Institut für Wirtschaftsforschung Halle *Halle Institute for Economic Research*

Flight Patterns and Yields of European Government Bonds

Gregor von Schweinitz

July 2013 No. 10

IWH-Diskussionspapiere

IWH Discussion Papers

Author: *Gregor von Schweinitz*
Martin-Luther-University Halle-Wittenberg and
Halle Institute for Economic Research,
Department of Macroeconomics
E-mail: gregorvon.schweinitz@iwh-halle.de
Phone: +49 345 7753744

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-Econpapers and in ECONIS.

Editor:
HALLE INSTITUTE FOR ECONOMIC RESEARCH – IWH
The IWH is a member of the Leibniz Association.

Address: Kleine Maerkerstrasse 8, D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61, D-06017 Halle (Saale), Germany
Phone: +49 345 7753 60
Fax: +49 345 7753 820
Internet: http://www.iwh-halle.de

ISSN 1860-5303 (Print)
ISSN 2194-2188 (Online)

Flight Patterns and Yields of European Government Bonds

Abstract

The current European Debt Crisis has led to a reinforced effort to identify the sources of risk and their influence on yields of European Government Bonds. Until now, the potentially nonlinear influence and the theoretical need for interactions reflecting flight-to-quality and flight-to-liquidity has been widely disregarded. I estimate government bond yields of the Euro-12 countries without Luxembourg from May 2003 until December 2011. Using penalized spline regression, I find that the effect of most explanatory variables is highly nonlinear. These nonlinearities, together with flight patterns of flight-to-quality and flight-to-liquidity, can explain the co-movement of bond yields until September 2008 and the huge amount of differentiation during the financial and the European debt crisis without the unnecessary assumption of a structural break. The main effects are credit risk and flight-to-liquidity, while the evidence for the existence of flight-to-quality and liquidity risk (the latter measured by the bid-ask spread and total turnover of bonds) is comparably weak.

Keywords: sovereign bonds, sovereign risk premiums, sovereign debt crisis, semiparametric regression

JEL Classification: C14, G01, G12

Fluchtbewegungen und Effektivzinsen von europäischen Staatsanleihen

Zusammenfassung

Die aktuelle Staatsschuldenkrise in Europa hat zu verstärkten Bemühungen geführt, die verschiedenen Risikoquellen und deren Einfluss auf die Effektivzinsen von europäischen Staatsanleihen zu identifizieren. Bisher wurde dabei der potenziell nichtlineare Einfluss dieser Quellen weitgehend vernachlässigt. Auch Fluchtbewegungen in hochqualitative (*flight-to-quality*) und hochliquide (*flight-to-liquidity*) Anleihen, obwohl theoretisch fundiert, werden in den neueren Studien nur selten berücksichtigt. In der vorliegenden Arbeit werden die Effektivzinsen von Staatsanleihen der Euro-12-Länder ohne Luxemburg von Mai 2003 bis Dezember 2011 geschätzt. Mit Hilfe von *penalized splines* kann dargelegt werden, dass die meisten erklärenden Variablen einen hochgradig nichtlinearen Einfluss nehmen. Diese Nichtlinearitäten sowie *flight-to-quality* und *flight-to-liquidity* erklären sowohl gleichgerichtete Entwicklung der Zinsen bis September 2008 als auch die starke Differenzierung zwischen den einzelnen Staaten während der Finanz- und der europäischen Staatsschuldenkrise. Dabei ist es nicht notwendig, auf einen möglicherweise problematischen Strukturbruch zurückzugreifen. Als Haupteffekte werden Kreditrisiko und *flight-to-liquidity* identifiziert. Die Evidenz für einen starken Einfluss von *flight-to-quality* und Liquiditätsrisiko (letzteres gemessen durch den Bid-Ask-Spread und das Umschlagsvolumen der Anleihen auf dem Sekundärmarkt) ist hingegen begrenzt.

Schlagwörter: Staatsanleihen, Risikoprämien, Staatsschuldenkrise, Semiparametrische Regression

JEL-Klassifikation: C14, G01, G12

1. Introduction

The current European Debt Crisis has led to a reinforced effort to identify the determinants of the risk premium of European government bonds. Since the crash of Lehman Brothers, bond yields of some European countries demanded on secondary markets increased dramatically. Greek bonds where traded at yields of up to 37% by the end of 2011 (see Figure 1). Germany, on the other hand, actually managed to issue low-maturity bonds with a near 0% coupon in the first half of 2012. Until now, this development was mostly attributed to a „wake-up-call“, that is, a discretionary increase in the reaction of markets to credit and liquidity risk of government bonds (Aizenman et al., 2013; Beirne and Fratzscher, 2013). This paper argues instead that variables associated with credit and liquidity risk have a highly nonlinear influence on bond yields. Furthermore, the existence of flight-to-quality and flight-to-liquidity (Vayanos, 2004) also plays a significant role in explaining the yields.[1] These two flight patterns describe an increased demand for high-quality and high-liquidity bonds in times of increased market uncertainty.

The differentiation process of bond yields, started by increased global risk after the crash of Lehman Brothers, is visible in Figure 1. Exploding uncertainty on global financial markets in September 2008 preceded the well-known divergence of European government bond yields. This differentiation increased further, while uncertainty dropped after the first shock. However, global uncertainty cannot be the only explanation for the observed divergence. Uncertainty, measured by the US corporate bond spread, was at similar levels in 2010 and 2011 as shortly after the Dotcom-bubble. However, while yields diverged strongly during the European debt crisis, there was practically no differentiation in 2001 and 2002. Hugely increased imbalances in the Euro area (Knedlik and von Schweinitz, 2012) explain, why highly uncertain markets had a stronger incentive to differentiate between countries in 2008 than they had in 2001. That is, interactions between global risk and other risk factors can explain yield levels and their different development. Estimating monthly benchmark bond yields of the Euro-12 countries without Luxembourg from May 2003 to December 2011, I find that credit risk and flight-to-liquidity strongly influence yields, while the evidence for the existence of flight-to-quality and an effect of liquidity risk is comparably weak.

Yield spreads over a risk-free interest rate of an asset with identical maturity are normally attributed to credit risk, liquidity risk and a global risk component (von Hagen et al., 2011). In an international context, exchange rate risk has to be taken into account as well. For Euro-denominated bonds of countries within the European Monetary Union, the last aspect may be ignored, at least until the very recent past. This paper is – to my knowledge – the first allowing for an unknown nonlinear relationship between yields and variables associated with these different risk premium components. Most papers used linear regressions instead, motivated sometimes as variations of the standard Capital-Asset-Pricing-Model (CAPM) (Schuknecht et al., 2009, see for example). The development of research on European government bonds is shortly described in the following.

The strong convergence of bond yields after the introduction of the Euro (compared to disparate levels before 1999) can largely be explained by the emergence of a common global factor (Geyer et al., 2004). The remaining differences could only partly be explained by different levels of government debt sustainability (Codogno et al., 2003). Liquidity risk did not disappear completely despite the liquidity gains due to the introduction of a common market (Gómez-Puig, 2006, 2008). Furthermore the importance of a global risk component capturing overall investor uncertainty is recognized: Codogno et al. (2003) finds strongly differing linear coefficients for this variable in country specific regressions. As fundamentals differ across countries, such a difference may point to the need of interactions of the variable capturing global risk and other explanatory variables. Gómez-Puig (2008) finds significant effects for interactions terms. Interaction terms with global risk can be seen

[1] A similar argument is also put forward in the media, see for example „To strive, to seek, to find, and not to yield“, *The Economist*, June 30, 2012.

as „flight-to-quality“ and „flight-to-liquidity“, described in a theoretical CAPM-model by Vayanos (2004).[2] To detect these flight patterns, Beber et al. (2009) split the sample in periods of high and low global risk (a split that almost completely coincides with a breakpoint at the crash of Lehman Brothers). They find that credit risk plays a major role in the determination of yields, and that investors seek liquid bonds during turmoil.

The unfolding financial and European debt crisis offered the alternative argument of a „wake-up call“ on financial markets. It is said that the no-bail-out clause of the Stability and Growth Pact was incredible before the crisis. That is, markets were convinced that European countries would help each other in case of an imminent default. Later during the crisis – the reasoning continues – the no-bail-out clause became plausible. Yields increased despite a series of rescue packages.[3] By this argument, yield estimations should include a structural break in order to account for the increasing differentiation of government bond yields after the financial crisis. For example, von Hagen et al. (2011) extend their analysis of US-Dollar- and Euro-denominated bond spreads (Schuknecht et al., 2009, previously until 2005,) and include the European debt crisis. They find that variables associated with debt sustainability have much higher coefficients after the crash of Lehman Brothers. This result points to a strong need of sound fiscal positions. It is further confirmed by Bernoth et al. (2012). Already in an early paper after the outbreak of the financial crisis, Dötz and Fischer (2010) find an increased market reaction to debt sustainability and competitiveness.

In a related context, sample splits are also used by Beirne and Fratzscher (2013) to detect regional contagion during the European debt crisis. In the case of Favero and Missale (2012), contagion is defined as a fundamentally unjustified premium on yields due to spillovers from other crisis countries. The existence of these adverse effects are used to advocate the need for Eurobonds.

This short literature review made clear, that market reactions to credit and liquidity variables is complex. The stronger reaction to explanatory variables during times of high global risk (found with sample splits) also points to the need to model flight-to-quality and flight-to-liquidity. However, a simple sample split as used in most of the previous works is problematic in two respects. First, results can only be interpreted in view of the current crisis situation. That implies, that for example advice as to what measures could reduce yields to a given level is impeded by the unknown future economic regime. Second, a structural break has the implicit assumption, that the empirical distribution of explanatory variables is comparable in both subsamples, while the distribution of the explained variable is not. That is, in different times markets are expected to react differently to the same values of the explanatory variables. Only under this assumption is a structural break really valid. If also the distribution of explanatory variables changes, markets are expected to react differently to different value of the explanatory variables. Then, the model should not contain a structural break, but rather be nonlinear. As will be shown in Subsection 3.6, all explanatory variables experienced a strong shift towards a more adverse distribution during the crisis. To overcome the strong assumption behind the inclusion of the dummy, Bernoth and Erdogan (2012) estimate a time-varying coefficient approach, i.e., they explain increasing yields by continuous behavioral changes. Thus, a wake-up call may not have a sudden, but only gradually developing effects. However, such an estimation needs a much higher number of observations in general. Therefore, coefficients in this application are mostly insignificant. Furthermore, their estimation does not include interaction effects, possibly

[2] In his model, investors withdraw their money from investment funds if returns fall below a certain given threshold. Increased market uncertainty increases the probability of withdrawal. In order to avoid this event, fund managers will rebalance portfolios towards safer assets (flight-to-quality as interaction of global and credit risk). However, portfolio shifts can only partly offset increased withdrawal probability. Therefore, fund managers also seek more liquid assets (flight-to-liquidity as interaction of global and liquidity risk) in order to limit their losses in case of a withdrawal.

[3] The frequent references of the German government to the no-bail-out clause, discussions about a Greek default, austerity and political unrest in crisis countries, and the size of rescue packages (often assumed to be too small) might have played its part in this process. Lane (2012) accordingly describes *„Europe's efforts to address its sovereign debt problem as makeshift and chaotic“*.

disregarding important influence channels.

The semi-parametric method applied in this paper, penalized spline regression (Ruppert and Carroll, 1997), approximates any unknown, but possibly highly non-linear, functional form of the influence of different risk sources on yields. First, it avoids the problematic issue of a structural break while giving in general significant results. Second, it is also easily capable of including flight patterns as interactions of a variable associated with global risk with variables associated with credit and liquidity risk. Using this method, the previous results are confirmed: credit risk is found to be the main determinant of bond yields, while flight-to-liquidity appears in periods of high global uncertainty. Liquidity risk and flight-to-quality, on the other hand, only play a limited role. Liquidity risk is approximated by MTS data on bid-ask spreads and traded volumes for different bond issues.[4] As these measures are not publicly available, they have only seldom been used (Beber et al., 2009; Favero et al., 2010). Most other authors employ the nominal amount of outstanding debt instead, thereby assuming a high correlation of potential and actual market size.

The basic idea of penalized splines is given in Section 2 (more details can be found in the Appendix). A difference to most former works lies in that I explain spreads over the money market rate (Beirne and Fratzscher, 2013, following) and not spreads over German Bunds. This point as well as the employed explanatory variables are described in detail in Section 3. The results are presented in Section 4, with robustness checks in Section 5. Section 6 concludes.

2. Penalized Splines

The two main weaknesses of most of the existing literature are an insufficient regard for possible nonlinearities and the often missing inclusion of interactions reflecting flight-to-quality and flight-to-liquidity. In order to overcome these two shortcomings, the (structurally unknown) function that describes the influence of the three sources of risk on government bond spreads has to be estimated.[5] These sources are credit risk, liquidity risk and a global risk component, but also possible interactions between them. I use a semiparametric method, penalized spline regression (Ruppert and Carroll, 1997), in order to estimate the European government bond yields. The method builds on a semiparametric generalized linear model (Green and Silverman, 1994), designed to assign as much explanatory power as possible to a standard regression with linear, quadratic and cubic terms. Splines are used to correct possible local estimation bias. It has been applied in diverse scientific fields (see Ruppert et al. (2009) for a literature review), but only seldom to economic problems: For example, Jarrow et al. (2004) use it to determine the yield curve at different maturities, Eisenbeiß et al. (2007) estimate the risk aversion of investors on the stock market, Flaschel et al. (2007) jointly estimate different Phillips curves, while Berlemann et al. (2012) determine different factors of U.S. presidential approval ratings.

Penalized spline regression aims at estimating the function $y = f(x_1, \ldots, x_N) + \varepsilon$, where $y = (y_1, \ldots, y_T)'$ is the explained and $x_i = (x_{i,1}, \ldots, x_{i,T})'$ are the explanatory variables. It is assumed that f can be additively decomposed into univariate functions $f_i(x_i)$ and bivariate functions $f_{i,j}(x_i, x_j)$, capturing interaction effects. Univariate functions are either linear (constants, dummies or lags) or of a higher polynomial order. In this application, every higher-order univariate function is modelled by a third-order polynomial and five spline terms with optimized knot locations, resulting in eight regressors. Bivariate functions are modeled by cross-multiplication of the respective

[4] MTS is the main trading platform for the secondary market of Euro-denominated government bonds. For a description of the platform, see for example Cheung et al. (2005).

[5] Different sources of risk may themselves be a nonlinear function of different approximating variables. This may actually well be main channel for nonlinearities. For example, the ratio of government debt to GDP does not relate linearly to credit risk, a fact reflected by the frequent use of quadratic terms in the above cited works. I do not differentiate between these two possible sources of nonlinearity.

univariate terms, resulting in a total of 34 regressors (nine polynomial and 25 spline terms). Further details concerning this and other aspects of the estimation method can be found in the Appendix.

