.

D. J. Bolder. Affine term-structure models: theory and implementation. Bank of Canada WP 2001-15, 2001.

A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis*. Oxford Statistical Science Series. Clarendon Press, Oxford, 1997.

A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models. *Annals of Statistics*, 17:453–555, 1989.

J. M. Chambers and T.J. Hastie. *Statistical models in S*. Chapman and Hall, New York, 1992.

F. Couderc, O. Renault, and O. Scaillet. Business and financial indicators: What are the determinants od default probability changes. Cahier de Recherches, HAC, Université de Genève, 2008.

J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interets rates. *Econometrica*, 53:385–407, 1985.

Q. Dai and K. J. Singleton. Specification analysis of affine term structure models. *The Journal of Finance*, LV:1943–1978, 2000.

A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.

J. Driessen. Is default risk even priced in corporate bonds? *Review of Financial Studies*, 18:165–195, 2005.

G. R. Duffee. The relation between treasury yields and corporate bond yield spread. *The Journal of Finance*, 53:2225–2241, 1998.

G. R. Duffee. Estimating the price of default risk. *The Review of Financial Studies*, 12: 197–226, 1999.

D. Duffie and K. J. Singleton. An econometric model of the term structure of interest-rate swap yields. *The Journal of Finance*, LII:1287–1321, 1997.

D. Duffie and K. J. Singleton. Modeling term structures of defaultable bonds. *The Review of Financial Studies*, 12:687–720, 1999.

D. Duffie and K.J. Singleton. *Credit Risk: Pricing, Measurement and Management.* Princeton Series in Finance. Princeton, Princeton, New Jersey, 2003.

D. Duffie, L. Saita, and K. Wang. Multiperiod corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 2007), volume = 83, pages = 635-666,.

E. J. Elton, M. J. Gruber, D. Agrawal, and C. Mann. Explaining the rate spread on corporate bonds. *The Journal of Finance*, 56:247–277, 2001.

J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87:998–1004, 1992.

J. Fan and J. Jiang. Nonparametric inference with generalised likelihood ratio tests. *TEST*, 16:409–444, 2007.

J. Fan and J. Jiang. Nonparametric inference for additive models. *Journal of the American Statistical Association*, 100:890–907, 2005.

J. Fan and Q. Yao. *Nonlinear time series. Nonparametric and parametric methods.* Springer series in statistics. Springer, 2003.

J. Fan, C. Zhang, and J.Zhang. Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics*, 23:153–93, 2001.

S. Figlewski, H. Frydman, and W. Liang. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. 2008.

J. Gil-Bazo and G. Rubio. A nonparametric dimension test of the term structure. *Studies in Nonlinear Dynamics and Econometrics*, 8, 2004.

J. G. De Gooijer and R. J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22:443–473, 2006.

C. W. J. Granger and T. Teräsvirta. *Modelling nonlinear economic time series.* O.U.P., 1993.

W. Härdle. *Applied nonparametric regression.* Number 19 in Econometric Society Monographs. Cambridge University Press, 1989.

W. Härdle and E. Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21:1926–47, 1993.

W. Hardle and J. S. Marron. Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 13:1465–81, 1985.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models.* Chapman and Hall, 1990.

J. Huang and W. Kong. Explaining credit spread changes: Some new evidence from option-adjusted spreads of bond indices. *Journal of Derivatives*, 11:30–44, 2003.

J. Huang and W. Kong. Macroeconomics news announcement and corporate bond credit spread. 2005.

R. J. Hyndman, D. M. Bashtannyk, and G. K. Grunwald. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5:315–336, 1996.

J. Jiang, H. Zhou, X. Jiang, and J. Peng. Generalised likelihoo ratio tests for the structure of semiparametric additive models. *The Canadian Journal of Statistics*, 35:381–398, 2007.

J. Knight, F. Li, and M.Yuan. A semiparametric two-factor term structure model. 4:204 – 237, 2006. Journal of Financial Econometrics.

