Modelling Macroeconomic Risk: The Genesis of the European Debt Crisis

Dissertation

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)
der Juristischen und Wirtschaftswissenschaftlichen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von
Gregor von Schweinitz
aus Tübingen
August 2013

Gutachter:	Prof. Dr. Oliver Holtemöller
	Priv.-Doz. Dr. Makram El-Shagi
Tag der Promotion:	19. Dezember 2013

List of Original Papers

This thesis consists of the following three papers:

1. Tobias Knedlik and Gregor von Schweinitz (2012):
   Macroeconomic Imbalances as Indicators for Debt Crisis in Europe
   Journal of Common Market Studies, Vol. 50(5), 726-745.

2. Makram El-Shagi, Tobias Knedlik and Gregor von Schweinitz (2013):
   Predicting Financial Crises: The (Statistical) Significance of the Signals Approach
   Journal of International Money and Finance, Vol. 35(1), 67-103.

3. Gregor von Schweinitz (2013):
   Flight Patterns and Yields of European Government Bonds
   IWH-Discussion Paper, 10/2013.

Contents

List of Figures

List of Tables

Overview

Granted that we face a world crisis which leaves us standing so often amid the surging murmur of life's restless sea. But every crisis has both its dangers and its opportunities. It can spell either salvation or doom.

Martin Luther King (1964)

The causes and effects of the U.S. subprime crisis and the European debt crisis have been one of the primary topics of economic research and the public debate for most of the past six years. The global financial crisis in 2007 and 2008 and its aftermath has been compared with the Great Depression that started with the Black Friday, October 24, 1929. Today we know that this comparison might be overstretched. However, given the severe economic consequences, political unrest, and rise of extremist parties in those days, even a similarity is particularly troubling. Understanding the development of the sequence of crises[1] that hit the global and European economy since 2007 as well as a thorough development of adequate policy responses is therefore of great importance (Brunnermeier 2009, Reinhart & Rogoff 2009, Gourinchas & Obstfeld 2011, Tsoukalis 2011). This dissertation contributes to the current debate. First, it analyzes the early-warning capabilities of an early-warning system that has been introduced as a consequence of the European debt crisis. The early-warning system is built on a set of indicators that are supposed to give early warnings of a looming

[1] Four main types of crises are usually distinguished: currency crises, banking crises, systemic financial crises, and sovereign debt crises (IMF 1998). A closer look at the developments of the last five years shows that the current situation can be described by a combination of these types.

crisis, defined as periods of unusually high government bond yields. That is, in a strict sense the system gives a warning of possibly increasing yields. Second, serious methodological shortcomings of the most intuitive method to set up this early-warning system are alleviated. Third, this dissertation not only aims at issuing warnings of a binary crisis variable derived from European government bond yields (as a measure of the extent of a crisis). It also explains the continuous levels of yields directly. To this end, yields are decomposed into different risk premium components. The model that is used to estimate the contribution of these components is much more flexible than previously used models.

In the following these three contributions will be framed in the context of current research. To do this, this introduction combines two different threads, a chronological report of events and a description of crisis types. As the European debt crisis may be seen as a convolution of different types of crises that first appeared at different points in time, it lends itself to such a structure. The historic development is divided into sections describing these different crisis types, accompanied by short reviews of the theoretical literature on the respective type of crisis under discussion. The historical description mostly focuses on the developments in the first twelve member states of the European Monetary Union (EMU), because these are the countries analyzed in all parts of this dissertation. Necessarily, the exposition will be painted with a broad brush where only the general outline is needed, while going into greater detail near the actual contributions. These contributions are shortly summarized at the end of the introduction.

The great moderation

In 2007, a long period of uninterrupted growth ended rather abruptly in most advanced countries. There are mainly three different hypotheses on the origin of the great moderation since the early 1980s (Stock & Watson 2003, Ahmed, Levin & Wilson 2004). According to the „good policy“-hypothesis, credible monetary policies during the Volcker-Greenspan-era stabilized inflation expectations and thereby the whole economy (Clarida, Galí & Gertler

2000, Cogley & Sargent 2005).[2] The shifts of monetary policy in that time are described in detail by Goodfriend (2007). The „good-practice"-hypothesis explains reduced volatility by improved business practices leading to stable income flows (McConnell & Perez-Quiros 2000). The third hypothesis invokes „good luck" as the explanation of falling volatilities (Sims & Zha 2006).[3] The first two explanations describe processes that facilitate a stable development in normal times. Due to good policy and good practice, the economy could withstand smaller shocks with greater ease.

It was thought to be good policy to lower interest rates after the burst of the Dotcom-bubble in 2001. Similarly, the development of a mortgage-backed securities market was supposed to be good practice (Ryding 1990). The flip side of the coin is that low interest rates led to increased risk-taking and a build-up of imbalances (Taylor 2012). A similar conclusion can be drawn from a disaggregated view at the economy (Gabaix 2011). If a single sector of the economy grows over-proportionally, then overall volatility increases.[4] Before the great moderation, the largest sector of the US-economy was manufacturing. Its decline reduced volatility since the 1980s. The rise of the financial sector in the 2000s, explained by larger risk taking and larger (expected) returns, led to increased overall volatility of the economy (Carvalho & Gabaix 2013).

Systemic crises in the financial network

The rise of the financial sector was accompanied by a real estate bubble in several advanced economies, for example the United States, Ireland or Spain (Brunnermeier 2009, Laeven & Valencia 2010). Among other causes, financial innovation and weak regulation (Keys, Mukherjee, Seru & Vig 2010), expansionary monetary policy (Taylor 2012), global imbalances (Obstfeld & Rogoff 2009) and (with the benefit of hindsight) overly optimistic ex-

[2] The Volcker-disinflation came at the cost of two recessions between 1980 and 1983 (Goodfriend & King 2005).

[3] The last hypothesis is rejected by many authors as over-simplistic (Giannone, Lenza & Reichlin 2008, Galí & Gambetti 2009).

[4] This result can be derived from Zipf's Law for firm sizes (Axtell 2001). Due to the power-law properties, idiosyncratic volatility does not cancel out in the aggregate.

pectations (as described by Minsky 1986) contributed to the development of this boom. The boom of the U.S. subprime market came to a stop in early 2007. Due to the involvement of many European banks on this market, for example via special purpose vehicles (Gorton 2008, Gorton 2009), losses then quickly spread to Europe. Thus, a shock on a comparably small market became global and systemic.

Systemic financial crises happen when an initial shock is transmitted to large parts of the densely-weaved network of financial institutions (Allen & Gale 2000, Allen & Babus 2009). Networks – as a theoretical and mathematical concept – are used in many scientific fields in order to model connections and dependencies among entities. The close resemblance to economic structures and the possibility to transfer gained insights is demonstrated by many theoretical studies (May, Levin & Sugihara 2008, Haldane & May 2011, Gai, Haldane & Kapadia 2011). The general findings can be summarized as follows: the network of financial institutions can absorb idiosyncratic or small systemic shocks quite easily. However, when a shock becomes contagious, it may quickly affect the largest part of the network. This property is described as *robust-yet-fragile* (Gai & Kapadia 2010). The potentially most stable system is therefore one with only medium levels of risk diversification (Battiston, Delli Gatti, Gallegati, Greenwald & Stiglitz 2012). The potential for contagion strongly depends on the affected institutions. Thus, current research tries to identify institutions that are too-big-to-fail or too-interconnected-to-fail (Brownlees & Engle 2012, Acharya, Engle & Richardson 2012). In parallel to the search for systemically important financial institutions, macroprudential policies are discussed and developed (for a literature survey, see Galati & Moessner 2013). These policies are aimed at the financial system as a whole as well as its link to the real economy and government finances.

Banking crises: bank runs and credit crunches

In Europe, Northern Rock, a U.K.-based bank and mortgage lender, was one of the first banks hit by the subprime crisis and aftershocks. It experienced a full-scale bank run in September 2007. A *bank run* can be triggered when depositors question the ability of a bank to service its obligations (Diamond & Dybvig 1983). Doubts may for example arise when the share of non-performing assets is high (Demirgüç-Kunt & Detragiache 1998), when reserves are low or when the liquidity transformation (lending money short-term to borrow long-term) does not function as usual (Diamond & Rajan 1999). The possibility that obligations may not be served becomes self-fulfilling when a critical mass of depositors starts to withdraw its accounts, depleting the liquidity reserves of the bank in question. Such a process is much likelier, if banks invest speculatively and if liquidity reserves are already low (Bucher, Dietrich & Hauck 2013).[5] A bank run may create spillover effects and thereby affect large parts or even the whole banking system. Thus, a fast reaction by authorities is required. Government measures aimed at stabilizing the whole banking sector include liquidity support, restructuring, government guarantees for deposits and nationalizations (Laeven & Valencia 2010).[6]

In the case of Northern Rock, guarantees were issued to calm depositors. Later, the bank was nationalized. However, shoring up only this individual bank was clearly not enough. Other banks were affected as well. Ireland was then the first country in the EMU to announce national deposit guarantees in September 2008. Laeven & Valencia (2010) find that among the first twelve members of the EMU, only Italy had no (borderline) systemic banking crisis between 2007 and 2009 (which may be seen as a hybrid of a banking and a systemic financial crisis).

However, even fast reactions by authorities may not be decisive enough to avoid that the interbank lending channel breaks down (Rochet & Tirole 1996).[7] The reason for a disturbance or even a breakdown may be for example an increase of distrust among banks or an increased need of own liquidity to withstand a bank run. The longer the lending channel

[5] Minsky (1986) describes different types of financing: hedge financing, speculative financing and Ponzi-scheme financing. As the memory of past crises fades, trust and income expectations rise. This leads to a shift from safer financing structures to ones that severely fail when incomes are below expectations.

[6] Yet another measure, withdrawal restrictions, was applied for example in Cyprus during March 2013. Cash withdrawals and electronic transactions were strongly restricted for ten days while a bank recapitalization program was negotiated.

[7] A decentralized interbank lending channel may at first transmit shocks to other banks (with the potential of producing a systemic financial crisis). In order to avoid such a spillover, banks may stop lending in general. This then negatively affects not only troubled, but all banks.

is disturbed, the stronger it affects firm-lending and thereby also the real economy. That is, problems in the interbank lending channel may lead in time to a *credit crunch* (Bernanke & Lown 1991).

Low central bank interest rates and other measures (such as the fixed-rate tender procedure of the European Central Bank (ECB))[8] increase money supply and thereby market liquidity (Trichet 2010). Thus, they are aimed at enhancing the lending channel (Laeven & Valencia 2013). However, if – as during the current European debt crisis – uncertainty on the markets remains high, central bank measures provide only a partial or temporary fix. In such a situation, the relationship between the monetary base and broad monetary aggregates breaks down (Carpenter & Demiralp 2012, El-Shagi & Giesen 2013).

Sovereign debt crises

From fall 2008 to spring 2009, European governments issued guarantees and rescue packages to avoid a meltdown of the financial system comparable to the one of the Great Depression. The rescue packages heavily burdened public finances. Ireland, as an extreme case, had a deficit of 14.5% of GDP in 2009, and 32% of GDP in 2010. Around 2.5 percentage points (in 2009) and a stunning 20 percentage points (in 2010) of these deficits were due to bank recapitalizations (Lane 2011). However, government spending could only partially ease the Irish recession. Decreased bank lending, increasing uncertainty, and a shrinking economy affected – among others – the construction sector strongly. This lead to a burst of the real estate bubble that had been built up in the previous boom. Both deficits and falling GDP pushed also public debt over the official 60%-threshold of the Stability and Growth Pact (SGP). A comparable development hit Spanish public finances: a construction and real estate boom as well as increasing domestic demand led to decreasing competitiveness and growing private debt between 1999 and 2007. With growing uncertainty on international capital markets the real estate boom in Spain ended. A sharp drop in GDP and strongly increasing unemployment implied higher public deficits. This in turn pushed public debt from

[8] Press release, April 8, 2009. `http://www.ecb.int/press/pr/date/2009/html/pr090305_3.en.html`

36% of GDP in 2007 to 84% of GDP in 2012. The important element in these developments is that, before 2007, fiscal positions in Ireland and Spain were not unsustainable per se. Rather, the economic developments called for decisive government intervention that pushed government finances into unsustainable regions. Due to the lower degree of diversification in Ireland, it's downturn and public deficits were much stronger than in Spain. In November 2010, the group of EU treasury secretaries decided on a rescue package of 85 billion Euro for Ireland. In Spain the slope of the downturn was flatter. It did not need a rescue package until July 2012, when 100 billion Euro were allocated to support Spanish banks.

In Greece, the situation was different. After an election, the new government announced in November 2009 that government deficits for the past years had been much higher than initially estimated. Hence, it became apparent that Greece had violated the statutes of the SGP for most of the time since the introduction of the Euro in January 2001. Additionally, Greece suffered from a competitiveness problem far worse than Spain. Therefore, markets started to demand higher yields for Greek government bonds. This effectively started the European debt crisis. At the beginning, the increase in yields was only mild. Several factors might have contributed to that. First, the full scale of the problems might not have been apparent. Second, markets may have thought that European governments would help each other in times of a crisis, although the no-bail-out clause of the SGP explicitly forbids any financial transfers (Arghyrou & Tsoukalas 2011). During the negotiations of the first rescue package for Greece, however, especially Germany signaled its reluctance. This may have contributed to a higher credibility of the no-bail-out clause. At the same time, it increased political uncertainty strongly.[9] So far, Greece needed two rescue packages. In March 2010, a first package of 110 billion Euro was set up. When this first package proved to be insufficient, a second package was negotiated between July 2011 and February 2012. Originally set up with a size of 100 billion Euro, it grew to 130 billion Euro until it was finally decided in February 2012. This rescue package was accompanied by a haircut for Greek government bonds held by private investors. The package had also unforeseen political consequences:

[9] As discussed in chapter three, the adverse effects of (political) uncertainty may be well responsible for a large part of the growing differentiation between European government bonds.

the Greek premier minister at that time, Georgios Papandreou, wanted to put the rescue package to a referendum. He resigned in November 2011 after being put under pressure by Germany and France to abandon this idea.

Being uncertain about the prospects and future developments in Greece, markets also started to demand higher yields from Ireland, Spain and Portugal. Portugal did not face a bursting real estate bubble like Ireland or Spain. It became a member of the crisis countries nonetheless due to competitiveness problems and high public debt levels. When its deficit reduction plans were not approved in March 2011, rating agencies downgraded Portuguese bonds and interest rates soared (Hodson 2012). In the following two months Portugal was the third country to receive a rescue package (of 78 billion Euro), in exchange for further austerity measures. Belgium and Italy, having equally high debt levels, have not been excluded from capital markets. One reason may be their higher competitiveness levels (measured for example by the current account or the real effective exchange rate).

Sovereign debt crises have long been theoretically investigated. In the first modern publications, they have been analyzed through the lenses of time inconsistency (Kydland & Prescott 1977, Barro & Gordon 1983, Calvo 1988).[10] It may not be the optimal fiscal policy to keep past promises on debt repayments. Instead, governments have an incentive to dishonor their debt obligations. This incentive can be kept at bay, if access to capital markets is strongly limited after a default (Eaton & Gersovitz 1981) or if there are other severe consequences such as direct sanctions (Bulow & Rogoff 1989). These two types of costs are imposed by foreign creditors. Borensztein & Panizza (2009) suggest two additional cost types of a sovereign default, originating from the domestically held share of debt: an economic downturn following problems in the financial system, and political consequences for the governments responsible for the default. For example, the reduction of output growth due to a public debt crisis (controlling for the effect of a simultaneously occurring banking or currency crisis) is expected to be around 10% over the course of the following eight years

[10] Classical economists have also been concerned with public finances and debt. Famous among these are the contributions of Ricardo (1888). For a formal development of the Ricardian equivalence theorem (although this is not strictly related to a sovereign debt crisis), see Barro (1974). However, O'Driscoll (1977) aptly summarizes that Ricardo was well aware of the fact that the equivalence might not hold because of „fiscal illusion".

(Furceri & Zdzienicka 2012). However, even if the consequences are severe enough to prevent an unforced default, there may still be circumstances that make a default unavoidable or at least much more likely. Many of them are quite closely related to other types of crises. Besides unsustainable levels of sovereign debt, there may be debt rollover problems (Cole & Kehoe 1996, Cole & Kehoe 2000). A government may decide to default rather than rollover its debt if current interest rates demanded by markets are too high. Such a crisis is self-fulfilling in its nature (De Grauwe & Ji 2013). Consequently, models inspired from the literature on currency crises have been used to explain sovereign debt crises (see for example Arghyrou & Tsoukalas 2011). Following this argument, the European debt crisis may be seen as a currency crisis in the periphery countries. Because devaluation is not possible without exiting the Euro, the (speculative) attacks are aimed at bond yields instead of exchange rates. In the context of the European debt crisis, Lane (2012) finds that such a development was aggravated by insufficient institutions and a chaotic crisis management.

Currency and Balance-of-payment crises

The first generation of *currency crisis* or *balance-of-payment* models is founded on the seminal work of Krugman (1979). In these models, speculators rationally attack an unsustainable currency peg. The peg is unsustainable because the exchange rate is kept at uncompetitive levels. The low competitiveness of the country leads to low exports and high imports, that is, a persistently negative current account balance. Markets expect that the currency exchange rate has to fall (i.e. that the peg needs to be abandoned) in order to make the external balance sustainable. If currency reserves are not sufficient to defend the peg, a sudden devaluation happens. With this simple mechanism, first-generation models could describe crises in Latin America in the 1980s very well. However, the experiences of the European exchange rate mechanism in 1992 and 1993 showed that such attacks can also happen to a previously sustainable peg. This lead to the development of second generation models of currency crises (Obstfeld 1996). While the speculation against a currency in the models of the first generation is rational *ex ante* in the sense that fundamentals point to an unsustainable peg, speculation in second generation models become only rational *ex post* due to the existence of multiple equilibria. That is, a sustainable peg may become non-defensible if only the attack is strong enough.

Twin and multiple crises

The term *twin crisis* was coined by Kaminsky & Reinhart (1999) for the joint occurrence of a banking and a balance-of-payment crisis. Both crises may happen jointly because of common causes. The size of a downturn is stronger and the time to recover is longer for a twin crisis than for the individual types alone (Bordo, Eichengreen, Klingebiel & Martinez-Peria 2001). First- and second-generation models for currency crises had problems explaining the Asian crisis in the 1990s, which was (in several countries) an example of a twin crisis. Accordingly, in her literature review Breuer (2004) classifies third-generation currency crisis models as models explaining twin crises. Bauer, Herz & Karb (2003) develop a model for a different sort of twin crisis, combining sovereign debt and currency crises. Depending on the level of debt and macroeconomic fundamentals, a government simultaneously decides if it is optimal to default on its debt and/or leaving a peg and devaluing strongly. Empirically, the link between currency and sovereign debt crises is tested for example by Reinhart (2002). She finds that a credit rating downgrade seldom precedes a currency crisis. Instead, rating agencies seem to see currency instability as a sign of increased default risk. This also builds a link between banking crises and sovereign debt crises, as banking crises are in turn often a precedent of currency crises (Kaminsky & Reinhart 1999).

This is close to what has been observed in Europe. Broadly speaking, a banking crisis and a balance-of-payments crisis triggered (together with already high debt levels due to bank bailouts) a sovereign debt crisis (Reinhart & Rogoff 2011). Ejsing & Lemke (2011) document how bank rescue packages transformed the banking to a sovereign debt crisis. CDS-premiums for sovereign debt increased and became more susceptible to further shocks after a rescue package, while CDS-premiums for banks decreased and became more stable. That is, the results of Ejsing & Lemke (2011) suggest that markets saw a transfer of risks from the banking to the sovereign sector.

Policy reaction and contribution of this dissertation

In 2010, it became apparent that government bond yields could not only be reduced by bringing down government deficit and debt. The European debt crisis had (at least) three economic dimensions: fiscal positions, competitiveness issues and the interdependence of bank and government balance sheets.

Besides drastic direct measures such as rescue packages and the Greek haircut, medium-term measures include institutional reforms (EC 2010). One of the instruments, introduced in November 2011, is the so-called sixpack. It includes a reform of the SGP and the introduction of a macroeconomic imbalances procedure. The macroeconomic imbalances procedure combines a (quantitative) early-warning system with a (qualitative) in-depth analysis of identified imbalances. Within the early-warning system, called scoreboard, a set of indicators is analyzed and published yearly. If an indicator of the scoreboard is above a predefined threshold (to be determined by the European Commission), a warning is issued. First recommendations as to what indicators should be included appeared during 2010 and 2011 (European Commission 2010a, BMWi 2010, ECB 2010, Heise 2011).

The first part of this dissertation tests the ability of the four proposed indicator sets to issue timely early-warning signals.[11] Given the political decision to create the scoreboard, we test if this instrument would at least have issued a warning of the current debt crisis. To answer this question, the signals approach (Kaminsky & Reinhart 1999) is employed. The signals approach uses thresholds to transform indicators into binary early-warning signals. These signals should appear inside an early-warning window before a crisis. In our case, a crisis is defined by unusually high government bond yields. The thresholds are set such that a quality measure combining the share of type-I errors (missing signals) and type-II errors (false signals) is optimized. Therefore, the signals approach is in concept quite

[11] The article, jointly written with Tobias Knedlik, is published in the Journal of Common Market Studies (Knedlik & von Schweinitz 2012).

close to the setting of a scoreboard.[12] We find that broader indicator sets have higher early-warning capabilities. This leads us to propose a new indicator set, combining all four original proposals. A composite indicator based on this broad set of individual indicators would have issued warnings of rising problems up to two years before the European debt crisis, well before the subprime crisis. We also find that macroeconomic imbalances built up slowly for most of the time since the introduction of the Euro. That is, the depth and length of the European debt crisis is probably more associated with the size of imbalances which need to be adjusted. The subprime crisis acted only as a trigger, unraveling the imbalances.

The signals approach used in the first part is a very suitable method to test the scoreboard because of its conceptual proximity. However, it has a severe disadvantage. Given the optimal threshold, only type-I and type-II-errors are counted to obtain a quality measure. But the signals approach does not provide an assumption on the distribution of these errors. Hence, it is unclear if the relation between an indicator (or rather, its transformed binary signal) and a looming crisis is significant or not. In the second part of this dissertation, this disadvantage is alleviated. We develop bootstrap methods for three important past applications of the signals approach.[13] These applications are: Kaminsky & Reinhart (1999) for banking and currency crises in developing countries; Alessi & Detken (2011) for costly asset price booms in OECD-countries; and the first part of the dissertation, Knedlik & von Schweinitz (2012) for the European debt crisis. The bootstrap methods are tailored to each application. Thereby, the simulated time series (crises and indicators) show similar statistical properties as the original ones. This allows to calculate quality measures under the null hypothesis that there is no relation between early-warning signals and crises. The resulting distribution is then compared to the true quality measure. We find that in-sample the signals approach significantly outperforms a random relation in five out of eight cases at the 1% level.[14] In one further case, the signals approach yields results that are significant

[12] The main difference is the setting of thresholds. The signals approach uses assumptions on the costs of type-I and type-II errors to optimize quality measures. The official thresholds of a scoreboard, on the other hand, are the result of political negotiations (Knedlik 2012).

[13] This article, jointly written with Makram El-Shagi and Tobias Knedlik, is published in the Journal of International Money and Finance (El-Shagi, Knedlik & von Schweinitz 2013).

[14] Kaminsky & Reinhart (1999) build an early warning system for two types of crises. Thus, we test four

at the 5% level. However, as the signals approach would in practice be used for issuing warnings out-of-sample (similar to every other early-warning system), the significance needs also to be established in that case.[15] In the out-of-sample analysis, we consider only one quality measure. The results are significant at the 1% level in three out of four cases.

As summarized in different publications, European (political) institutions did not only decide on the sixpack (Hodson 2012, Hodson 2013). They established lending facilities providing liquidity to European governments, first the European Financial Stability Facility (EFSF) in July 2010, then the European Stability Mechanism (ESM) in September 2012. Furthermore, the strong interdependence of the stability of the financial system and fiscal sustainability called for a reform of banking supervision. As a first step, the European System of Financial Supervision (ESFS) was founded in January 2011. The ESFS also contains the European Systemic Risk Board (ESRB) that gives advice on macro-prudential policies. The ESFS works in close cooperation with the ECB. For example, the chairs of the ESRB have been Jean-Claude Trichet and Mario Draghi, the former and current president of the ECB.

In addition to its involvement in the set up of up a more robust financial system in the medium run, the reaction of the ECB to the crisis also encompassed short-run measures. When short-term money market rates soared in August 2007, the ECB soothed markets by providing overnight liquidity funds. After the crash of Lehman Brothers in September 2008, it quickly lowered interest rates below 1%. However, the transmission mechanism of interest rates was not working as expected. Therefore, the ECB adopted measures beyond pure interest rate decisions to reduce spreads and restore a proper transmission of monetary policy impulses (see for example Trichet 2010, Cour-Thimann & Winkler 2012). While the first non-standard measures tried to ensure a properly working financial system, a new focus of the ECB emerged in May 2010 with the launch of the Securities Markets Program (SMP). Under the SMP, the ECB has the right to buy government bonds on the secondary market in order to reduce yields to acceptable levels.[16] However, at first it did not succeed in doing

different early-warning systems. For each of them, two different quality measures are considered.

[15] This is in itself an extension of the original papers which only partially consider out-of-sample analysis.

[16] The SMP (and the following OMT) has drawn wide-spread and controversial critique. On the one hand, it should be seen as both a monetary and a fiscal policy tool and could therefore endanger the central banks'

so. Only the announcement of Mario Draghi that *„[w]ithin our mandate, the ECB is ready to do whatever it takes to preserve the euro.“* (Draghi 2012) in July 2012 could finally initiate a slow convergence process of government bond yields.[17] A month later, the ECB announced the introduction of the Outright Monetary Transactions (OMT) program. Under the OMT, the ECB may buy limitless bonds of governments, with two caveats. The purchases are limited to secondary markets; and they are conditional on the existence of financial support from the EU and the International Monetary Fund (IMF) (Hodson 2013). The OMT has been the subject of wide-spread debate: It raises concerns about the independence of the ECB, inflationary monetary policy and the unconstitutionality of monetary financing of government debt. However, there is also considerable agreement that the pure existence of the OMT (which has not been used until now) eased uncertainty and lowered government bond yields. It thereby helped countries to regain access to financial markets.[18]

This soothing effect could not be attributed to all other previously undertaken measures. Before the OMT, Lane (2012) describes *„Europe's efforts to address its sovereign debt problem as makeshift and chaotic“*. This is the concern of the third part of this dissertation.[19] In this chapter, I estimate determinants of yields of European government bonds. In the first chapter of the dissertation, yields are used to determine if a country is in a crisis or not. In the third chapter, not binary crisis variables derived from yields, but the yields itself are the variables of interest. Contrary to most papers that try to estimate determinants of yields (take only Schuknecht, von Hagen & Wolswijk (2009) and Beirne & Fratzscher (2013) as an example), I do not employ a breakpoint at the crash of Lehman Brothers. A breakpoint leads to two sets of different parameter estimates. The reasoning for not using such an easy switch is twofold. First, from a practical point of view it does not necessarily make

independence. This point was – and is still – mostly made by members of the German Bundesbank and ordo-liberal economists (see for example Issing 2013). On the other hand, the volume of the SMP (211 billion Euro in 2011) was deemed to be too small to be of real significance (Hodson 2012).