It can be observed that in every estimation without lags, errors are heavily autocorrelated. To account for this, one can either assume a corresponding error process or include the lag of the explained variable. I choose to adopt the latter (Favero and Missale, 2012), as autocorrelated error processes would necessitate an adaptation of the estimation process of penalized splines. The yield process itself is nonstationary due to the exploding yields during the crisis. Collinearity reduces the autocorrelation coefficient. In all estimations, it is significantly smaller than unity, and estimation errors are not autocorrelated. I estimate the spread of benchmark government bond yields (maturity ten years) over the three-months money market rate by

$$y_{t,c} = \rho y_{t-1,c} + f(x_{1,t,c}, \dots, x_{N,t,c}) + D_{t,c}\delta + \varepsilon_t. \tag{1}$$

The matrix D contains dummy variables, explained below, δ is the associated parameter vector. The index c denotes different countries, introduced for notational reasons. The univariate and bivariate functions are independent of c. Combining all polynomial terms (including D) in the matrix X and all spline terms in the matrix Z, the objective function to be optimized is

$$\min_{\beta,b,\Lambda} \|y - X\beta - Zb\|^2 + b'(\Lambda\Lambda')^{-1}b, \tag{2}$$

where the spline parameters b are random parameters, $b \sim \mathcal{N}(0, \sigma_\varepsilon^2\Lambda\Lambda')$, opposed to the unknown, but fixed parameters β. That is, penalty splines assume $\mathrm{E}(y - X\beta) = Zb = 0$ globally, while random spline parameters are only used to correct a locally biased estimation. Λ is a matrix with the penalty parameters on the diagonal (Bates, 2012), as explained in the Appendix.

The objective function (2) contains the squared sum of residuals in the first and the sum of penalty-weighted quadratic spline parameters in the second term. The reason for the inclusion of penalties is twofold: first, splines tend to produce an overfit in the estimation. The assumption of random parameters b offers a second interpretation of the penalty parameter Λ. Effectively, the penalty parameter is used to balance residuals and spline parameters. In the Appendix, the linear transformation $b = \Lambda u$ is motivated. This transformation reduces the second term of the objective function (2) to $\|u\|^2$ with $u \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. As both ε and u are assumed to be identically distributed, the penalty parameter Λ can be seen as the transformation parameter that guarantees the fulfillment of this assumption. Equation (2), the assumption for b and its transformation allows to use estimation techniques developed for linear mixed models (Bates, 2012).

I adapt the standard estimation in two ways. First, every function f_i and $f_{i,j}$ (or the associated spline parameters) has an individual penalty parameter. This is (although already described in Ruppert and Carroll (1997)) uncommon in the literature due to larger computational necessities, but allows for much more freedom in the overall contribution of splines. The same holds for the optimization of knot locations (Spiriti et al., 2013), which is all but absent in empirical applications (to my knowledge, the only papers with such features have a methodological focus). It can be observed that penalty parameters are not independent from knot locations, therefore predetermining their locations is problematic. Concerning the number of spline terms, Ruppert and Carroll (1997) propose five to forty, depending on the number of datapoints. I use only five spline terms, as bivariate functions then already have 25 spline terms. Given the optimized knot location and optimized penalty parameters Λ, the comparably low number of splines should suffice (Ruppert and Carroll, 1997).

The low number of spline terms in every function should help to avoid a problem of nonparametric estimations in general. While these methods are very useful to uncover complex and unknown relationships between different variables, they come at a cost. First, results are often fluctuating strongly, as the local fit of a function is improved by accounting for specific features of small groups of datapoints. Therefore, they are sometimes hard to explain on theoretic grounds that mostly assume monotonic development. Second, the large number of regressors used in a spline regression carries the risk of an overfit. Both problems should be smaller if the number of splines terms is not excessive. Still, the importance of the spline terms has to be evaluated by examining the optimal penalty parameters. For that reason, the result section will also contain regression results without spline terms. It provides evidence, that penalized splines provide similar results with much higher confidence.

3. Data and description of interactions

Bond yields can usually be decomposed in a risk-free interest rate and different risk premium components. This section is intended to describe the dependend variable and motivate the choice of explanatory variables used to approximate the true „risk“ components. All data start in May 2003 and end, if available, in December 2011. The availability of the liquidity measures determines the starting date. The final date is chosen in order to avoid estimation problems due to the Greek haircut in 2012. Only a near-full dataset is available for the Euro-12 countries without Luxembourg, mostly because of the limited availability of the current account as a ratio of GDP for Greece. Roughly 65% of my sample are before the crash of Lehman Brothers in September 2008, the remaining 35% are during the financial and the ensuing European debt crisis.

3.1. Explained Variable

The explained variable is the monthly spread of benchmark government bond yields with a maturity of ten years over the three-months money market rate, given by the Euribor. The choice of bond yields is standard in the literature. The money market rate is used less often as a risk free interest rate (Beirne and Fratzscher, 2013).[6] Most authors employ the yields of benchmark German Bunds or US Treasury bonds as a risk-free interest rate instead for bonds denominated in Euro or US-Dollar, respectively (Schuknecht et al., 2009, see for example). The selection of German Bunds, however, would render the explanation of low German bond yields impossible. This implies that the main asset towards which investors should flock – if flight patterns are observed – would be missing from the estimation. The use of US Treasury bond yields as a risk-free rate for Euro-denominated bonds is inadvisable as well, as one would have to account for exchange-rate risk, since bonds of European governments denominated in US-Dollar are increasingly rare (Bernoth et al., 2012). Figure 2 shows the time series for the yields of German benchmark bonds and the money market rate. The correlation between the two time series is quite high (78%). Furthermore, the money market rate is below the German Bund yields for most of the time, indicating that it is closer to the true risk-free rate than German Bund yields. The exception is a short period between mid 2007 and end of 2008. It could be argued that in that period, the money market rate was not entirely risk free as it was partly driven by emerging problems on the interbanking market. These problems increased rollover risk, reversing the yield curve.[7] Overall, the money market rate is a better measure of a risk-free interest rate than German Bund yields. The yields of benchmark bonds are interpolated by Thompson Reuters, the Euribor is provided by the European Banking Federation. For reasons of simplicity and in order to distinguish between bid-ask spreads and yield spreads, I will talk generally of yields in the rest of the paper, even though in a strict sense, yield spreads over the money market rate are meant.

[6]The money market rate is often used as a risk free rate in consumption Euler equations in many neoclassical models (Woodford, 2003). However, it should be noted, that it differs strongly from estimated discount factor in Euler equations (Canzoneri et al., 2007).

[7]An alternative argument, that Germany managed to get yields below a risk-free interest rate due to strongly increased market uncertainty (like they did for short-term bonds in the beginning of 2012) does not hold, as market uncertainty in the period under question was not yet extremely high.

3.2. Credit Risk

One component of bond yields is credit risk, i.e. the probability of a default multiplied by the expected loss in that case. Depending on the debtor, this may in many cases even be the largest risk premium component. Most studies use variables that describe debt sustainability. Debt sustainability should be comparable to the long-term credit risk, and may be described by government debt (%GDP) (*DebtGDP*) and the government budget balance (%GDP) (*DefGDP*) (Bernoth et al., 2012, as for example in), both drawn from Eurostat (ES). Beirne and Fratzscher (2013), among others, also include interest payments and GDP growth. However, interest payments are strongly lagging yield development, as government debt is only partially rolled over every period (Knedlik and von Schweinitz, 2012). That is, interest payments describe past rather than current debt sustainability. Because of the sometimes long maturities of government bonds, GDP growth should also be used as a medium-term forecasted variable, since only future growth can ease the weight of heavy debt. However, medium-term forecasts are subject to large uncertainty. Therefore, both variables are not included.

Debt sustainability may differ from the market perception of the probability (and severity) of a credit event. This is contained (in principle) in the prices of Credit Default Swaps (CDS). However, markets for CDS may also be subject to distorting liquidity risk similar to the government bond market (Fontana and Scheicher, 2010). Additionally, at least for daily frequencies, government bond spreads are found to lead CDS spreads in emerging markets (Ammer and Cai, 2011). Aizenman et al. (2013) therefore conclude that

> *taken together, both studies suggest that sovereign interest rates and CDS spreads have common underlying causes rather than one driving the other.*

In light of increasing imbalances in the Euro-zone, visible already long before the subprime crisis (Knedlik and von Schweinitz, 2012), the sustainability of debt might not be completely captured by fiscal variables alone. It could be argued, that high deficits are much more sustainable if they are „insured“ by a high current account surplus. Therefore, I also employ the current account balance (*CurrAcc*, ES) as a measure of competitiveness (Sgherri and Zoli, 2009; Beirne and Fratzscher, 2013).

The three variables *GovDef*, *GovDebt* and *CurrAcc* are interpolated linearly from quarterly data (each captured at the end of the quarter) (Dell'Ariccia et al., 2006; Hauner et al., 2010; Beirne and Fratzscher, 2013).[8]

3.3. Liquidity Risk

The discussion of the existence of a liquidity risk premium in European government bonds emerged after the introduction of the Euro, when bond yields did not converge fully (Gómez-Puig, 2008). Liquidity risk exists when, due to small market size, sellers and buyers both have to place offers at a discount in order to achieve an exchange. Strictly speaking, liquidity risk is only dependent on actual trading (and less on potential market size). The difference can be seen very clearly for Greek government bonds. While the total amount of debt increased significantly during the European debt crisis, trading of bonds on the secondary market nearly ceased to happen. Due to this difference, the bid-ask spread (that is, the difference between buy and sell offers) or actually traded volumes are closer to the concept of liquidity risk (Codogno et al., 2003; Beber et al., 2009; Favero et al., 2010) than, for example, the total amount of issued government debt (Gómez-Puig, 2006; Bernoth et al., 2012, used for example by).

[8] In light of the expected nonlinear influence on yields, interpolation is not desirable. However, data at higher frequencies would have to be built as a real-time variable, including also preliminary estimates, to account for publication issues. These issues are avoided by using interpolated quarterly data.

MTS provides for every single government bond traded on their platform the average daily bid-ask spread and total daily turnover. For every country and trading day, the individual bond measures of benchmark bonds are aggregated in order to obtain the mean daily bid-ask spread and the total daily turnover of the respective country. For Spain, bond selection was additionally restricted to maturities between seven and fifteen years, as short term debt experienced exploding bid-ask spreads after November 2009 (even surpassing the effective interest rate of its ten-year bonds).[9] Monthly data are simple averages of daily data. If not a single trade via MTS is recorded in a given month, the turnover is zero. If the bid-ask spread is unobserved, such a natural benchmark does not exist. Therefore, I interpolate missing bid-ask spreads linearly if less than three months of data are missing, and choose the last available bid-ask spread, if the gap is smaller than half a year. Such a procedure is acceptable, since bid-ask spreads do not fluctuate too widely in the periods under question. A total of 31 months was interpolated in this way, while the longest consecutive period is five months from April to August 2011 in Ireland. In Portugal, bid-ask spreads are not observed from April 2011 onwards. Due to the length of the missing period, these datapoints are not interpolated. Therefore an estimation of Portuguese yields including the effects of the bid-ask spread is not feasible at the end of sample. This calculation results in monthly averages of the mean daily bid-ask spread (*BidAsk*) and the total daily turnover (*Turnover*).

It should be noted, that Italian bonds are by far the most liquid bonds in the market, if measured by *Turnover*. Their turnover is in more than 95% of the periods between five and ten times larger than that of the second-most traded bonds. Therefore, I also present the results of an estimation without Italy in the robustness section. There are multiple reasons for the higher turnover of Italian bonds. First, Italy has the highest nominal amount of outstanding debt. Second, before the introduction of the Euro, MTS was the trading platform for Italian bonds. After the introduction, it quickly became the main trading platform for other bonds as well (Cheung et al., 2005). The share of trades via MTS might therefore still be lower for other countries than it is for Italy. However, as MTS is the main trading platform, the liquidity risk should be lowest on MTS, providing a benchmark for liquidity risk variables.

3.4. Global Risk

Global risk describes the general risk-aversion of investors. A higher risk-aversion implies that investors react stronger to sources of risk. Most often, the American corporate bond spread (the spread between the yields of AAA- and BBB-rated corporations) is used (Favero et al., 2010; Schuknecht et al., 2009; von Hagen et al., 2011; Bernoth and Erdogan, 2012). Alternatives are the VIX or the VSTOXX (the American/European equity market volatility index) (Beber et al., 2009), or the European corporate bond spread (Geyer et al., 2004; Dötz and Fischer, 2010).[10] Like most of the literature, I use the American corporate bond spread (*CorpSpr*, provided by the Bank of America and Merrill Lynch via Datastream).

All these measures are essentially a description how strongly investors differentiate between assets of different quality (and possibly liquidity). Global risk is therefore a measure of differentiation between asset classes (government bonds and other classes) and assets in the same class. Therefore, it should be included both individually and in interaction with other risk variables in the estimation.

[9] The high correlation between the maturity restricted and the unrestricted series before November 2009 (98%) completely breaks down afterward (-28%). I don't restrict maturities in the other countries. Although such a restriction would correspond more closely to the estimation of yields of bonds with a maturity of ten years, trades become scarcer, leading to more missing datapoints.

[10] Bernoth and Erdogan (2012) find the European corporate bond spreads to be highly correlated with the American AAA-BBB-spread and use the latter due to its slightly more „global“ character.

3.5. Further variables

Since mid 2010, several rescue packages for crisis countries were implemented. Assuming, that a rescue package affects only the yields of the receiving country, a dummy variable is set to one in that country from the moment of decision until the end of sample (as none of the packages was completely depleted at the end of sample). Two such dummy variables are introduced for Greece, starting July 2010, and for Ireland, starting November 2010.[11] As described above, I also use a lagged term of the explained variable.

Country fixed effects are not included in the estimation. One could argue that yield differentials are dependent on the individual history of debt defaults. However, events of the pre-Euro era should not have a huge influence on markets valuation of debt sustainability during the Euro era. That is, the „history" of the different countries should restart with their entry into the European monetary union. This implies, that different countries in the Euro area should not have persistently different yields. Information criteria provide indecisive evidence on the question, which of the two models should be preferred. While the Akaike criterion (adjusted for finite samples) suggests using country fixed effects, the Bayesian information criterion suggests the superiority of the model without country fixed effects. The robustness section includes results for an estimation including country fixed effects as a comparison.

3.6. Stability of variables and endogeneity

The distribution of exogenous variables is not the same in times of high and in times of low global risk. Table 1 shows statistics for two subsamples defined by periods where the corporate spread exceeds its median or stays below it. I report the mean as well as the 5% and 95% quantile in the two subsamples. Every variable associated with credit and liquidity risk shows a more adverse development during periods of high global uncertainty. The periods of high global risk coincide with periods where government and current account deficits are higher, government debt and bid-ask spreads increase and total turnover of government bonds decreases. This holds both for mean values as well as for the extremes of the respective distributions. The case can most clearly be seen for the bid-ask spread, where the highest spreads in times of low global risk are only marginally higher than the lowest spreads in times of high global risk. That is, the distributions in the two subsamples are nearly separated. Consequently, a Kolmogorov-Smirnoff-test strongly rejects the assumption of equal distribution for all five variables.