F.A. Longstaff, S. Mithal, and E. Neis. Corporate yield spreads: default risk or liquidity? UCLA Working Papers, 2004.

H. Lütkepohl and M. Krätzig, editors. *Applied Time Series Econometrics.* Themes in Modern Econometrics. Cambridge University Press, 2004.

R. Luukkonen, P. Saikkonen, and T. Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499, 1988.

J. Opsomer and D. Ruppert. Asymptotic properties of back-fitting estimators. *Journal of Multivariate Analysis*, 73:166–179, 2000.

J. Opsomer and D. Ruppert. Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25:186–211, 1997.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2007. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

C.J. Stone. Additive regression and other nonparametric models. *Annals of Statistics*, 10: 1040–1053, 1985.

T. Teräsvirta. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218, 1994.

D. van Dijk, T. Teräsvirta, and P.H. Franses. Smooth transition autoregressive models — a survey of recent developments. *Econometric Reviews*, 21:1–47, 2002.

S. B. Wadhwani. Inflation, bankruptcy, default premia and the stock market. *The Economic Journal*, 96:120–138, 1986.

Appendix

A The spread formula

This appendix illustrates the derivation of the linear spread equations of section 1 (equations 6 and 7).

Duffie and Singleton (1997) showed that the price of a defaultable zero-coupon bond is given by the expectation of a cumulative risk-adjusted interest rate:

$$P_t = E_t^Q[\exp\{-\int_t^T R_s ds\}], \tag{37}$$

where T denotes the maturity date and R the risk-adjusted rate; the risk-adjusted rate equals the risk-free rate r plus a *premium*, λ, which captures default risk:

$$R = \lambda + r; \tag{38}$$

In the financial literature, λ is often referred to as the *mean-loss rate*, because it depends on the probability of default and on the fraction of market value of the asset lost upon default. Here, for simplicity, I assume that the recovery rate is zero so that λ depends only on the probability of default. (This is why, in this article, λ is referred to as the default rate.)

Alternative approaches to the econometric modelling of spreads on risky securities follow from equation 37 above. These are based either on the direct parameterisation of R, or on the parameterisation of its components λ and r. Duffie and Singleton (1997) pursued the first approach, and modelled the risk-adjusted rate as the sum of unobservable factors, or state variables, described by square-root diffusion processes (Cox et al., 1985). In contrast, Duffee (1999) specified dynamic processes for both r and λ as follows:

$$\begin{aligned} r &= y_1 + y_2, \\ \lambda &= b_1 y_1 + b_2 y_2 + y_3; \end{aligned}$$

One can see that this model allows the default rate to depend on the risk-free rate via the common factors y_1 and y_2 and the "correlation" coefficients b_1 and b_2. (Factors y_is follow independent square-root stochastic process.)

From the models above, one can obtain an analytic solution for the price of the risky bond as a function of the underlying variables (Bolder, 2001; Duffie and Singleton, 2003). This function represents the link between the factors' dynamics and the term-structure of the risk-adjusted rate, the latter being represented as a deterministic static relation between yields and state variables. Once the bond price is known, the spread is easily calculated as the difference between the yield on the risky bond and the yield on the risk-free bond.

Consider again the pricing equation 37. Substituting equation 38 in 37, one can see that the price of the zero-coupon zero-recovery risky bond can be written as the product of the risk-free bond price Δ and a price component that depends on the default rate D:

$$P_t^{(\tau)} = D^{(\tau)}(\lambda)\Delta^{(\tau)}(r); \tag{39}$$

(Here, $\tau \equiv T - t$ denotes time to maturity.) The two price components have the following analytical forms:

$$D^{(\tau)}(\lambda) = A_D(\tau)\exp\{-B_D(\tau)\lambda\}, \tag{40}$$
$$\Delta^{(\tau)}(r) = A_\Delta(\tau)\exp\{-B_\Delta(\tau)r\}; \tag{41}$$