[17] Perhaps, the soothing effect should instead be attributed to the afterthought: *„And believe me, it will be enough.“* (Draghi 2012)

[18] At the monthly press conference in June 2013, Mario Draghi said that *„it's really very hard not to state that OMT has been probably the most successful monetary policy measure undertaken in recent time“* (Draghi 2013).

[19] The article is published as an IWH-Discussion Paper (von Schweinitz 2013).

sense. If results are to be used to provide an argument for certain policies, then the future economic regime would have to be known. Otherwise, one would not know if the future is best described by the „normal“ or the „crisis“ set of parameters. Second, statistically a breakpoint should be used if there is evidence that the explained variable (yields) reacts differently to more or less equal values of the explanatory variables, including a constant. In the current case, these are public debt and deficit, liquidity measures and the corporate bond spread, among others. However, during the crisis, the support of all main explanatory variables shifted to the worse. Debt and deficits increased and market liquidity was reduced substantially. That is, we observe a stronger relationship for more extreme values of the explanatory variables. Such a relationship should, however, rather be modeled nonlinearly. Using penalized splines (Ruppert, Wand & Carroll 2003), I am able to estimate unknown nonlinear relationships between explanatory variables (approximating credit default risk, liquidity risk and a global risk component) and yields. Additionally, the method allows for the inclusion of interaction terms between the corporate bond spread (a measure of market uncertainty) and credit risk or liquidity risk variables. Using these interaction terms, I am able to identify the possible existence of flight-to-quality or flight-to-liquidity (Vayanos 2004). These two patterns describe the stronger incentive of investors to buy assets of higher quality and liquidity in times of heightened market uncertainty. The results confirm findings in the previous literature, namely a strong effect of credit risk and of flight-to-liquidity. Liquidity risk and flight-to-quality seem to have played only a minor role for the yields of government bonds.

The main focus of this dissertation is technical. Different non-standard measures are used to test and enhance an early-warning system for sovereign debt crises and estimate the complex relationship between different explanatory variables and government bond yields. However, the dissertation also contributes to the political debate. The (political and economical) conclusions that can be drawn from the technical analysis can be recapitulated in two key messages:

First, the European debt crisis was mostly home-made. The subprime crisis only acted as a trigger of a crisis that, in its essence, is due to macroeconomic imbalances. These imbalances would have issued clear (and significant) warnings long before the crisis. Thus, there was ample time to react to and reduce the imbalances within the Euro area. In the end, the development is quite close to the one famously described in the book of Reinhart & Rogoff (2009): piling-up debt and losing competitiveness dramatically increases the probability of a crisis. Such a crisis can then be a banking crisis, a currency crisis, a sovereign debt crisis – or all these types together. In any case, the economic consequences are severe.

Second, there are at least two strategies governments and policymakers can now pursue in order to reduce government bond yields to sustainable levels and thereby go a long way to end the current crisis. Governments should surely reduce deficits and debt levels in order to reach again regions of sustainable government finances. However, given the general unfavorable economic environment and low growth rates in crisis countries, this will be a process of several years. Thus, European governments and policymakers should strive jointly to lower market uncertainty and reduce macroeconomic imbalances in the Euro area, in order to reduce risk premiums. The OMT is certainly successful in the first respect. However, other ongoing projects need to be decided before uncertainty about future development continues to harm the usually beneficial role of financial intermediation.

Furthermore, although this argument is not further pursued in the following chapters, European politicians should be extremely concerned about the growing „Euro-fatigue“ and other troubling developments in the European societies (Tsoukalis 2011). These developments include growing unrest, rise of extremist parties (for example, the Golden Dawn in Greece) and large-scale protests and riots. They not only undermine the faith of financial markets in the European project but most importantly the project itself. The Chinese scripture knows only one symbol for „crisis“ and „opportunity“. The introductory statement of Martin Luther King as well points to the fact, that crises offer opportunities. However, these opportunities need to be grabbed. In order to do this, politicians need to develop and convince the electorate of their visions. Visual driving may instead only produce local optima and dramatically fail in the long run.

Bibliography

Acharya, V., Engle, R. & Richardson, M. (2012). Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks, *American Economic Review* **102**(3): 59–64.

Ahmed, S., Levin, A. & Wilson, B. A. (2004). Recent US Macroeconomic Stability: Good Policies, Good Practices, or Good Luck?, *Review of Economics and Statistics* **86**(3): 824–832.

Alessi, L. & Detken, C. (2011). Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity, *European Journal of Political Economy* **27**(3): 520–533.

Allen, F. & Babus, A. (2009). Networks in Finance, *in* P. Kleindorfer, Y. Wind & R. Gunther (eds), *The Network Challenge. Strategy, Profit and Risk in an Interlinked World*, Prentice Hall, pp. 367–382.

Allen, F. & Gale, D. (2000). Financial Contagion, *Journal of Political Economy* **108**(1): 1–33.

Arghyrou, M. G. & Tsoukalas, J. D. (2011). The Greek Debt Crisis: Likely Causes, Mechanics and Outcomes, *World Economy* **34**(2): 173–191.

Axtell, R. L. (2001). Zipf Distribution of US Firm Sizes, *Science* **293**(5536): 1818–1820.

Barro, R. J. (1974). Are Government Bonds Net Wealth?, *Journal of Political Economy* **82**(6): 1095–1117.

Barro, R. J. & Gordon, D. B. (1983). Rules, Discretion and Reputation in a Model of Monetary Policy, *Journal of Monetary Economics* **12**(1): 101–121.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. & Stiglitz, J. E. (2012). Liaisons Dangereuses: Increasing Connectivity, Risk Sharing, and Systemic Risk, *Journal of Economic Dynamics and Control* **36**(8): 1121–1141.

Bauer, C., Herz, B. & Karb, V. (2003). The Other Twins: Currency and Debt Crises, *Jahrbuch für Wirtschaftswissenschaften/Review of Economics* **54**(3): 248–267.

Beirne, J. & Fratzscher, M. (2013). The Pricing of Sovereign Risk and Contagion During the European Sovereign Debt Crisis, *Journal of International Money and Finance* **34**(1): 60–82.

Bernanke, B. S. & Lown, C. S. (1991). The Credit Crunch, *Brookings Papers on Economic Activity* **1991**(2): 205–247.

Bordo, M., Eichengreen, B., Klingebiel, D. & Martinez-Peria, M. S. (2001). Is the Crisis Problem Growing More Severe?, *Economic Policy* **16**(32): 51–82.

Borensztein, E. & Panizza, U. (2009). The Costs of Sovereign Default, *IMF Staff Papers* **56**(4): 683–741.

Breuer, J. B. (2004). An Exegesis on Currency and Banking Crises, *Journal of Economic Surveys* **18**(3): 293–320.

Brownlees, C. T. & Engle, R. F. (2012). Volatility, Correlation and Tails for Systemic Risk Measurement, *Working Paper*, New York University.

Brunnermeier, M. K. (2009). Deciphering the Liquidity and Credit Crunch 2007-2008, *Journal of Economic Perspectives* **23**(1): 77–100.

Bucher, M., Dietrich, D. & Hauck, A. (2013). Business Cycles, Bank Credit and Crises, *Economics Letters* **120**(2): 229–231.

Bulow, J. I. & Rogoff, K. S. (1989). Sovereign Debt: Is to Forgive to Forget?, *American Economic Review* **79**(1): 43–50.

Bundesministerium für Wirtschaft und Technologie (BMWi) (2010). Ein neues Verfahren für die wirtschaftspolitische Koordinierung in Europa, *Monatsbericht*, Schlaglichter der Wirtschaftspolitik.

Calvo, G. A. (1988). Servicing the Public Debt: The Role of Expectations, *American Economic Review* **78**(4): 647–661.

Carpenter, S. & Demiralp, S. (2012). Money, Reserves, and the Transmission of Monetary Policy: Does the Money Multiplier Exist?, *Journal of Macroeconomics* **34**(1): 59–75.

Carvalho, V. M. & Gabaix, X. (2013). The Great Diversification and its Undoing, *American Economic Review* **103**(5): 1697–1727.

Clarida, R., Galí, J. & Gertler, M. (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Quarterly Journal of Economics* **115**(1): 147–180.

Cogley, T. & Sargent, T. J. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US, *Review of Economic Dynamics* **8**(2): 262–302.

Cole, H. L. & Kehoe, T. J. (1996). A Self-Fulfilling Model of Mexico's 1994–1995 Debt Crisis, *Journal of International Economics* **41**(3): 309–330.

Cole, H. L. & Kehoe, T. J. (2000). Self-Fulfilling Debt Crises, *Review of Economic Studies* **67**(1): 91–116.

Commission of the European Communities (European Commission) (2010). Economic Governance Package (2): Preventing and Correcting Macroeconomic Imbalances, *MEMO/10/454*, `http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/10/454&format=HTML&aged=0&language=EN&guiLanguage=en`. 29. September 2011.

Cour-Thimann, P. & Winkler, B. (2012). The ECB's Non-Standard Monetary Policy Measures: the Role of Institutional Factors and Financial Structure, *Oxford Review of Economic Policy* **28**(4): 765–803.

De Grauwe, P. & Ji, Y. (2013). Self-fulfilling Crises in the Eurozone: An Empirical Test, *Journal of International Money and Finance* **34**(1): 15–36.

Demirgüç-Kunt, A. & Detragiache, E. (1998). The Determinants of Banking Crises: Evidence from Developed and Developing Countries, *IMF Staff Papers* **45**(1): 81–109.

Diamond, D. W. & Dybvig, P. H. (1983). Bank Runs, Deposit Insurance, and Liquidity, *Journal of Political Economy* **91**(3): 401–419.

Diamond, D. W. & Rajan, R. G. (1999). Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking, *Journal of Political Economy* **109**(2): 287–327.

Draghi, M. (2012). Speech by Mario Draghi, President of the European Central Bank at the Global Investment Conference in London.

Draghi, M. (2013). Monthly Press Conference by Mario Draghi, President of the European Central Bank at the ECB.

Eaton, J. & Gersovitz, M. (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis, *Review of Economic Studies* **48**(2): 289–309.

Ejsing, J. & Lemke, W. (2011). The Janus-Headed Salvation: Sovereign and Bank Credit Risk Premia During 2008–2009, *Economics Letters* **110**(1): 28–31.

El-Shagi, M. & Giesen, S. (2013). A Note on the Hidden Risk of Inflation. mimeo.

El-Shagi, M., Knedlik, T. & von Schweinitz, G. (2013). Predicting Financial Crises: The (Statistical) Significance of the Signals Approach, *Journal of International Money and Finance* **35**(1): 76–103.

European Central Bank (ECB) (2010). Reinforcing Economic Governance in the Euro Area, `http://www.ecb.int/pub/pdf/other/reinforcingeconomicgovernanceintheeuroareaen.pdf`. 10. June 2011.

European Council (EC) (2010). Strengthening Economic Governance in the EU: Report of the Task Force to the European Council (Van Rompuy Task Force), `http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/117429.pdf`. 21. October 2011.

Furceri, D. & Zdzienicka, A. (2012). How Costly are Debt Crises?, *Journal of International Money and Finance* **31**(4): 726–742.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations, *Econometrica* **79**(3): 733–772.

Gai, P., Haldane, A. & Kapadia, S. (2011). Complexity, Concentration and Contagion, *Journal of Monetary Economics* **58**(5): 453–470.

Gai, P. & Kapadia, S. (2010). Contagion in Financial Networks, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **466**(2120): 2401–2423.

Galati, G. & Moessner, R. (2013). Macroprudential Policy – a Literature Review, *Journal of Economic Surveys* . forthcoming.

Galí, J. & Gambetti, L. (2009). On the Sources of the Great Moderation, *American Economic Journal: Macroeconomics* **1**(1): 26–57.

Giannone, D., Lenza, M. & Reichlin, L. (2008). Explaining the Great Moderation: It is not the Shocks, *Journal of the European Economic Association* **6**(2-3): 621–633.

Goodfriend, M. (2007). How the World Achieved Consensus on Monetary Policy, *Journal of Economic Perspectives* **21**(4): 47–68.

Goodfriend, M. & King, R. G. (2005). The Incredible Volcker Disinflation, *Journal of Monetary Economics* **52**(5): 981–1015.

Gorton, G. B. (2008). The Panic of 2007, *NBER Working Paper 14358*.

Gorton, G. B. (2009). Information, Liquidity, and the (Ongoing) Panic of 2007, *American Economic Review* **99**(2): 567–572.

Gourinchas, P.-O. & Obstfeld, M. (2011). Stories of the Twentieth Century for the Twenty-First, *NBER Working Paper 17252*.

Haldane, A. G. & May, R. M. (2011). Systemic Risk in Banking Ecosystems, *Nature* **469**(7330): 351–355.

Heise, M. (2011). Notwendigkeit und Ausgestaltung makroökonomischer Überwachung im Euroraum, *Wirtschaftsdienst* **91**(1): 22–30.

Hodson, D. (2012). The Eurozone in 2011, *Journal of Common Market Studies* **50**(S2): 178–194.

Hodson, D. (2013). The Eurozone in 2012: 'Whatever It Takes to Preserve the Euro'?, *Journal of Common Market Studies* **51**(S1): 183–200.

International Monetary Fund (IMF) (1998). Financial Crises: Causes and Indicators, *IMF World Economic Outlook May 1998*.

Issing, O. (2013). A New Paradigm for Monetary Policy?, *CFS Working Paper 2013/02*.

Kaminsky, G. L. & Reinhart, C. M. (1999). The Twin Crises: the Causes of Banking and Balance-of-Payments Problems, *American Economic Review* **89**(3): 473–500.

Keys, B. J., Mukherjee, T., Seru, A. & Vig, V. (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans, *Quarterly Journal of Economics* **125**(1): 307–362.

Knedlik, T. (2012). The European Commission's Scoreboard of Macroeconomic Imbalances: The Impact of Preferences on an Early Warning System, *IWH Discussion Paper 10/2012*.

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

Krugman, P. (1979). A Model of Balance-of-Payments Crises, *Journal of Money, Credit and Banking* **11**(3): 311–325.

Kydland, F. E. & Prescott, E. C. (1977). Rules Rather Than Discretion: The Inconsistency of Optimal Plans, *Journal of Political Economy* **85**(3): 473–491.

Laeven, L. & Valencia, F. (2010). Resolution of Banking Crises: The Good the Bad and the Ugly, *IMF Working Paper 10/146*.

Laeven, L. & Valencia, F. (2013). The Real Effects of Financial Sector Interventions During Crises, *Journal of Money, Credit and Banking* **45**(1): 147–177.

Lane, P. R. (2011). The Irish Crisis, *CEPR Discussion Paper 8287*.

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

Martin Luther King (1964). Nobel Lecture, delivered in the Auditorium of the University of Oslo.

May, R. M., Levin, S. A. & Sugihara, G. (2008). Complex Systems: Ecology for Bankers, *Nature* **451**(7181): 893–895.

McConnell, M. M. & Perez-Quiros, G. (2000). Output Fluctuations in the United States: What has Changed Since the Early 1980's?, *American Economic Review* **90**(5): 1464–1476.

Minsky, H. P. (1986). *Stabilizing an Unstable Economy*, Yale University Press.

Obstfeld, M. (1996). Models of Currency Crises with Self-Fulfilling Features, *European Economic Review* **40**(3): 1037–1047.

Obstfeld, M. & Rogoff, K. S. (2009). Global Imbalances and the Financial Crisis: Products of Common Causes, *CEPR Discussion Paper 7606*.

O'Driscoll, G. P. (1977). The Ricardian Nonequivalence Theorem, *Journal of Political Economy* pp. 207–210.

Reinhart, C. M. (2002). Default, Currency Crises, and Sovereign Credit Ratings, *World Bank Economic Review* **16**(2): 151–170.

Reinhart, C. M. & Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*, Princeton University Press.

Reinhart, C. M. & Rogoff, K. S. (2011). From Financial Crash to Debt Crisis, *American Economic Review* **101**(5): 1676–1706.

Ricardo, D. (1888). *The Works of David Ricardo. With a Notice of the Life and Writings of the Author*, John Murray.

Rochet, J.-C. & Tirole, J. (1996). Interbank Lending and Systemic Risk, *Journal of Money, Credit and Banking* **28**(4): 733–762.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric Regression*, Vol. 12 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge Univ Press.

Ryding, J. (1990). Housing Finance and the Transmission of Monetary Policy, *Federal Reserve Bank of New York Quarterly Review* **15**(2): 42–55.

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

Sims, C. A. & Zha, T. (2006). Were There Regime Switches in US Monetary Policy?, *American Economic Review* **96**(1): 54–81.

Stock, J. H. & Watson, M. W. (2003). Has the Business Cycle Changed and Why?, *NBER Macroeconomics Annual 2002*, Vol. 17, MIT Press, pp. 159–230.

Taylor, J. B. (2012). Monetary Policy Rules Work and Discretion Doesn't: A Tale of Two Eras, *Journal of Money, Credit and Banking* **44**(6): 1017–1032.

Trichet, J.-C. (2010). State of the Union: The Financial Crisis and the ECB's Response between 2007 and 2009, *Journal of Common Market Studies* **48**(S1): 7–19.

Tsoukalis, L. (2011). The JCMS Annual Review Lecture. The Shattering of Illusions – And What Next?, *Journal of Common Market Studies* **49**(S1): 19–44.

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

von Schweinitz, G. (2013). Flight Patterns and Yields of European Government Bonds, *IWH Discussion Paper 10/2013*.

Chapter 1

Macroeconomic Imbalances as Indicators for Debt Crises in Europe*

Abstract

European authorities and scholars published proposals on which indicators of macroeconomic imbalances might be used to uncover risks for the sustainability of public debt in the European Union. In this article the ability of four proposed sets of indicators to send early warnings of debt crises is tested using a signals approach for the study of indicators and the construction of composite indicators. It is found that a broad composite indicator has the highest predictive power. This fact still holds true if equal weights are used for the construction of the composite indicator in order to reflect the uncertainty about the origin of future crises.

* The published version of this chapter contained a coding error in one indicator (share of world trade). The resulting changes minor: the results in Tables 1.2 and 1.3, Figures 1.1 and 1.2, and Footnote 13 are changed accordingly. Further changes in the text were not necessary.

1.1 Introduction

The current debt crises in the economic and monetary union (EMU) have exposed the limitations of existing policy instruments for the prevention of excessive public debt. In particular, the focus on the ratio of a country's public deficit to its gross domestic product (GDP) in the Stability and Growth Pact (SGP) has proved to be insufficient for an accurate analysis of debt sustainability.[1] Because of these insufficiencies, the consideration of longer-term fundamental developments and risks underlying the sustainability of public debt have been discussed under the keywords 'Macroeconomic Surveillance' and 'Economic Governance', and an evaluation of macroeconomic imbalances using an indicator approach has been proposed.[2] In September 2011, the European Parliament (EP) adopted two regulations regarding the prevention and correction of macroeconomic imbalances (EP 2011). The central principle underpinning these regulations is a scoreboard that includes macroeconomic and macro-financial variables. The basic characteristics of components of the scoreboard are described, but it has been left to the discretion of the European Commission to come up with a set of indicators and a method to determine critical values. Several indicators have been proposed, for example, by the European Commission (European Commission 2010a), the European Central Bank (ECB 2010), the German Ministry of Economics and Technology (BMWi 2010) and Heise (2011).

In this article we analyze, for the first time, whether the proposed indicators are indeed able to signal possible risk for excessive public debt in EMU countries.[3] The question is explored focusing on the early-warning capability of the proposed indicators using the signals approach. This approach has been used by Kaminsky & Reinhart (1999) and Kaminsky (1999) for currency and banking crises, and by Reinhart (2002) to link currency crises to sovereign debt crises in emerging markets. Sovereign debt crises in emerging markets have been the focus of many studies – for example, by Manasse, Roubini & Schimmelpfennig

[1] For a detailed discussion of the SGP, see Heipertz & Verdun (2010).

[2] See BMWi (2010), European Commission (2010a) European Commission (2010b) European Commission (2010c) ECB (2010), EC (2010), European Council (EC) (2011). For a discussion, see Hodson (2011).

[3] We analyze the Euro-12 without Luxembourg – namely Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal and Spain.

(2003), Manasse & Roubini (2009) and Pescatori & Sy (2004). However, in our article, the signals approach is applied for the first time to signal debt crises in developed markets.

Another pioneering approach in this article is that we apply this debt crisis forecasting technique to the case of a currency union, where only five out of the 11 countries we analyzed have experienced a crisis. Our results show that a combination of all the proposed indicators could have sent strong advance signals of upcoming public debt crises.

The rest of the article is organized as follows: in the next section, we discuss how the existence of macroeconomic imbalances might lead to public debt crises. This section also presents indicators for an early-warning system taken from four different proposals. The methods of the signals approach – especially the conflation of various indicators into a composite indicator and the introduction of different quality measures – are then set out. We go on to describe the data we used and present the results, and follow up with the conclusions.

1.2 Macroeconomic Imbalances, Public Debt Crises and Indicators

Before and since the introduction of the euro as the common currency for EMU, it has been argued that a monetary union should ideally institute some form of fiscal policy coordination (von Hagen & Hammond 1998). In the absence of such coordination, excessive debt in some countries in the monetary union may cause costs to other members in the form of increased interest rates, or may even lead to higher inflation and risks for the external stability of the common currency if the central bank is not willing or able to counteract expansionary fiscal policies by members (Beetsma & Uhlig 1999). Thus, the minimum form of coordination should be the imposition of limits to public debt and deficits. In the case of EMU, member states agreed on the SGP, which stipulates that public debt should not exceed a sustainable level (60 per cent of a member country's GDP) and that government budgets should be balanced, allowing for deficits of up to 3 per cent of the GDP. The SGP has, however, demonstrably failed to cope with sudden changes in the financial sector, their real economy effects and their fiscal effects in the aftermath of the global financial crisis. Even countries that were able to fulfill all the criteria of the SGP were not immune to the turbulent times and found that they could not sustain their public finances. It became evident that reacting only once there already was a fiscal deficit that exceeded the limits of the SGP might be too late to avoid excessive debt, refinancing difficulties and default pressures. Stylized facts about the countries with such crises and the ones without led to the hypothesis that gradually evolving imbalances in macroeconomic and financial parameters – for example, private debt and foreign trade deficits – may underlie sudden changes in public finances (Eichengreen 2012).

Public institutions and scholars have therefore proposed sets of indicators to be observed to detect imbalances that might eventually result in fiscal distress. The argument is that if imbalances emerge, European institutions might request a deeper analysis of the potential vulnerabilities of economies and – if countries fail to react – might impose sanctions on non-compliant countries. Thus far, the literature does not provide an analysis of which indicators are useful as early indications of unsustainable public debt and ultimately a looming or actual public debt crisis. This article provides an analysis of the ability of different proposed indicators and sets of indicators to forecast public debt crises. In addition, we consider a broad set of all indicators.

Four different sets of indicators have been proposed: the European Commission (2010a), the ECB (2010), the BMWi (2010) and Heise (2011). The Commission set contains seven indicators: government debt, the current account balance, real effective exchange rates based on unit labor costs and the GDP deflator, increases in real house (property) prices, the net foreign asset position, and the ratio of private sector credit to the GDP. The ECB set contains ten indicators, subdivided into four main indicators and six qualitative control indicators. The main indicators are competitiveness indices based on consumer prices, on a GDP-deflator and on unit labor costs, as well as deviations from stability-oriented wage developments. The ECB's qualitative control indicators include indebtedness of the public sector, external imbalances, degree of convergence, asset prices, indebtedness of the private sector, and credit booms. The BMWi set contains five indicators: unit labor costs, the current account balance, price level developments, the unemployment rate, and private debt. Heise's set contains 15 indicators: government debt, government deficits, government interest payments as a share of government expenditure, required adjustments in primary balances due to demographic change, unit labor costs, the current account balance, export shares in global trade, domestic demand, the unemployment rate, the employment ratio, labor productivity, consumption of energy, private debt-to-GDP ratios for households and non-financial corporations, and the net international investment position.

To discuss their economic relevance, the indicators can be grouped into five categories: fiscal, competitiveness, asset prices, labor market, and private and foreign debt. For the matching of our indicators with the four sets of indicators listed above and a description of sources and transformations of the indicators, see the Appendix.

The variables in the first group – fiscal indicators – are obvious candidates to signal potential debt crises. They include measures of government debt, government deficits and interest payments as a share of government expenditure. It is proposed that extraordinarily high levels of debt, large deficits and a relatively high share of interest payments might signal a threat to the sustainability of debt. All sets, except for the BMWi one, include fiscal indicators.

The second and largest group of indicators is comprised of competitiveness indicators. The rationale for choosing these is that, in a currency union without the possibility of nominal devaluations, the emergence of disadvantages in price competitiveness cannot be easily corrected and may therefore, by the erosion of the tax base and ever-increasing spending on social security systems, increase the risk of a debt crisis. This group of indicators includes current account balances (in all sets) and trade shares (in Heise), whereas high deficits and declining shares would be interpreted as the results of low or declining competitiveness. Heise includes longer-term developments of domestic demand, arguing that relying on trade surpluses may not be an adequate long-term strategy since growth and the tax base also depend on local developments. The competitiveness indicators include constructed measures of competitiveness – namely real effective exchange rates based on different prices (consumer prices, GDP-deflators and unit labor costs). The constructed competitiveness measures are proposed by the Commission and the ECB, whereas Heise and the BMWi suggest looking directly at unit labor costs. In addition, the BMWi suggests considering consumer prices. Hence, high increases in unit labor costs or consumer prices are expected to reduce competitiveness.

The third group of indicators consists of asset prices as indicators of macroeconomic imbalances. The Commission recommends looking at house prices, whilst the ECB considers asset prices in general. Extraordinary increases in asset prices could be a valuable predictor of debt crises because they may indicate the emergence of a bubble that, if it bursts, may not only cause other forms of financial crises, such as banking crises, but could also lead to debt crises (for example, because of government bail-outs of banks, growth effects resulting from banking crises or reduced collateral).

labor market indicators constitute the fourth group of indicators. They can be found in the sets proposed by Heise and the BMWi. Both these sets use the unemployment rate as an indicator. Heise adds labor force participation and labor productivity. labor market indicators might demonstrate the flexibility of an economy to cushion shocks and to increase its longer-term growth prospects. Thus, extraordinarily high increases in unemployment, decreases in labor force participation and labor productivity may signal risks regarding government budgets.

The final group of indicators consists of private and foreign debt. All sets include a private debt indicator, but Heise uses household debt and the debt of non-financial institutions instead of an aggregate private debt indicator. We include all three. Extraordinarily high levels of private debt may indicate a risk for public debt because the government might have to bail out private debt in the case of crises, and because high levels of private debt may indicate a position in a financial cycle that is related to speculative investment (Minsky 1972). In addition to private debt, Heise and the Commission include net foreign assets. This indicator is derived from accumulated current account deficits and reflects a country's foreign indebtedness. It is therefore not only an indicator of long-term competitiveness, but also one of vulnerability to withdrawals and forced reversals in the balance of payments.

When one compares the characteristics of the different sets, one finds that Heise's is the largest, but does not cover asset prices and real exchange rates. The BMWi set is the smallest, but includes a relatively wide range of indicators, although it excludes fiscal indicators. The Commission set is small, but it covers all categories except labor. The ECB set is similar to that of the Commission, but has a stronger focus on competitiveness indices. Only two variables are included in all sets: current account balances and private debt measures. In this article we test the performance of the four proposed sets as well as that of a set containing all the proposed indicators.