Most of the previous literature uses sample splits (as a time split before and after the financial crisis – nearly equivalently – along the lines of global risk or) in their estimation of government bond yields. They find, that the parameters of their linear models are significantly different in the two subsamples. Mostly, coefficients are larger (in absolute terms) during the crisis than before. Therefore, one could conclude that markets demanded higher premia for the same level of risk during the crisis than before. However, as most sources of risk themselves became worse during the crisis (i.e., debt sustainability and liquidity worsened), this interpretation is not fully correct. Rather, the marginal reaction of markets to an increase in a variable became stronger while the variable itself changed. That is, the different parameters can be seen as evidence for the need of nonlinear models (and not a sudden shift in market reactions). A second problem of sample splits (in crisis and non-crisis periods) arises when estimation results are used for policy advice: In the future, the economic regime of Europe is unknown, i.e., which of the two estimated parameter sets actually describes yields best. Therefore, even when values of all explanatory variables are known

[11] Two further rescue packages for Greece were decided in July and October 2011, well inside my estimation range. The same holds for the Portuguese rescue package in May 2011. However, the limited availability of the current account for Greece and the bid-ask spread for Portugal makes the estimation of the rescue package effects infeasible.

with certainty, future yields are driven by the uncertainty about the state of the economy.[12] That is, when governments seek council as to what measures would be best suited to lower yields to a desirable level, such an advice cannot be given. Nonlinear methods, in this case penalized splines, provide consistent estimations of yields without the need of breakpoints.

In addition to the limited amount of collinearity between global risk and other explanatory variables, there might be endogeneity in contemporaneous variables. For example, higher yields might induce a high deficit in the same period if debt has to be rolled over. Market liquidity measures are surely influenced by market yields in the same month. Similarly, uncertainty on the government bond market triggered by widening spreads may have strong repercussions on global risk. Because of such endogeneity, the six variables for the different risk components (corporate bond spread, government deficit and debt, total turnover, bid-ask spread and the current account) all enter with a lag of one month.

3.7. Interactions

One of the main objectives of this paper is to identify patterns of *flight-to-quality* and *flight-to-liquidity*. The effect of such flights should be higher yields for bonds of low quality and low liquidity, and lower yields for highly liquid high-quality bonds, if market uncertainty is high and investors differentiate more strongly between bonds. Therefore, both flight patterns are best described by interactions of the global risk variable with variables for credit and liquidity risk. A total of four interactions are included in the estimation. I interact the government deficit, total turnover and the bid-ask spread with the corporate spread. Furthermore, the current account is interacted with the government deficit, as current account surpluses might work as an insurance against the negative effects of high government deficits. I do not include the interaction of government debt with the corporate spread, as that interaction decreases the explanatory power of the model.[13] If flight-to-quality exists on the European government bond market, then bonds from countries with high deficits should be traded at excess premia when global risk is high. At the same time, countries with low deficits or even a surplus should be able to benefit from lower risk premia. If flight-to-liquidity exist in times of high risk, lower yields for bonds with a high turnover and low bid-ask spreads are expected.

4. Results

In this section, the influence of the different variables are mainly presented in the form of plots showing the direct effect of an individual explanatory variable (corporate spread, government debt and deficit, current account, total turnover and the bid-ask spread) or the joint direct effect of two interacted variables on yields. The linear lag parameter, the constant as well as rescue dummies are jointly given in Table 2 for all regressions presented in this and the following robustness section. For a first impression of the nonlinear influence on yields, estimation results without splines (and without a structural break) are reported in Figure 3. The results of the main estimation are split between different risk components. Figure 6 (for global risk), Figure 7 (credit risk and flight-to-quality) and Figure 8 (liquidity risk and flight-to-liquidity) show the results for the nonlinear univariate and bivariate components of the regression function f. They do not show marginal reactions (comparable to a table of parameters), which would be hard to interpret for interactions. Instead, the immediate direct effect of a variable on yields is displayed. For example, subfigure (1) of Figure 7 shows that a government surplus of 5% implies ceteris paribus 0.1% lower yields, while a deficit of around

[12]Strictly speaking, this argument applies only to exogenously defined breakpoints (Bernoth et al., 2012, for example in) and not to to the same extent to endogenous ones (Beber et al., 2009, used by).

[13]The model without the interaction effect is selected by all standard information criteria.

20% implies around 0.2% higher yields.[14] Univariate effects, for example in Figure 6, are centered at the sample mean and given with 95% confidence bands.[15] The latter are shown only for the observed datapoints in order to give an impression of the scarcity of observations in some areas (like, for example, the high deficits in Ireland in the first subplot of Figure 7). Confidence bands are influenced by two sources of uncertainty: estimation errors and spline parameters. If splines are comparably unimportant, their parameters decrease. The resulting confidence band in such a case only depends on estimation errors. Therefore, it has a width of near zero at the sample mean of the variable, as is visible exemplary for the bid-ask spread in subfigure (2) of Figure 8 or for all results of the estimation without splines in Figure 3.[16] Interaction plots are contour plots, for example in subplots (3) and (4) of Figure 8, where the strength of the effect is indicated by the colorbar to the side of the plot. As implied by the scarcity of observations, some large areas in the interaction plots are defined by very few datapoints. Areas without any surrounding datapoints are kept blank.

4.1. Estimation without splines

First, I present the results for an estimation with all the polynomials, but no spline terms. That is, I estimate a regression with an autocorrelated term, cubic functions for the six explanatory variables, the four interaction terms described above, and a constant. The ten functions are graphically given in Figure 3. Linear parameters are reported in column (1) of Table 2. This leads to a total of 58 estimated parameters, of which four are linear parameters (lag, constant and two rescue dummies). The pure polynomial regression is basically a higher order Taylor expansion of the unknown function f. Therefore, it could be well sufficient to avoid the problem of structural breakpoints.

The univariate functions are largely as expected: worse fundamentals lead to higher yields. Variables associated with credit risk, namely the government deficit and debt in subplots (1) and (2), have a strong effect on yields. However, the effect is insignificant for all fiscal positions that are justified under the Maastricht criteria (at most 3% government deficit and 60% government debt). The effects of the current account and variables linked to liquidity risk are even insignificant at all levels. Both results are consistent with previous research finding credit risk to be important and liquidity risk insignificant and only existing in periods of higher global uncertainty (Beber et al., 2009). Interaction effects are almost identical to the effects found in the main estimation. In short, they indicate a strong pattern of flight-to-liquidity, but only partly flight-to-quality. While the estimation results are in general comparable to those of the main regression, they are often insignificant. Only small regions of the corporate spread, the government deficit and government debt significantly increase yields. This is reflected by the estimated parameters: only 13.8% of the 58 parameters are significantly different from zero at the 5% level.[17] As the effects are not unreasonable, a pure polynomial estimation provides evidence that structural breakpoints are not needed in order to capture the nonlinear relationship between explanatory variables and yields. However, the large width of the confidence bands are probably attributable to small, local estimation biases that are avoided by the inclusion of splines, as presented in the next subsections.

[14] I do not show long-run effects (dividing every parameter by $1-\rho$), because of the interdependence of exogenous variables. For example, the government deficit affects government debt. Because of these dependencies and the high autocorrelation of yields, a vicious cycle might appear whereby large deficits today do not only influence future yields, but also affect future debt levels (and possibly trading volume and the corporate spread). This would lead to further rising yields. As the detailed description of such a disequilibrating process is out of the scope of the current paper, only immediate effects are analyzed.

[15] For numerical stability, all variables with splines are standardized with mean zero and variance one.

[16] Similar plots in a linear regression would show an estimated line bx as well as two confidence bands $(b+c)x$ and $(b-c)x$, where $2c$ is the width of the confidence band around the parameter b. Due to the standard-normalization of variables for estimation, confidence bands are centered around the sample mean.

[17] All measures of significance are obtained from a block bootstrap, performed 1'000 times. I choose a blocklength of twelve months in order to preserve the statistical properties of the residual data series. In estimations including splines, Λ and regression parameters are optimized in every bootstrap iteration, and knot locations are kept fixed. Therefore, confidence bands are a conservative estimate.

4.2. General results

Before turning to the individual results, I present the estimates of the yields and the corresponding residuals. The observed yields (blue) and estimated yields (red) nearly coincide in Figure 4, a fact that is mirrored in the extremely low standard deviation of 0.288% (see also Table 2) and the correspondingly small residuals in Figure 5.[18] Errors are not autocorrelated (with a p-value of 0.29 for the autocorrelation parameter), suggesting no stationarity issues. These results suggest that the estimation results are very satisfying.

The main problem of the polynomial regression presented above was the low significance of parameters and therefore insignificant effects. This problem is mostly reduced by using penalized splines. Around 77.6% of the polynomial regression parameters are significant at the 5% level. A similar level of significant parameters (73.1%) is reached for splines. The importance of splines for the estimation of the different functions is shown by the penalty parameters λ (together composing the diagonal penalty matrix Λ. As the standard deviation of the random spline parameters b is given by $\lambda\sigma_\varepsilon$, a larger λ implies that the uncertainty of splines outweighs otherwise larger estimation errors. To be exact, the likelihood puts equal weight on every single error term and every penalty-corrected spline parameter. Therefore, splines get relatively more important in the estimation, if the penalty term exceeds the ratio of spline terms over total observations. The last column of Table 3, presenting the relative importance of the splines for every term, shows the multiple of this ratio with λ. It therefore accounts for the size issue. The higher the relative importance, the more splines are needed to improve the estimation. Splines are unimportant for the government deficit, and all the interaction terms. The individual influence of the government deficit on yields is close to linear, while the larger number of polynomial regressors for the interaction functions may explain the reduced need for splines there. On the other hand, spline terms are strongly needed to improve the estimation and capture the existing nonlinearities in the individual effect of the corporate spread, the government debt and the total turnover of bonds. For the two remaining functions, splines are of moderate importance. Together with the strongly increased share of significant parameters, this shows that splines improve the estimation of government bond yields.

4.3. Global Risk

Figure 6 shows the individual effect of the corporate bond spread on government bond yields. This effect is small and nearly constant around zero. Moreover, confidence bands are quite wide. Therefore, a clear individual influence of (past) global risk on (current) bond yields cannot be found, as also reported previously. Bernoth et al. (2012) find the (contemporaneous) effect of global risk to be significant only for the subsample from August 2007 to May 2009. Similar results are reported by Bernoth and Erdogan (2012). The insignificance of global risk is reasonable: Markets could react to higher global risk by shifting portfolios towards (generally safer) government bonds. Alternatively, they could react by demanding higher risk premia for all possible assets in the market. Accordingly, the main effect global risk exerts should be in interaction with the other risk variables.

4.4. Credit Risk, Flight-to-Quality

Three variables and two interactions are used to determine the influence of credit risk and flight-to-quality on bond yields. Figure 7 shows the individual effects of government deficit (1), government debt (2), and the current account (3), as well as the two interactions of government deficit with the corporate spread (4) and with the current account (5).

The effects are generally as expected. The result is particularly striking for the individual effect of government deficit (1). It has an almost linear negative effect on yields, also reflecting the limited

[18] The residuals show a very small amount of heteroscedasticity: the outliers of the residuals are mostly concentrated in crisis countries at the end of sample. However, the size of errors and the amount of heteroscedasticity is small enough to be of little concern.

need for splines for this variable. The linear effect contradicts estimations with breakpoints: They find strongly increasing parameters for deficits in periods when deficits where higher in general. In this estimation, this would imply a strictly convex instead of a slightly concave function in subplot (1) of Figure 7. The immediate effect is not particularly strong: increasing the deficit by one percentage point in this month leads to 0.016 percentage points higher yields in the next. The long-run effect of permanently rising deficits on yields (without taking collinearity of variables into account) is about 0.27 percentage points, quite close to the one found by Bernoth et al. (2012).

The effect of government debt (2) is slowly increasing with a kink at very high debt levels. However, the debt levels in that area have only been experienced by Greece. Therefore, the results may not be valid in general. The existence of such a kink results in a strong need for nonlinear elements. In particular the splines are essential for a suitable description of that effect. That comes at the cost of generally wider confidence bands for this variable. The width of these bands makes the effect of government debt on yields insignificant at all levels of debt. The European Commission put a much stronger focus on government deficits than debt levels, even though the original Stability and Growth Pact saw both measures complementary. That may have led markets to ignore debt levels. Another explanation for the insignificance is the persistence of the debt levels. Because of the high autocorrelation of yields, explanatory variables should explain only slightly more than the changes. Strong movements of yields may be due to deficits rather than debt levels. Even though confidence bands make a strict inference impossible, the debt level at which the average reaction picks slightly up is around 100% of GDP. That is near the level of government debt that is said to have an increasingly negative effect on GDP growth (Reinhart and Rogoff, 2010; Reinhart et al., 2012). This finding is much stronger in the previous estimation without splines, see subplot (3) of Figure 3.

The individual effect of a current account deficit (3) is close to zero. Only a surplus leads to dropping yields of the same order of magnitude as for the government deficit. This result is partly consistent with the theory, as the current account describes the competitiveness of the real economy. A current account surplus implies lessened risk for the future sustainability of government debt, while high deficits are an indicator for future risks (Knedlik and von Schweinitz, 2012). The main effect of the current account comes from its interaction with the government deficit (5). Overall, yields increase with increasing deficits, both in the current account and the governments budget, as shown by the plateau in the lower left corner of the plot. Furthermore, a positive current account cannot serve as insurance anymore, if government deficits exceed a threshold of around -10%. For such high deficits, yields increase nearly unilaterally.

Subplot (4) in Figure 7 shows the interaction of government deficits and the corporate bond spread. It should display flight-to-quality in the market, if existing. In theory, differentiation between countries with different budget balances increases with higher global uncertainty. Due to the stronger differentiation, countries with higher deficits should be punished, while bonds of countries with surplusses are traded at lower yields than in tranquil times. Subplot (4) of Figure 7 suggests, that the expected flight process only exists for medium levels of the corporate spread between 1.5 and 2. That level corresponds to the time directly before the bankruptcy of Lehman Brothers between February and August 2008, as well as most periods during the European Debt Crisis. These are periods when there is either some (not-yet acute) fear for the real economy, or when public finances are known to be at the heart of the greatest economic problems. There, the subplot shows a slightly stronger influence on yields for high deficits compared to the effect of low deficits. Higher global uncertainty arose during the financial crisis, when a global recession loomed. In those times, governments are expected to run large deficits to stabilize the otherwise fragile economy. This demand for expansionary fiscal policy during a recession can be seen by the yield-decreasing effect of higher deficits in periods of high global risk. That is, for a corporate bond spread above 2, not flight-to-quality, but a reward for rescue packages is the dominating effect in the interaction of the corporate spread and government deficit.