Here, the coefficients A_Δ, B_Δ, A_D, B_D are deterministic functions of underlying diffusion parameters and time to maturity. (These functions are given in Bolder, 2001, , pg 42-46.). Substituting equations 40 and 41 back into the price equation 39 yields an analytic expression for the risky-bond price as a function of factors. Yields are then easily computed by taking logs. The resulting observable spread, obtained by subtracting the risk-free yield from the yield on the risky bond, is linear in the default rate λ:

$$s = -\frac{\log A_D(\tau)}{\tau} + \frac{B_D(\tau)}{\tau}\lambda; \tag{42}$$

(Note that this expression correspond to equation 6 in the main text.)

Let us consider a simplified version of Duffee's model above:

$$r = y_2, \tag{43}$$
$$\lambda = y_1 + b y_2, \tag{44}$$

Where ys are independent state variables following, say, square-root processes. (Note that λ depend on the risk-free rate r, so that the risk-adjusted rate has the general form $R = \lambda(r) + r$.) Using 43 and 44, the risk-adjusted rate can be written as $R = y_1 + (1+b)y_2$, and the pricing equation becomes:

$$\begin{aligned} P_t &= E_t^Q\left[\exp\left\{-\int_t^T y_1 + (1+b)y_2 ds\right\}\right] \\ &= E_t^Q\left[\exp\left\{-\int_t^T y_1 ds - \int_t^T (1+b)y_2 ds\right\}\right]; \end{aligned}$$

The price of the risky bond is, once again, the product of a default and risk-free component, and is given by:

$$P_t^{(\tau)} = A_D(\tau)\exp\{-B_D(\tau)y_1\}A_\Delta(\tau)\exp\{-B_\Delta(\tau)(1+b)y_2;\} \tag{45}$$

(As above, B and A denotes deterministic coefficients depending on underlying diffusions' parameters.) The price of a risk-free bond can be written as follows:

$$\Delta = A_\Delta(\tau)\exp\{-B_\Delta(\tau)y_2;\} \tag{46}$$

The spread is computed by subtracting the yield of the risk-free asset from the yield on the risky bond:

$$\begin{aligned} s &= -logD/\tau + log\Delta/\tau \\ &= -\frac{logA_D(\tau)}{\tau} + \frac{B_D(\tau)y_1}{\tau} - \frac{logA_\Delta(\tau)}{\tau} + \frac{B_\Delta(\tau)(1+b)y_2}{\tau} + \frac{logA_\Delta(\tau)}{\tau} - \frac{B_\Delta(\tau)y_2}{\tau}; \end{aligned}$$

Which gives:

$$s = -\frac{\log A(\tau)}{\tau} + \frac{bB_\Delta(\tau)}{\tau} y_2 + \frac{B_D}{\tau} y_1; \tag{47}$$

Using the fact that $y_1 = \lambda - by_2$, some algebraic manipulations lead to a spread equation linear in default rate and risk-free rate:

$$s = -\frac{\log A(\tau)}{\tau} + \frac{b(B_\Delta(\tau) - B_D(\tau))}{\tau} r + \frac{B_D}{\tau}\lambda; \tag{48}$$

By setting $bB_\Delta = \tilde{B}_\Delta$ and $bB_D = \tilde{B}_D$, one can see that equation 48 above is equivalent to equation 7 in Section 1. (Note that the spread equation derived in this section can be inverted to write the mean-loss rate as a linear function of observable variables.) One should also note that, in terms of implications on yield-spread, the models proposed by Duffee (1999) and Duffie and Singleton (1997, 1999) are equivalent (this is not proved here for reasons of space).