1.3 Data

Although we tried to include as many variables as possible from the proposed sets in our sample, we had to exclude a few due to a lack of available data at a sufficiently high frequency (at least quarterly) or because the description was too broad. The variables that were excluded were two variables from Heise's set (the required adjustment of primary balances due to demographic change and energy consumption) and three variables from the ECB set (deviations from stability-oriented wage developments, degree of convergence and credit booms). Data on all the remaining 20 variables were available either monthly or quarterly in the period 1 January 1997–1 November 2011.[4] We drew the data from Eurostat, the ECB and Datastream (including original data from the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD), Morgan Stanley Capital International (MSCI) and the CPB Netherlands Bureau for Economic Policy Analysis). Most of the variables were transformed (for example, expressed in terms of the GDP) or expressed as year-over-year changes. For the details of the variable transformations, see the Appendix.

To analyze the explanatory power of the potential determinants of changes in the sustainability of public debt, it is necessary to define a dependent variable – in other words, what is meant by 'unsustainable public debt', 'fiscal stress' or a 'public debt crisis'. We choose the criterion of extraordinarily high default risk premiums, which can be measured as the difference between the yield of a country's bond and a proxy for a safe investment.

[4] Monthly data were obtained from quarterly data using cubic interpolation. Where necessary, we calculated seasonally adjusted data using the Berlin procedure.

A high spread of government bond yields is interpreted as a sign of serious mistrust in a government's capacity to service its debt in the future (Pescatori & Sy 2004).[5] In the rest of the article, we refer to extraordinarily high bond spreads as 'public debt crises'. Because we wanted to exclude exchange rate risks from our analysis, we use the yield of a country's government bond with a maturity of ten years, minus the average yield of government bonds of EMU countries with an AAA rating.[6] We use monthly data on the yields of government bonds of all Euro-12 countries except Luxembourg from the introduction of the euro onward (1 January 1999 for the majority, and 1 January 2001 for Greece) until 1 November 2011. We calculate the mean μ_S and standard deviation σ_S of all spreads. As in Knedlik & Scheufele (2008), a crisis occurs in a country k at time t when $S_t^k \geq \mu_S + 1.65\sigma_S \cong 275bp$. Using this definition, we note five countries with debt crises in our sample: Greece, starting in February 2010; Ireland and Portugal, starting in September 2010; and Italy and Spain, starting in August 2011. Since this definition identifies both the countries and first manifestations of debt servicing difficulties that are in line with common knowledge regarding crisis countries in the eurozone, we use these crisis dates for testing the goodness of indicator sets to signal debt crises in the eurozone.

1.4 Empirical Method

The framework used to translate the values of the different indicators of macroeconomic imbalances into an early-warning system for public debt crises is the signals approach. We use – with small adaptations – the methodology presented by Kaminsky & Reinhart (1999). We employ the signals approach because of its simplicity and its ability to detect the individual causes of the risk even when a composite indicator is used. The idea behind the signals approach is that an indicator is transformed into a binary signal indicator, which sends an early-warning signal (equal to 1) when its value exceeds a given threshold. If the indicator is below the threshold, the binary signal indicator takes the value 0. The signals of the

[5] De Grauwe & Ji (2012) also discuss the relation between yields on government bonds and their default risk.

[6] The yields are provided by Datastream.

different indicators are then combined and condensed into one composite indicator that can be translated into pseudo-probabilities for the occurrence of a crisis. This idea of the signals approach is in line with the ECB's argument, which emphasizes that a 'single criterion or at best a limited number of criteria, encompassing key aspects of competitiveness should serve as trigger values to classify countries and determine the surveillance needs' (ECB 2010, p. 9).

Table 1.1: Indicator state definition

	Crisis within the next 24 months	*No crisis within the next 24 months*
Signal issued	*A*	*B*
No signal issued	*C*	*D*

Source: Kaminsky & Reinhart (1999)

We use an early-warning horizon of 24 months to allow for policy reactions (Kaminsky & Reinhart 1999). Hence, the search for signals begins in January 1999 in Greece, and January 1997 in all other countries in the sample. Like Detragiache & Spilimbergo (2001), we exclude periods up to four years after the onset of a crisis from our analysis because during a crisis indicators behave differently from the way they behave during tranquil periods. For countries that did not experience a crisis in the observation period, we exclude the last 24 months of our sample from the calibration because we do not know whether or not the currently crisis-free countries will perhaps experience a crisis in the near future. We use the same threshold for a given indicator for all countries because we believe that countries in a monetary union should be evaluated (at least quantitatively) against a common benchmark. Thresholds are calculated from data of the whole sample.[7] We express the threshold as a quantile of the indicators' empirical distribution. The determination of the threshold is crucial for the predictive power of the indicator. The prediction of an indicator is regarded as correct if it is above the threshold before a crisis or, if it is below the threshold, it is not followed by a

[7] We conduct an in-sample analysis to gain as much information about the different crises as possible. Since all crisis episodes are concentrated at the end of the sample, a standard out-of-sample analysis is not applicable. We present results of a 'panel out-of-sample analysis' below.

crisis in the next 24 months. In line with Kaminsky & Reinhart (1999), we list four possible outcomes (states) of the indicator for every point in time, as defined in Table 1.1.

A is thus the number of months in which there was a signal followed by a crisis in the next 24 months; B corresponds to the number of periods where a signal was not followed by a crisis period; C contains all periods without a signal, despite a following crisis; and D is the number of months where the absence of a signal is followed by 24 tranquil (non-crisis) periods. States with correct signals are thus A and D, while those with erroneous (Type II error) or missing (Type I error) signals are B and C, respectively. The ratio of Type I errors to pre-crisis periods is expressed as $\frac{C}{A+C}$, while the ratio of Type II errors to tranquil periods is expressed as $\frac{B}{B+D}$. It is clear that a low threshold will result in many signals (both correct and incorrect) and may thus possibly lead to a high probability of Type II errors, while a high threshold will result in few signals, potentially missing crisis periods and thus a high Type I error probability.

The specific weights of Type I and Type II errors in a combined measure of the quality of the forecast depends on the preferences of the political decision-makers who base their actions on an early-warning system.[8] We use a utility function that is a linear combination of Type I and Type II errors to determine the optimal thresholds (Alessi & Detken 2011) – namely $U(\theta) = \min(\theta, 1-\theta) - \left(\theta \frac{C}{A+C} + (1-\theta)\frac{B}{B+D}\right)$. Because a crisis without any warning (a Type I error) is normally more costly than preemptive action that is not followed by a crisis (a Type II error), one could argue that a higher weight θ should be attributed to Type I errors. However, a decisionmaker may prefer a higher weight on Type II errors because it is problematic to justify costly preemptive actions if no crisis follows. Because of these contradictory tendencies, we work with a balanced risk aversion ($\theta = 0.5$). For every indicator, the optimal threshold is chosen as the quantile which maximizes the utility.[9]

From the discussion above, it is clear that the utility of a single indicator is an appropriate measure of the quality of that indicator for signalling purposes. We therefore use the (normalized) utilities of the different indicators (setting the weight of an indicator with a

[8] For a discussion of this topic, see, for example, Bussière & Fratzscher (2008).

[9] Quantiles are evaluated between the 50 and 95 per cent quantiles in increments of 1 per cent.

negative utility to zero) as the method to define the composite indicator as the weighted average of all binary signal indicators. However, we also consider the case of equal weights for all single indicators since the calculated utility depends on past realizations, but future crises might have another combination of causes.

In addition, we use the components of utility – namely the empirical probability of Type I and Type II errors – as quality measures. However, it is not only these measures that should be discussed, since the probability of Type I and Type II errors depends on the (non)existence of a crisis. This can be seen only *ex post facto*. However, we may be interested in the *ex ante* probability of a crisis conditional on having a signal $\left(\frac{A}{A+C}\right)$. Admittedly, this measure of *ex ante* probability may still be imprecise in some ways, because policymakers want to know what an actually observed composite indicator value means. There is probably a considerable difference between a value of 0.4 and a value of 1.0 for the composite indicator (a value of 1.0 would mean that all the components of the indicator are sending a signal of existing imbalances). Because of this, we also look at the probability of having a crisis within the next 24 months if the value of the composite indicator falls into a predefined interval, as discussed by Edison (2003), Kaminsky (1999), Brüggemann & Linne (2002).

Using the estimated probabilities and the actual realizations of a crisis, we can calculate the quadratic probability score, as described by Brier (1950), Diebold & Rudebusch (1989) and Brüggemann & Linne (2002). A lower quadratic probability score means a lower mismatch between probabilities and actual outcomes. As a last quality measure, we analyze the probability of a correctly signalled period (measured as the proportion of correct signals/no signals over all periods).

Applying the signals approach and the different measures of quality for single and composite indicators as described above, we are able to present results for the four proposed indicator sets and a broad indicator set.

1.5 Results

We present the results of our analysis in two parts. We first set out the results of the quality tests of single indicators, and then do the same for the composite indicators.

Single Indicators

The results for the single indicators are summarized in Table 1.2. The analysis relies on the performance measures described above. We can see that higher optimal thresholds tend to produce more missed signals, whereas lower thresholds tend to be found with a higher share of Type II errors. We also see that indicators with a higher utility tend to do better in terms of the other quality measures, too. Some indicators perform very well, whilst others do not yield even a positive utility and hence have no value in forecasting a crisis in our sample. Next, we discuss the performance of the indicators by looking at the five groups defined above.

The first group comprises the fiscal indicators. Government deficit (as a share of GDP) works best among the three indicators for upcoming crises and is also the best indicator in the whole set of indicators considered. Government debt (as a share of GDP) lags behind government deficit. Interest rates declined and converged during most of the time since the introduction of the euro, leading to declining shares of interest payments in almost all countries until the beginning of 2008. Because of that, interest payments as a share of total government expenditure in general send far more signals (exceed the optimal threshold) directly after the introduction of the euro than in the early-warning window before the crisis. The slightly positive utility derives from the cases of Italy and Greece, where signals are sent over the whole period.

The second group of indicators comprises competitiveness indicators. The current account indicator is a very good indicator, sending wrong signals (Type II errors) almost only in the crisis countries, whereas countries that did not experience a crisis have positive or only slightly negative balances over the entire observation period. This supports the argument that a slow build-up of imbalances can also be a warning: Greece, Italy, Portugal and Spain

Table 1.2: Quality measures for individual indicators, sorted by utility

Indicator	*Utility*	*Optimal threshold*	*Type I error*	*Type II error*	*Probability of correct crisis forecast* $\left(\frac{A}{A+C}\right)$	*Probability of correct crisis/non-crisis period forecast* $\left(\frac{A+D}{A+B+C+D}\right)$
Government deficit	0.4191	80%	0.0248	0.1370	36.65%	87.14%
Unemployment rate	0.3265	85%	0.2400	0.1069	35.45%	88.35%
Current account	0.3042	66%	0.0851	0.3064	14.85%	70.58%
Domestic demand	0.2754	88%	0.3740	0.0753	42.54%	90.03%
Non-MFI debt	0.2563	88%	0.4050	0.0825	36.55%	89.36%
Household debt	0.2478	84%	0.4050	0.0995	45.57%	86.29%
Private debt	0.2340	83%	0.3967	0.1353	26.26%	84.54%
Foreign asset	0.2010	65%	0.2810	0.3170	16.96%	68.60%
Labor-force participation	0.2007	71%	0.3415	0.2571	18.71%	73.59%
Government debt	0.1724	68%	0.3636	0.2916	16.59%	70.23%
Unit labor cost	0.1182	87%	0.6504	0.1132	19.46%	84.79%
Trade share	0.1152	59%	0.3760	0.3936	10.89%	60.77%
Labor productivity	0.0995	54%	0.3577	0.4434	11.79%	56.39%
Interest payments	0.0284	73%	0.6777	0.2655	9.11%	70.30%
ULC-Competitiveness	0.0182	91%	0.8760	0.0876	9.55%	85.76%
Inflation	0.0021	94%	0.9360	0.0598	7.62%	87.74%
GDP-Competitiveness	-0.0245	88%				
HICP-Competitiveness	-0.0324	94%				
Asset prices	-0.0324	94%				
Property prices	-0.0331	94%				

Source: Own calculations

signalled high current account deficits for nearly the entire observation period. Domestic demand is nearly as good an indicator as the current account in terms of utility. EMU countries that experienced the strongest increase in domestic demand over most of the period under review have been Ireland, Spain and Greece (in that order). However, starting with the financial crisis, the development slowed down significantly, leading to the three lowest growth rates (declines, especially for Ireland) by the end of 2010. Changes in unit labor costs and the change in the share of world trade (exports) are average indicators in terms of their utility. Unit labor costs send signals in the most volatile periods. These periods occurred mostly in Greece and Ireland at all times during the period under review. The problem with trade shares as an indicator is that there were two main periods when all countries lost a significant share of world trade: from after the introduction of the euro to 2002 and during the financial crisis, where trade as a whole declined in developed countries. It therefore seems that trade shares depend more on the general competitiveness of the eurozone in the world and less on the country-specific imbalances within the eurozone. However, the utility of these two indicators is still positive.

This does not hold true for all competitiveness indices based on different price indices. The indicators are calculated from effective exchange rates that are deflated by either consumer prices (HICP), the GDP deflator or unit labor costs (ULC), and are provided by the ECB. The development of these indicators seems to depend strongly on the development of the euro. This is especially true for the indicators that are deflated by HICP and GDP. A nominal depreciation of the euro from 1999 to 2001, and an appreciation afterwards, can be observed in the data until 2007. This dependence leads to a strong comovement of indicator development across countries. Before the current fiscal crisis, only a few signals were sent by the indicator deflated by ULC, and none by the two other indicators, leading to a negative utility for the indicators that are deflated by HICP and GDP. The final indicator in the group of competitiveness indicators is inflation. Most signals of high inflation can be seen in and after 2001, when inflation spiked due to a variety of price shocks (Gregoriou, Kontonikas & Montagnoli 2011), and only some can be seen in 2007 and 2008 in Greece. On the basis of this finding, it appears that inflation is not a good indicator.

The third group of indicators contains asset prices. We look at high asset and property price increases as signs of developing bubbles. These indicators are unable to predict the current crises within a time horizon of 24 months for two reasons. First, the highest increases were observed before and during the new economy bubble (as well as the highest declines of asset prices afterwards). Second, the second period of high increases in asset and property prices ended with the financial crisis at the start of 2008. This means that if there is a connection through different transmission channels between the burst of a bubble and a public debt crisis, it takes more than two years to unfold. Asset prices might also be a bad indicator for problems in single countries because of their high colinearity.

The fourth group of indicators consists of three measures of the labor market. We looked at the absolute year-on-year change of the unemployment rate. A strong increase in the unemployment rate leads to problems for public finances. This effect could especially be observed in Ireland and Spain, where declines in construction investment led to recessions and large increases in the unemployment rate. Thus, the unemployment rate is the second-best performing single indicator for debt crises in our sample. The labor force participation rate and labor productivity did not perform as well as the unemployment rate, but still yielded positive utility. labor productivity, which has given warning of increasing imbalances for Italy and Belgium for a long time, is less useful than labor force participation, which sends signals more or less equally distributed over all countries and times (at least up to the financial crisis) and a nearly continuous signal for France, starting in 2000.

The final group of indicators comprises private and foreign debt indicators. For almost all countries in EMU, the share of private debt to GDP has risen since the introduction of the euro. This share is particularly high in Ireland, Spain, Portugal and the Netherlands, while Greece and Italy have the lowest ratios of private debt to GDP (together with Finland). There, the ratios only increased strongly after the emergence of the crisis (mainly due to the corresponding drop in output). Warnings of possible problems generated by the transmission from private to public debt were issued much earlier in some countries (Ireland and the Netherlands, since 2004). All three private debt indicators performed well. The performance of the foreign asset indicator is also above average. This indicator represents the accumulation of the current account balances and shows the divergence in EMU caused by consistently different current accounts. Whereas the German net foreign asset position increased from 10 per cent of GDP in 1999 to 70 per cent in 2010, Italy stayed close to 0 per cent (between -1 and 5 per cent). The strongest negative development was noted for Portugal, where foreign assets valuing 10 per cent of the GDP in 1999 developed to a foreign debt of almost 30 per cent of the GDP at the start of 2007. Most countries (with the exception of Portugal) having current account surpluses started with a higher net foreign asset position in 1999. Because of that, the indicator either sends a signal over the entire period (in the crisis countries, except for Ireland) or it does not send a signal at all, thus not offering much new insight, compared to the current account balance.

As can already be noted from the discussion above, our data suggest three main causes for false signals. The burst of the new economy bubble led to a period of macroeconomic imbalances. The corresponding indicators (for example, asset prices and trade shares) normally display two blocks of time during which most signals were sent.[10] Another reason for false signals may be that some macroeconomic imbalances may require more than two years to materialize as problems with public debt, if these imbalances are not too large. A good example for this type of false signal is given by government deficits, current account balances and foreign assets. They could thus be used as 'very-early-warning indicators', while other indicators, such as inflation, sent their signals only for a very short time directly before and during a crisis. The last source of wrong signals concerns the competitiveness indicators whose development shows a lot of comovement that depends on the exchange rate of the euro. The signals of these indicators may therefore reflect the strength of the euro rather than imbalances in EMU.

Composite Indicators

Here we report the performance of the composite indicators, including the proposals of the BMWi, the Commission, the ECB and Heise, and also a broad-based indicator, which we

[10] One major difference between the new economy bubble and the asset bubbles in the current financial crisis is that the earlier has been mainly equity financed, while the later has been mainly debt financed.

also test in a version using equal weights.[11] From the analysis of individual indicators, it seems logical to conclude that the success of a composite indicator depends on the number and diversity of good individual indicators used in its construction. This becomes even more evident when one looks at quality measures of the composite indicators.

Table 1.3: Quality measures for composite indicators

Composite indicator	*EC*	*ECB*	*BMWi*	*Heise*	*Broad (utility weights)*	*Broad (equal weights)*
Utility	0.3155	0.3445	0.3589	0.4183	0.4244	0.3704
Probability of correct crisis forecast	0.1683	0.1936	0.3036	0.3750	0.4000	0.5833
Probability of a correct crisis/ non-crisis period forecast	0.6567	0.7107	0.8477	0.8826	0.8940	0.9444
Quadratic probability score	0.1054	0.1043	0.0753	0.0523	0.0552	0.0732

Source: Own calculations

Like every other indicator, a composite indicator can be evaluated in terms of the correct signals it gives of an upcoming crisis or tranquil periods. We deduce that 0.4 is the optimal threshold by optimizing the utilities over different thresholds.[12] Table 1.3 shows the summarized scores of the different sets regarding the above described quality measures. It also shows that the broad indicator from Heise (2011) is very good at all different performance measures. The correct crisis forecast measure is of particular interest to policymakers: in our sample, and given that the composite indicator exceeds the threshold of 0.4, a crisis within the next 24 months occurred with a probability of only 17 per cent in the Commission set, but with a probability of 40 per cent for the Heise composite indicator. Our broad indicator including all single indicators based on utility weights performs slightly better than Heise's indicator in most of the measures.[13] When equal weights are used for the broad indicator, we

[11] We use all the individual indicators discussed in the previous section, excluding only the competitiveness indicator based on the GDP deflator because of its similarity to the competitiveness indicator based on the CPI.

[12] Considered thresholds are between 0 and 1 in increments of 0.1 due to low variability of composite indicator values. The optimal threshold for the broad composite indicator based on equal weights would be 0.3. However, this indicator is specifically not designed to optimize past utility. In order to ensure comparability with the other indicators, we report the results for a 0.4 threshold.

[13] It might be that the good performance of these indicators is just a technical artifact. Because of that, we

would expect fewer correct signals and more incorrect ones. As expected, utility goes down, but it is still as good as the utility of the three smaller sets. Other quality measures are even better: equal weights for all indicators lead to lower values of the composite indicator, compared to utility weights. This leads to the highest probability of a correct crisis forecast at a threshold of 0.4 and the highest probability of a correct crisis/non-crisis forecast of all indicators.

Figure 1.1 contains the probability of a crisis given the specific value of an indicator. Here, we see very clearly the effect of lower values for the composite indicator based on equal weights in the left shift of the curve. In addition, we see that only the broadest indicators and the BMWi manage to predict a crisis in our sample with complete certainty if the indicator reaches a certain level. The other indicators proposed by the Commission and the ECB are too unstable because of the low number of usable indicators (they rely to a large extent on competitiveness indicators that have no weight in the utility combined composite indicator). A closer look at the composition and performance of the different composite indicators provides the following insights.

The EC's composite indicator consists of only five indicators because the competitiveness indicator (based on the CPI) and property prices have negative utility. Government and private debt, the current account balance and the net foreign asset position contribute a total of 98 per cent to the indicator, in more or less equal shares. These four indicators are good ones, capturing three out of our five categories. Until 2009, high values are only found in countries that experienced a crisis later. After that, signals from France and Austria occurred as well. Greece stands out because signals based on indicator values of around 0.7 are sent for most of the period after the introduction of the euro. This hints to the constant

evaluate the robustness of the results. We calculate the optimal threshold of indicators excluding the data of one country. Given that threshold, signals and the value of the composite indicator are then calculated for all countries. This 'panel out-of-sample' calculation, done for each of the 11 countries, leads to an average utility of 0.42. The exclusion of Italy (with a rather specific cause of crisis) leads to the lowest utility value, 0.397. In all crisis countries, the upcoming crisis is also signalled if its own data are excluded from the determination of thresholds. El-Shagi, Knedlik & von Schweinitz (2012) test the significance of in-sample results of the signals approach for various studies, including the present one, and show that there is a statistically relevant relation between the values of the composite indicators and the timing of crises.

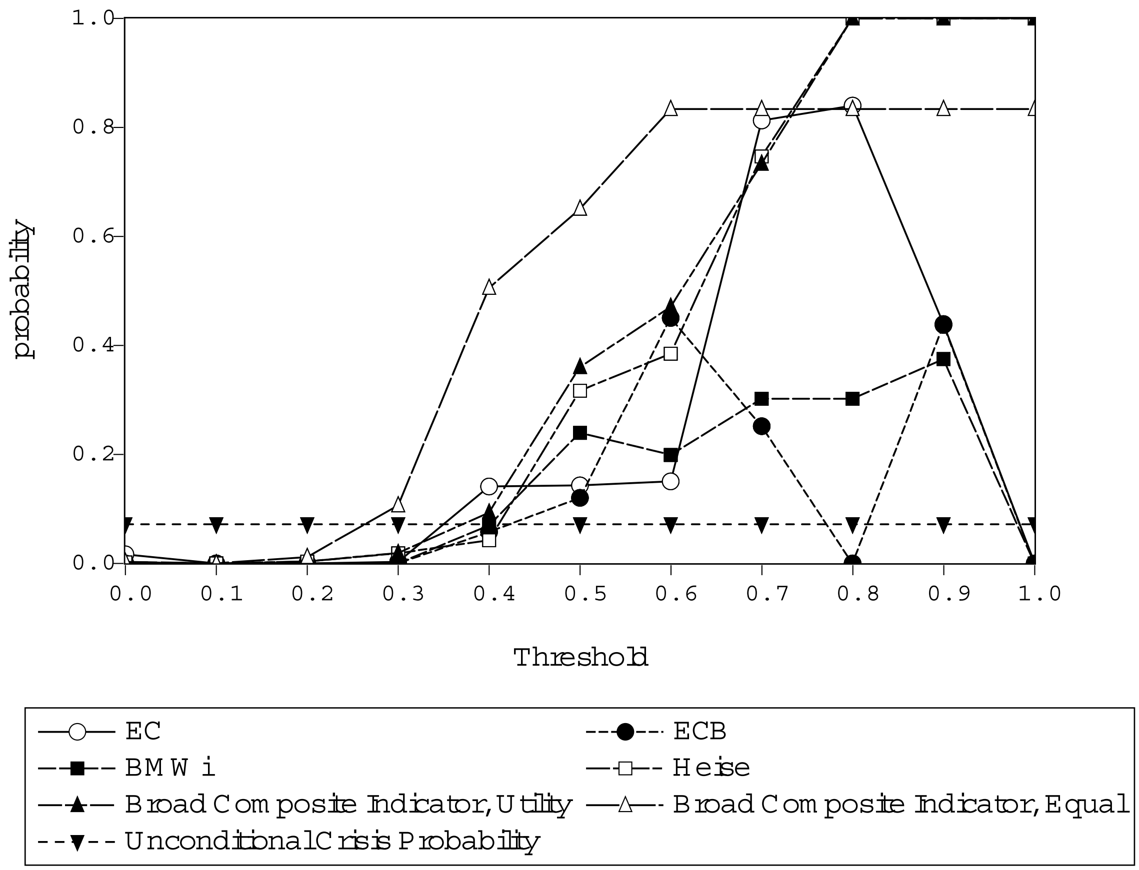


Figure 1.1: Probability of having a crisis within the next 24 months given the value of the composite indicator is in a predefined interval of width 0.1

Source: Own calculations.

imbalances in Greece as described in Featherstone (2011). Ireland is the only crisis country where the composite indicator did not exceed the threshold until 2005.

The composite indicator of the ECB consists of all three competitiveness indicators, public and private debt, the current account and asset prices. Four of these indicators have a weight of 0 or close to it. This leaves nearly equal shares for the three other indicators: 42 per cent for current accounts, 32 per cent for private debt and 24 per cent for public debt. This indicator is similar to the one proposed by the Commission (foreign assets are not included), which is also reflected in the results. The main difference from the Commission indicator can be seen in the case of Italy, where the ECB indicator stays close to but below 0.4 from 1999 to 2006, while this threshold is marginally crossed by the Commission indicator.

The composite indicator of the BMWi is fairly balanced over the different groups of indicators. We can observe low values for the composite indicator in many countries. These values only rose before and during the financial crisis. However, one main feature stands out: there is a long history of higher values (0.4 to 0.7) for Greece, Portugal and Ireland, while the values for Spain only increased in 2007 and Italy shows frequent ups and downs after 2008.

Out of the four proposed indicator sets, Heise's composite indicator is the broadest. This leads to a wide range of possible values and nearly no 0 values of the indicator. Until 2007, the values of the indicators (with the exception of Greece and Portugal) stayed under 0.4. Comparable to the indicator of the BMWi, we can observe low indicator values until mid-2001 (with persistent deviations in Greece, Portugal and possibly Italy and Spain). After a turbulent period following the burst of the new economy bubble (with most countries staying under a value of 0.3), we observe the long tranquil period until 2007. In this period, Ireland starts to build up imbalances, whereas, for example, Germany and the Netherlands reduce them. During and after the financial crisis, all countries have increasing indicator values, but the benchmark of 0.4 is only crossed persistently by Greece, Spain, Ireland, Portugal and Italy (in that order).