Taken together, I find the expected individual effect of government deficit and current account on yields. The current account serves as an insurance against government deficits if they are not too high. Government debt seems to be only of limited relevance for government bond yields. The interaction of government deficit with the corporate bond spread points to a demand for expansionary fiscal policy in times of high risk rather than flight-to-quality. The latter is only slightly visible for average levels of global risk.

4.5. Liquidity and Flight-to-Liquidity

Liquidity risk is reflected by two terms in the estimation: the individual effect of the total turnover, shown in subplot (1) of Figure 8 and the individual effect of the bid-ask spread (2). The interaction of these two variables with the corporate spread, given in subplots (3) and (4), is expected to show patterns of flight-to-liquidity. From theory, the existence of liquidity risk would imply that lower turnovers and higher bid-ask spreads lead to higher yields. Flight-to-liquidity would imply that this effect becomes stronger in times of higher global risk.

Liquidity risk is not found as predicted. The effect of the bid-ask spread, in subplot (2), is opposite to what would be expected. The use of a lag of one month (as in all other variables) to avoid otherwise existing endogeneity issues might offer an explanation: high bid-ask spreads in the previous month might have an overshooting effect on autocorrelated yields, that needs to be at least partly corrected in the following month. The individual effect of total turnover, in subplot (1), is mostly insignificant. There is, however, a discernible downward trend in the effect. Higher turnovers are a sign of increased market liquidity, lowering yields. Taken together, liquidity risk, measured by lagged variables, does not have a clear effect on yields. This fits to the literature that finds no strong influence of liquidity risk (Beber et al., 2009).

Flight-to-liquidity, on the other hand, exists. For most of the time, the interaction of the bid-ask spread and global risk, in subplot (4), has no effect on bond yields. For high bid-ask spreads and a high level of global risk, however, there is a strongly increasing effect on the yields. A similar tendency, although not as pronounced, is visible for the interaction of the corporate bond spread with total turnover, in subplot (3). For medium and high levels of global risk, small turnovers lead to higher yields, while large turnovers get a discount. At first contradicting are the increasing yields at the frontier of high turnovers and high yields. However, the whole border is obtained from the interpolation of two datapoints: the highest turnover of around 6E+09 and the period with the highest global risk. As both these values are much higher than the respective second highest values (see subplot (1) for the total turnover and Figure 6 for the corporate spread), this border should not be interpreted.

To conclude, I find that liquidity risk does not influence bond yields with a lag of one month. Therefore, I am unable to detect liquidity risk at that frequency. Flight-to-liquidity, on the other hand, is detected and plays a major role for yield development,. This result is consistent with the one reported by Beber et al. (2009) and especially Favero et al. (2010).

4.6. Linear variables and constants

Besides the variables employed in polynomial and spline functions, the model includes a lag ρ, dummies for rescue packages in Greece and Ireland and a constant. These variables contribute only linearly to yields. The results of this estimation are reported in Table 2 in column (2). The lag and the dummies for the two rescue packages are highly significant, while the constant is not.

The lag-term of 0.941 is close to, but significantly below one ($p = 0$). Its estimate is comparable to autocorrelation coefficients identified in other studies (Beber et al., 2009; Favero et al., 2010; Favero and Missale, 2012) and does not vary much between different specifications reported in Table 2. Such a high autocorrelation is somewhat sobering when one considers possibilities for reducing yields in crisis countries. This will only be possible over the medium run. Accordingly, even the decisive announcements of the ECB in August 2012 could only start a slow convergence process of yields.

The first rescue package for Greece in April 2010 had a strong increasing effect on yields. Apparently, the approval process and the size of rescue packages could not completely soothe markets. A fear that the package would be too small was realized when further rescues had to be approved in July and October 2011. Furthermore, the terms and conditions under which funds were granted prove to be hard to implement, as successive prolonged investigations and re-negotiations of the so-called „Troika" show.[19] The difficulties in the implementation of measures in Greece also point to moral hazard problems of rescue packages, since they diminish incentives for the strongest austerity measures. In Ireland, the effect of the rescue package is insignificant. This reflects both the smaller scale of the problems and the fact that the package was approved long after the bail-outs of Irish banks had pushed Ireland into a crisis. That is, Greece needed a rescue package even before implementing the much-needed measures in order to avoid a far worse downturn than later experienced. Ireland, on the other hand, had already implemented austerity measures and needed the funds to bridge the way out of the deep trough that had already been hit.

5. Robustness checks

5.1. Panel out-of-sample

The stability of estimates and results should be – to a certain extent – independent of the countries included in the estimation. A panel out-of-sample estimation can show if that is truly the case. One of the countries is left out of the estimation. The parameters calculated from the ten remaining countries are then applied to calculate estimated yields for the missing country (El-Shagi et al., 2013).[20]

Figure 9 displays the result of the panel out-of-sample estimation for the eleven countries. Visually, the estimate is about as good as in the baseline estimation in eight out of eleven countries and has only average outliers in two more countries (Greece and Ireland), while it is off-scale for Italy.

Table 4 contains the standard deviation of the out-of-sample errors for the eleven countries in the first column. The second column shows the share of that standard deviation over the baseline (in-sample) standard deviation. According to this estimate, the panel out-of-sample estimation surpasses the baseline estimation in Germany, Belgium, France, Finland, the Netherlands and Austria.[21] In Spain and Portugal, the estimation is only slightly worse than the baseline scenario, with shares of up to 1.22.

In Greece and Ireland, the share of panel out-of-sample standard deviation over baseline standard deviation is roughly 3.3:1 and 7.1:1. The reason is, that Greece was the only country to surpass a government debt of 125% of GDP in 2009, while Ireland experienced the highest deficits of all countries during the financial crisis. That is, for both countries a significant number of datapoints in one variable is outside the range observed in-sample. The exclusion of those extreme values leads to changing parameters (compared to the baseline estimation). Extreme values are particularly affected by parameters in the case of penalized splines, because they are reinforced by quadratic and cubic terms. Therefore, estimation errors increase in the affected periods. This fact also explains, why the out-of-sample performance of Italy is abysmal. Only 4% of the traded volumes of Italian bonds is inside the observed range of the ten other countries. Therefore, already slight parameter changes

[19]The Troika consists of members of the European Commission, the European Central Bank and the IMF.

[20]I use the optimal knot locations from the baseline estimation. The reason is, that changing knot locations imply a change in the spline variables. Therefore, parameter estimates would not be comparable anymore.

[21]The estimated standard deviations of the baseline scenario accounts for the large numbers of parameters. Out-of-sample standard deviations do not include them in their degrees of freedom, as previously estimated parameters are taken as exogenously given. This reduces the out-of-sample standard deviation.

for total turnovers will have an extreme effect on the Italian estimates, although they might not strongly affect the estimates of the ten other countries.

The estimation errors show that the assumption of stable parameters for previously unobserved values of exogenous variables is not reasonable for the applied method. That is, penalized splines should not be used to deduct consequences of an unfavorable development, if such a development has never been observed before. For example, one could not predict yields in a scenario with Japanese debt levels (around 200% of GDP). Instead, an as-if analysis (such as a panel out-of-sample estimation) is only valid inside previously observed ranges of variable values. While this caveat is in general applicable to every estimation, it should be taken into special account when higher polynomials are used, as that implies potentially stronger effects of extreme values.

5.2. Excluding Italy

The previous check made clear that Italy, due to the much higher turnover of its bonds, might have a strong influence on estimation results. Therefore, the estimation is done again with a smaller dataset excluding Italy. Opposite to the out-of-sample estimation, now also knot locations are reoptimized. For the different explanatory variables, mostly the observed values of the total turnover change. The highest turnover is reduced from around 6E+09 to 1.25E+09.

The marginal results presented in Figure 10 show that changes to the baseline estimation (Figures 6 to 8) mostly affect the corporate spread, subplot (1), the government deficit, subplot (2), and the interaction of the two variables, subplot (7). These differences are mostly driven by the exploding effect of the corporate spread. This effect is counterbalanced by an equally high negative effect in the interaction of the corporate spread and government deficits, making in turn adjustments in the effect of the government deficit necessary. All other effects are quite close to the ones in the baseline estimation. This also holds for most linear parameters, shown in column (3) of Table 2. Only the effect of the Irish rescue package is now strongly significant. This is explained by the excessively negative effect of the high Irish government deficits, that need to be counteracted to account for Irish peculiarities. The effect of turnovers can now be estimated with a much higher precision, such that the decreasing effect of higher turnover for non-Italian bonds is significant in subplot (4).

Especially the strongly increasing effect of the corporate spread is extremely unreasonable on that scale. Italy is special in comparison to other EMU countries as its government has achieved a primary surplus even during the crisis. From 2010 to 2012, its primary surplus was the highest in the group of the Euro-12 countries (even including Luxembourg), while in 2009 it was only outperformed by Germany and Luxembourg. That is, the public finances of Italy are only problematic due to interest payments. Removing Italy of the sample greatly increases the relation between global risk, government deficit, and yields, especially for crisis countries. The stronger relation leads to unstable parameter estimates. Therefore, one has to conclude that the model specifications are to be chosen with great care.

5.3. Estimation with country fixed effects

The baseline estimation did not include country fixed effects. The results of an estimation including these effects is presented in Figure 11. As in the previous robustness test, there is an extreme (negative) effect of high global risk, balanced by an equally positive effect of an interaction (in this case the interaction with the bid-ask spread) and a slight adjustment by the individual effect of the bid-ask spread. Also, as in the previous case, the individual effect of the government deficit is now insignificant. The reason for the instability is in this case that bid-ask spreads differ between countries and are very persistent from the beginning of the estimation period until the financial crisis. That is, the use of country fixed effects and the bid-ask spread is nearly equivalent in the first 65% of the sample. As in the previous case, such an equivalence reduces parameter stability and leads to the partly unreasonable results.

Despite this instability, the other effects are close to the ones in the baseline estimation. The individual effect of the government debt, subplot (3), shows a slow and insignificant increase as well as the characteristic kink. The individual effect of Italy is now captured by a country dummy, producing a clear yield-decreasing effect of higher turnovers in subplot (4). The curve for the current account, subplot (6), is a reinforced version of the one in the baseline estimation. The interaction effects show the need for bailouts in subplot (7), flight-to-liquidity in subplot (8) and the insurance effect of the current account balance in subplot (10).

Table 2 shows in column (4) the estimated values for linear parameters in the estimation with country fixed effects. The autocorrelation is slightly reduced. The coefficients of the rescue packages are nearly identical as the baseline parameters. All country fixed effects are insignificant at the 10% level. A joint F-test of the significance of the country dummies can reject significance at the 1% level, but it scarcely fails to do so at the 5% level. This ambiguity is also reflected by model selection criteria, where the Akaike criterion prefers the model including country fixed effects, while the Bayesian information criterion prefers the model without these effects.

In a common currency area, country differences should be not be existent by definition, but only as a consequence of different levels of exogenous economic variables. This, taken together with the unstable parameter estimates of the bid-ask spread and global risk, as well as individually insignificant country fixed effects, leads to the selection of the model without country fixed effects as the baseline model.

6. Conclusion

Previous estimations of European government bond yields found highly different parameters for credit, liquidity and global risk variables in normal times and during the current European debt crisis. As explanatory variables show more extreme behavior in the current crisis, this points first to a nonlinear reaction of government bond yields to these variables. Second, it emphasizes the need for an interaction of variables for credit and liquidity risk with global risk, the latter being much more elevated in the current crisis. In this paper, I use penalized spline regression to explain the yields of government bonds of the Euro-12 countries without Luxembourg. The method is fit to incorporate both unknown nonlinear behavior and complex interaction effects without the need to refer to an arbitrarily set structural breakpoint. This is a clear advantage over previous studies, as policy advice is normally not possible when breakpoints are included. This impossibility is due to the unknown future parameter regime. The results indicate that non-linearities are indeed strong for some variables, thereby justifying the use of such a complex method.

The estimation also allows to identify flight patterns described by Vayanos (2004). In this estimation, I find that the European debt crisis and the starting differentiation of yields was mostly driven by increasing credit risk and by flight-to-liquidity. Clear patterns of flight-to-quality could not be identified, mostly because markets do not punish governments for bank bailouts and other extremely costly measures when a financial crisis occurs and global risk is high. That is, in those times, governments are requested to stabilize the economy rather than put a strong focus on their own debt sustainability. Liquidity risk, opposed to flight-to-liquidity, could not be identified in the yields. This result is again consistent with previous results.

Robustness checks indicate that most of the results are stable over different model specifications. Differences mainly stem from reduced variation in exogenous variables, leading to unstable parameters estimates. Therefore, estimation models have to be set up carefully, reflecting underlying data structures, in order to provide reliable results. Moreover, the nonlinear effect of explanatory variables on yields implies that results should not be used in as-if analyses, if the scenario includes variable values well outside those observed in the estimation.

Autocorrelation of yields is found to be very high. A consequence of this is that it may take crisis countries a long time to reduce their current yield levels and return to capital markets. This is true even if high government and current account surplusses are achieved. The long process can be shortened if global risk is reduced. Flight-to-liquidity, mainly affecting scarcely traded bonds, is largely reduced in that case. Given the high correlation between global risk and political uncertainty (Pástor and Veronesi, 2011), this calls for decisive joint actions of the EMU member countries suitable to reduce political uncertainty.

Acknowledgments

The author is highly indebted to Claudia Buch, Makram El-Shagi, Oliver Holtemöller and Willi Semmler for their valuable comments, help and support. I also thank Jin Cao, Jesus Crespo-Cuaresma, Sebastian Giesen, Ludger Linnemann, James Ramsey, Peter Tillmann, and participants of the SEEK-workshop on „Nonlinear Economic Modelling: Theory and Applications" for their comments and suggestions.

Agarwal, G. and Studden, W. (1980). Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines, *The Annals of Statistics* **8**(6): 1307–1325.

Aizenman, J., Hutchison, M. and Jinjarak, Y. (2013). What is the Risk of European Sovereign Debt Defaults? Fiscal Space, CDS Spreads and Market Pricing of Risk, *Journal of International Money and Finance* **34**: 37–59.

Ammer, J. and Cai, F. (2011). Sovereign CDS and Bond Pricing Dynamics in Emerging Markets: Does the Cheapest-to-Deliver Option Matter?, *Journal of International Financial Markets, Institutions and Money* **21**(3): 369–387.

Bates, D. (2012). Computational Methods for Mixed Models, *Technical report*.

Beber, A., Brandt, M. W. and Kavajecz, K. A. (2009). Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market, *Review of Financial Studies* **22**(3): 925–957.

Beirne, J. and Fratzscher, M. (2013). The Pricing of Sovereign Risk and Contagion During the European Sovereign Debt Crisis, *Journal of International Money and Finance* **34**: 60–82.

Berlemann, M., Enkelmann, S. and Kuhlenkasper, T. (2012). Unraveling the Complexity of US Presidential Approval: A Multi-Dimensional Semi-Parametric Approach, *Research Paper*, HWWI.