B Back-fitting of GAMs

The *back-fitting algorithm*, used for fitting additive models such as $y = \alpha + m_1(x_1) + \ldots\ldots + m_D(x_D) + \epsilon$ consists of the following steps:

- Step 1. Initialisation ($i = 0$):
  $\hat{\alpha} = \sum y/n$; $m_j^0 = 0, \quad j = 1, .., d$;

- Step 2. $i = i + 1$;
  for each $j = 1, .., D$ define partial residuals $\epsilon_j = y - \hat{\alpha} - \sum_{d \neq j} \hat{m}_d^{(i)}(x_d)$;
  compute $\hat{m_j}^{(i+1)} = S_j(\epsilon_j)$, where S denotes a univariate smoother (usually, this is a local-linear smoother or a spline smoother);

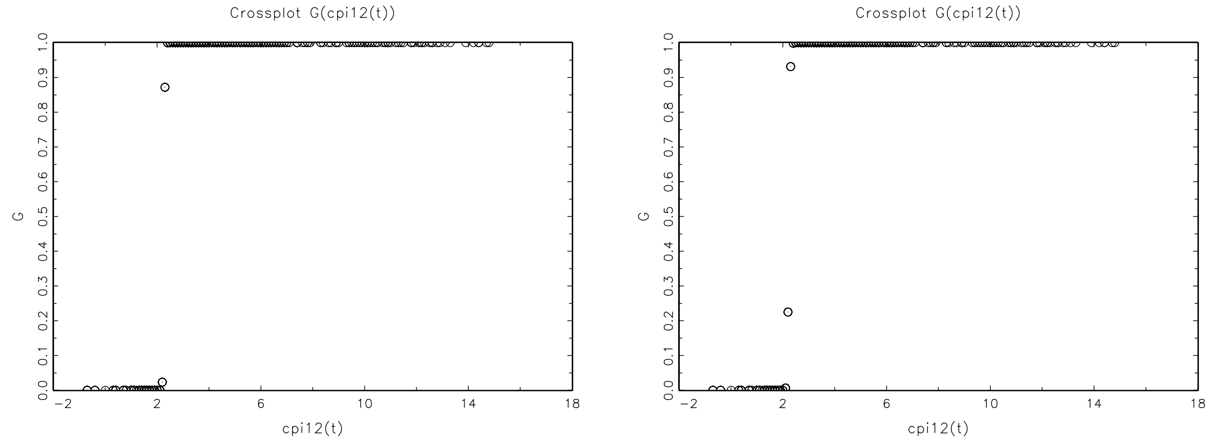
- Step 3. Keep cycling step 2 until convergence is reached.

The idea is to produce a fit, compute partial residuals, re-fit, until some convergence criterion is satisfied. One can see that the basic building block of the algorithm is the linear smoother S, which estimates the individual functions. (This article uses the local-linear smoother.) Intuitively, a justification to this approach is provided by the following fact: if the additive model is correct, then

$$E[Y - \alpha - \sum_{d \neq j} m_d(X_d)|X_j] = m_j(X_j); \tag{49}$$

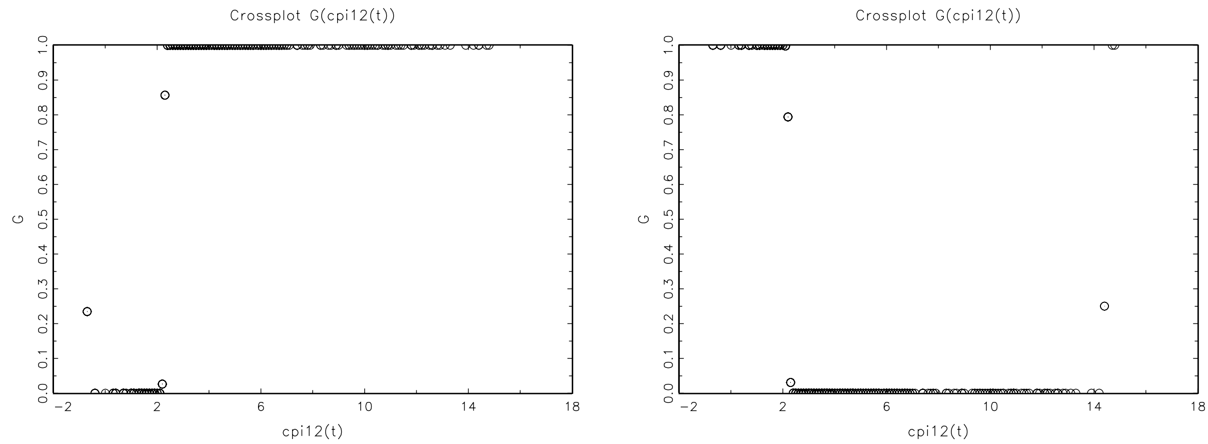
The back-fitting algorithm is detailed in Buja et al. (1989). One can also see Hastie and Tibshirani (1990), and Chambers and Hastie (1992) for its implementation in S.