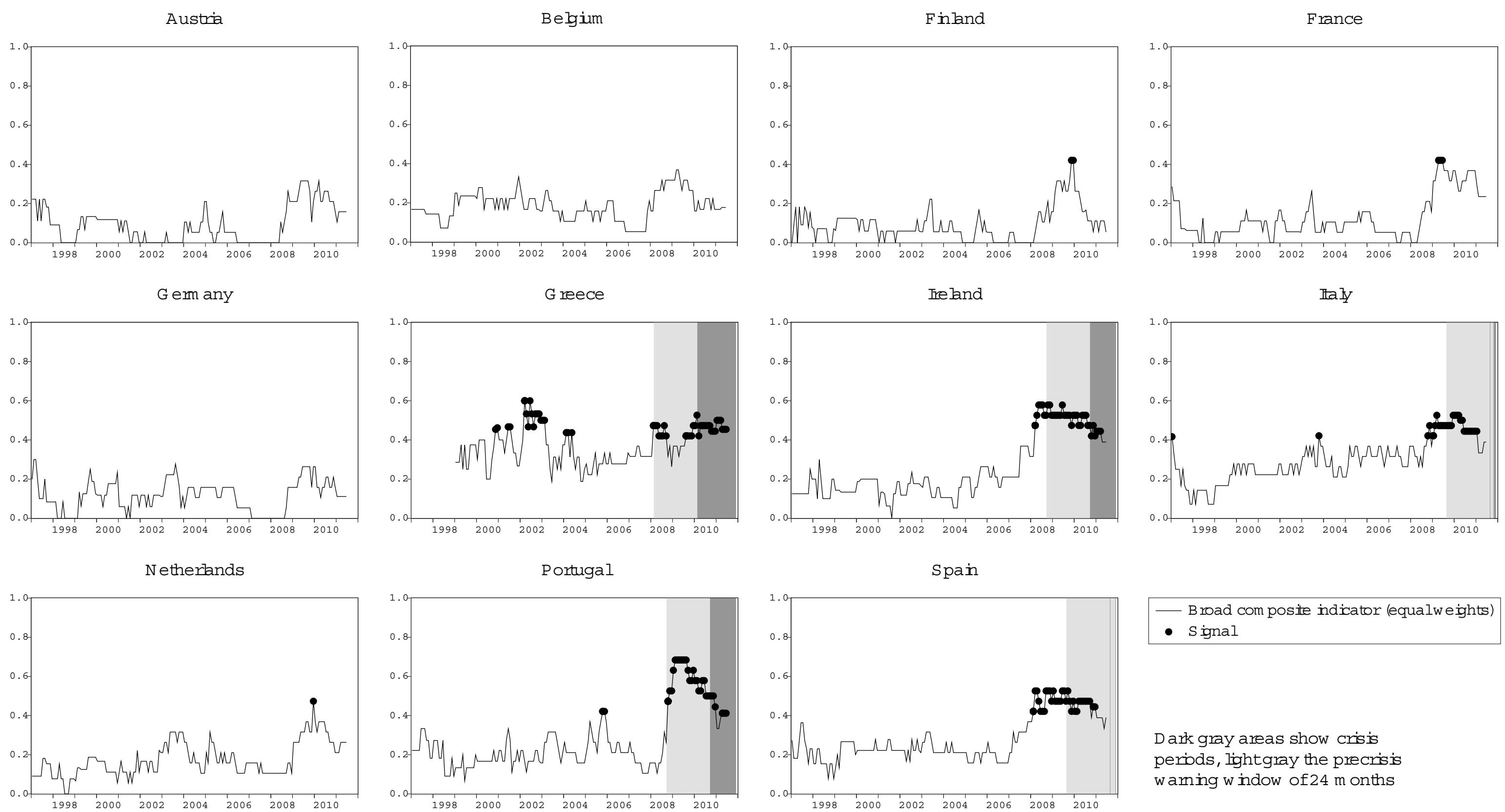


Figure 1.2: Values for our broad composite indicator, calculated using equal weights. Crisis periods and early-warning window (where the indicator is supposed to exceed 0.4) shaded.

Source: Own calculations.

Comparing our broad composite indicator with the best performing composite indicator proposed by other authors – namely Heise's indicator – we see that the broad indicator based on utility performs slightly better than Heise's on all quality measures except on the quadratic probability score due to the slightly larger set of indicators. Moving from utility weights to equal weights, the values of the indicator converge, while quality measures remain surprisingly stable or improve, as described above. This suggests both the high quality of the equally weighted indicator and the stability of the result because the broad indicator based on equal weights avoids (at least partly) the problem that its measured performance is only a past performance. Because the equally weighted indicator still obtains similar results to those obtained by its utility-weighted counterpart, we propose this indicator as a system to monitor macroeconomic imbalances in EMU. The composite indicator values for the broad indicator based on equal weights are reported in Figure 1.2.

1.6 Summary and Conclusions

We find that – consistent with economic theory – most indicators in the proposed sets were helpful in signalling the current public debt crises. For the indicators that did not provide valid signals before the current crisis, the missing signals could mostly be explained by the new-economy bubble that did not result in a public debt crisis. All other indicators showed a build-up of macroeconomic imbalances in very different areas of the economies. The cumulative imbalances led to debt crises in five crisis countries (Greece, Ireland, Italy, Portugal and Spain).

Not surprisingly, the in-sample forecast quality (as measured in terms of utility) of the composite indicators is better than that of almost all single indicators by themselves. A further improvement can be achieved by combining all the indicators of all the different sets into one single composite indicator. In doing this, we had to be careful to maintain a balance between the five different categories. The broad composite indicator outperformed (or has performed almost as well as) all other composite indicators not only in terms of utility, but also in terms of various other quality measures. This recommended indicator also reflects the desire of the ECB (2010) ideally to use only one indicator to acquire condensed information about the extent of macroeconomic imbalances in Europe, while still capturing as many different developing imbalances as possible.

In addition to the optimal construction (in terms of utility) of composite indicators, we also showed the results obtained by using equal weights. Doing so is necessary because we observed debt crises only once in only five out of 11 countries since the introduction of the euro. This means that there may as well be other sorts of imbalances leading to public debt crises in future. The theoretical rationale for equal weights is strongly supported by the high quality of the results.

Because we do not know exactly where the next crisis will start to develop and unfold, we propose to combine as many meaningful single indicators as possible into one composite indicator, using equal weights. We also suggest undertaking an in-depth quantitative and qualitative analysis of country cases if a risk is signalled. This will help to issue warnings and will allow some time to address potential problems before they culminate in fully grown crises such as the current ones.

Acknowledgments

The authors are indebted to Henry Dannenberg, Makram El-Shagi, Andrew Hughes Hallet, Oliver Holtemöller and Axel Lindner for valuable comments and suggestions.

Bibliography

Alessi, L. & Detken, C. (2011). Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity, *European Journal of Political Economy* **27**(3): 520–533.

Beetsma, R. & Uhlig, H. (1999). An Analysis of the Stability and Growth Pact, *The Economic Journal* **109**(458): 546–571.

Brüggemann, A. & Linne, T. (2002). Are the Central and Eastern European Transition Countries Still Vulnerable to a Financial Crisis? Results from the Signals Approach, *IWH Discussion Paper 157*.

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability, *Monthly Weather Review* **78**(1): 1–3.

Bundesministerium für Wirtschaft und Technologie (BMWi) (2010). Ein neues Verfahren für die wirtschaftspolitische Koordinierung in Europa, *Monatsbericht*, Schlaglichter der Wirtschaftspolitik.

Bussière, M. & Fratzscher, M. (2008). Low Probability, High Impact: Policy Making and Extreme Events, *Journal of Policy Modeling* **30**(1): 111–121.

Commission of the European Communities (European Commission) (2010a). Economic Governance Package (2): Preventing and Correcting Macroeconomic Imbalances, *MEMO/10/454*, `http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/10/454&format=HTML&aged=0&language=EN&guiLanguage=en`. 29. September 2011.

Commission of the European Communities (European Commission) (2010b). Economic Governance Package (3): Chronology and Overview of the New Framework of Surveillance and Enforcement, *Memo/10/456*, `http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/10/456&format=HTML&aged=0&language=EN&guiLanguage=en`. 29. September 2011.

Commission of the European Communities (European Commission) (2010c). EU Economic Governance: the Commission Delivers a Comprehensive Package of Legislative Measures, *Ip/10/1199*, `http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1199&format=HTML&aged=0&language=EN&guiLanguage=en`. 29. September 2011.

De Grauwe, P. & Ji, Y. (2012). Mispricing of Sovereign Risk and Macroeconomic Stability in the Eurozone, *Journal of Common Market Studies* **50**(6): 866–880.

Detragiache, E. & Spilimbergo, A. (2001). Crises and Liquidity – Evidence and Interpretation, *IMF Working Paper 02/2001*.

Diebold, F. X. & Rudebusch, G. D. (1989). Scoring the Leading Indicators, *Journal of Business* **62**(3): 369–391.

Edison, H. J. (2003). Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System, *International Journal of Finance and Economics* **8**(1): 11–53.

Eichengreen, B. (2012). European Monetary Integration with Benefit of Hindsight, *Journal of Common Market Studies* **50**(S1): 123–136.

El-Shagi, M., Knedlik, T. & von Schweinitz, G. (2012). Predicting Financial Crises: The (Statistical) Significance of the Signals Approach, *IWH Discussionpaper 3/2012*.

European Central Bank (ECB) (2010). Reinforcing Economic Governance in the Euro Area, `http://www.ecb.int/pub/pdf/other/reinforcingeconomicgovernanceintheeuroareaen.pdf`. 10. June 2011.

European Council (EC) (2010). Strengthening Economic Governance in the EU: Report of the Task Force to the European Council (Van Rompuy Task Force), `http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/117429.pdf`. 21. October 2011.

European Council (EC) (2011). State of Play concerning the Economic Governance Package, *7579/11*, `http://register.consilium.europa.eu/pdf/en/11/st07/st07579.en11.pdf`. 11. March 2012.

European Parliament (EP) (2011). Texts Adopted at the Sitting of Wednesday 28 September 2011, `http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+20110928+SIT+DOC+PDF+V0//EN&language=EN`.

Featherstone, K. (2011). The Greek Sovereign Debt Crisis and EMU: A Failing State in a Skewed Regime, *Journal of Common Market Studies* **49**(2): 193–217.

Gregoriou, A., Kontonikas, A. & Montagnoli, A. (2011). Euro Area Inflation Differentials: Unit Roots and Nonlinear Adjustment, *Journal of Common Market Studies* **49**(3): 525–540.

Heipertz, M. & Verdun, A. (2010). *Ruling Europe: The Politics of the Stability and Growth Pact*, Cambridge University Press.

Heise, M. (2011). Notwendigkeit und Ausgestaltung makroökonomischer Überwachung im Euroraum, *Wirtschaftsdienst* **91**(1): 22–30.

Hodson, D. (2011). The EU Economy: The Eurozone in 2010, *Journal of Common Market Studies* **49**(S1): 231–249.

Kaminsky, G. L. (1999). Currency and Banking Crises: The Early Warnings of Distress, *IMF Working Paper 99/178*.

Kaminsky, G. L. & Reinhart, C. M. (1999). The Twin Crises: The Causes of Banking and Balance-of-Payments Problems, *American Economic Review* **89**(3): 473–500.

Knedlik, T. & Scheufele, R. (2008). Forecasting Currency Crises: Which Methods Signaled the South African Crisis of June 2006?, *South African Journal of Economics* **76**(3): 367–383.

Manasse, P. & Roubini, N. (2009). "Rules of Thumb" for Sovereign Debt Crises, *Journal of International Economics* **78**: 192–205.

Manasse, P., Roubini, N. & Schimmelpfennig, A. (2003). Predicting Sovereign Debt Crises, *IMF Working Paper 03/221*.

Minsky, H. (1972). The Financial Instability Hypothesis, *Reprinted as Working Paper 74 (1992) by the Jerome Levy Economics Institute of Bard College*.

Pescatori, A. & Sy, A. N. R. (2004). Debt Crises and the Development of International Capital Markets, *IMF Working Paper 04/44*.

Reinhart, C. M. (2002). Default, Currency Crises, and Sovereign Credit Ratings, *World Bank Economic Review* **16**(2): 151–170.

von Hagen, J. & Hammond, G. W. (1998). Regional Insurance Against Asymmetric Shocks: An Empirical Study for the European Community, *The Manchester School* **66**(3): 331–353.

Annex: List of Individual Indicators

Cat.	*No.*	*Heise*	*BMWi*	*EC*	*ECB*	*Indicator employed (construction, source, original frequency)*
Fiscal	1	Gross government debt as a percentage of the GDP		Government debt	Indebtedness of public sector (qual.)	Government debt (gross general government debt as a percentage of the GDP, level, Eurostat, quarterly)
	2	Gross government deficit/surplus as a percentage of the GDP				Government deficit (general government net lending/net borrowing as a percentage of the GDP, level, Eurostat, quarterly)
	3	General government interest payments as a percentage of the total government expenditure				Interest payment (general government interest payments as a percentage of general government expenditure, level, Eurostat, quarterly)
	4	Required adjustment in the primary balance due to demographic ageing in percentage points				X
	5	Unit labor costs, total economy, deviation from the target path of a 1.5 percent rise per year in index points	Unit labor cost development (Lohnstückkostenentwicklung)			Unit labor costs (as a percentage change, year-on-year, Eurostat, quarterly)

Competitiveness and domestic demand	6	Current account balance as a percentage of the GDP	Current account balance (Leistungs- und Handelsbilanzsalden)	Current account balances	External imbalances (qual.)	Current account (current account balance as a percentage of the GDP, level, OECD, quarterly)
	7	Global merchandise trade shares, exports, deviation from base year 2000 as a percentage				Trade share (year-on-year percentage change of the share of exports – IMF-IFS, monthly – in world trade – CPB export volume index, world, monthly)
	8	Domestic demand, average annual change over the last five years				Domestic demand (average annual change over the last five years, Eurostat, quarterly)
	9				Competitiveness index based on HICP	HICP-Competitiveness (competitiveness index based on HICP, as a percentage change, year-on-year, ECB, monthly)
	10			Real effective exchange rate based on the GDP deflator	Competitiveness index based on the GDP-deflator	GDP-deflator-Competitiveness (competitiveness index based on the GDP-deflator, as a percentage change, year-on-year, ECB, quarterly)
	11			Real effective exchange rate based on unit labor costs	Competitiveness index based on the ULC	ULC -Competitiveness (competitiveness index based on the ULC, as a percentage change, year-on-year, ECB, quarterly)

	12				Deviations from stability-oriented sectoral/national wage developments	X
	13				Degree of convergence (qual.)	X
	14		Price level development (Preisniveau-entwicklung)			Inflation (HICP, percentage change, year-on-year, ECB, monthly)
Asset prices	15			Increases in real house prices		Property prices (residential property price index, percentage change, year-on-year, ECB, quarterly)
	16				Asset prices (qual.)	Asset prices (MSCI equity price index, percentage change, year-on-year, monthly)
	17	Harmonized unemployment rate as a percentage	Unemployment rate (Arbeitslosenquote)			Unemployment rate (harmonized unemployment rate, absolute change, year-on-year, Eurostat, monthly)
	18	Employment ratio, change over five years in percentage points				Labor-force participation (percentage change, year-on-year, Eurostat, quarterly)

Labor	19	Labor productivity per person employed, average annual change over the last five years				Labor productivity (average annual change over the last five years, Eurostat, quarterly)
	20	Gross inland consumption of energy divided by the GDP (kilogram of oil equivalent per EUR 1000)				X
Private and foreign debt	21		Private debt (Private Verschuldung)	Ratio of private sector credit to GDP	Indebtedness of private sector (qual.)	Private debt (loans to MFI and non-MFI as a percentage of the GDP, level, ECB, monthly)
	22	Debt-to-GDP ratio of households, change over five years in percentage points				Household debt (loans to households as a percentage of the GDP, level, ECB, monthly)
	23	Debt-to-GDP ratio of non-financial corporations, change over five years in percentage points				Non-MFI debt (loans to non-MFI members as a percentage of the GDP, level, ECB, monthly)
	24	Net international investment position as a percentage of the GDP		Net foreign asset position		Foreign assets (net foreign assets position as a percentage of the GDP, level, IMF-IFS, monthly)
	25				Credit booms (qual.)	X

Chapter 2

Predicting Financial Crises: the (Statistical) Significance of the Signals Approach

Abstract

The signals approach as an early-warning system has been fairly successful in detecting crises, but it has so far failed to gain popularity in the scientific community because it cannot distinguish between randomly achieved in-sample fit and true predictive power. To overcome this obstacle, we test the null hypothesis of no correlation between indicators and crisis probability in three applications of the signals approach to different crisis types. To that end, we propose bootstraps specifically tailored to the characteristics of the respective datasets. We find (1) that previous applications of the signals approach yield economically meaningful results; (2) that composite indicators aggregating information contained in individual indicators add value to the signals approach; and (3) that indicators which are found to be significant in-sample usually perform similarly well out-of-sample.

2.1 Introduction

The use of the signals approach or indicators approach to construct early warning systems for financial crises goes back to the seminal papers of Kaminsky, Lizondo & Reinhart (1998) and Kaminsky & Reinhart (1999). The signals approach is widely used to forecast financial crises. Current applications include early warning systems for currency crises (Edison 2003), banking crises (Borio & Drehmann 2009), asset price bubbles (Alessi & Detken 2011), and debt crises (Knedlik & von Schweinitz 2012). It has also been used by the International Monetary Fund and other policy makers. The approach has the advantage of simplicity and traceability, but it has been criticized for being a nonparametric approach that does not allow users to derive crisis probabilities directly and to judge the significance of individual indicators in a composite indicator, or the significance of a composite indicator as such. The analysis reported in this paper focuses on the in-sample and out-of-sample statistical significance of signals and indicators derived by means of the signals approach. In particular, we test whether the calculated maximum utilities or minimum noise-to-signal ratios of individual indicators and/or composite indicators could be the result of a random process or actually have statistical power. We use datasets from previous studies that apply the signals approach. To test the significance of the signals approach, we include studies addressing different types of crises. We use data from Kaminsky & Reinhart (1999) to cover currency and banking crises, from Alessi & Detken (2011) to cover asset price bubbles, and from Knedlik & von Schweinitz (2012) to cover sovereign debt crises. By decomposing the statistical components of the time series and bootstrapping the sample, we simulate random time series with similar characteristics as the original data. Thereby, we are able to reproduce the distribution of the performance of both individual indicators and the composite indicator under the null hypothesis of no predictive power of the considered indicators. The results show that the previous applications of the signals approach yield significant results. Most notably, composite indicators can have high (and highly significant) predictive power even when they are based on individual indicators with moderate levels of utility.

The remainder of the paper is structured as follows: the next section gives an introduction to the literature on early warning systems and relates the signals approach to this literature, describes the signals approach with the focus on the optimization methods used in the calibration of thresholds beyond which signals are issued by the indicators, and introduces the papers that we consider to test the significance of the signals approach; in Section 3, we present our methodology and testing procedure, and in Section 4 the results are set out, before recapping the conclusions in the final section.

2.2 The signals approach and alternatives

2.2.1 Early warning systems

Approaches to early warning systems vary with regard to the techniques employed. The signals approach is among the first and most widely used techniques. Two other prominent approaches are binary-choice models and the Markov-switching approach.[1]

The signals approach, introduced in the seminal paper by Kaminsky & Reinhart (1999), is a nonparametric threshold approach. An indicator is read as sending a warning signal for an (upcoming) crisis if it exceeds a threshold that is chosen to optimize the predictability of a crisis. Various indicators can be summarized to form a composite indicator (Kaminsky 1999). Applications in the literature include Berg & Pattillo (1999); Brüggemann & Linne (2002); Edison (2003).

Similar to the signals approach, binary choice models aim to predict a binary crisis variable (see e.g. Frankel & Rose 1996, Berg & Pattillo 1999, Kamin, Schindler & Samuel 2001, Kumar, Moorthy & Perraudin 2003, Bussière & Fratzscher 2006). Binary choice models fit a specific stable relation between the variation in a set of indicator variables and a latent variable that translates to crisis probability at different points in time. However, establishing a stable relation between indicators and crises requires quite rigid assumptions, including a latent variable that is both linearly dependent on the indicators and strictly monotonically related to crisis probability.

[1] See Abiad (2003) for a more detailed survey on Early-Warning Systems.

Unlike the other approaches, the Markov-switching approach does not depend on an a priori definition of crises. Instead, it assumes that there are two regimes of the economy that exhibit different characteristics with regard to the model parameters, where one regime is interpreted as a crisis state and the other as a non-crisis state. The identification of crisis periods is done simultaneously with the parameter estimation of the regimes. The probability of a regime switch can be directly interpreted as crisis probability. Early papers on regime-switching models (including the seminal paper by Hamilton (1989)) deal with time-constant crisis probabilities, whilst the applications for crisis forecasts usually employ versions with time-varying crisis probabilities (Filardo 1994, Diebold, Lee & Weinbach 1994). In some studies, the Markov-switching approach yields a better forecasting performance than binary choice models or the signals approach (Abiad 2003, Mariano, Gultekin, Ozmucur, Shabbir & Alper 2004, Kittelmann, Tirpak, Schweickert & Vinhas De Souza 2006, Knedlik & Scheufele 2008). However, since crises are not defined, the economic interpretation of the regimes is arbitrary. In addition to this drawback, applications of the Markov-switching approach can only deal with a limited number of (representative) variables. In view of the fact that in policy advice, the objective of crisis prediction tools is usually the monitoring of a broad set of macroeconomic risks, the applicability of Markov-switching models is fairly limited.

Given the limitations of the refinements in crisis prediction, the signals approach still offers substantial added value. First, it allows for strong nonlinearities in the relation between indicators and the probability of a crisis, without any additional need for highly restrictive assumptions, such as the specific functional form of the nonlinearity that is required for many nonlinear techniques, such as binary choice models.[2]

Second, the signals approach does not only allow for the identification of relevant in-

[2] The flexibility of the signals approach stems from the fact that no specific link function needs to be defined. However, this flexibility comes at a cost – the results of the signals approach cannot be interpreted directly in terms of crisis probability. It merely distinguishes between high and low risk periods. Thus, the interpretation of an individual indicator is subject to considerable uncertainty. If the risk of a crisis is only increasing gradually over a broad range of indicator values, this problem is particularly severe. Therefore the signals approach is especially apt, when there is a strong nonlinear relation between an indicator and crisis probability.

dicators, but also delivers thresholds directly where indicator values are considered to be critical. Because these thresholds can easily be (and are often) interpreted as boundaries beyond which political action is required, the signals approach is widely used in policy circles. Essentially, the signals approach is the econometric technique that is most closely related to the idea of a scoreboard, as for example used in the "Macroeconomic Imbalances Procedure" of the European Commission.

However, due to the simplicity of this approach, it is difficult to judge to what extent the in-sample predictive power of the individual indicators (and thus that of the composite indicator) is driven by chance (Berg & Pattillo 1999). Moreover, due to the rare occurrence of crises, out-of-sample analysis is limited. Thus, it is even more important to assess the in-sample performance of the signals approach properly in order to guarantee that the results reflect a relationship between indicators and crises that can be assumed to persist in future.

In the current paper, we overcome this drawback and show that the in-sample predictive power of the signals approach is actually highly significant for several types of crisis in varying samples. As far as data availability permits, we add a test of out-of-sample performance. This allows us to evaluate whether indicators that perform well and significantly in-sample are also useful in predicting crises out-of-sample. This is particularly relevant, because it has been found that many indicator systems perform poorly out-of-sample (see Rose & Spiegel 2011).

2.2.2 The signals approach

The signals approach builds on the observation that there is a risk-free range of fluctuation for most macroeconomic time series around their long-run equilibrium, but that strong deviations from that average often indicate looming crises.

Let C denote a binary crisis variable,[3] which equals one in crisis periods, and $\overline{C}$ be a corresponding matrix indicating the periods where indicators should issue a signal (warning variable), that is:

[3] We generally denote a stochastic process by X and its value taken in time t in country n by $X_{t,n}$.

$$\overline{C}_{t,n} = \begin{cases} 1, & \text{if } \exists\, k \in \{0, \dots, h\} : \; C_{t+k,n} = 1 \\ 0, & \text{else} \end{cases}, \tag{2.1}$$

where $t \in \{1, \dots, T\}$ is the time index, $n \in \{1, \dots, N\}$ the country index, and $h \in \mathbb{N}$ the early warning horizon inside which a signal should be sent. The early warning horizon is chosen by economic reasoning, taking into account the type of crisis and the time that is necessary for any action taken to have an effect on the economy. Since it is assumed that the relationship between the warning variable $\overline{C}$ and the indicator I is extremely nonlinear, the signals approach aims to identify a threshold Q_I for every indicator I where a sudden increase in crisis probability is observed.

Using the threshold,[4] every indicator is translated into a binary signal S_I:

$$S_{I,t,n} = 1_{I_{t,n} \geq Q_I}. \tag{2.2}$$

The combination of $\overline{C}$ and the signaling variable S_I results in a division of the observations into four subsets A, B, C, and D as described in Table 2.1 for every indicator (Kaminsky & Reinhart 1999).[5]

$S_{I,t,n}$ \ $\overline{C}_{t,n}$	1	0
1	A	B
0	C	D

Table 2.1: Combination of crisis and signaling variable. The letters A, B, C, and D denote both the sets themselves and the number of elements in the respective sets.

From this division, different quality measures can be derived. Most common are the noise-to-signal ratio (Kaminsky & Reinhart 1999) and a utility function based on Type I

[4] The equation describes the standard application of the signals approach where the thresholds are constant. There are, however, applications where the threshold varies over time, across countries, or both.

[5] In some applications of the signals approach, the periods immediately after a crisis – when indicator values have not yet returned to normal behavior – are excluded from the sample. In our analysis, we follow the original applications with regard to the treatment of post-crisis periods. Moreover, we generally exclude periods at the end of sample, when it is impossible to determine whether or not a signal should be issued without considering crises occurring after the sample period.

and Type II errors (Bussière & Fratzscher 2008, Alessi & Detken 2011):

$$NSR = \frac{B}{B+D} \Bigg/ \frac{A}{A+C}\,, \tag{2.3}$$

$$U(\theta) = \min(\theta, 1-\theta) - \theta\frac{C}{A+C} - (1-\theta)\frac{B}{B+D}. \tag{2.4}$$

The first papers on the signals approach use the noise-to-signal ratio (NSR) to assess the indicators' quality, but more recent papers employ utility. The advantage of utility optimization over NSR optimization is that utility optimization clearly defines a marginal rate of substitution between Type I and Type II errors. Since the preferred marginal rate of substitution remains unrestricted in NSR optimization, it sometimes produces a corner solution where only noise is minimized, almost ignoring the failure in signaling, as shown in Figure 2.10. The additional parameter $\theta \in (0, 1)$ in equation (2.4) captures decision-makers' preference for avoiding either Type I or Type II errors (Bussière & Fratzscher 2008, Alessi & Detken 2011). The threshold Q_I is set either to minimize the NSR or to maximize utility. An indicator with $NSR > 1$ is ignored, since a better NSR can be achieved by choosing the lowest possible value as a threshold and letting the indicator send a signal in every period. An indicator with $U(\theta) < 0$ is ignored for similar reasons.

While the quality measure of an individual indicator assesses how important instabilities in a particular area of the economy are, policymakers are frequently interested in the overall assessment of the current risk in the economy, which can be achieved by combining individual signals into a composite indicator $CI = \sum_{m=1}^{M} \omega_m S_{I_m}$, as proposed by Kaminsky (1999), where M refers to the number of indicators.[6]

In the literature, a number of different weighting schemes for w_m have been employed, most notably, equal weights, normalized inverse noise-to-signal ratios and utilities.[7] The resulting composite indicator can be treated (almost) as if it were a normal indicator. There is, however, one caveat. Because the composite indicator is a linear combination of only

[6] In our out-of-sample study, the composite indicator is calculated from in-sample weights combined with out-of-sample signals from individual indicators.