Bernoth, K. and Erdogan, B. (2012). Sovereign Bond Yield Spreads: A Time-Varying Coefficient Approach, *Journal of International Money and Finance* **31**(3): 639–656.

Bernoth, K., von Hagen, J. and Schuknecht, L. (2012). Sovereign Risk Premiums in the European Government Bond Market, *Journal of International Money and Finance* **31**(5): 975–995.

Canzoneri, M., Cumby, R. and Diba, B. (2007). Euler Equations and Money Market Interest Rates: A Challenge for Monetary Policy Models, *Journal of Monetary Economics* **54**(7): 1863–1881.

Cheung, Y. C., Rindi, B. and De Jong, F. (2005). Trading European Sovereign Bonds: the Microstructure of the MTS Trading Platforms, *ECB Working Paper 432*, European Central Bank.

Codogno, L., Favero, C. and Missale, A. (2003). Yield Spreads on EMU Government Bonds, *Economic Policy* **18**(37): 503–532.

Davidson, R. and MacKinnon, J. (2004). *Econometric Theory and Methods*, Oxford University Press New York.

Dell'Ariccia, G., Schnabel, I. and Zettelmeyer, J. (2006). How Do Official Bailouts Affect the Risk of Investing in Emerging Markets?, *Journal of Money, Credit, and Banking* **38**(7): 1689–1714.

Dötz, N. and Fischer, C. (2010). What can EMU Countries' Sovereign Bond Spreads Tell Us About Market Perceptions of Default Probabilities During the Recent Financial Crisis?, *Discussion Paper 11*, Deutsche Bundesbank.

Eisenbeiß, M., Kauermann, G. and Semmler, W. (2007). Estimating Beta-Coefficients of German Stock Data: A Non-Parametric Approach, *The European Journal of Finance* **13**(6): 503–522.

El-Shagi, M. (2011). An Evolutionary Algorithm for the Estimation of Threshold Vector Error Correction Models, *International Economics and Economic Policy* **8**(4): 341–362.

El-Shagi, M., Knedlik, T. and von Schweinitz, G. (2013). Predicting Financial Crises: The (Statistical) Significance of the Signals Approach, *Journal of International Money and Finance* **35**: 76–103.

Favero, C. and Missale, A. (2012). Sovereign Spreads in the Eurozone: Which Prospects for a Eurobond?, *Economic Policy* **27**(70): 231–273.

Favero, C., Pagano, M. and von Thadden, E.-L. (2010). How Does Liquidity Affect Government Bond Yields?, *Journal of Financial and Quantitative Analysis* **45**(01): 107–134.

Flaschel, P., Kauermann, G. and Semmler, W. (2007). Testing Wage and Price Phillips Curves for the United States, *Metroeconomica* **58**(4): 550–581.

Fontana, A. and Scheicher, M. (2010). An Analysis of Euro Area Sovereign CDS and their Relation with Government Bonds, *Working Paper 1271*, European Central Bank.

Geyer, A., Kossmeier, S. and Pichler, S. (2004). Measuring Systematic Risk in EMU Government Yield Spreads, *Review of Finance* **8**(2): 171.

Gómez-Puig, M. (2006). Size Matters for Liquidity: Evidence from EMU Sovereign Yield Spreads, *Economics Letters* **90**(2): 156–162.

Gómez-Puig, M. (2008). Monetary Integration and the Cost of Borrowing, *Journal of International Money and Finance* **27**(3): 455–479.

Green, P. J. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Vol. 58 of *Monographs on Statistics and Applied Probability*, Chapman & Hall.

Hauner, D., Jonas, J. and Kumar, M. (2010). Sovereign Risk: Are the EU's New Member States Different?, *Oxford Bulletin of Economics and Statistics* **72**(4): 411–427.

Hinterding, R., Michalewicz, Z. and Peachey, T. C. (1996). Self-Adaptive Genetic Algorithm for Numeric Functions, *Parallel Problem Solving from Nature – PPSN IV*, Vol. 1141 of *Lecture Notes in Computer Science*, Springer, pp. 420–429.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.*, The University of Michigan Press.

Jarrow, R., Ruppert, D. and Yu, Y. (2004). Estimating the Interest Rate Term Structure of Corporate Debt with a Semiparametric Penalized Spline Model, *Journal of the American Statistical Association* **99**(465): 57–66.

Kauermann, G. and Opsomer, J. (2011). Data-Driven Selection of the Spline Dimension in Penalized Spline Regression, *Biometrika* **98**(1): 225–230.

Knedlik, T. and von Schweinitz, G. (2012). Macroeconomic Imbalances as Indicators for Debt Crises in Europe, *Journal of Common Market Studies* **50**(5): 726–745.

Krivobokova, T. and Kauermann, G. (2007). A Note on Penalized Spline Smoothing with Correlated Errors, *Journal of the American Statistical Association* **102**(480): 1328–1337.

Lane, P. R. (2012). The European Sovereign Debt Crisis, *The Journal of Economic Perspectives* **26**(3): 49–67.

Pástor, L. and Veronesi, P. (2011). Political uncertainty and risk premia, *NBER Working Paper 17464*, National Bureau of Economic Research.

Reinhart, C. M., Reinhart, V. R. and Rogoff, K. S. (2012). Public Debt Overhangs: Advanced-Economy Episodes Since 1800, *Journal of Economic Perspectives* **26**(3): 69–86.

Reinhart, C. M. and Rogoff, K. S. (2010). Growth in a Time of Debt, *American Economic Review: Papers & Proceedings* **100**(2): 573–578.

Ruppert, D. and Carroll, R. J. (1997). Penalized Regression Splines, *Technical report*, Cornell University.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*, Vol. 12, Cambridge Univ Pr.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2009). Semiparametric Regression During 2003–2007, *Electronic Journal of Statistics* **3**: 1193–1256.

Schuknecht, L., von Hagen, J. and Wolswijk, G. (2009). Government Risk Premiums in the Bond Market: EMU and Canada, *European Journal of Political Economy* **25**(3): 371–384.

Sgherri, S. and Zoli, E. (2009). Euro Area Sovereign Risk During the Crisis, *IMF Working Paper 09/222*, International Monetary Fund.

Spiriti, S., Eubank, R., Smith, P. W. and Young, D. (2013). Knot selection for least-squares and penalized splines, *Journal of Statistical Computation and Simulation* **83**(6): 1020–1036.

Vayanos, D. (2004). Flight to Quality, Flight to Liquidity, and the Pricing of Risk, *NBER Working Paper 10327*, National Bureau of Economic Research.

von Hagen, J., Schuknecht, L. and Wolswijk, G. (2011). Government Bond Risk Premiums in the EU Revisited: The Impact of the Financial Crisis, *European Journal of Political Economy* **27**(1): 36–43.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press (Princeton, New Jersey).

Appendix A: Tables and Figures

Table 1: Descriptive statistics for the credit and liquidity risk variables

	GovDef	*GovDebt*	*Turnover*	*BidAsk*	*CurrAcc*
Mean	-3.54%	71.84%	5.32E+08	0.10%	-0.90%
Mean (low risk)	-1.69%	66.84%	6.42E+08	0.04%	-0.34%
Mean (high risk)	-5.42%	76.90%	4.21E+08	0.15%	-1.47%
$q_{0.05;low}$	-6.80%	29.37%	3.83E+07	0.02%	-10.10%
$q_{0.95;low}$	2.86%	108.17%	3.59E+09	0.07%	7.33%
$q_{0.05;high}$	-15.27%	35.74%	1.68E+07	0.06%	-12.68%
$q_{0.95;high}$	0.84%	121.70%	1.94E+09	0.24%	7.14%
p(equal dist)	0.00	0.00	0.00	0.00	0.00

Note: The Table lists descriptive statistics for the used exogenous variables, split in two subsamples. These measures are the mean as well as the 5%- and 95%-quantile of the distribution (named $q_{0.05}$ and $q_{0.95}$). The subindices *high* and *low* denote the subsamples where global risk is above or below its median value (0.936%). The last row reports the probability, that the distributions in the two subsamples are equal, as given by a Kolmogorov-Smirnoff-Test.

Table 2: Lag term, rescue dummies and constants for the different estimations

	(1) No Splines		(2) Baseline		(3) Excl. Italy		(4) Country	
Parameter	Value	p	Value	p	Value	p	Value	p
ρ	0.944***	0	0.941***	0	0.939***	0	0.901***	0
Rescue Greece	0.303**	0.021	1.002***	0	1.052***	0	1.116***	0
Rescue Ireland	0.026	0.435	-0.008	0.613	0.265***	0	0.008	0.842
Constant	0.034	0.185	0.126	0.538	3.983	0.448		
Germany							0.114	0.831
Belgium							-0.165	0.143
Spain							0.464	0.775
Finland							0.335	0.803
France							0.09	0.832
Greece							-0.259	0.136
Ireland							0.4	0.78
Italy							-0.256	0.139
Netherlands							0.295	0.809
Austria							0.087	0.832
Portugal							0.325	0.805
F-Test							2.279**	0.038
σ_ε	0.278		0.288		0.285		0.284	
n	1128		1128		1024		1128	

Note: **,*** displays significance at the 5% and the 1%-level. Results are given in the following order: linear coefficients for the preliminary estimation without splines, Subsection 4.1 (1); the baseline estimation, Subsections 4.2 to 4.6 (2); the estimation excluding Italian data, Subsection 5.2 (3); and the estimation with country dummies, Subsection 5.3 (4). An F-Test for the joint significance of the country dummies is reported for the last estimation.

Table 3: Penalty parameters for the different functions

	λ	No. of splines	Relative importance
CorpSpr	30.22	5	6817.398
GovDef	1.74E-09	5	3.94E-07
GovDebt	34.10	5	7.69E+03
Turnover	8.57	5	1934.422
BidAsk	8.85E-04	5	0.200
CurrAcc	6.98E-04	5	0.157
CorpSpr; GovDef	1.88E-05	18	0.001
CorpSpr; VolTrade	1.43E-04	25	0.006
CorpSpr; BidAsk	9.37E-05	25	0.004
GovDef; CurrAcc	2.60E-05	21	0.001

Note: Relative importance is calculated as $\lambda \frac{q_i}{n}$, where q_i are the number of splines in that function, and $n = 1128$ are the total number of observations. A relative importance above unity shows that splines improve the estimation strongly.

Table 4: Standard deviation of errors for the panel out-of-sample estimation

	σ_ε	Share over baseline
Baseline	0.288	
Germany	0.209	0.704
Belgium	0.225	0.760
Spain	0.353	1.190
France	0.278	0.937
Finland	0.210	0.709
Greece	0.963	3.250
Ireland	2.107	7.108
Italy	26.256	88.589
Netherlands	0.219	0.739
Austria	0.237	0.799
Portugal	0.360	1.216

Note: In the panel out-of-sample estimation, one country is excluded from the estimation. The parameters obtained from the ten other countries are used to estimate the yields of the missing country. The first column reports the standard deviation of out-of-sample errors. The second column contains the share of these standard deviations over the standard deviation of the baseline estimation.

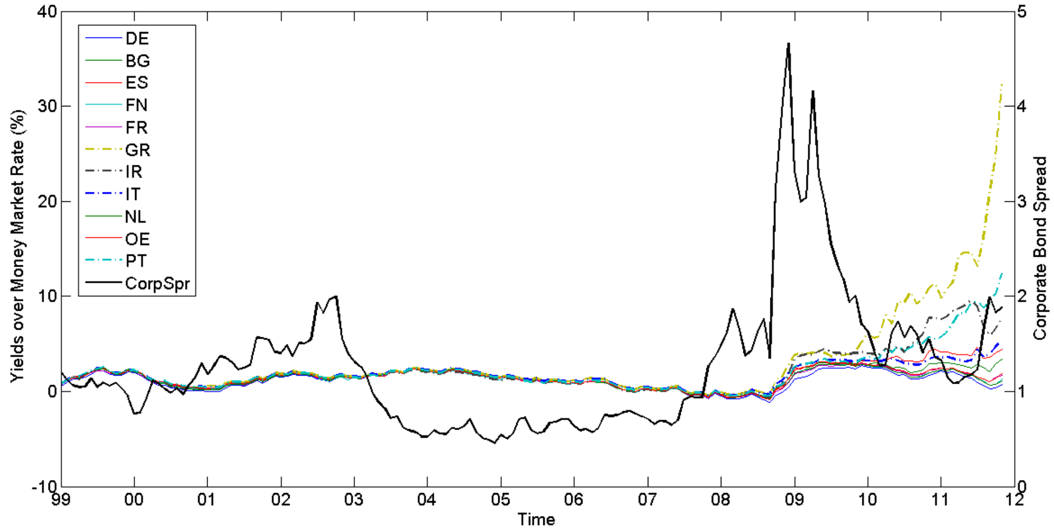

Figure 1: European government bond yields over money market rates and the corporate spread (right scale) between Januar 1999 and December 2012. Global uncertainty is measured by the corporate spread (the spread between corporate bond yields of AAA- and BBB-rated US-companies).

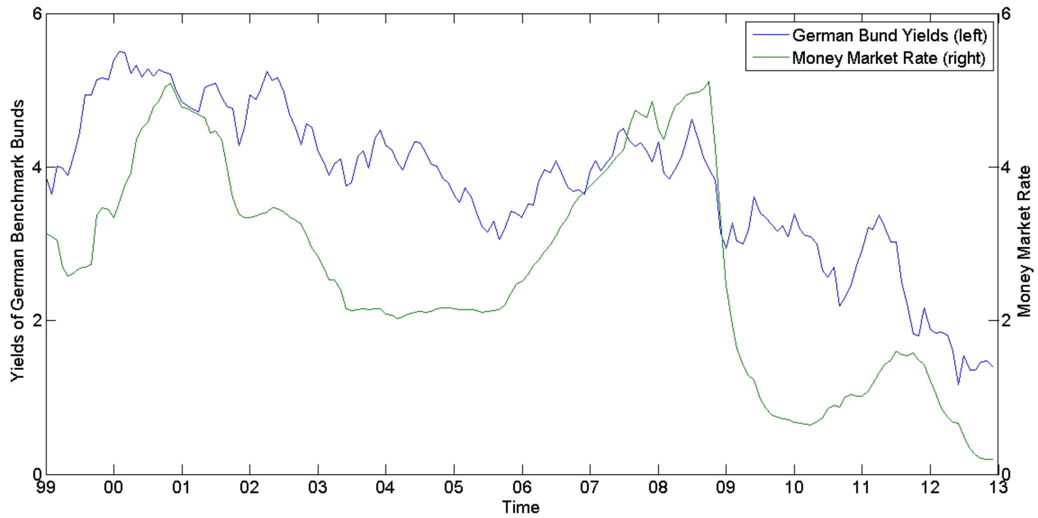

Figure 2: Yields of German benchmark Bunds (maturity ten years) (left axis) and three-months money market rate Euribor (right axis).