C Smooth transition model of the spread



(a) (b)

Figure 12: LSTR(1) models for Aaa (a) and Baa (b) spread: transition function (G) against the transition variable (inflation).



(a) (b)

Figure 13: LSTR(2) models for Aaa (a) and Baa (b) spread: transition function (G) against the transition variable (inflation).

model coefficient	LSTR(1) estimates	t value	LSTR(2) estimates	t value
linear part				
int	-0.777 (0.110)	-7.031^{***}	-0.822 (0.115)	-7.127^{***}
y	0.592 (0.044)	13.467^{***}	0.586 (0.0442)	13.261^{***}
r	0.176 (0.0284)	6.203^{***}	0.175 (0.028)	6.161^{***}
π	0.132 (0.048)	2.745^{***}	0 .171 (0.056)	3.0587^{***}
nonlinear part				
int	1.888 (0.126)	14.947^{***}	1.932 (0.130)	14.813^{***}
y	-0.535 (0.047)	-11.479^{***}	-0.529 (0.047)	-11.296^{***}
r	-0.235 (0.030)	-7.853^{***}	-0.233 (0.030)	-7.815^{***}
π	-0.137 (0.048)	-2.804^{***}	-0.176 (0.057)	-3.110^{***}
transition parameters				
γ	166.683 (369.268)	0.451	158.791 (330.180)	0.481
c_1	2.266 (0.080)	28.625^{***}	-0.722 (0.049)	-14.851^{***}
c_2			2.267 (0.071)	31.904^{***}
goodness-of-fit measures				
se	0.381		0.381	
AIC	-1.911		-1.911	
R^2	0.412		0.414	

Table 10: **Determinants of Aaa Spread: LSTR models.** For coefficients' estimates, values in parentheses are coefficients' standard errors; significance code: *** corresponds to 1% significance.

model	STR(1)		STR(2)	
coefficient	estimates	t value	estimates	t value
linear part				
int	-0.822 (0.147)	-5.589^{***}	1.195 (0.082)	14.525^{***}
term-spread	0.935 (0.058)	16.026^{***}	0.227 (0.021)	10.853^{***}
inflation	0.144 (0.064)	$2.235{*}{*}$	0.036 (0.012)	-1.131^{***}
short rate	0.284 (0.038)	7.500^{***}	0.024 (0.012)	2.005^{**}
nonlinear part				
int	1.989 (0.168)	11.845^{***}	-1.976 (0.166)	-11.871^{***}
term-spread	-0.707 (0.062)	-11.430^{***}	0.265 (0.039)	6.757^{***}
inflation	-0.101 (0.065)	-1.529^{*}	0.038 (0.037)	1.030^{***}
short rate	-0.262 (0.039)	-6.574^{***}	0.731 (0.059)	12.352^{***}
transition parameters				
γ	113.297 (109.801)	1.032	34.288 (58.232)	0.589
c_1	2.232 (0.042)	52.506^{***}	2.228 (0.055)	40.370^{***}
c_2			14.423 (0.043)	332.546^{***}
goodness-of-fit measures				
se	0.505		0.504	
AIC	-1.350		-1.351	
R^2	0.493		0.495	

Table 11: **Determinants of Baa Spread: STR models.** For coefficients' estimates, values in parentheses are coefficients' standard errors; significance code: *** corresponds to 1% significance.