[7] There are also some measures that try to capture either the duration or the severity of a signal (Kaminsky 1999, Brüggemann & Linne 2002). However, this partly removes the key advantage of the signals approach, namely that it does not rely on a specific model.

few binary variables, it is discrete rather than continuous. Hence, it is often impossible to distinguish between different thresholds that are close to each other. Although an optimum threshold could be derived, it is more common to set a threshold arbitrarily close to 0.4 or to provide conditional crisis probabilities for particular ranges of the composite indicator. Correspondingly, we use a threshold of 0.4 for all applications.

2.2.3 Three contributions on the signals approach

We add an evaluation of statistical significance to three previous contributions that cover a broad range of possible applications of the signals approach. They capture both different types of crises and countries, showing that our approach leads to a general reassessment of the signals approach. Below, we briefly summarize the data used in those papers. For more information, refer to Table 2.2 in the annex.

The first paper we analyze is that by Kaminsky & Reinhart (1999), where the signals approach is first proposed and applied to a total of 76 currency and 26 banking crises in 15 developing and 5 industrial countries between 1970 and 1995. Currency crises are identified by an index of currency market turbulence, while banking crises are identified qualitatively by means of an assessment of public support of the banking sector. Their study considers only the beginning of the crises, thereby limiting the length of a crisis (as considered for prediction) to one month. According to Kaminsky & Reinhart, currency crises should be indicated starting 24 months before the crisis, up to the first month of the crisis. Banking crises should be indicated starting 12 months before the crisis and ending 12 months after the crisis. While there is limited cross-country correlation, the crises are scattered over time and regions. 16 different monthly indicators capturing monetary stress, vulnerability of the banking sector, prices and competitiveness are employed as potential indicators. The paper has been reproduced several times with wider datasets (for example by Kaminsky et al. 1998, Edison 2003, Peng & Bajona 2008). We use the dataset prolonged until June 2003, which is publicly available for Kaminsky (2006), where only currency crises are examined. This dataset contains 112 currency and 41 banking crises. Following Kaminsky (2006), we employ banking crises as an additional indicator for currency crises and extend this principle by also using currency crises as a possible indicator for banking crises in our application. The test of in-sample performance is based on the entire Kaminsky (2006) dataset. To assess out-of-sample performance we produce recursive crisis forecasts starting in January 1996 and running until June 2003. Thus, our out-of-sample period equals the period not covered by Kaminsky & Reinhart (1999)

The second paper, by Alessi & Detken (2011), applies the signals approach to costly asset price booms in 18 OECD countries. Asset price booms are identified by high growth of asset prices in at least four consecutive quarters. High- and low-cost booms are distinguished by using deviations from potential growth in the years following the boom. Using quarterly data ranging from 1970 to 2007, the authors identify 29 high-cost and 16 low-cost booms until 2002. These booms occur in three waves, and consist mainly of high-cost booms in the last two of these waves. They identify an additional wave in ten countries – the precursor of the recent financial crisis – at the end of their sample, but they do not use it in their analysis because at the time of writing, it was unclear whether the boom before the financial crisis would have to be classified as a high-cost or low-cost boom.[8] Because Alessi & Detken predict an asset price boom rather than the subsequent drop in GDP, and political action might still be taken after the boom has started, they do not only consider the beginning of the boom but also its first few quarters. Thus, they allow for the possibility that early warning signals can be issued both in the six quarters before the event (i.e. the boom) and the first three boom quarters. For the prediction of one of those quarters, 89 indicators constructed from different transformations of 18 underlying quarterly indicators are used. Because of the confidentiality of equity, housing and aggregate asset prices provided by the BIS to Alessi & Detken, we are only able to use 15 of these 18 variables. Since some transformations are based on multivariate systems, we are limited to a total of 50 instead of 89 different indicators.[9] Most of the 15 underlying indicators are available from the OECD

[8] Alessi & Detken (2011) use BIS data to classify the booms. Although the BIS asset price index includes housing, the weight of real estate was too low to enable them to recognize the real estate bubble preceding the financial crisis in the US as an asset price boom.

[9] We also do not consider their “global” indicators, as these indicators do not differentiate between single countries.

Economic Outlook and Main Economic Indicators; domestic credit can be found in the IMF's International Financial Statistics (IFS); private credit, corrected for structural breaks, was provided by the authors. As it is now apparent that the booms preceding the financial crisis can in fact be classified as "high-cost booms", we extend the analysis to cover the entire sample until 2007 for our in-sample study. The periods excluded by Alessi & Detken (Q2 2002 until Q4 2007) are used in our recursive out-of-sample study.

The third paper, by Knedlik & von Schweinitz (2012), is on the European debt crisis. It analyzes 11 countries in the European Monetary Union (EMU), from the introduction of the Euro until November 2011. A public debt crisis is defined using the spreads of government bond yields over the yield of the average AAA-rated country in the EMU. Using this criterion, a crisis is identified in five countries at the end of the observation period. Like Kaminsky & Reinhart (1999), Knedlik & von Schweinitz only consider the outbreak of a crisis, using an early warning horizon of 24 months. However, in contrast to Kaminsky & Reinhart (1999), the crises Knedlik & von Schweinitz identify are not scattered over time. A total of 20 monthly indicators out of five different categories (fiscal indicators, competitiveness and domestic demand, asset prices, labor, private and foreign debt) are used. All of them are publicly available from EuroStat, ECB, the OECD Main Economic Indicators, the IMF's IFS and Morgan Stanley Capital International (MSCI). Because the data availability requirements for our bootstrap analysis are slightly higher than those of the signals approach, we can only use 17 of the originally proposed indicators. Since the beginnings of all crises are clustered at the end of sample, it is not possible to split the sample for an out-of-sample analysis in such a way that a) the out-of-sample portion of the data is sufficiently large for meaningful analysis and b) the support period contains enough crisis periods for calibration. Therefore, following Knedlik & von Schweinitz (2012), we exploit the cross-country dimension for out-of-sample analysis instead.

2.3 Dealing with uncertainty in the signals approach

The quality of the signals approach is hard to judge from an econometric perspective. Because thresholds are adjusted to create optimal fit of individual indicators, even randomly chosen indicators that are not causally related to the crisis probability may produce considerable in-sample fit by chance. In some cases this in-sample fit may translate to out-of-sample performance that does not in fact reflect any causal impact by the indicator. If both crisis risk and some indicators contain a substantial global component, the signals approach will (inter alia) select indicators where the cycle of the global component resembles the cycle of global crisis risk. Since current risk affects future risk and the current level of such an indicator affects its future level, in-sample correlation thus implies some degree of out-of-sample correlation, even if the in-sample fit is driven by chance. However, the spurious predictive power of these indicators is based on a prediction of the global component of crisis risk. Thus, they would not be helpful in the identification of endangered countries. Moreover, it is impossible to interpret signals derived from these indicators since there is merely a statistical correlation and no economic causality. Therefore, this kind of predictive power can hardly be exploited in policy making.

Whether or not a set of indicators performs well albeit not driving crisis probability, depends strongly on the data-generating process underlying the chosen indicators. For example, it is unlikely that a white noise process or (more generally) a strongly mean reverting process produces high utility by chance, since the signals approach usually requires a good indicator to send a signal for a number of consecutive periods. To account for the data structure, we thus propose a bootstrap to evaluate the distribution of the performance under the null hypothesis that the indicators do not truly explain crises. To do so, we can either resample the indicators and match them with the original crisis data, or resample the occurrence of crises, matching the counterfactual crises with the original indicator data. Depending on the structure of the indicator data and the crisis, one of these approaches might be more feasible than the other.

Because Kaminsky & Reinhart (1999) mostly cover emerging markets that display highly volatile macroeconomic indicators, it is hard to find a statistical model that can reproduce the original indicator structure. Thus, it seems more appropriate to resample the occurrence of crises. To capture the cross-country correlations and dynamics of crisis propagation, we propose an adapted moving block bootstrap (Künsch 1989, Fitzenberger 1998) for this purpose, as described in Subsection 2.3.1.

Alessi & Detken (2011), who cover an OECD sample, use indicators that are more stable than those used by Kaminsky & Reinhart, but their sample includes a high number of indicators that are missing for extended periods at the beginning and end of the sample in some countries. This makes the estimation of cross-country and cross-indicator correlations unfeasible. Thus, we again propose to resort to resampling the crises. However, this requires a different approach from that indicated above, since crises continue over several periods. A moving block bootstrap would produce more crises that begin or end in the same period in a large number of countries than found in the true data, thereby causing inappropriately high cross-country correlation. Instead, we develop a conditional probability bootstrap, as described in Subsection 2.3.1.

Opposed to the applications described above, Knedlik & von Schweinitz (2012) use a sample where crisis resampling is inappropriate. Their data only covers a single block of crises close to the end of the sample. However, because they only consider highly developed members of the European Union, most data are easily available for the entire sample and the data-generating process of the indicators is relatively stable. We are thus able to resample the indicators rather than the crises, again providing counterfactual simulations where indicators and crises are independent, while still retaining similar statistical features concerning cross-country correlations and dynamics. We use the econometric model described in Subsection 2.3.2.[10]

We always use 10 000 bootstrap samples of either crisis or indicator data that are subsequently used with the remaining original data (i.e. either original indicators or crises) in

[10] Since we are resampling indicators, one could argue that we do not capture the true model (in terms of parameter estimates and included variables). However, for our purpose, this is of limited importance. We are more interested in the true statistical properties when we want to test our null hypothesis of no correlation between crises and indicators.

the signals approach. Both the thresholds of individual indicators and the weights of the individual indicators in the composite indicator are determined separately for each counterfactual sample, producing a set of 10 000 quality measures (utility or noise to signal ratio) under the null hypothesis. This results in a simulated probability distribution for the quality measures that is then used to assess the significance of the signals approach.

2.3.1 Resampling crises

Our resampling aims to preserve the unconditional crisis probability in the whole sample, and the crisis probability conditional on current and recent crises in all countries. Our first resampling method achieves this objective by roughly following a conventional moving blocks bootstrap. However, since we do not want to carry over country-specific crisis probabilities, we add a second layer of resampling. The second resampling method relies on a set of conditional probabilities obtained from the true data to resample counterfactual crises with similar statistical properties. Both the in-sample and out-of-sample evaluations by Kaminsky & Reinhart (1999) and Alessi & Detken (2011) are based on those methods.

In the out-of-sample studies, we update the signals recursively. That is, the thresholds are recalibrated in every out-of-sample period based on data available until the previous period. When computing significance, we face the problem that a signal does not predict an event at a specific time but within the early warning horizon h. However, since the spurious out-of-sample fit is mostly due to the persistence of indicators and crisis probability, it makes a considerable difference whether a signal is judged as a one-period-ahead forecast or as an h-period-ahead forecast. Thus, we simulate $h+1$ periods of counterfactual crises and derive the need for a signal one period ahead (in other words, if a crisis is simulated in a country in the next $h+1$ periods, a signal is already needed in the next period).

A moving block bootstrap

Let us recall our binary crisis $T \times N$ matrix C where

$$C_{t,n} = \begin{cases} 0, & \text{if there is no crisis in country n in period t} \\ 1, & \text{if there is a crisis in country n in period t} \end{cases}. \tag{2.5}$$

We generate $T + 1 - b$ blocks B of consecutive rows of blocklength b from our matrix C, resulting in a sequence of blocks defined by:

$$B_t = \begin{bmatrix} C_{t,1} & \dots & C_{t,N} \\ \vdots & \ddots & \\ C_{t+b-1,1} & & C_{t+b-1,N} \end{bmatrix}. \tag{2.6}$$

Resampling is achieved by randomly drawing a number of blocks (with replacement) that is sufficient[11] to create a new counterfactual version of matrix C, denoted by $\hat{C}$. We then add a second layer of resampling, drawing N columns from $\hat{C}$ (with replacement), to create a final bootstrap crisis sample C^*.

In the application to currency and banking crises, we choose a blocklength b of 12 months.

In the out-of-sample study, a link between the last observed periods and the bootstrapped crises is required. While it is well established, that a moving block bootstrap maintains the correlation properties of the original data on average, the true correlation over time is not enforced between observations at the end of one block and the beginning of the next. Therefore, to capture the information on future crisis probability contained in the current occurrence of crises, we have to adjust the selection mechanism for the first block of the out-of-sample bootstrap. To do this, we use a randomly drawn overlap range b_o (between 1 and 6 periods). We limit the selection of eligible blocks to those where the number of crises in the first b_o periods is not significantly different from the number of crises in the last b_o observed periods. Instead of adding a complete block of length b, we only add periods $b_o + 1$ to b of the selected block. We proceed as in-sample for the subsequent blocks.

[11] To be precise, $\lceil T/b \rceil$ blocks are needed. Surplus periods are cut.

The conditional probability bootstrap

Starting from our binary crisis matrix C, as defined in equation (2.5), we can compute the probability of a crisis in any country at time t, conditional on the occurrence of a crisis in this country at time $t-1$, and the number of countries being in the state of crisis in period $t-1$. Since we are more interested in the statistical properties of our resampled matrix C, rather than the question of which countries play a specific role in the transmission of crises, we assume that these probabilities are identical across countries.[12] That is, we have a set of $2(N+1)$ probabilities defined by:

$$p_{\chi,n} = p(C_{t,i} = 1 | C_{t-1,i} = \chi \wedge \sum_{j=1}^{N} C_{t-1,j} = n), \tag{2.7}$$

where χ is 0 or 1 and n is an integer between 0 and N, capturing the number of countries in the state of crisis in the previous period. If $p_{\chi,n}$ cannot be defined from the sample for a certain n, we replace $p_{\chi,n}$ by $p_{\chi,\tilde{n}}$, $\tilde{n}$ being the number of crises as close to n as possible, where a probability $p_{\chi,\tilde{n}}$ can be computed from the sample. Situations where this replacement is necessary occur if the simulation produces a number of crises in a single period that has never been observed in the data.

Each repetition of the bootstrap is initialized using a random row from C as the first period of the resampled crisis matrix. Starting from this counterfactual crisis situation in the first period, the following periods are randomly generated, using the probabilities defined in equation (2.7).[13] As above, this results in a bootstrap crisis sample C^*.

The conditional probability bootstrap is particularly adequate for out-of-sample analy-

[12] This assumption should capture the main dynamics if countries are highly integrated. If, however, bilateral financial or economic ties between countries prevail, we observe repetitive patterns of certain country groups that experience crises jointly. In these situations, our simplification might prove unfeasible.

[13] Because the sample used by Alessi & Detken (2011) does not include costly asset price booms of less than four periods, we adjust the simulation slightly to produce corresponding results. If a counterfactual boom (labeled a crisis in the technical description for the sake of consistency) would end after one period, the boom is immediately removed from the sample. Booms that continue over three periods are prolonged to cover a fourth period without randomization. Booms that continue for two periods and are about to end from the random drawing, are either removed from the sample (with probability q) or extended to the following two periods (with a probability of $1-q$), thereby overwriting the original random draw. The probability q is calibrated to produce an unconditional probability for boom periods that matches the true data.

sis. Using the current period for initialization, the counterfactual out-of-sample periods are simulated just as in the in-sample case.

2.3.2 Resampling the indicators – the panel VAR bootstrap

We assume that every indicator I_m, $m \in \{1, \dots, M\}$ has a global and a national component. Both follow an individual VAR process.

Roughly following the model structure that underlies the panel unit root test accounting for cross sectional dependencies proposed by Bai & Ng (2005), we define

$$f_t = \left[\begin{array}{cccc} f_{1,t} & f_{2,t} & \dots & f_{M,t} \end{array} \right], \tag{2.8}$$

where $f_{m,t}$ is the first principal component of the matrix of indicator m:

$$I_m = \left[\begin{array}{ccc} I_{m,1,1} & \dots & I_{m,1,N} \\ \vdots & \ddots & \\ I_{m,T,1} & & I_{m,T,N} \end{array} \right] \tag{2.9}$$

at time t.

Defining a vector of values of all indicators at a given time t in country n,

$$I_{t,n} = [\begin{array}{cccc} I_{1,t,n} & I_{2,t,n} & \dots & I_{M,t,n} \end{array}], \tag{2.10}$$

we can then write:

$$I_{t,n} = f_t P_n + \eta_{t,n}, \tag{2.11}$$

where P_n is a diagonal matrix produced from the individual inverse loading matrices of the principal components decompositions and $\eta_{t,n}$ is the vector of national contributions to the indicator vector.

Both the global and individual components are assumed to follow a VAR process, i.e.

$$f_t = L(A) f_t + v_{t,n} \tag{2.12}$$

$$\eta_{t,n} = L(B)\eta_{t,n} + c_n + \varepsilon_{t,n}, \tag{2.13}$$

where c_n captures national effects; $v_{t,n}$ and $\varepsilon_{t,n}$ are the respective errors.

The national processes are estimated using a panel VAR. Since $T > 30$ for all examples, we can forgo GMM estimation and employ OLS, as shown by Maddala & Wu (1999). Their conclusion, originally for the application in single equations, can be directly applied to this VAR because the VAR can be estimated blockwise.

Lag orders of the global VAR and the national level panel VAR are determined separately using the Schwartz criterion. The residuals used when resampling are drawn from all the residuals, because we want to exclude cross-country heteroscedasticity. We simulate $T + 100$ periods of data in each iteration, thereby allowing to discard the first 100 periods to eliminate the impact of the starting values. The principal components analysis is based on standardized data. We reverse the standardization when computing the counterfactual series in the bootstrap, i.e. each simulated series has roughly the same mean and standard deviation as those found in the true time series of that country. This results in a bootstrap indicator sample I^*, containing all indicators, which is used together with the original crisis matrix C in order to test our null hypothesis. Since the cross-sectional heterogeneity is reproduced in resampling, the null hypothesis for an indicator is only rejected if the dynamic behavior of that indicator is related to crises. If, however, persistent level differences between crisis countries and stable countries result in high utility of an indicator, the bootstrapped levels of this indicator's utility are similarly high and the null hypothesis is not rejected.

The shortest subsample in this exercise that allows calibration ends in February 2010 when the first crisis (Greece) started. However, this split is not feasible for an out-of-sample analysis, since we do not know whether or not we would like to have signals in the six countries that did not experience a crisis in the remainder of the sample (see also Footnote 5). That is, in the complete subsample that would be available for out-of-sample analysis over the given time frame, we do not know the observations where no signal should be issued and thus cannot compute B and D (see Table 2.1.) Therefore, we follow Knedlik & von Schweinitz (2012) and perform the out-of-sample analysis along the cross-section dimension of the panel. That is, the signals for each country are obtained using thresholds based on the ten other countries. To assess the significance of those results, we treat the global component as known when simulating counterfactual distributions of the data. As with the in-sample test, we reverse the standardization. This implies a very restrictive null hypothesis, because persistent country effects are treated as known.

2.3.3 Evaluation techniques

The above described methodology of bootstrapping crisis dates and indicators allows us to test the significance of the signals approach. To do this we rely on three criteria. First, we test the null hypothesis that the utility or noise-to-signal ratio for individual indicators could have been achieved by chance. Second, we employ a Fisher test (based on bootstrapped distribution functions)[14] to test the null hypothesis that, in a set of indicators, the utility or noise-to-signal ratio of all the individual indicators could have been the result of a random process. Third, we test for the null hypothesis that a composite indicator (based on the noise-to-signal ratio or utility optimization) of the whole set of indicators in one application could have been the result of a random process.

2.4 Results

The discussion of the results is divided into four subsections. In the first three, we reevaluate the economic findings of Kaminsky & Reinhart (1999), Alessi & Detken (2011) and Knedlik & von Schweinitz (2012) and assess the statistical significance of their results, concentrating on specific findings. Against this background, we then present general results on the application of the signals approach in the fourth subsection. The results of the different tests for four applications are presented in Tables 2.3 to 2.9 in the annex.[15]

[14] The original test (Fisher 1932) has been rediscovered by Maddala & Wu (1999), who proposed the application to panel unit-root testing.

[15] The utilities are normally calculated for $\theta = 0.5$ and – for the composite indicators – with a threshold of 0.4. Only for the results of Alessi & Detken (2011), we choose $\theta = 0.4$ in line with these authors. The row of the "Fisher test statistic" contains the value of the Fisher test and its p-value for the two optimization methods.

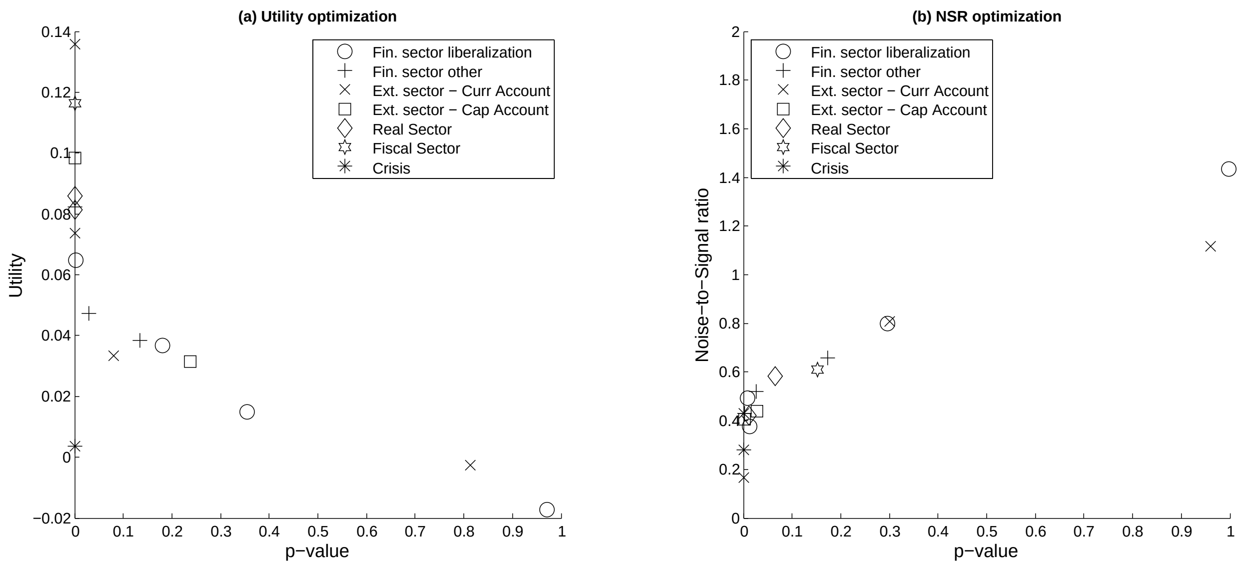
2.4.1 Currency and banking crises (Kaminksy and Reinhart, 1999)

Currency crises Our in-sample results concerning the application of the signals approach to currency and banking crises are shown in Tables 2.3 and 2.5 in the annex. Unlike Kaminsky & Reinhart (1999), who use NSR optimization in their paper, we show results for both utility and NSR optimization.

Our findings mostly confirm the results reported by Kaminsky & Reinhart (1999). The seven best performing indicators in the application to currency crises found by Kaminsky & Reinhart (1999) are the real exchange rate, stock prices, exports, output, the deficit (% GDP), M2 (% reserves) and reserves. The same indicators are also among the best and most significant when using utility optimization. However, when we apply NSR optimization as Kaminsky & Reinhart do, we find neither deficit nor output to be significant at the five percent level. Deficit and output also have a considerably higher NSR than in the original paper. However, other indicators improve substantially compared to the original paper, for example, real interest rates and the real interest rate differential are highly significant, both displaying a substantially lower NSR than in the original sample, which ends in 1995.[16]

As noted above, we also include banking crises as a possible indicator. Although the utility of banking crisis is close to zero, it is highly significant, as can also be seen from the outlier in Figure 2.1(a). This low utility is mostly due to the rare occurrence of banking crises in the sample. If there were as many banking crises as currency crises, the ratio of potential signals to required signals (24 periods before the crisis and the crisis itself) would be 1/25. That is, even if there was no noise at all, but every upcoming currency crisis was announced by a banking crisis, the utility would still only be $0.02 = 0.5 - 0.5\left(\frac{24}{25} + 0\right)$. Even though banking crises have low utility as an indicator, this result (especially the significance of the indicator) can be seen as a confirmation of the findings of Kaminsky & Reinhart, who argue that banking crises frequently cause currency crises. Nevertheless, banking crises would only play a negligible role in a composite indicator using utility optimization.

[16] The difference in results between the original paper and our recreation is only partly due to our extension of the sample. Even when we apply the signals approach to the original sample (ending in 1995), we get slightly different results, probably due to revisions included in the data.



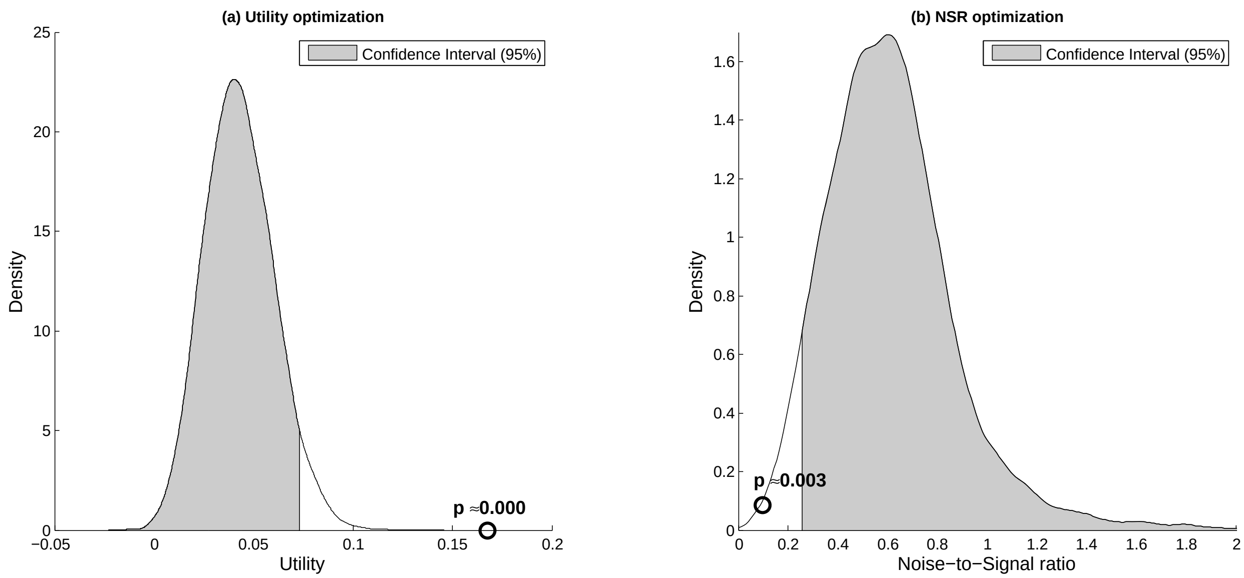
Note: The 16 indicators are grouped according to Kaminsky & Reinhart.

Figure 2.1: Significance and quality measure of individual indicators in the currency crises application

Since utility optimization seems to be superior to NSR optimization as described in Section 2.2.2, it might be a promising approach to include the significance of indicators in the weighting scheme of a composite indicator in order to overcome the abovementioned problem. In addition, a moving average of indicators might be used in the composite indicator in order to compensate for the very short duration of signals sent by the banking crisis indicator.

Even without this possible improvement, the composite indicator obtained by both optimization schemes is better (in terms of the same measure as used in the scheme) than all the individual indicators, and is highly significant, as shown in Figure 2.2. We see that the same does not necessarily hold true when looking at the respective other measure in Table 2.3: the utility of the composite indicator obtained by NSR optimization is both low and insignificant. This is explained by the high thresholds of individual indicators' leading to low signal ratios (and correspondingly high Type I error probabilities).[17]

[17] See also Figure 2.10(b) in the annex.