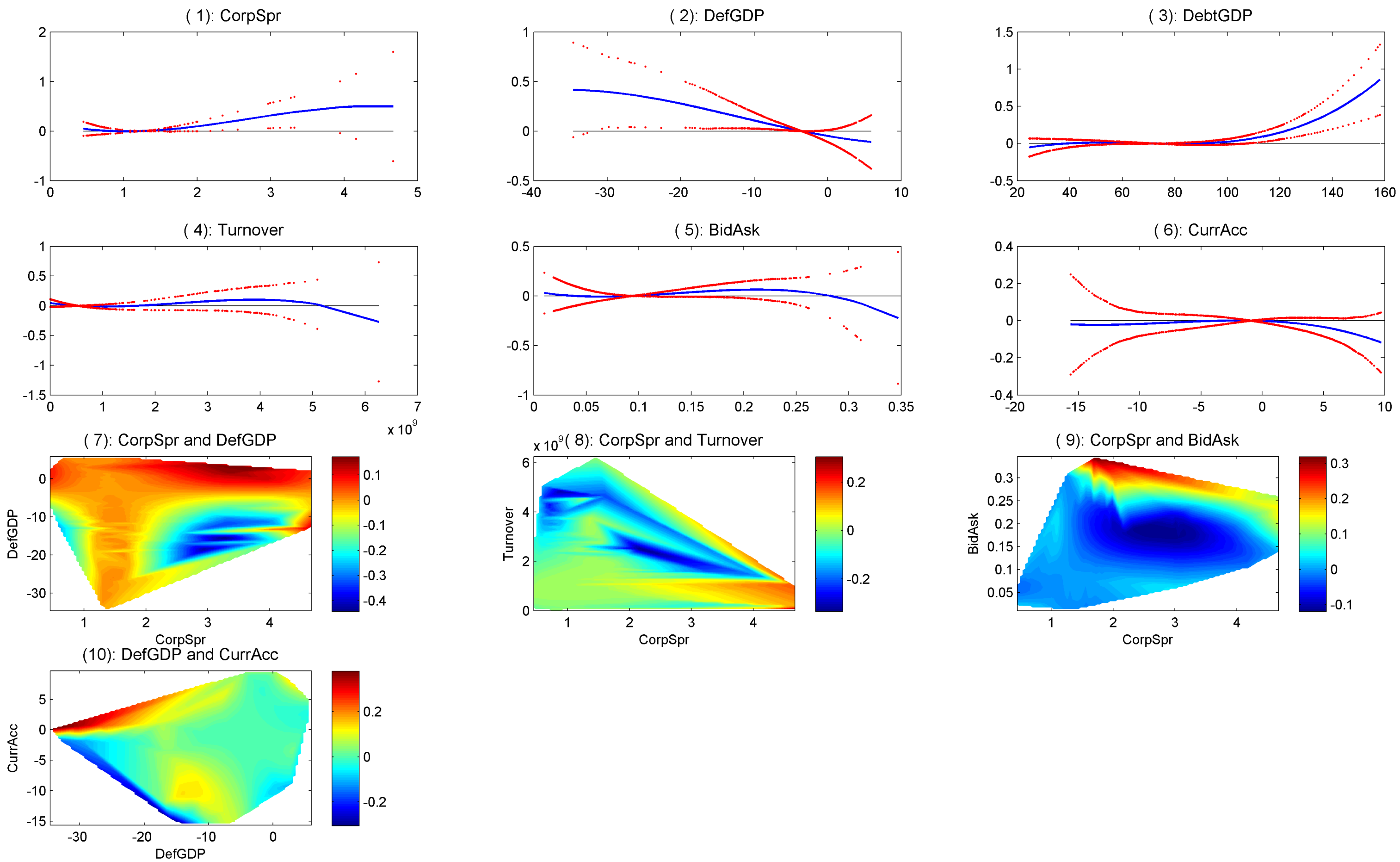

Figure 3: Individual and interaction effects for explanatory variables, estimation excluding splines.

Note: The y-axis or the height of the color bar show the effect on yields (in percentage points). Confidence bands for univariate functions are given in red, while the estimated function is blue.

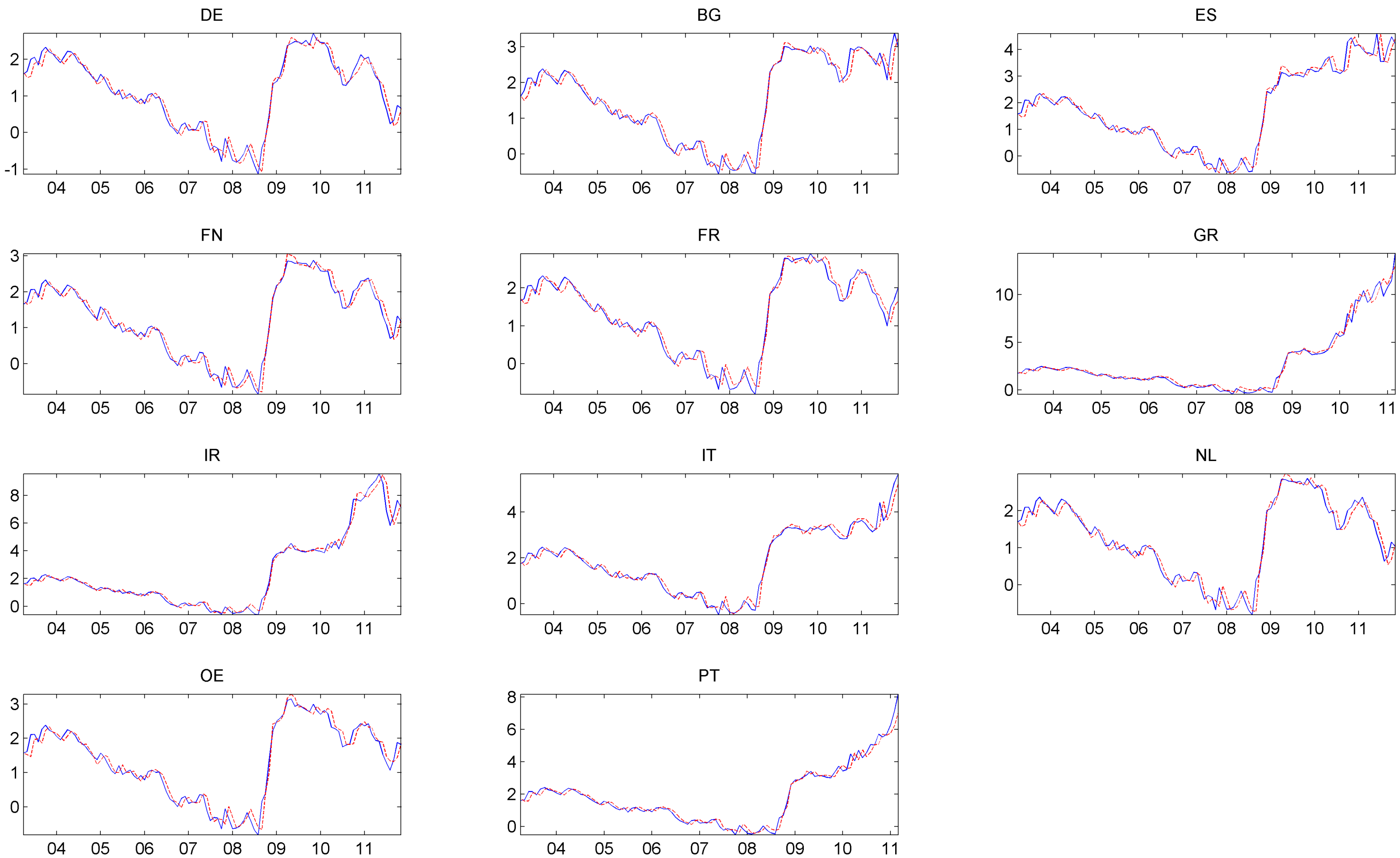

Figure 4: Observed and estimated yield spreads (blue/red) for the Euro-12 countries without Luxembourg, baseline estimation.

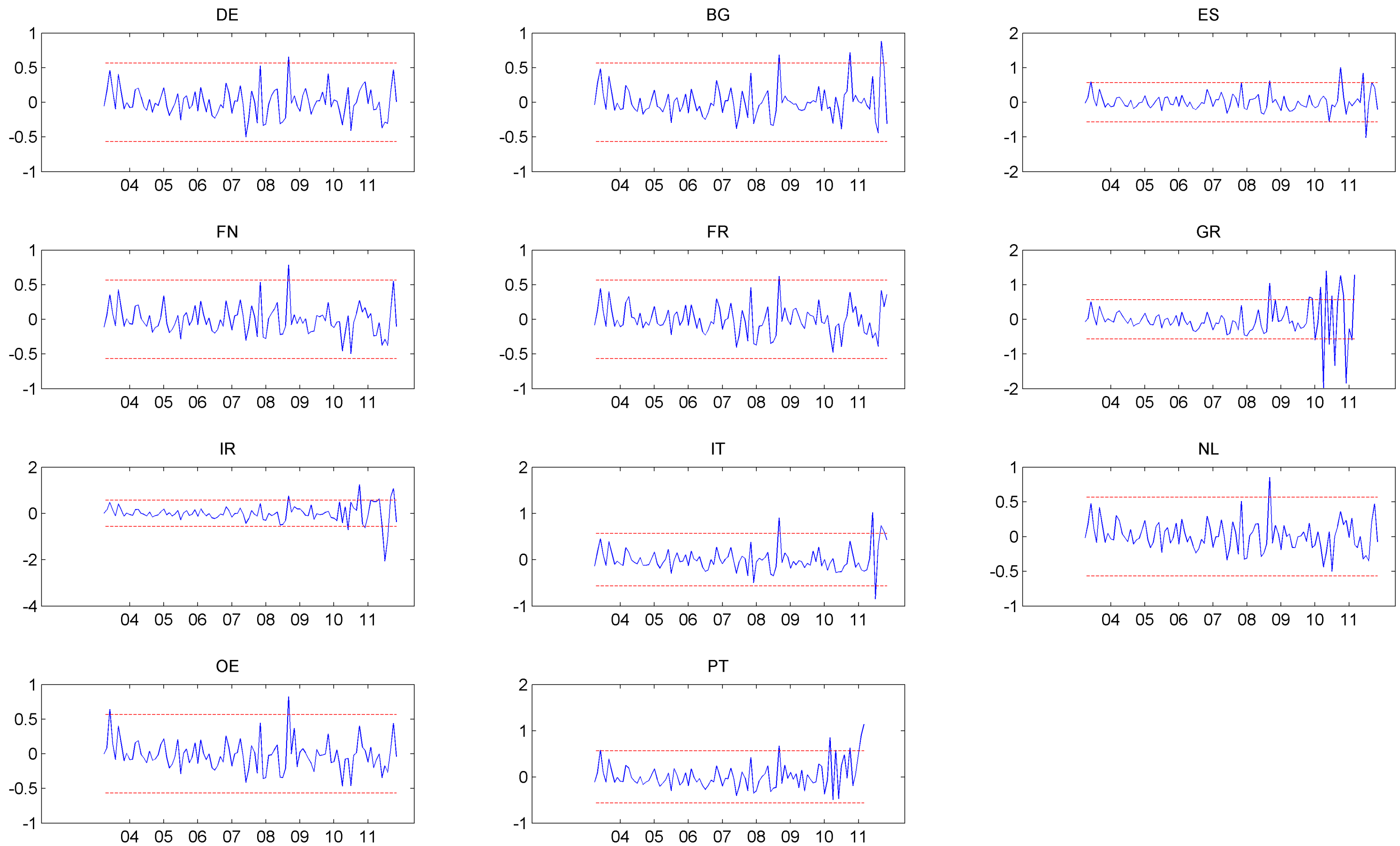

Figure 5: Residuals and 95% confidence bands (not adapted to sample size) of the baseline estimation.

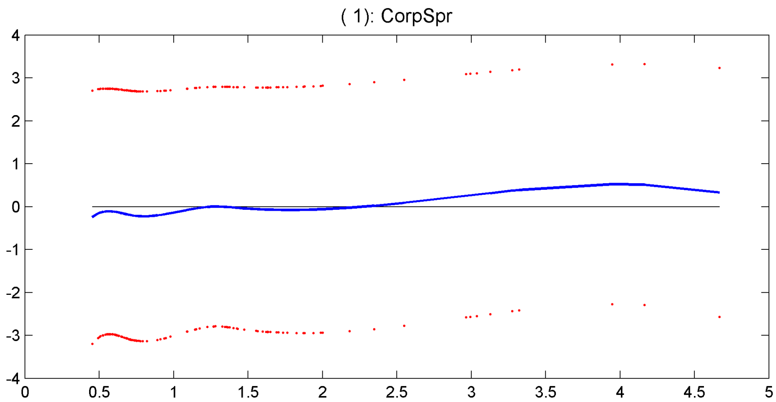

Figure 6: Individual effect of global risk, baseline estimation.

Note: The y-axis shows the effect on yields (in percentage points). Confidence bands for univariate functions are given in red, while the estimated function is blue.