Note: The density is derived under the null hypothesis, the circle indicates the respective quality measure of the original sample.

Figure 2.2: Density of utility and NSR in the currency crises application

The out-of-sample utility of the utility-weighted composite indicator is substantially lower than its in-sample utility. However, due to low cross-sectional and cross-time correlation in the occurrence of crises, the out-of-sample utility under the null hypothesis is very close to zero in this application. Thus, the utility of the composite indicator (0.041) is still highly significant (see Table 2.4). Six of the individual indicators are significant at the 1% level. Four of them provide a higher utility than the composite indicator (see Figure 2.11). This is partly explained by our short out-of-sample period. Whereas Kaminsky & Reinhart (1999) explain a variety of types of currency crises in the in-sample analysis, our out-of-sample period is dominated by one wave of crises in emerging markets. Thus, the performance of the indicators that play a particularly strong role in that wave does not necessarily imply the general superiority of those indicators. Similarly, the weak performance of some other indicators may be explained by the short sample, rather than disproving the economic theory behind indicator selection.

In the case of NSR-optimization, the out-of-sample composite indicator has a lower NSR than the in-sample composite indicators. However, this does not reflect good performance in this case, as the composite indicator strongly underestimates crisis risk, issuing only eight signals (both true or false) in the whole out-of-sample period. These are distributed over three out of the total of 24 crises. While this corresponds to a low signal ratio, noise is also almost reduced to zero, causing the low NSR (see also Figure 2.10.)

Banking crises In the utility optimization in the case of banking crises, the three indicators that perform best and are the most significant in-sample are the deficit (% GDP), the real interest rate and the real interest rate differential. In the NSR optimization, however, the best indicator from the original sample period (real exchange rate) performs comparably poorly, while the real interest rate and the real interest rate differential become the two indicators that perform best when the full sample (ending in 2003) is used. This is partly explained by their good performance in the Asian crisis, which is only included for out-of-sample analysis by Kaminsky & Reinhart (1999).

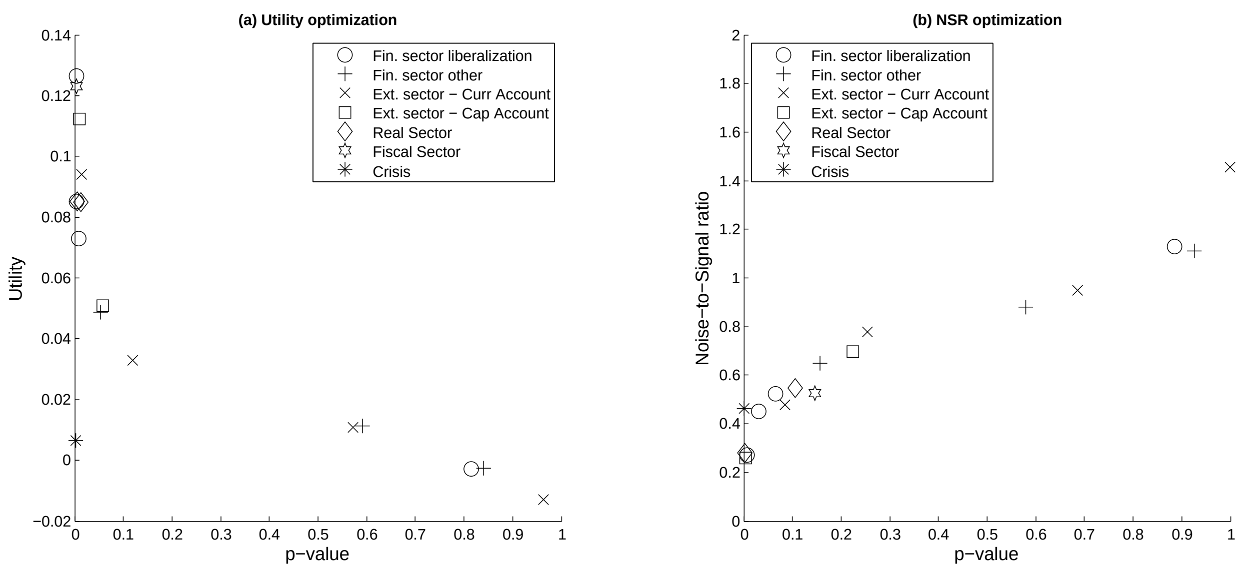
Currency crises seem to "predict" banking crises about as well as banking crises predict currency crises. However, this is mostly due to the fact that the "early warning" horizon continues until 12 months after the crisis breakout in the banking crisis application.[18] Generally, we find a substantially different distribution of p-values of the considered indicators when we look at utility and NSR optimization respectively. Dividing the indicators into groups, "good" (p-value smaller than 5%), "moderate" (p-value between 5% and 20%) and "bad" (p-value above 20%), the utility optimization results in a large group of good (9/17) and a smaller one of "bad" indicators (5/17), while only M2, reserves and exports belong to the "moderate" group (see Figure 2.3(a)). By contrast, the NSR optimization yields three groups of nearly equal size. This, combined with the fact that the indicators in the "moderate" group have only average NSR and a combined weight of 31% in the composite indicator, leads to the insignificance of the composite indicator using NSR optimization (and a worse NSR than that of its best components) as presented in Figure 2.4(b). The same result holds

[18] While 24 banking crises announce upcoming currency crises, only ten currency crises actually precede banking crises within a 24 month window. However, 21 currency crises occur in a 25 month window centered at the outbreak of a banking crisis.

for the utility of the composite indicator in the NSR optimization. Real interest rates, real interest rate differentials and stock prices produce better NSRs than the composite indicator does when forecasting banking crises. However, the utility of the utility-optimized composite indicator is better than that of the best individual indicator by itself.

This is no longer true in the out-of-sample test, where three (one) individual indicators in the case of utility (NSR) optimization outperform the composite indicator (see Table 2.6 and Figure 2.12). Nevertheless, the out-of-sample performance of the utility-optimized composite indicator is much closer to the in-sample performance compared to the case of currency crises.[19]

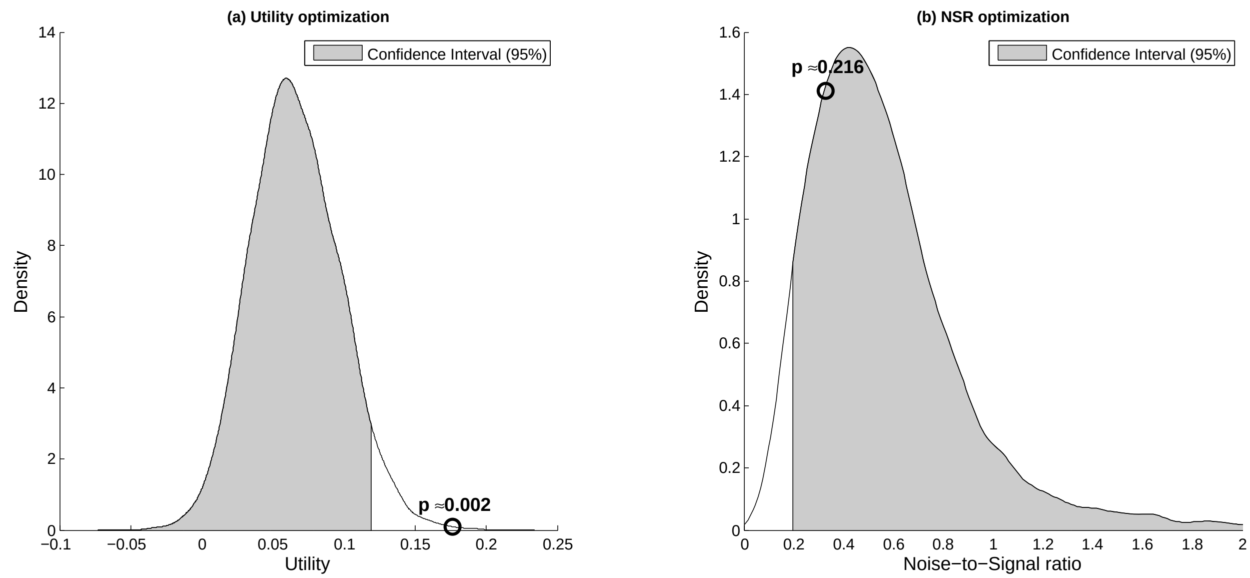
Given the large number of individual indicators in the applications to currency and banking crises that are highly significant both in-sample and out-of-sample, it is hardly surprising that the Fisher test strongly rejects the null hypothesis that no indicators are truly correlated to crises.



Note: The 16 indicators are grouped according to Kaminsky & Reinhart.

Figure 2.3: Significance and quality measure of individual indicators in the banking crisis application

[19] As in the case of currency crises, the out-of-sample composite indicator obtained from NSR optimization has both low noise and low signal ratios.



Note: The density is derived under the null hypothesis, the circle indicates the respective quality measure of the original sample.

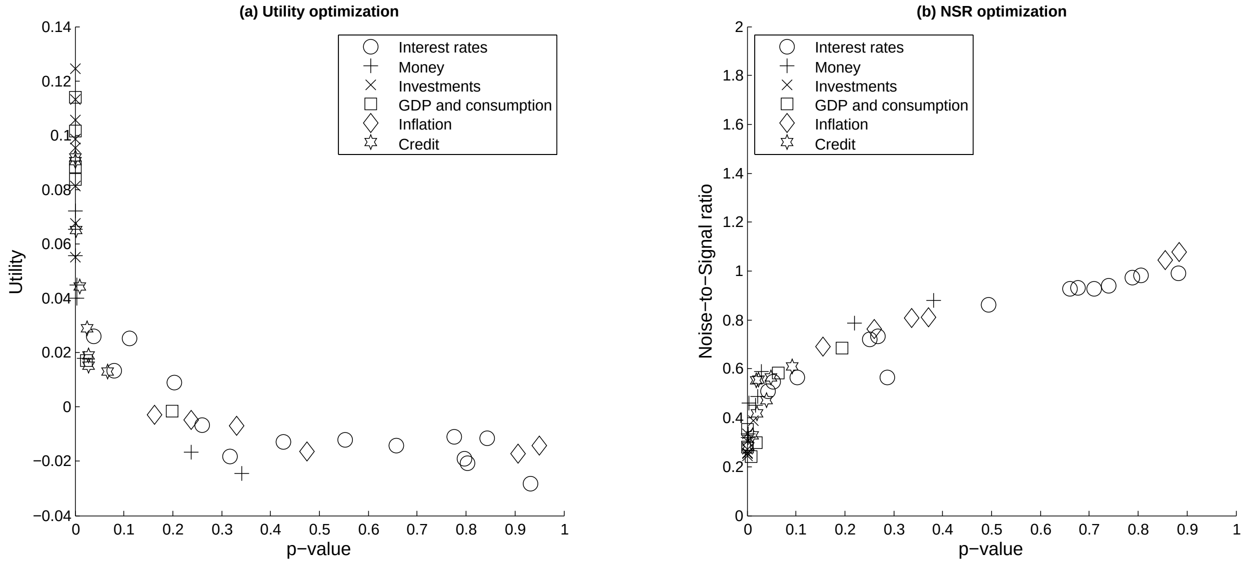
Figure 2.4: Density of utility and NSR in the banking crisis application

2.4.2 Costly asset price booms (Alessi and Detken, 2011)

One of the most interesting features of the study by Alessi & Detken (2011) is that the authors do not select one preferred transformation for each macroeconomic variable, but use every variable in a number of different transformations, producing groups of highly related indicators. In-sample some groups of indicators have near-equal performance (e.g. investment and consumption), while others display a large variety of performance measures and p-values (e.g. money indicators such as M3 or interest rate variables), as can be seen in Figure 2.5. As in the original paper, we find a large number of indicators that perform poorly. However, their low utility (and high NSR) is partly due to the long early warning horizon before the contraction after a boom. This also has an effect on the significance of the individual indicators: using the classification of significance introduced in Section 2.4.1, 18 out of 50 indicators are “bad” for both optimization rules, while five (six in the NSR case) of them are “moderate”.

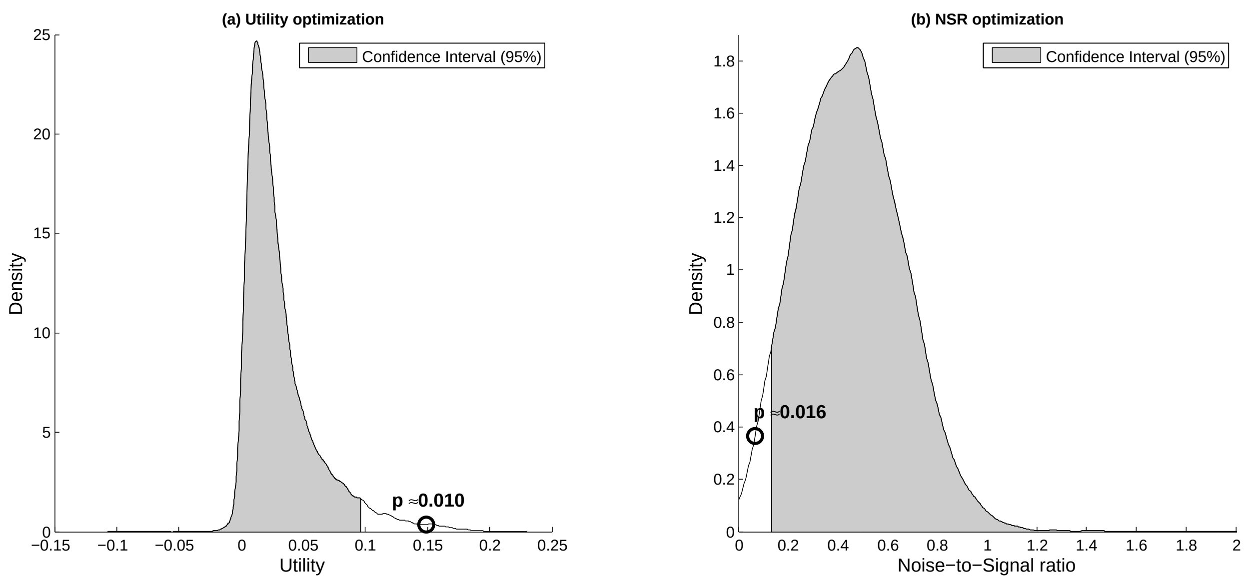
Using different transformations of a single variable increases the number of indicators, and hence the number of significant indicators, substantially, but it adds little to the significance of the whole set in the Fisher test and to the significance of the composite indicators. Given that one third of all indicators (17/50) are significant at the 0.1% level, the p-values of the Fisher test (enhanced to account for cross-series correlation) and the composite indicators (see Figure 2.6) are surprisingly high. This is due to two characteristics of the indicator set. In the large group of insignificant indicators, many have a utility below zero (19 indicators) or an NSR near or above one (13 indicators have an NSR above 0.8). Thus, the composite indicator is constructed from relatively few individual indicators: Transformations of housing investment, investment, consumption and the real GDP (that is, four out of 18 underlying variables) have a share of 61% and 44% in the utility and NSR-optimized composite indicator, respectively.

In the case study by Alessi & Detken (2011), both individual and composite indicators have higher and more significant utility out-of-sample than in-sample (see Tables 2.7 and 2.8). This is partly due to the long boom before the recent financial crisis, that dominates the out-of-sample period. The share of periods in which a signal is required almost triples compared to the whole sample. Therefore, noise is strongly reduced. Since the utility function of Alessi & Detken (2011) penalizes noise more severely than missing signals, this leads to a general improvement of indicator utility. The long boom is also reflected by our bootstrap, which captures the time persistence and joint occurrence of booms, leading to a rightshift of the utility distribution under the null. However, the null hypothesis is strongly rejected, with a p-value substantially lower than the in-sample p-value, see Figure 2.13. As in the other out-of-sample cases, NSR optimization produces a corner solution, in which booms are only correctly identified in two out of ten countries.



Note: The 50 indicators have been grouped to enhance readability.

Figure 2.5: Significance and quality measure of individual indicators in the costly asset price boom application



Note: The density is derived under the null hypothesis, the circle indicates the respective quality measure of the original sample.

Figure 2.6: Density of utility and NSR in the costly asset price boom application

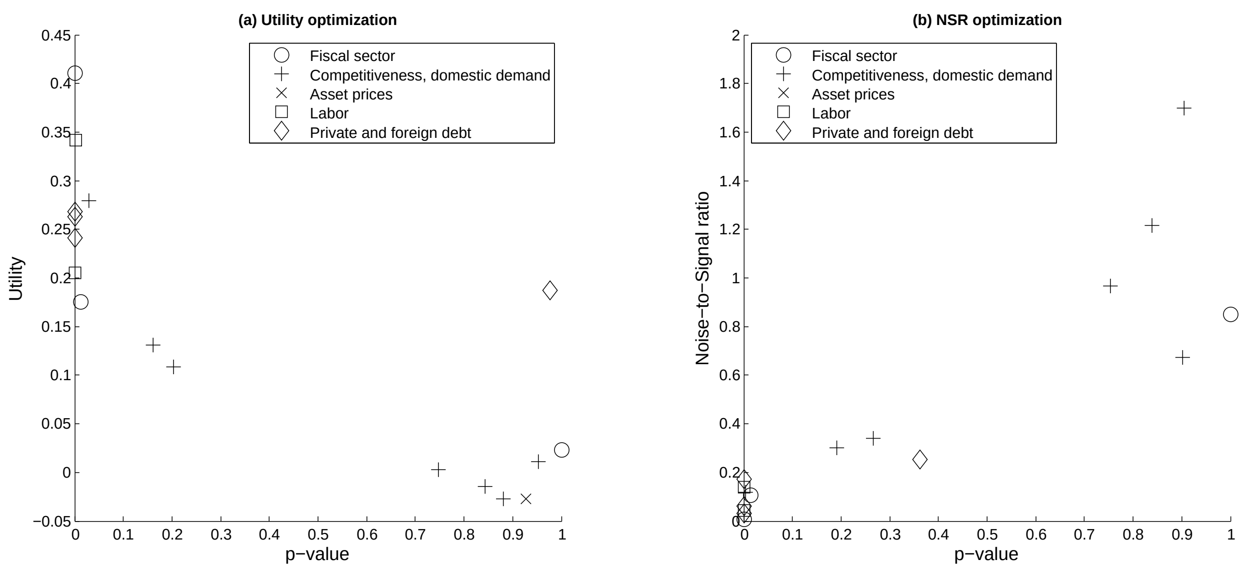
2.4.3 Public debt crises in Europe (Knedlik and von Schweinitz, 2012)

The in-sample results of the fourth application – public debt crises in a monetary union – are shown in Table 2.9. We were unable to exercise the bootstrap for three out of the original 20 indicators, due to the low number of data points available for these variables, as all indicators need to be available at all times in all countries to estimate the VAR.

Independent of whether we use utility or NSR optimization we find eight “good”, one “moderate” and eight “bad” indicators. The best indicator (government deficit) performs extremely well, with a utility level of 0.41 (an NSR of 0.07) (see Figure 2.7). Nevertheless, despite the excellent performance of this individual indicator (bearing in mind that a utility of 0.5 and an NSR of 0 are the best results that can be achieved), the composite indicator outperforms even government deficit using both utility and NSR optimization. However, even the extremely low NSR of the composite indicator turns out to be insignificant under the null hypothesis due to corner solutions produced by the bootstrap (see Figure 2.8).

While we generally observe a strong negative (positive) correlation between utility (NSR) and the corresponding p-value there are some exceptions. Particularly foreign assets provide a utility of 0.18 at a p-value of 0.97. That is, changes in foreign assets are actually outperformed by a random indicator variable with similar time series properties. The reason for this is, that the predictive ability of foreign assets is entirely based on persistent cross country differences rather than on the dynamics over time. Since our bootstrap reproduces country specific features (such as mean and variance of indicators), these differences are not considered significant.

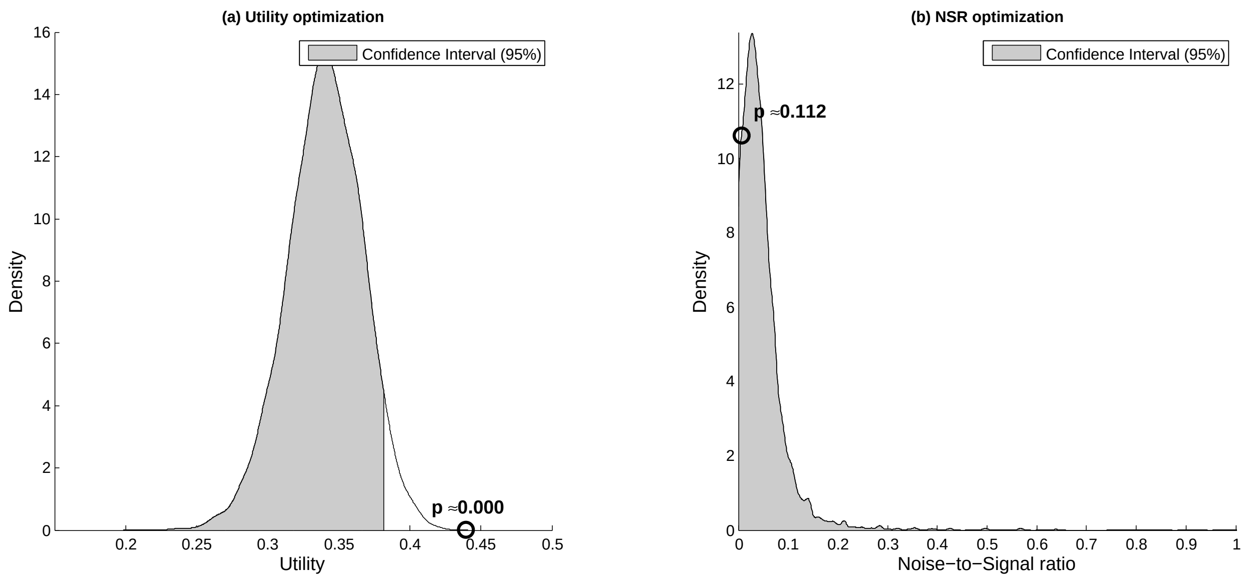
While the utility levels achieved out-of-sample are still high in absolute terms compared to those of all other applications of the signals approach that are considered in our paper, we find one single significant effect when using utility optimization, as shown both in Table 2.10 and Figure 2.14. The results are similar for NSR optimization where another single indicator is significant. However, as indicated by the Fisher test, this may equally well be explained as a statistical artifact due to the number of indicators. The reason for this counterintuitive result is, that our out-of-sample bootstrap treats both the global dynamics of the indicators and the persistent cross-country differences as given.[20] The crises in the European countries were not independent, but were caused by a global shock that could not be mitigated by some economies that were already vulnerable before the shock. Thus, global dynamics (the shock) and the persistent differences (reflecting the different levels of vulnerability to crises) do indeed explain most of what occurred. Therefore, the bootstrap produces utility (NSR) distributions under the null hypothesis, that are close to what we observe in reality. If we relax the highly restrictive assumptions of the bootstrapped null hypothesis and no longer keep cross-country differences constant, we find that most of the indicators found to be significant in-sample, including the composite indicator, are also significant out-of-sample.



Note: The 17 indicators are grouped according to Knedlik & von Schweinitz.

Figure 2.7: Significance and quality measure of individual indicators in the debt crisis application

[20] Conversely, the global component is resampled in the in-sample bootstrap.



Note: The density is derived under the null hypothesis, the circle indicates the respective quality measure of the original sample.

Figure 2.8: Density of utility and NSR in the debt crisis application

2.4.4 General results

First, our results show that the key findings of the three applications of the signals approach that we recreated in this paper are significant. This applies to both the general predictive power of the approach, for example, through a composite indicator, and the most important individual indicators (compare the last row “Composite Indicator” in Tables 2.3 to 2.9 and Figures 2.2, 2.4, 2.6 and 2.8. Considering the four applications in two versions each (utility maximization and the noise-to-signal ratio minimization) we find that in five out of eight cases the composite indicator is significant at a one percent error probability level, one further version is significant at a five percent level. In only two cases (noise-to-signal ratio minimization applied to banking and debt crises) do we have to accept higher error probabilities. When applying utility optimization – which generally yields more stable results – we consistently find highly significant results. This finding indicates that the signals approach can provide a major contribution to early warning systems.

Second, we find that using composite indicators adds substantial value to the signals approach. The composite indicator can be a particularly good crisis predictor, since the simultaneous movement of a large number of economic indicators provides additional information that can be exploited. Accordingly, we find the utility of the composite indicator to exceed the utility of the best individual indicator in all in-sample cases.

Third, while the best indicators in each considered application are highly significant, there is no perfect correlation between the chosen quality measure (NSR or utility) and significance, especially in the out-of-sample case. Due to different statistical properties of the indicators, the extent of in-sample fit that can be produced by chance differs widely. This becomes clear when one looks at the quality measures and significance levels in Tables 2.3 to 2.10. To demonstrate this point, we show that the cumulative distribution functions (CDFs) for the performance of the individual indicators under the null hypothesis may cross, as shown, for example, in Figure 2.9 for three indicators from Alessi & Detken (2011).[21] Since the CDFs are not identical even in the same application, the sequential arrangement of utility (or NSR) and significance is not necessarily the same. Comparing the applications with each other provides even stronger evidence of the fact that we cannot define a generally valid threshold for a "good" indicator. For example, several insignificant indicators from the debt crisis application provide a level of utility matching the best performing indicators of other applications.

Fourth, indicators that are significant in-sample tend to perform better out-of-sample than indicators with a lower level of in-sample significance, as can be seen from Figures 2.11, 2.12, 2.13 and 2.14. Due to the relatively stable performance of most indicators, the out-of-sample performance of utility optimized composite indicators is good and (in three out of four cases) is significant at the 1% level. However, in rare cases, significant indicators with high utility do not perform well out-of-sample. Therefore, it seems advisable to use the composite indicator, even though we find in a few cases in two of our applications that individual indicators outperform the composite indicator out-of-sample.

[21] Actually, we show the countercumulative distribution functions because they show directly the declining p-value for increasing utility.

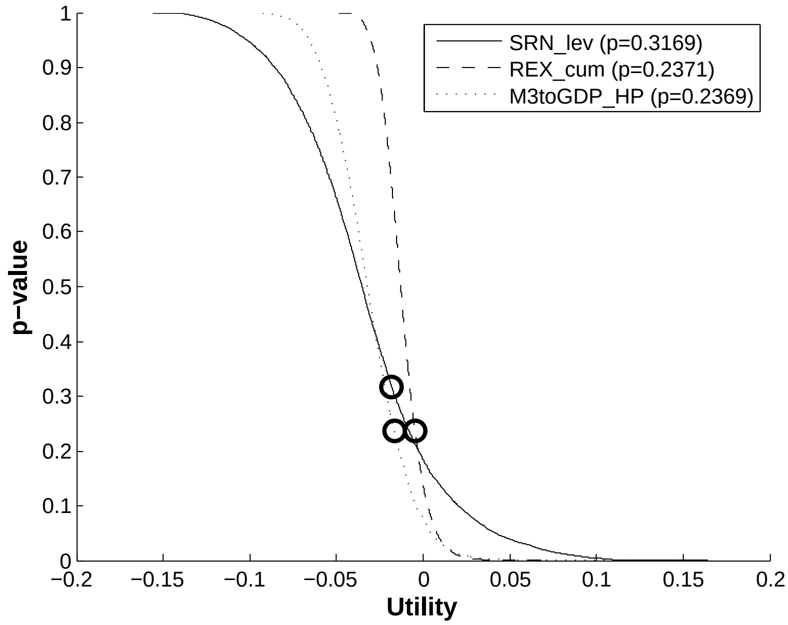


Figure 2.9: Utility and significance for three different indicators. Circles show the real values.