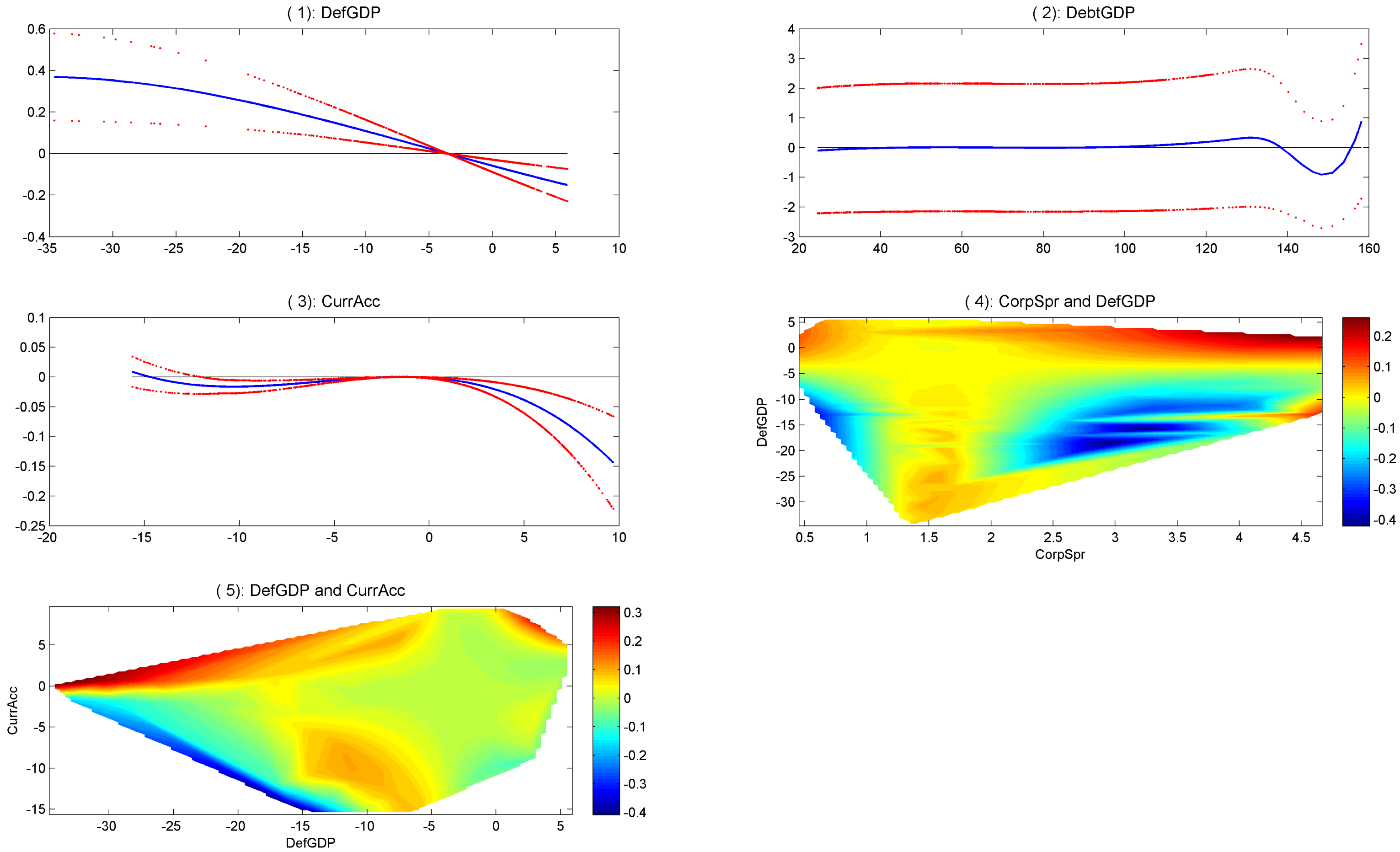

Figure 7: Individual and interaction effects for the credit risk variables *GovDef*, *GovDebt* and *CurrAcc*, baseline estimation.

Note: The y-axis or the height of the color bar show the effect on yields (in percentage points). Confidence bands for univariate functions are given in red, while the estimated function is blue.

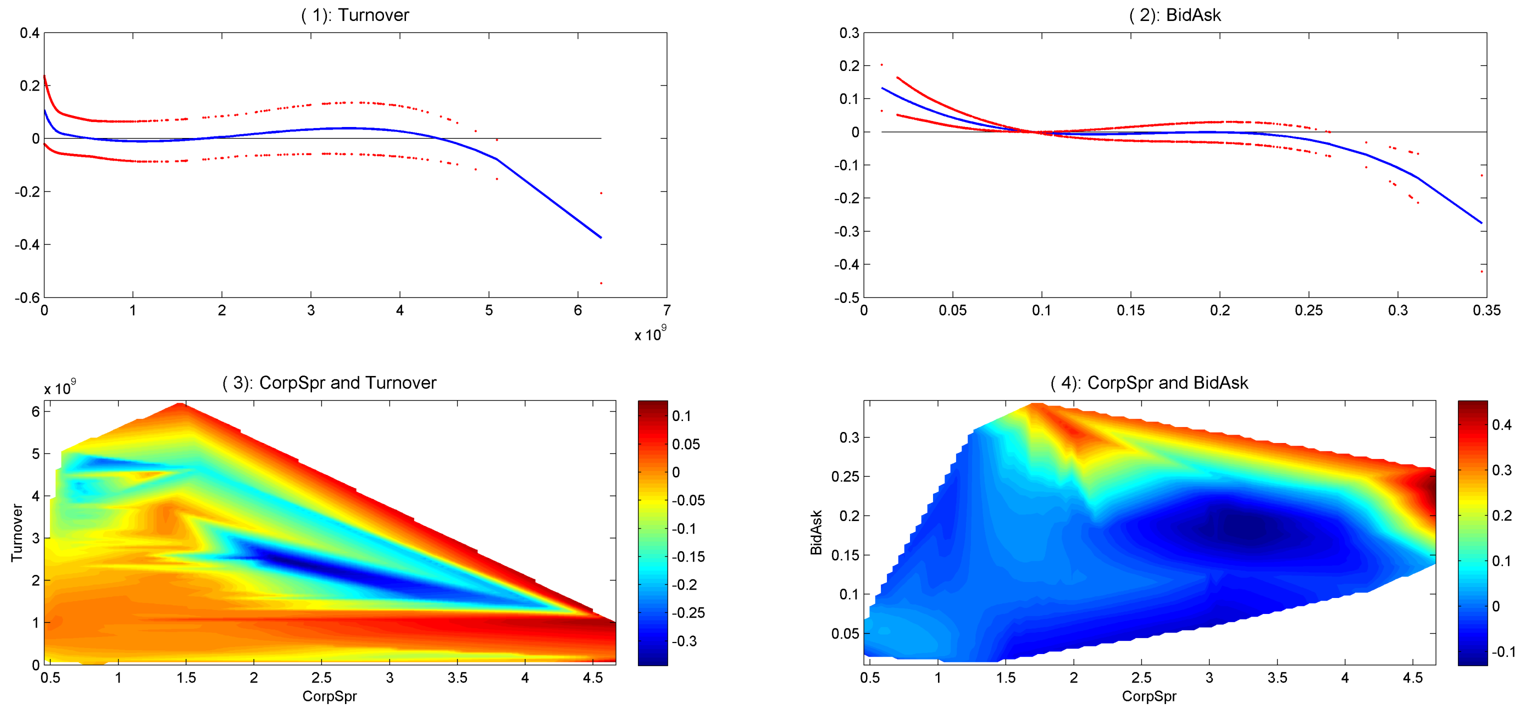

Figure 8: Individual and interaction effects for the liquidity risk variables *Turnover* and *BidAsk*, baseline estimation.

Note: The y-axis or the height of the color bar show the effect on yields (in percentage points). Confidence bands for univariate functions are given in red, while the estimated function is blue.

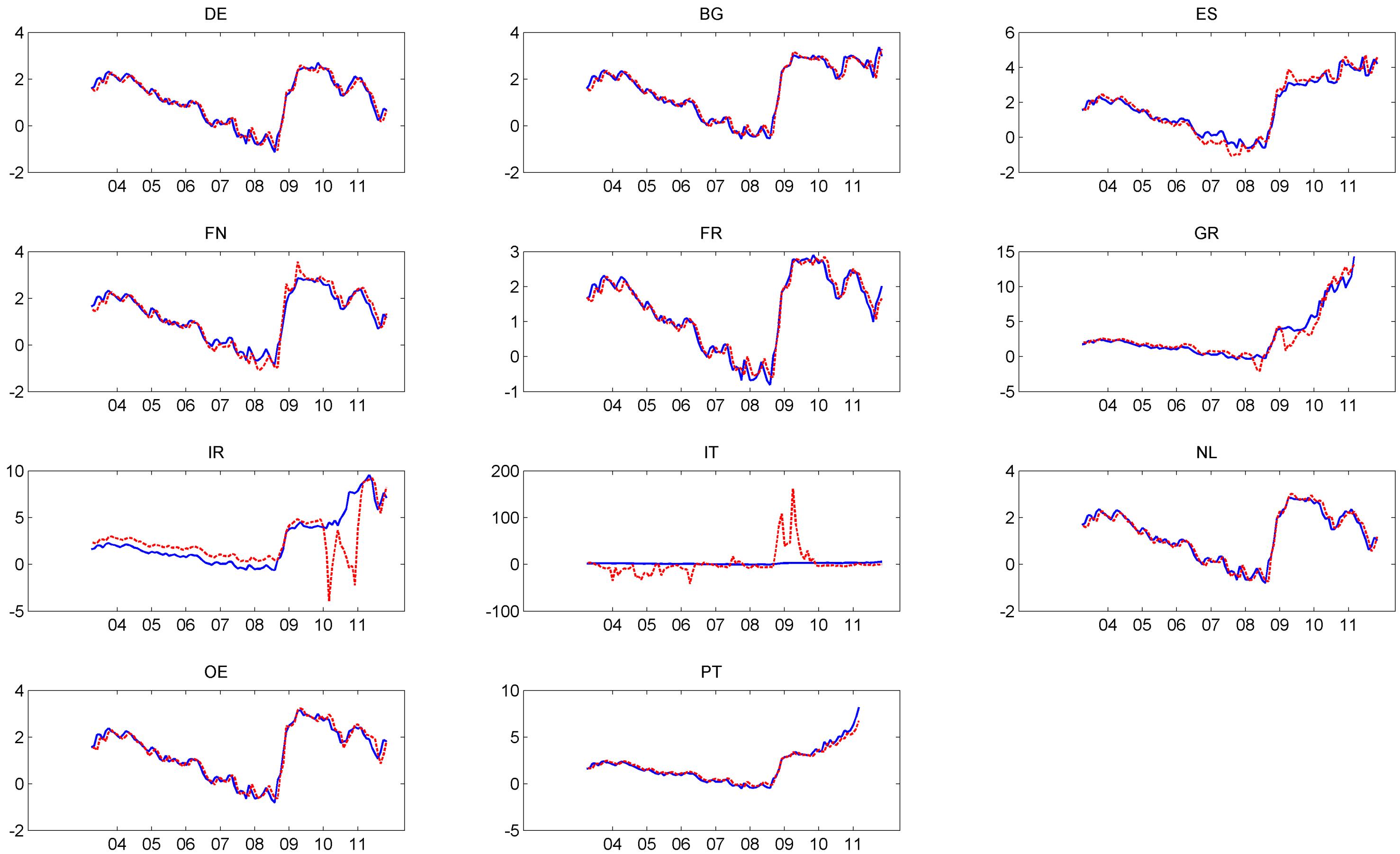

Figure 9: Yield spreads in the panel-out-of-sample estimation.

Note: The blue series is the observed yields, the red series results from panel out-of-sample estimation. In that estimation, one country is excluded from the estimation. The parameters obtained from the ten other countries are used to forecast the yields of the missing country. Note also the different scaling for Italy.

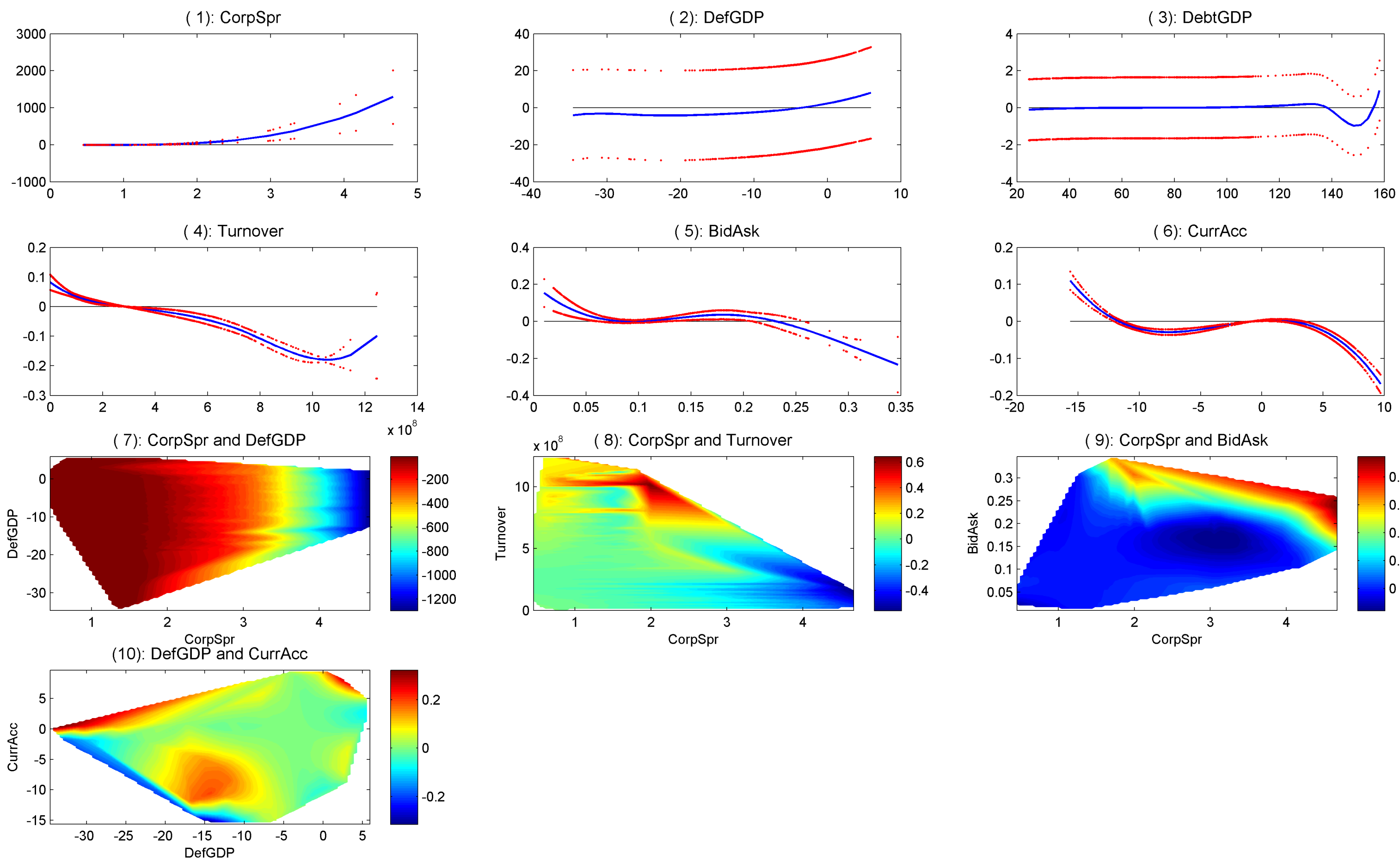

Figure 10: Individual and interaction functions for the estimation excluding Italy.

Note: The y-axis or the height of the color bar show the effect on yields (in %). Confidence bands for univariate functions are given in red, while the estimated function is blue.