2.5 Conclusions

We augment four previous applications of the signals approach by adding an evaluation of significance based on a bootstrap approach. For all applications considered, we find that the major findings are indeed significant both in-sample and out-of-sample. This justifies a general reassessment of the usefulness of the signals approach. The signals approach holds a number of advantages and features favorable characteristics for early warning systems compared to alternative approaches. In particular, the signals approach's ability to provide both a composite crisis indicator and a more detailed picture when one looks at the individual indicators enables policy makers to get an impression of general risk and to identify areas where there might be a need for action to be taken.

For those interested in an early warning system rather than in just the best individual indicator, we propose using a broad set of indicators, covering a wide range of economic issues, to create a composite indicator based on utility optimization. The results achieved in this study when using this method produce composite indicators that are stable over time.

Acknowledgements

The authors are indebted to Andrew Hughes Hallet, Carsten Detken and two anonymous referees for their valuable comments and suggestions. We also thank Katja Drechsel and Herbert Buscher for comments and discussions.

Bibliography

Abiad, A. (2003). Early Warning Systems: a Survey and a Regime-Switching Approach, *IMF Working Paper 32*.

Alessi, L. & Detken, C. (2011). Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity, *European Journal of Political Economy* **27**(3): 520–533.

Bai, J. & Ng, S. (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data, *Journal of Business and Economic Statistics* **23**(1): 49–60.

Berg, A. & Pattillo, C. (1999). Predicting Currency Crises: The Indicators Approach and an Alternative, *Journal of International Money and Finance* **18**(4): 561–586.

Borio, C. & Drehmann, M. (2009). Assessing the Risk of Banking Crises–Revisited, *BIS Quarterly Review* (March 2009): 1–18.

Brüggemann, A. & Linne, T. (2002). Are the Central and Eastern European Transition Countries Still Vulnerable to a Financial Crisis? Results from the Signals Approach, *IWH Discussion Paper 157*.

Bussière, M. & Fratzscher, M. (2006). Towards a New Early Warning System of Financial Crises, *Journal of International Money and Finance* **25**(6): 953–973.

Bussière, M. & Fratzscher, M. (2008). Low Probability, High Impact: Policy Making and Extreme Events, *Journal of Policy Modeling* **30**(1): 111–121.

Diebold, F. X., Lee, J.-H. & Weinbach, G. C. (1994). Regime Switching with Time-Varying Transition Probabilities, *in* C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, Advanced Texts in Econometrics, Oxford University Press., pp. 283–302.

Edison, H. (2003). Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System, *International Journal of Finance and Economics* **8**(1): 11–53.

Filardo, A. (1994). Business Cycle Phases and Their Transitional Dynamics, *Journal of Business and Economic Statistics* **12**: 299–308.

Fisher, S. (1932). *Statistical Methods for Research Workers*, Vol. 5, Genesis Publishing Pvt Ltd.

Fitzenberger, B. (1998). The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions, *Journal of Econometrics* **82**(2): 235–287.

Frankel, J. A. & Rose, A. K. (1996). Currency Crashes in Emerging Markets: An Empirical Treatment, *Journal of International Economics* **41**(3-4): 351–366.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* **57**: 357–384.

Kamin, S., Schindler, J. & Samuel, S. (2001). The Contribution of Domestic and External Factors to Emerging Market Devaluation Crises: An Early Warning Systems Approach, *FRB International Finance Discussion Paper 711*.

Kaminsky, G. L. (1999). Currency and Banking Crises: The Early Warnings of Distress, *IMF Working Paper 99/178*.

Kaminsky, G. L. (2006). Currency Crises: Are They All the Same?, *Journal of International Money and Finance* **25**(3): 503–527.

Kaminsky, G. L., Lizondo, S. & Reinhart, C. M. (1998). Leading Indicators of Currency Crises, *IMF Staff Papers* **45**(1): 1–48.

Kaminsky, G. L. & Reinhart, C. M. (1999). The Twin Crises: the Causes of Banking and Balance-of-Payments Problems, *American Economic Review* **89**(3): 473–500.

Kittelmann, K., Tirpak, M., Schweickert, M. & Vinhas De Souza, L. (2006). From Transition Crises to Macroeconomic Stability? Lessons from a Crises Early Warning System for Eastern European and CIS Countries, *Comparative Economic Studies* **48**(3): 410–437.

Knedlik, T. & Scheufele, R. (2008). Forecasting Currency Crises: Which Methods signaled the South African Crisis of June 2006?, *South African Journal of Economics* **76**(3): 367–383.

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

Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations, *The Annals of Statistics* **17**(3): 1217–1241.

Kumar, M., Moorthy, U. & Perraudin, W. (2003). Predicting Emerging Market Currency Crashes, *Journal of Empirical Finance* **10**(4): 427–454.

Maddala, G. & Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* **61**(S1): 631–652.

Mariano, S., Gultekin, B., Ozmucur, S., Shabbir, T. & Alper, C. (2004). Prediction of Currency Crises: Case of Turkey, *Review of Middle East Economics and Finance* **2**(2): 87–107.

Peng, D. & Bajona, C. (2008). China's Vulnerability to Currency Crisis: A KLR Signals Approach, *China Economic Review* **19**(2): 138–151.

Rose, A. K. & Spiegel, M. M. (2011). Cross-Country Causes and Consequences of the Crisis: An Update, *European Economic Review* **55**(3): 309–324.

Annex

Table 2.2: Detailed description of surveyed applications

Category \ Paper	*Kaminsky & Reinhart (1999)*	*Alessi & Detken (2011)*	*Knedlik & von Schweinitz (2012)*
Countries	Argentina, Bolivia, Brazil, Chile, Colombia, Denmark, Finland, Indonesia, Israel, Malaysia, Mexico, Norway, Peru, Philippines, Spain, Sweden, Thailand, Turkey, Uruguay, Venezuela	Australia, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Spain, Sweden, Switzerland, United Kingdom, United States	Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, The Netherlands, Portugal, Spain
Time Period	Jan 1970 – Jun 1995	1970Q1 – 2007Q4	Jan 1999 – Nov 2011
Crisis Index	Currency crises: Exchange Rate Pressure Index $I = \frac{\Delta_e}{e} - \frac{\sigma_e}{\sigma_R}\frac{\Delta_R}{R}$, Banking crises: Expert judgment based on bank runs and government aids	Boom: Asset price index exceeds its filtered average + 1.75 times the standard deviation up to that time for 3 periods in a row: $I_t > HP_{I,t} + 1.75\sigma_{I,t}$	Spreads between 10 year government bond yields and the yield of the EMU-average 10 year AAA-rated government bond: $I = R_c - R_{AAA}$
Crisis Occurrence	$I > \mu_I + 3\sigma_I$	Costly boom: A boom with growth of real GDP 3% below potential growth	$I > \mu_I + 1.65\sigma_I$
Indicators	M2 multiplier, domestic credit (% GDP), real interest-rate, lending-deposit rate ratio, excess M1 balances, M2/reserves, bank deposits, exports, imports, terms of trade, real exchange rate, reserves, real interest-rate differential, output, stock returns, government deficit (% GDP)	Real GDP (GDPR), real consumption (CONS), real investment (INV), real housing investment (HINV), equity prices (QEPR), housing prices (QRPR), aggregate asset prices (QAAPR), term spread (SPREAD), real effective exchange rates (REX), real and nominal 3-month interest rates (SRR, SRN), 10-year bond yields (LRR, LRN), real M1 (M1), real M3 (M3), real private credit (PCR), real domestic credit (DCR). Variables in different transformations: yearly growth rates (_yoy), 6q-cumulated growth rates (_cum), level (_lev), % GDP (toGDP), detrended (_detr, _HP)	Government debt (% GDP), government deficit (% GDP), government interest payments (% government expenditure), unit labor costs, current account (% GDP), trade share, domestic demand, competitiveness indicators (deflated with HICP, GDP deflator and ULC), inflation, property prices, asset prices, unemployment rate, labor-force participation rate, labor productivity, private debt (% GDP), household debt (% GDP), non-MFI debt (% GDP), foreign assets (% GDP)
Warning horizon	24 months (before currency crises and 12 month before and after banking crises)	6 quarters	24 months
Comments	Data set until June 2003 used instead.	Quasi real-time approach: optimal threshold based on all data, but signals in t sent only, if the quantile based on data until t has been crossed.	

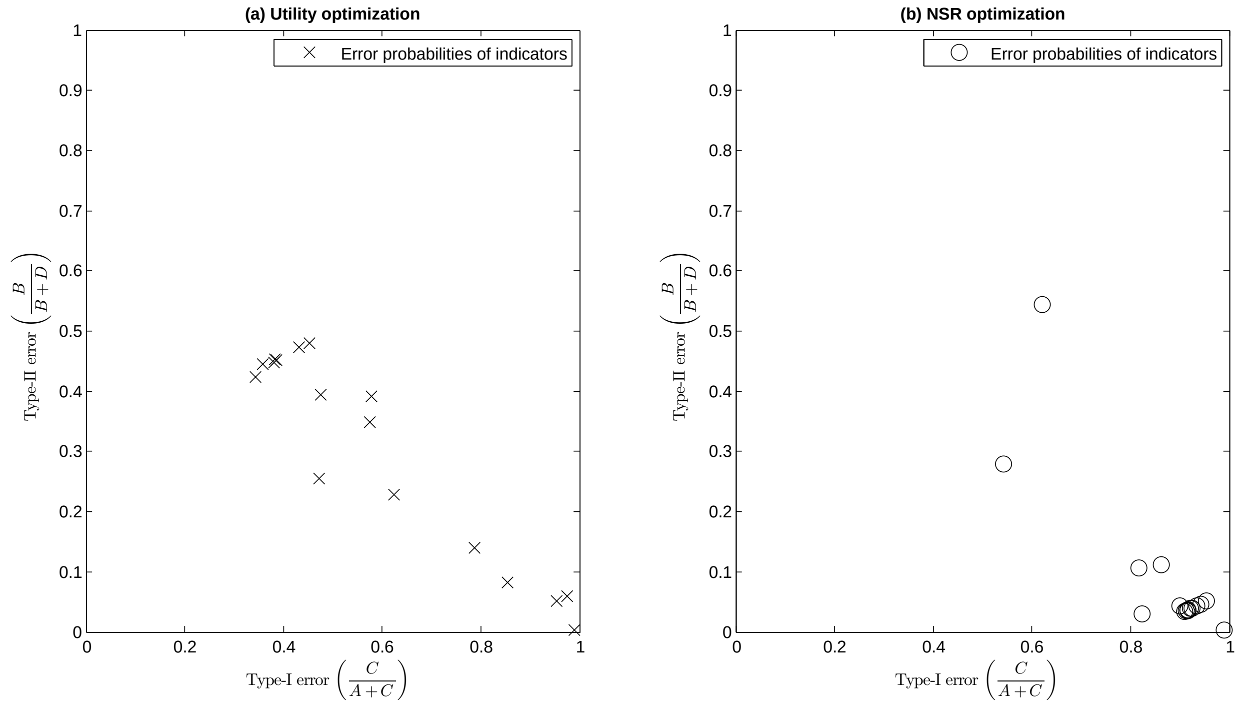


Figure 2.10: Type I and Type II error probabilities of individual indicators in the application to currency crises (Kaminsky and Reinhart, 1999)

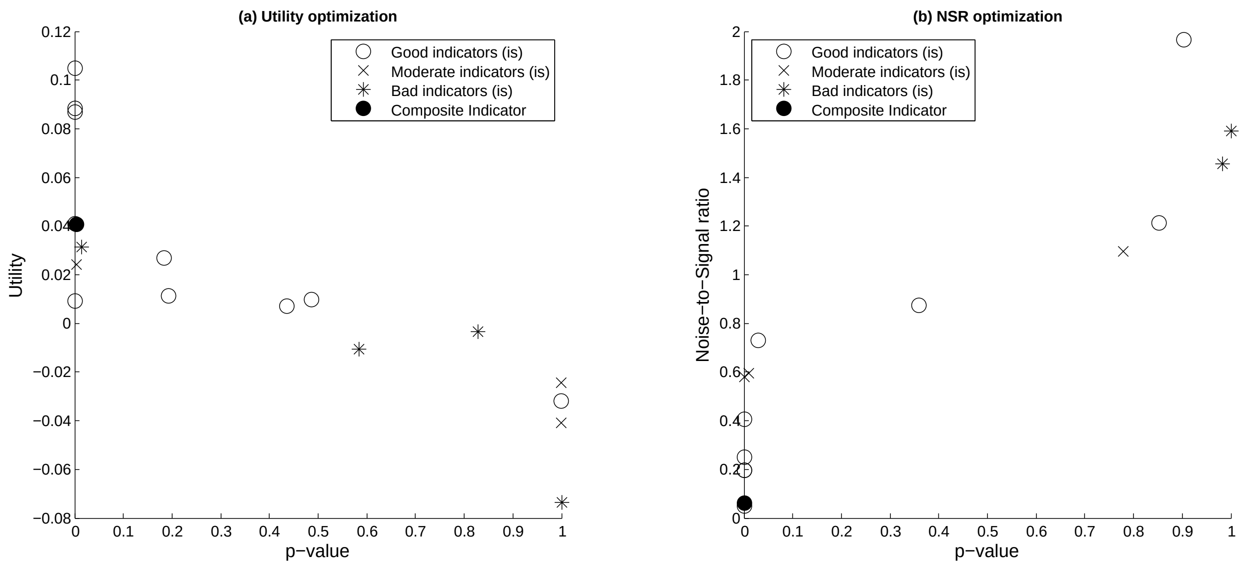
Table 2.3: Results for the currency crisis application

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
M2 multiplier	0.015	0.355	0.929	0.516	0.006	0.547	0.800	0.296
Domestic credit (%GDP)	0.065	0.002	0.752	0.075	0.020	0.124	0.494	0.008
Real interest-rate	0.037	0.181	0.655	0.073	0.028	0.132	0.376	0.012
Lending-deposit rate ratio	-0.017	0.970	2.393	0.976	-0.083	0.998	1.435	0.998
Excess M1 balances	0.038	0.134	0.819	0.269	0.011	0.457	0.657	0.173
M2 (%reserves)	0.082	0.000	0.734	0.051	0.024	0.052	0.430	0.002
Bank deposits	0.047	0.029	0.834	0.244	0.018	0.174	0.521	0.027
Exports	0.074	0.000	0.608	0.001	0.024	0.029	0.431	0.000
Imports	-0.003	0.813	1.118	0.862	-0.003	0.791	1.118	0.960
Terms of trade	0.033	0.081	0.878	0.326	0.013	0.263	0.808	0.300
Real exchange rate	0.136	0.000	0.484	0.006	0.074	0.001	0.167	0.000
Reserves	0.098	0.000	0.694	0.025	0.026	0.051	0.407	0.002
Real interest-rate differential	0.032	0.238	0.566	0.026	0.028	0.125	0.439	0.027
Output	0.086	0.000	0.722	0.059	0.038	0.009	0.584	0.066
Stock prices	0.081	0.000	0.735	0.118	0.024	0.101	0.421	0.011
Deficit (%GDP)	0.116	0.000	0.645	0.060	0.089	0.000	0.610	0.152
Banking crisis	0.004	0.000	0.279	0.000	0.004	0.000	0.279	0.000
Fisher test statistic	451.430	0.000					247.266	0.000
Composite Indicator	0.168	0.000	0.475	0.020	0.008	0.307	0.097	0.003

Note: This table and the following tables show the quality measures of the original samples as well as the significance obtained by means of the bootstrap. The first columns show the values obtained from utility optimization, the last columns those obtained from noise-to-signal ratio optimization.

Table 2.4: Results for the currency crisis application, out-of-sample

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
M2 multiplier	-0.011	0.584	1.063	0.585	-0.035	1.000	3.172	1.000
Domestic credit (%GDP)	0.087	0.000	0.693	0.000	0.012	0.032	0.731	0.029
Real interest-rate	0.024	0.003	0.868	0.004	0.033	0.000	0.195	0.000
Lending-deposit rate ratio	-0.074	1.000	3.985	1.000	-0.138	1.000	1.590	1.000
Excess M1 balances	-0.024	1.000	1.179	1.000	-0.004	0.790	1.097	0.779
M2 (%reserves)	0.011	0.192	0.952	0.195	-0.004	0.914	1.964	0.903
Bank deposits	-0.032	1.000	1.144	1.000	-0.010	1.000	10.038	1.000
Exports	0.007	0.435	0.973	0.435	0.003	0.341	0.873	0.360
Imports	-0.003	0.829	2.619	0.833	-0.004	0.863	2.837	0.867
Terms of trade	-0.041	1.000	1.232	1.000	-0.012	0.983	1.455	0.983
Real exchange rate	0.041	0.000	0.835	0.000	0.052	0.000	0.196	0.000
Reserves	0.010	0.487	0.967	0.481	-0.002	0.866	1.212	0.852
Real interest-rate differential	0.031	0.014	0.674	0.014	0.028	0.000	0.249	0.000
Output	0.088	0.000	0.724	0.000	0.031	0.010	0.596	0.010
Stock prices	0.027	0.184	0.885	0.159	0.036	0.000	0.405	0.000
Deficit (%GDP)	0.105	0.000	0.572	0.000	0.058	0.000	0.580	0.000
Banking crisis	0.009	0.000	0.048	0.000	0.009	0.000	0.048	0.000
Fisher test statistic	391.710	0.000					398.468	0.000
Composite Indicator	0.041	0.003	0.871	0.002	0.007	0.002	0.062	0.001



Note: The 16 indicators are grouped according to their in-sample significance.

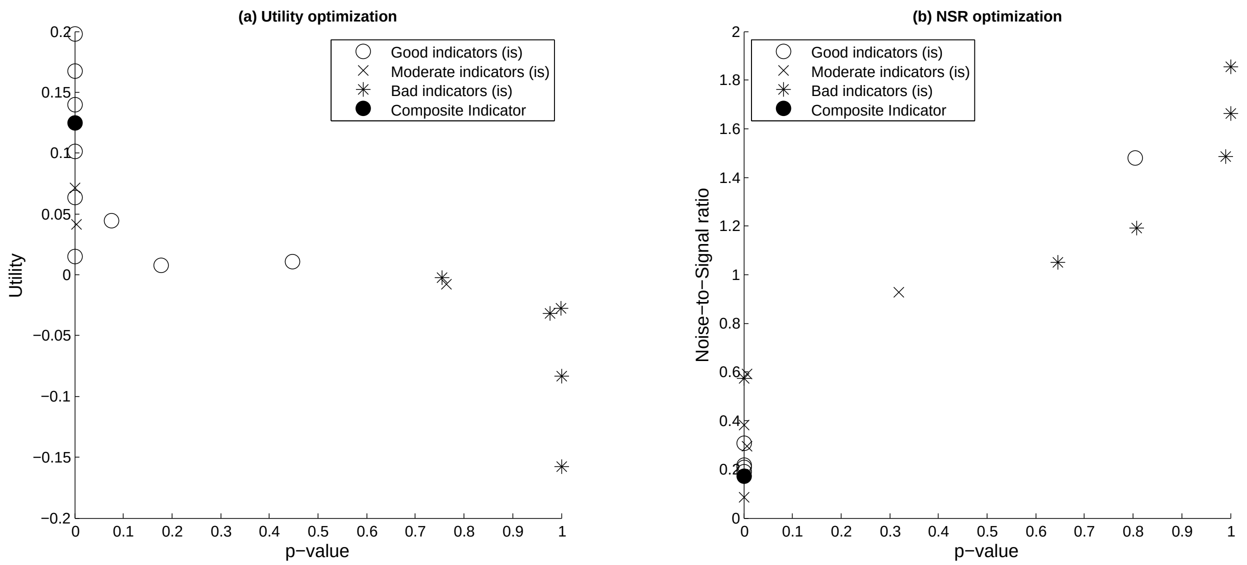
Figure 2.11: Significance and quality measure of individual indicators in the currency crisis application, out-of-sample

Table 2.5: Results for the banking crisis application

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
Currency crisis	0.007	0.002	0.463	0.001	0.007	0.002	0.463	0.001
M2 multiplier	0.073	0.008	0.682	0.096	0.026	0.143	0.452	0.031
Domestic credit (%GDP)	0.085	0.004	0.645	0.060	0.021	0.264	0.522	0.065
Real interest-rate	0.126	0.004	0.594	0.093	0.047	0.110	0.271	0.007
Lending-deposit rate ratio	-0.003	0.816	1.128	0.858	-0.003	0.813	1.128	0.886
Excess M1 balances	-0.003	0.841	1.112	0.893	-0.003	0.838	1.112	0.926
M2 (%reserves)	0.049	0.054	0.833	0.308	0.013	0.417	0.650	0.157
Bank deposits	0.011	0.592	0.881	0.511	0.011	0.510	0.881	0.580
Exports	0.033	0.120	0.882	0.359	0.015	0.233	0.777	0.255
Imports	-0.013	0.963	1.953	0.970	-0.081	1.000	1.456	1.000
Terms of trade	0.011	0.572	0.949	0.665	0.011	0.482	0.949	0.686
Real exchange rate	0.094	0.014	0.547	0.053	0.072	0.014	0.477	0.085
Reserves	0.051	0.058	0.738	0.130	0.010	0.495	0.697	0.224
Real interest-rate differential	0.112	0.009	0.631	0.130	0.058	0.049	0.261	0.004
Output	0.085	0.005	0.737	0.167	0.067	0.002	0.547	0.106
Stock prices	0.085	0.013	0.479	0.019	0.047	0.044	0.279	0.002
Deficit (%GDP)	0.123	0.004	0.653	0.137	0.028	0.255	0.528	0.146
Fisher test statistic	111.718	0.001					84.281	0.005
Composite Indicator	0.176	0.002	0.467	0.090	0.017	0.352	0.328	0.216

Table 2.6: Results for the banking crisis application, out-of-sample

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
Currency crisis	0.015	0.001	0.209	0.001	0.015	0.001	0.209	0.001
M2 multiplier	0.008	0.177	0.955	0.176	-0.014	0.813	1.479	0.804
Domestic credit (%GDP)	0.167	0.000	0.471	0.000	0.164	0.000	0.381	0.000
Real interest-rate	0.140	0.000	0.610	0.000	0.040	0.000	0.217	0.000
Lending-deposit rate ratio	-0.158	1.000	1.835	1.000	-0.093	1.000	1.854	1.000
Excess M1 balances	-0.028	0.999	3.079	0.999	-0.028	0.999	3.079	0.999
M2 (%reserves)	0.071	0.000	0.774	0.000	0.016	0.002	0.296	0.007
Bank deposits	-0.031	0.976	1.284	0.971	-0.015	0.827	1.191	0.808
Exports	-0.008	0.764	1.025	0.765	-0.005	0.645	1.050	0.645
Imports	-0.002	0.755	2.302	0.755	-0.047	0.990	1.485	0.991
Terms of trade	-0.084	1.000	1.730	1.000	-0.072	1.000	1.663	1.000
Real exchange rate	0.064	0.000	0.677	0.000	0.003	0.323	0.929	0.319
Reserves	0.041	0.004	0.883	0.005	0.100	0.000	0.575	0.000
Real interest-rate differential	0.101	0.000	0.732	0.000	0.065	0.000	0.189	0.000
Output	0.045	0.075	0.844	0.072	0.054	0.006	0.591	0.006
Stock prices	0.011	0.447	0.949	0.440	0.069	0.000	0.306	0.000
Deficit (%GDP)	0.198	0.000	0.427	0.000	0.059	0.000	0.086	0.000
Fisher test statistic	468.062	0.000					471.806	0.000
Composite Indicator	0.125	0.000	0.683	0.000	0.036	0.000	0.173	0.000



Note: The 16 indicators are grouped according to their in-sample significance.