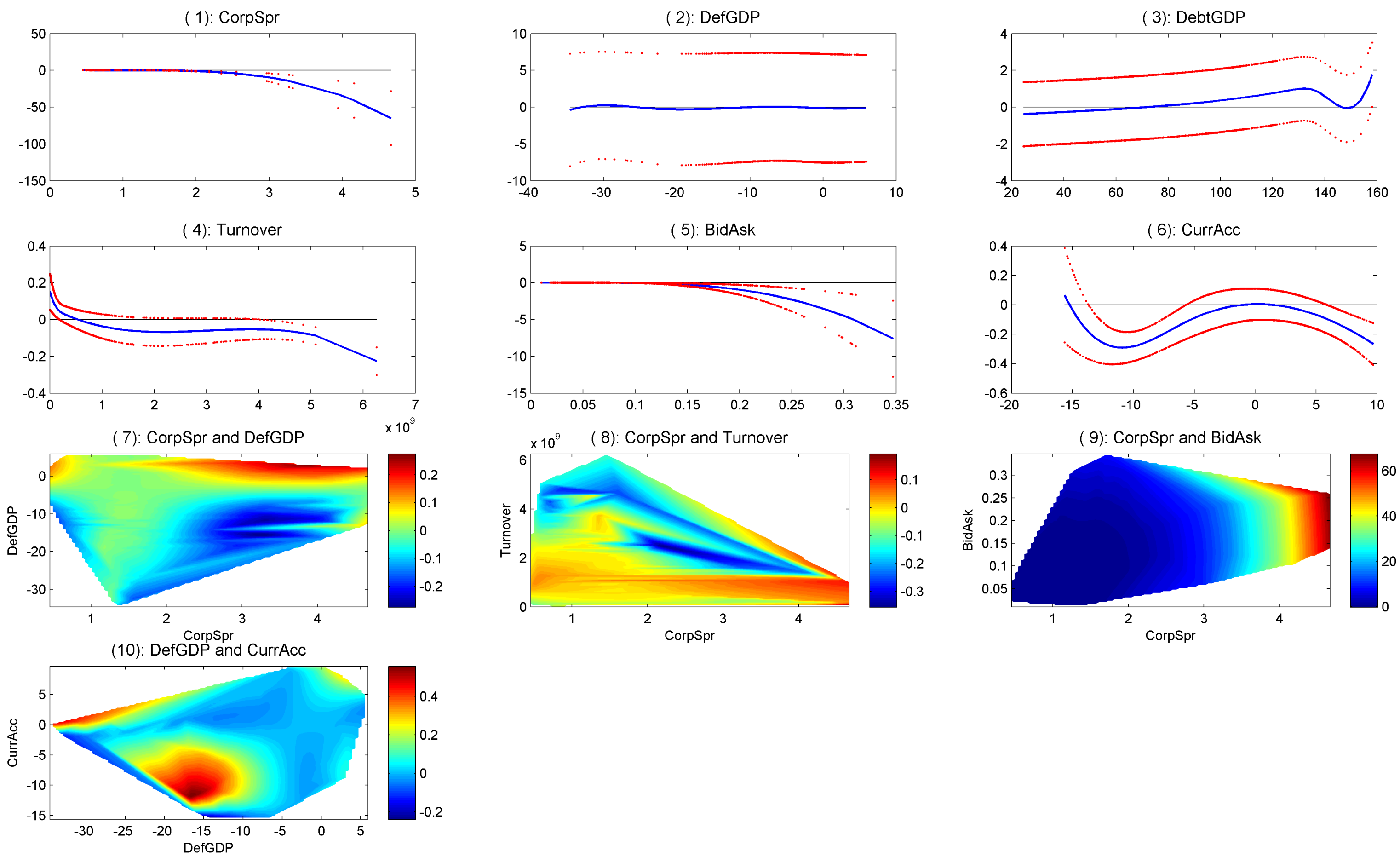

Figure 11: Marginal results for an estimation with country fixed effects.

Note: The y-axis or the height of the color bar show the effect on yields (in %). Confidence bands for univariate functions are given in red, while the estimated function is blue.

Appendix B: Penalized Splines

Penalized Splines are used to estimate an arbitrary, unknown function f with $y = f(x_1, \dots, x_N) + \epsilon$. It is assumed that f is an additive function that can be split into univariate components $f_i(x_i)$ and bivariate components $f_{i,j}(x_i, x_j)$ (Ruppert and Carroll, 1997).[22] Both univariate and bivariate components can be approximated by a third-order Taylor expansion (around x_0):[23]

$$f_i(x_i) = f(x_0) + \frac{\partial f}{\partial x_i}(x_i - x_0) + \frac{\partial f}{2!\partial^2 x_i}(x_i - x_0)^2 + \frac{\partial f}{3!\partial^3 x_i}(x_i - x_0)^3 + \mathcal{O}(x_i^4) \tag{A.1}$$

$$= \sum_{l=0}^{3} \beta_l x_i^l + \varepsilon_i, \tag{A.2}$$

where the equation (A.1) is the normal Taylor expansion and (A.2) an estimation of that expansion (reordered), such that the error ε_i is of order x_i^4. However, while ε_i is globally unbiased, this does not rule out strong local errors of the Taylor expansion. These local errors can be reduced by local polynomials of the form $(x_i - \kappa_{i,k})_+^3$. These functions are also called splines and have a value of zero below a certain threshold (also called knot) $\kappa_{i,k}$. Above that threshold, they are a third-order polynomial. Thus, the function f_i is estimated by

$$f_i(x_i) = \sum_{l=0}^{3} \beta_l x_i^l + \sum_{k=0}^{K_i} b_k (x_i - \kappa_{i,k})_+^3 + \varepsilon_i, \tag{A.3}$$

where K_i are the number of knots on the space covered by the values of x_i. The parameters β_p are assumed to be unknown, but fixed, while parameters b_k are normally distributed random parameters with expectation 0 and standard deviation $\lambda_i \sigma_\varepsilon$ with an unknown penalty parameter λ_i.

From the structure of f_i and f_j, we can develop the structure of the bivariate function $f_{i,j}$ by multiplication of all polynomial and all spline terms (without cross-multiplication). That is, $f_{i,j}$ is given by

$$f_{i,j}(x_i, x_j) = \sum_{l_i=0}^{3} \sum_{l_j=0}^{3} \beta_{l_i,l_j} x_i^{l_i} x_j^{l_j} + \sum_{k_i=0}^{K_i} \sum_{k_j=0}^{K_j} b_{k_i,k_j} (x_i - \kappa_{i,k_i})_+^3 (x_j - \kappa_{j,k_j})_+^3 + \varepsilon_{i,j}.^{24} \tag{A.4}$$

Let C be the total number of countries, T the number of datapoints for each country, p the total number of polynomial regressors and q the total number of spline regressors. Combining – for all univariate functions f_i and all bivariate functions $f_{i,j}$ – all polynomial terms into one variable X (of dimension $CT \times p$, that is, countries are appended below each other) and all spline terms into one variable Z (of dimension $CT \times q$), we get the model

$$y = X\beta + Zb + \varepsilon, \tag{A.5}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ are iid and $b \sim \mathcal{N}(0, \sigma_\varepsilon^2(\Lambda\Lambda'))$.[25] The objective function to be minimized is

$$\min_{\beta,b,\Lambda} \|y - X\beta - Zb\|^2 + b'(\Lambda\Lambda')^{-1} b, \tag{A.6}$$

[22] Interaction of three or more variables may be included in the same way as for two variables.

[23] The method is not restricted to third-order polynomials. However, a third-order polynomial has the advantage of a continuous second derivative and thus guarantees a certain smoothness of the estimate.

[24] It may be that for certain knot locations, the regressor $(x_i - \kappa_{i,k_i})_+^3 (x_j - \kappa_{j,k_j})_+^3$ has only few non-zero entries, making the estimation of b_{k_i,k_j} unstable. Therefore, I exclude those spline terms (interactions), that contain less than 20 non-zero datapoints.

[25] An extension to more complicated error structures (i.e. heteroscedasticity) is possible, but not considered here (Krivobokova and Kauermann, 2007)

where $b'(\Lambda\Lambda')^{-1}b$ is the penalty term giving the method its name.[26] Without penalties, it would be optimal to use as many spline terms as data points in order to produce a perfect fit. Penalties thus counteract the tendency of an overfit. To do this more efficiently, we use a different penalty term λ_i or $\lambda_{i,j}$ for every function f_i or $f_{i,j}$. This leads to the diagonal penalty matrix $\Lambda \in \mathbb{R}_+^{q\times q}$ having the penalty terms λ_i ($\lambda_{i,j}$) on the diagonal, each one K_i ($K_{i,j}$) times. All other elements of Λ are zero, as the elements of b are supposed to be independent.

In practice, not all variables x_i contribute splines or third-grade polynomials to the model (A.5). For example, dummies or fixed effects only enter as constants. Similarly, interaction effects should only be accounted for when they offer substantial added value (both statistically and economically), because the cross-multiplication of terms adds a large number of variables to the model and thereby increases runtime.

To solve the objective function (A.6), it is assumed that the parameters b are multivariate normally distributed (i.e., they are random parameters $b \sim \mathcal{N}(0, \sigma_\varepsilon^2\Lambda\Lambda')$), while β are unknown, but fixed parameters. Thus, the model (A.5) is equivalent to a Linear Mixed Model (Ruppert et al., 2003).[27] The linear transformation $b = \Lambda u$, where u is independent of Λ, transforms the objective function to

$$\min_{\beta,u,\Lambda} \|y - X\beta - Z\Lambda u\|^2 + \|u\|^2. \tag{A.7}$$

With this transformation, Bates (2012) achieves to integrate both β and u out of the likelihood function corresponding to the objective function (A.7), thereby producing a profiled likelihood function. The calculation of optimal parameters in a Linear Mixed Model is then straightforward using (*restricted*) maximum likelihood estimation. REML estimation (Davidson and MacKinnon, 2004, also known as *concentrated* maximum likelihood,) accounts for the possible bias of ML estimates, which can be shown to exist already in the estimation of a simple sinus curve. Therefore, I use the REML-variant of the algorithm proposed by Bates (2012).

The algorithm is quite simple. Let the discrepancy function $\tilde{d}(y|\Lambda)$ be the (transformed) objective function of y given Λ, i.e., the quadratic objective function (A.7) being only minimized over regression parameters β and u:

$$\tilde{d}(y|\Lambda) = \min_{\beta,u} \|y - X\beta - Z\Lambda u\|^2 + \|u\|^2 = \left\| \begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda & X \\ I_q & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2. \tag{A.8}$$

Let now $A = \begin{bmatrix} Z & X \end{bmatrix}' \begin{bmatrix} Z & X \end{bmatrix}$ be the matrix of squared regressors and $L(\Lambda)$ be the cholesky decomposition needed to solve the optimization problem given by the discrepancy function:

$$L(\Lambda)L(\Lambda)' = \begin{bmatrix} \Lambda & 0 \\ 0 & I_p \end{bmatrix} A \begin{bmatrix} \Lambda & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}. \tag{A.9}$$

The cholesky decomposition is used both for the determination of the discrepancy function and in the profiled likelihood (depending only on the penalty parameter Λ):

$$-2l_R(\Lambda|y) = 2\log(|L(\Lambda)|) + (n-p)\left(1 + \log\left(\frac{2\pi\tilde{d}(y|\Lambda)}{n-p}\right)\right), \tag{A.10}$$

which is minimized by standard minimization algorithms (Matlab *fminsearch*) to determine the optimal Λ. The appeal of this algorithm is evident: The calculation of the cholesky decomposition

[26] I use inverse penalties opposed to the original work of Ruppert and Carroll (1997) because I want to keep notation close to the one used by Bates (2012).

[27] A Linear Mixed Model is a linear model with fixed, unknown as well as random parameters.

is rather efficient. The dimension of the parameter space in the likelihood function (A.10) is reduced to the number of different penalty terms. The estimation of $p + q$ regression parameters is straightforward.

As shown by Bates (2012), the estimate for the error variance is

$$\sigma_\varepsilon^2 = \frac{\tilde{d}(y|\Lambda)}{n - p}. \tag{A.11}$$

This estimate takes into account that the *equivalent number of parameters* Ruppert et al. (2003, p. 81) is between the number of polynomial parameters p and the total number of estimated parameters $p + q$. The reduction of degrees of freedom by using additional splines depends on the weight they get in the estimation, that is, on the size of the individual penalty term λ.

It can be observed that Λ is not independent of the location (and possibly the number) of knots. Therefore, a joint optimization of Λ and knot specifications should be performed. This point has – to my knowledge – only been mentioned in passing in the literature so far, probably due to computational reasons: already the (rather low-dimensional) likelihood minimization with predetermined knot specifications uses some time. Performing this minimization for changing knot locations until a global optimum is reached is a tedious task. Therefore, knot numbers and positions have seldom been endogenized in practice. Instead, the number of knots is usually predetermined (between five and forty, depending on the number of datapoints) and they are evenly distributed over the quantiles of x_i (Ruppert and Carroll, 1997).

However, the theoretical literature observes that both the number of knots K_i and their placement at $\kappa_{i,k}$ should depend on the local regression errors of equation (A.2) and the density of the datapoints x_i (Agarwal and Studden, 1980; Spiriti et al., 2013). In a simple framework with only one exogenous variable, Spiriti et al. (2013) use a genetic algorithm (Holland, 1975) to determine knot location given the number of knots. Kauermann and Opsomer (2011) propose to select the number of knots that optimizes the likelihood function, based again on examples with only one explanatory variable. I only adopt the genetic algorithm of Spiriti et al. (2013) for optimizing the knot location, as this already provides a strong improvement. Their paper presents the general framework of the algorithm, leaving some specifications free to be selected by the user: my chosen minimization criterion is the profiled likelihood given in equation (A.10).[28] The algorithm stops if no further improvement can be found in one generation (which happens on average after 62 generations). I differ from the proposed algorithm in that I perform both the crossover and mutation algorithm proposed by Spiriti et al. (2013) for each explanatory variable x_i separately. Mutation is only possible on a predefined set of possible knots, selected to be all quantiles between the 5%- and 95%-quantile. The restriction to quantiles reduces runtime, as it effectively restricts possible variation, while still allowing enough freedom to obtain results that are close to the optimum. It is furthermore imposed that at least five datapoints are between adjacent knots. This ensures non-singularity of the matrix Z and thus stability of the estimation.

It can be observed that the optimal result of a single genetic run with random starting population is not necessarily stable. Therefore, Spiriti et al. (2013) propose to run 20 repetitions of the genetic algorithm with random starting generations. As they have only one variable, the number of repetitions necessary to find the global optimum with high probability is lower than in this case. Instead, 100 repetitions are used. It can be observed that there are multiple local optima with quite similar likelihood functions. However, a random sample of half of the repetitions contains a run

[28] Minimization is done by *fminsearch*, that uses different stopping criteria, including a minimum step size for both the value of variables ('TolX') and the value of the optimized function ('TolFun'). For the initial optimization, I set both minimum step sizes to 10^{-6}. To increase the speed of the genetic algorithm, I relax these restrictions, and set both function parameters to 1. In case no improvement is found in the current generation, the best members of the current generation are reoptimized with minimum step sizes set to 10^{-3}. If there is still no improvement, the genetic algorithm stops after recalculating the optimal solution with the highest precision.

ending in the optimum presented here in more than 90% of the cases. An alternative to several repetitions with random starting generations would be to allow for larger variation from one generation to the next. Following El-Shagi (2011), several possibilities were tested (individually and jointly), among them a preliminary parent generations selected by remainder stochastic sampling with different evaluation functions, higher mutation probabilities and a self adaptive genetic algorithm (Hinterding et al., 1996, SAGA,) with five subpopulations. However, more refined genetic algorithms failed to reproduce stable results and mostly arrived in a local optimum only.