Figure 2.12: Significance and quality measure of individual indicators in the banking crises application, out-of-sample

Table 2.7: Results for the costly asset price boom application

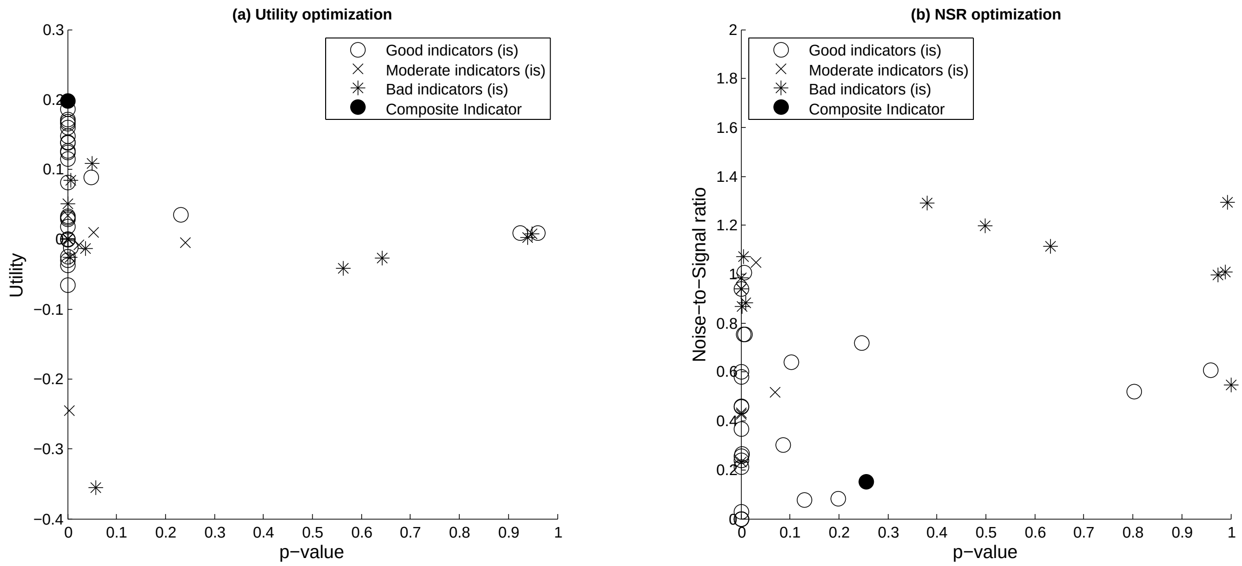
	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
LRN_lev	0.025	0.112	0.566	0.098	0.025	0.109	0.566	0.103
LRR_lev	0.013	0.080	0.551	0.059	0.012	0.078	0.508	0.043
SRN_lev	-0.018	0.317	0.733	0.261	-0.028	0.389	0.733	0.268
SRR_lev	0.009	0.202	0.564	0.256	0.009	0.164	0.564	0.287
SPREAD_lev	-0.011	0.776	2.052	0.819	-0.117	0.720	0.927	0.660
SPREAD_-lev	-0.028	0.932	1.973	0.891	-0.146	0.797	0.931	0.676
CPI_yoy	-0.014	0.950	Inf	0.897	-0.173	0.844	1.046	0.855
CONS_yoy	0.101	0.000	0.413	0.048	0.011	0.030	0.243	0.008
INV_yoy	0.106	0.000	0.362	0.000	0.049	0.000	0.305	0.000
HINV_yoy	0.081	0.000	0.384	0.003	0.037	0.001	0.333	0.001
REX_yoy	-0.007	0.330	0.843	0.336	-0.047	0.587	0.807	0.336
M1_yoy	0.056	0.001	0.488	0.014	0.056	0.000	0.488	0.021
M3_yoy	0.072	0.000	0.379	0.002	0.037	0.001	0.318	0.002
PCR_yoy	0.091	0.000	0.324	0.005	0.023	0.011	0.284	0.003
DCR_yoy	0.065	0.002	0.419	0.015	0.065	0.001	0.419	0.021
CPI_cum	-0.017	0.906	9.569	0.895	-0.180	0.874	1.078	0.884
CONS_cum	0.114	0.000	0.385	0.038	0.029	0.003	0.298	0.019
INV_cum	0.124	0.000	0.371	0.001	0.038	0.001	0.244	0.000
HINV_cum	0.095	0.000	0.375	0.007	0.053	0.000	0.256	0.001
REX_cum	-0.005	0.237	0.764	0.246	-0.005	0.231	0.764	0.260
M1_cum	0.040	0.004	0.531	0.044	0.030	0.005	0.452	0.018
M3_cum	0.065	0.001	0.398	0.004	0.054	0.001	0.334	0.002
PCR_cum	0.092	0.000	0.363	0.013	0.047	0.001	0.327	0.011
DCR_cum	0.044	0.009	0.476	0.032	0.036	0.010	0.472	0.040
GDPR_detr	0.084	0.001	0.355	0.000	0.079	0.001	0.279	0.000
LRN_detr	-0.007	0.259	0.737	0.260	-0.008	0.269	0.723	0.251
LRR_detr	-0.014	0.657	1.212	0.668	-0.113	0.724	0.973	0.788
SRN_detr	-0.021	0.804	1.476	0.757	-0.189	0.920	0.992	0.882
SRR_detr	-0.012	0.552	0.991	0.544	-0.185	0.849	0.983	0.805
REX_detr	-0.003	0.163	0.690	0.154	-0.003	0.162	0.690	0.155
CONStoGDP_detr	0.017	0.023	0.583	0.059	0.017	0.019	0.583	0.063
INVtoGDP_detr	0.099	0.000	0.362	0.001	0.066	0.000	0.257	0.000
HINVtoGDP_detr	0.055	0.001	0.254	0.000	0.055	0.001	0.254	0.000
M1toGDP_detr	0.018	0.019	0.589	0.028	0.018	0.018	0.589	0.029
M3toGDP_detr	-0.024	0.340	0.938	0.373	-0.039	0.385	0.879	0.382
PCNtoGDP_detr	0.015	0.028	0.554	0.018	0.015	0.026	0.554	0.019
DCNtoGDP_detr	0.013	0.067	0.611	0.093	0.013	0.064	0.611	0.093
GDPR_HP	0.088	0.001	0.409	0.001	0.058	0.002	0.352	0.001
LRN_HP	0.026	0.038	0.546	0.048	0.026	0.034	0.546	0.054
LRR_HP	-0.019	0.797	1.730	0.791	-0.063	0.659	0.928	0.710
SRN_HP	-0.013	0.426	0.861	0.418	-0.013	0.392	0.861	0.494
SRR_HP	-0.012	0.844	3.636	0.885	-0.080	0.719	0.941	0.739
REX_HP	-0.016	0.475	1.042	0.543	-0.034	0.506	0.812	0.372
CONStoGDP_HP	-0.002	0.198	0.686	0.195	-0.002	0.198	0.686	0.195
INVtoGDP_HP	0.113	0.000	0.399	0.005	0.083	0.000	0.272	0.000
HINVtoGDP_HP	0.068	0.001	0.411	0.015	0.058	0.001	0.389	0.013
M1toGDP_HP	0.045	0.004	0.461	0.003	0.045	0.004	0.461	0.003
M3toGDP_HP	-0.017	0.237	0.848	0.280	-0.018	0.235	0.788	0.220
PCNtoGDP_HP	0.019	0.028	0.600	0.046	0.014	0.041	0.555	0.024
DCNtoGDP_HP	0.029	0.024	0.564	0.047	0.029	0.021	0.564	0.049
Fisher test statistic	800.491	0.003					513.364	0.004
Composite Indicator	0.149	0.010	0.239	0.099	0.050	0.087	0.069	0.016

Note: acronyms and abbreviations of indicators are spelled out in Table 2.2.

Table 2.8: Results for the costly asset price boom application, out-of-sample

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
LRN_lev	-0.245	0.004	1.338	0.001	-0.176	0.045	1.048	0.030
LRR_lev	-0.004	0.240	0.719	0.246	-0.004	0.240	0.719	0.246
SRN_lev	-0.355	0.058	2.321	0.015	-0.280	0.618	1.290	0.380
SRR_lev	-0.041	0.562	2.157	0.463	-0.248	0.690	1.196	0.499
SPREAD_lev	0.000	0.000	NaN	NaN	0.036	1.000	0.546	1.000
SPREAD_-lev	-0.027	0.642	7.550	0.630	-0.226	0.006	1.073	0.005
CPI_yoy	0.000	0.000	NaN	NaN	-0.190	0.990	1.010	0.989
CONS_yoy	0.165	0.000	0.264	0.000	0.003	0.987	0.000	0.000
INV_yoy	0.126	0.000	0.255	0.000	0.022	0.117	0.302	0.086
HINV_yoy	-0.037	0.000	0.867	0.000	-0.005	0.009	0.755	0.009
REX_yoy	-0.013	0.037	0.944	0.036	-0.102	0.009	0.882	0.010
M1_yoy	0.018	0.000	0.602	0.000	0.018	0.000	0.602	0.000
M3_yoy	0.147	0.000	0.273	0.000	0.078	0.004	0.211	0.001
PCR_yoy	0.138	0.000	0.258	0.000	0.016	0.000	0.000	0.000
DCR_yoy	0.124	0.000	0.238	0.000	0.124	0.000	0.238	0.000
CPI_cum	0.000	0.000	NaN	NaN	-0.180	0.973	0.996	0.974
CONS_cum	0.168	0.000	0.252	0.000	0.024	0.592	0.000	0.000
INV_cum	0.138	0.000	0.146	0.000	0.021	0.264	0.084	0.199
HINV_cum	0.031	0.000	0.534	0.000	0.019	0.000	0.460	0.000
REX_cum	-0.026	0.005	1.107	0.004	-0.091	0.002	0.868	0.003
M1_cum	-0.065	0.001	0.845	0.001	-0.062	0.001	0.940	0.001
M3_cum	0.114	0.000	0.288	0.000	0.101	0.000	0.256	0.000
PCR_cum	0.160	0.001	0.259	0.000	0.001	0.103	0.639	0.103
DCR_cum	0.089	0.049	0.357	0.016	0.093	0.001	0.368	0.000
GDPR_detr	0.029	0.000	0.063	0.000	0.013	0.001	0.000	0.000
LRN_detr	0.000	0.000	NaN	NaN	0.000	0.000	NaN	NaN
LRR_detr	0.003	0.939	0.000	0.000	-0.037	0.630	1.115	0.633
SRN_detr	0.000	0.000	NaN	NaN	-0.234	0.803	4.341	0.408
SRR_detr	0.000	0.000	NaN	NaN	-0.030	1.000	Inf	0.517
REX_detr	-0.008	0.023	0.731	0.022	-0.164	0.000	0.940	0.000
CONStoGDP_detr	0.081	0.000	0.426	0.000	0.081	0.000	0.426	0.000
INVtoGDP_detr	0.171	0.000	0.033	0.000	0.047	0.377	0.076	0.130
HINVtoGDP_detr	-0.012	0.006	1.007	0.006	-0.012	0.006	1.007	0.006
M1toGDP_detr	-0.024	0.001	0.806	0.001	-0.011	0.005	0.755	0.005
M3toGDP_detr	0.085	0.006	0.147	0.000	-0.160	0.000	0.939	0.000
PCNtoGDP_detr	0.008	0.960	0.608	0.958	0.008	0.960	0.608	0.958
DCNtoGDP_detr	0.010	0.053	0.585	0.052	0.021	0.076	0.518	0.070
GDPR_HP	0.033	0.001	0.544	0.001	0.054	0.003	0.266	0.002
LRN_HP	0.000	0.000	NaN	NaN	0.000	0.000	NaN	NaN
LRR_HP	0.008	0.949	0.189	0.967	-0.017	0.998	1.294	0.994
SRN_HP	0.000	0.000	NaN	NaN	0.000	0.000	NaN	NaN
SRR_HP	0.000	0.000	NaN	NaN	-0.116	0.414	2.617	0.277
REX_HP	0.050	0.001	0.315	0.001	-0.191	0.000	0.985	0.000
CONStoGDP_HP	0.034	0.001	0.460	0.001	0.057	0.002	0.433	0.001
INVtoGDP_HP	0.186	0.000	0.059	0.000	0.129	0.000	0.030	0.000
HINVtoGDP_HP	-0.001	0.000	0.669	0.000	-0.054	0.457	2.181	0.392
M1toGDP_HP	-0.030	0.000	0.755	0.000	0.059	0.000	0.457	0.000
M3toGDP_HP	0.108	0.050	0.133	0.001	0.107	0.007	0.231	0.000
PCNtoGDP_HP	0.009	0.925	0.631	0.921	0.036	0.831	0.520	0.803
DCNtoGDP_HP	0.035	0.232	0.476	0.177	0.033	0.000	0.580	0.000
Fisher test statistic	1884.661	0.000					1388.387	0.000
Composite Indicator	0.197	0.000	0.095	0.000	0.010	0.240	0.151	0.256

Note: acronyms and abbreviations of indicators are spelled out in Table 2.2. A Noise-to-Signal ratio of NaN is given, if both noise B and true signals A are 0 (i.e., if the indicator never sends a signal).



Note: The 50 indicators are grouped according to their in-sample significance.

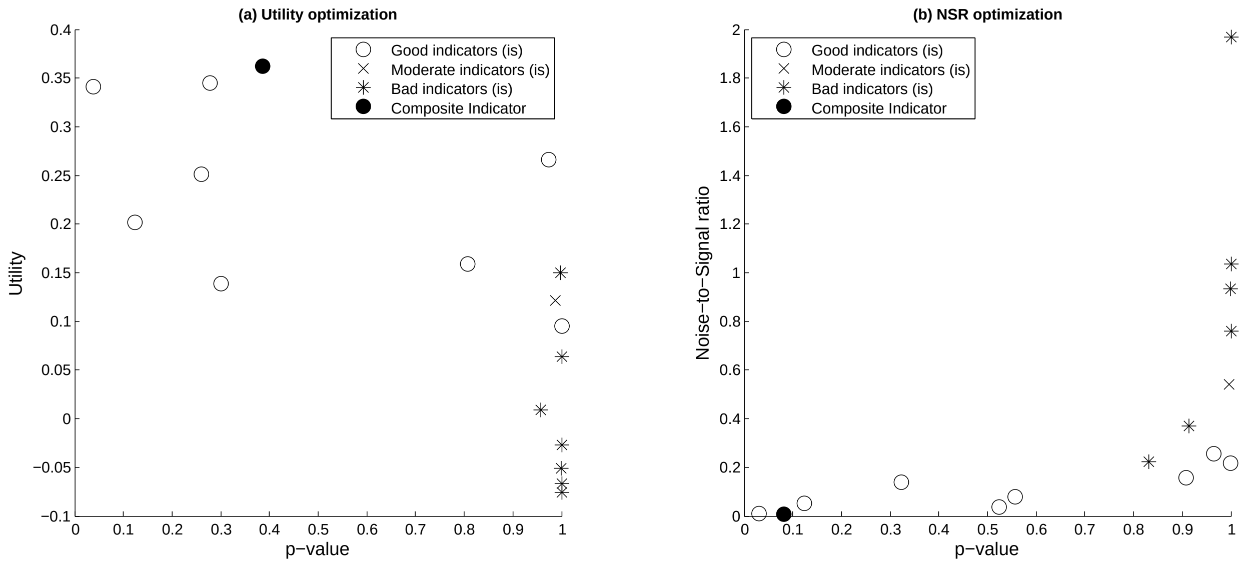
Figure 2.13: Significance and quality measure of individual indicators in the costly asset price boom application, out-of-sample

Table 2.9: Results for the debt crisis application

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
Government debt (%GDP)	0.175	0.013	0.463	0.385	0.124	0.031	0.107	0.014
Government deficit (%GDP)	0.410	0.000	0.165	0.019	0.319	0.000	0.007	0.000
Interest payment (%gov expenditure)	0.023	1.000	0.851	1.000	0.023	0.996	0.851	1.000
Unit labor costs	0.131	0.161	0.410	0.155	0.109	0.056	0.303	0.192
Current account	0.280	0.030	0.350	0.011	0.123	0.212	0.119	0.003
Share in world trade	0.109	0.203	0.662	0.480	0.052	0.348	0.339	0.267
Unemployment rate	0.341	0.002	0.097	0.000	0.221	0.004	0.043	0.000
Labor participation rate	0.205	0.000	0.395	0.057	0.105	0.008	0.139	0.001
Private debt	0.241	0.000	0.208	0.070	0.144	0.001	0.171	0.000
Non-MFI debt	0.263	0.000	0.137	0.001	0.184	0.000	0.062	0.000
Household debt	0.268	0.000	0.078	0.003	0.164	0.014	0.031	0.001
Foreign assets	0.187	0.977	0.469	0.945	0.143	0.252	0.255	0.363
Inflation	0.011	0.952	0.674	0.861	0.011	0.946	0.674	0.902
Asset prices	-0.027	0.928	Inf	0.928	-0.100	0.963	2.098	0.973
HICP-competitiveness	-0.027	0.881	Inf	0.881	-0.049	0.861	1.698	0.904
GDP-deflated competitiveness	-0.014	0.843	1.255	0.823	-0.022	0.841	1.215	0.839
ULC-competitiveness	0.003	0.747	0.967	0.753	0.003	0.747	0.967	0.754
Fisher test statistic	288.312	0.000					292.039	0.000
Composite Indicator	0.439	0.000	0.105	0.002	0.320	0.000	0.007	0.112

Table 2.10: Results for the debt crisis application, out-of-sample

	Utility optimized				*NSR optimized*			
	Utility	p	NSR	p	Utility	p	NSR	p
Government debt (%GDP)	0.096	1.000	0.585	1.000	0.116	0.797	0.217	0.999
Government deficit (%GDP)	0.345	0.278	0.164	0.951	0.374	0.000	0.011	0.031
Interest payment (%gov expenditure)	-0.075	1.000	2.084	0.997	0.037	1.000	0.761	1.000
Unit labor costs	0.122	0.987	0.509	1.000	0.038	0.995	0.541	0.996
Current account	0.266	0.973	0.349	1.000	0.146	0.332	0.158	0.908
Share in world trade	0.064	1.000	0.756	1.000	0.049	0.586	0.370	0.913
Unemployment rate	0.341	0.038	0.099	0.085	0.235	0.338	0.052	0.124
Labor participation rate	0.202	0.124	0.412	0.550	0.116	0.069	0.140	0.323
Private debt	0.139	0.301	0.292	0.302	0.081	0.972	0.255	0.965
Non-MFI debt	0.159	0.808	0.188	0.838	0.180	0.446	0.080	0.558
Household debt	0.251	0.260	0.083	0.822	0.188	0.550	0.038	0.525
Foreign assets	0.150	0.998	0.491	1.000	0.152	0.606	0.224	0.831
Inflation	0.009	0.957	0.736	0.975	-0.002	1.000	1.035	1.000
Asset prices	-0.027	1.000	Inf	0.247	-0.123	1.000	2.351	1.000
HICP-competitiveness	-0.027	1.000	Inf	0.193	-0.063	1.000	1.968	1.000
GDP-deflated competitiveness	-0.067	1.000	Inf	0.998	-0.053	1.000	2.016	0.984
ULC-competitiveness	-0.051	0.999	2.171	0.989	0.008	0.998	0.933	0.999
Fisher test statistic	18.988	0.977					16.737	0.986
Composite Indicator	0.362	0.386	0.114	0.784	0.337	0.007	0.006	0.082



Note: The 17 indicators are grouped according to their in-sample significance.

Figure 2.14: Significance and quality measure of individual indicators in the debt crisis application, out-of-sample

Chapter 3

Flight Patterns and the 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.

* In some formulations, this chapter differs from the IWH-Discussion Paper. Results, however, remain the same.

3.1 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 3.1). 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, Hutchison & Jinjarak 2013, Beirne & 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 3.1. 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. 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 & 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

[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.

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, Schuknecht & Wolswijk 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) (see for example Schuknecht et al. 2009). 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, Kossmeier & Pichler 2004). The remaining differences could only partly be explained by different levels of government debt sustainability (Codogno, Favero & Missale 2003). Liquidity risk did not disappear completely despite the liquidity gains due to the introduction of a common market (Gómez-Puig 2006, Gómez-Puig 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 of credit and liquidity risk variables with global risk can be seen as „flight-to-quality“ and „flight-to-liquidity“, described in a theoretical CAPM-model by Vayanos (2004).[2] To detect these

[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.

flight patterns, Beber, Brandt & Kavajecz (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 intuitive 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 (previously until 2005, Schuknecht et al. 2009) 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, von Hagen & Schuknecht (2012). Already in an early paper after the outbreak of the financial crisis, Dötz & Fischer (2010) find an increased market reaction to debt sustainability and competitiveness.

In a related context, sample splits are also used by Beirne & Fratzscher (2013) to detect regional contagion during the European debt crisis. Contagion may be defined as a fundamentally unjustified premium on yields due to spillovers from other crisis countries (Favero & Missale 2012). These authors use the existence of adverse effects 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

[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"*.

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 support of explanatory variables changes, markets are expected to react differently to different values of the explanatory variables. Then, the model should not contain a structural break, but rather be nonlinear. As will be shown in Subsection 3.3.6, all explanatory variables experienced a strong shift towards a more adverse distribution during the crisis. To overcome the strong assumption behind the inclusion of a dummy, Bernoth & 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 disregarding important influence channels.

A nonlinear model in general is inspired by the market discipline hypothesis (Bayoumi, Goldstein & Woglom 1995). This hypothesis assumes a convex relationship between different credit risk variables and yields. That is, the marginal effect on yields is stronger at higher levels of credit risk. Such a market behavior would provide a disciplining effect on the borrower. The use of quadratic terms in previous contributions (Gómez-Puig 2006, Gómez-Puig 2008, Bernoth et al. 2012) reflects this hypothesis.[4] The semi-parametric method applied in this paper, penalized spline regression (Ruppert & Carroll 1997), approximates any unknown, but possibly highly non-linear, functional form of the influence of different

[4] In a narrow sense, the hypothesis only refers to credit risk. However, the same argument also applies to other risk types.

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.[5] As these measures are not publicly available, they have only seldom been used (Beber et al. 2009, Favero, Pagano & von Thadden 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 3.2 (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 (following Beirne & Fratzscher 2013) and not spreads over German Bunds. This point as well as the employed explanatory variables are described in detail in Section 3.3. The results are presented in Section 3.4, with robustness checks in Section 3.5. Section 3.6 concludes.

3.2 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.[6] These sources are credit risk, liquidity risk and a global risk

[5] 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, Rindi & De Jong (2005).

[6] 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 and government deficit to GDP can hardly both relate linearly to credit risk at the same time. I do not differentiate between these two possible origins of nonlinearity.

component, but also possible interactions between them. I use a semiparametric method, penalized spline regression (Ruppert & Carroll 1997), in order to estimate the European government bond yields. The method builds on a semiparametric generalized linear model (Green & 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 a possible local estimation bias. It has been applied in diverse scientific fields (see Ruppert, Wand & Carroll (2009) for a literature review), but only seldom to economic problems: For example, Jarrow, Ruppert & Yu (2004) use it to determine the yield curve at different maturities, Eisenbeiß, Kauermann & Semmler (2007) estimate the risk aversion of investors on the stock market, Flaschel, Kauermann & Semmler (2007) jointly estimate different Phillips curves, while Berlemann, Enkelmann & Kuhlenkasper (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 modeled 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 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 & Missale 2012), as autocorrelated error processes would necessitate an adaptation of the estimation process of penalized splines.[7] In all estimations, the autocorrelation coefficient is significantly smaller

[7] A unit root for yields is rejected at the 10% level for the whole sample from January 1999 to December 2011. In the pre-crisis sample (January 1999 to September 2008), the rejection is even at the 1% level. The rejection probabilities are computed from the unit-root test by Chang (2002) that accounts for possible

than unity. I estimate the spread of benchmark government bond yields (maturity ten years) over the three-months money market rate by

$$y_{t,c} = \rho y_{t-1,c} + f(x_{1,t,c}, \ldots, x_{N,t,c}) + D_{t,c}\delta + \varepsilon_t. \tag{3.1}$$

The matrix D contains dummy variables, explained below, δ is the associated parameter vector. The index c denotes different countries and is introduced for notational reasons. The univariate and bivariate functions are independent of c. Combining all polynomial terms (including y_{t-1} and D) in the matrix X and all spline terms in the matrix Z, the objective function to be optimized is

$$\min_{\beta,b,\Lambda} \|y - X\beta - Zb\|^2 + b'(\Lambda\Lambda')^{-1}b, \tag{3.2}$$

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 (3.2) 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 (3.2) 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 (3.2), the assumption for b and its transformation allows to use estimation techniques developed for linear mixed models (Bates 2012).

cross-sectional dependency.

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 & 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, Eubank, Smith & Young 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 & 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 & 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.

3.3 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 dependent 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.

3.3.1 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 & Fratzscher 2013).[8] Most authors employ the yields of benchmark German Bunds or US Treasury bonds as a risk-free interest rate for bonds denominated in Euro or US-Dollar, respectively (see for example Schuknecht et al. 2009). 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. 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 3.2 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

[8] 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, Cumby & Diba 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.[9] 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.

3.3.2 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*) (as for example in Bernoth et al. 2012), both drawn from Eurostat (ES). Beirne & 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 & 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

[9] 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.

(CDS). However, markets for CDS may also be subject to distorting liquidity risk similar to the government bond market (Fontana & Scheicher 2010). Additionally, at least for daily frequencies, government bond spreads are found to lead CDS spreads in emerging markets (Ammer & 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 & 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 & Zoli 2009, Beirne & 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, Schnabel & Zettelmeyer 2006, Hauner, Jonas & Kumar 2010, Beirne & Fratzscher 2013).[10]

3.3.3 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

[10] 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.

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 (used for example by Gómez-Puig 2006, Bernoth et al. 2012).

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).[11] 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 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

[11] 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.

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 1999, 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.

3.3.4 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 & 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 & Fischer 2010).[12] 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 of 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.

3.3.5 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

[12] Bernoth & 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.

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.[13] 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.

3.3.6 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 3.1 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

[13] One further rescue package for Greece was negotiated between July 2011 and February 2012, partly 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.

spreads in times of high global risk. That is, the support regions 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 or – nearly equivalently – along the lines of global risk) 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 premiums 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., it is uncertain which of the two estimated parameter sets actually describes yields best. Therefore, even when values of all explanatory variables are known with certainty, future yields are driven by the uncertainty about the state of the economy.[14] 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

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

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.

3.3.7 Interactions

One of the main objectives of this paper is to identify patterns of *flight-to-quality* and *flight-to-liquidity*. Such flights should result in 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. That is, if flight-to-quality exists on the European government bond market, then bonds from countries with high deficits should be traded at excess premiums 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 premiums. 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. Both flight patterns describe a yield reaction to credit and liquidity risk conditional on the level of global risk. Therefore, they 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.[15]

3.4 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

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

as rescue dummies are jointly given in Table 3.2 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 3.3. The results of the main estimation are split between different risk components. Figure 3.6 (for global risk), Figure 3.7 (credit risk and flight-to-quality) and Figure 3.8 (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 3.7 shows that a government surplus of 5% implies ceteris paribus 0.1% lower yields, while a deficit of around 20% implies around 0.2% higher yields.[16] Univariate effects, for example in Figure 3.6, are centered at the sample mean and given with 95% confidence bands.[17] 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 3.7). 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 3.8 or for all results of the estimation without splines in Figure 3.3.[18] Interaction plots are contour plots, for example in subplots (3) and (4) of Figure 3.8, where the strength of the effect is indicated by the colorbar at the side of the plot. As implied by the scarcity of observations, some large areas in the interaction

[16] 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.

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

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

plots are defined by very few datapoints. Areas without any surrounding datapoints are kept blank.

3.4.1 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, rescue dummies and a constant. The ten functions are graphically given in Figure 3.3. Linear parameters are reported in column (1) of Table 3.2. 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 well be sufficiently nonlinear to avoid the problem of structural breakpoints.

The univariate functions in subplots (1) to (6) 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, 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.[19] As the effects are not

[19] 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.

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.

3.4.2 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 3.4, a fact that is mirrored in the extremely low standard deviation of 0.288% (see also Table 3.2) and the correspondingly small residuals in Figure 3.5.[20] 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 3.3, 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

[20] 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.

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.

3.4.3 Global Risk

Figure 3.6 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 & 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 premiums for all possible assets in the market. Accordingly, the main effect global risk exerts should be in interaction with the other risk variables.

3.4.4 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 3.7 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 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 3.7. 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 & Rogoff 2010, Reinhart, Reinhart & Rogoff 2012). This finding is much stronger in the previous estimation without splines, see subplot (3) of Figure 3.3.

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 & 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 3.7 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 surpluses should be traded at lower yields than in tranquil times. Subplot (4) of Figure 3.7 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.

3.4.5 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 3.8 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 only 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 3.6 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).

3.4.6 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 parameter estimates are reported in Table 3.2 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 & Missale 2012) and does not vary much between different specifications reported in Table 3.2. 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 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.[21] 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.

3.5 Robustness checks

3.5.1 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).[22]

Figure 3.9 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). It is off-scale for Italy.

Table 3.4 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,

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

[22] 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.

the Netherlands and Austria.[23] 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 the panel out-of-sample standard deviation over the 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 observed in-sample support. The effect of extreme values is particularly affected by changing parameter estimates 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 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 on the previously observed support 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.

[23] 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.

3.5.2 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 3.10 show that changes to the baseline estimation (Figures 3.6 to 3.8) 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 3.2. 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.

3.5.3 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 3.11. 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 3.2 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.

3.6 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 nonlinearities 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 analysis, 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 surpluses 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 & 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.

Bibliography

Agarwal, G. & 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. & 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. & 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*.

Bayoumi, T., Goldstein, M. & Woglom, G. (1995). Do Credit Markets Discipline Sovereign Borrowers? Evidence from US States, *Journal of Money, Credit and Banking* **27**(4): 1046–1059.

Beber, A., Brandt, M. W. & 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. & 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. & Kuhlenkasper, T. (2012). Unraveling the Complexity of US Presidential Approval: A Multi-Dimensional Semi-Parametric Approach, *HWWI Research Paper 118*.

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

Chang, Y. (2002). Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics* **110**(2): 261–292.

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

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

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

Dell'Ariccia, G., Schnabel, I. & 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. & Fischer, C. (2010). What can EMU Countries' Sovereign Bond Spreads Tell Us About Market Perceptions of Default Probabilities During the Recent Financial Crisis?, *Deutsche Bundesbank Discussion Paper 11*.

Eisenbeiß, M., Kauermann, G. & Semmler, W. (2007). Estimating Beta-Coefficients of German Stock Data: A Non-Parametric Approach, *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. & von Schweinitz, G. (2013). Predicting Financial Crises: The (Statistical) Significance of the Signals Approach, *Journal of International Money and Finance* **35**(1): 76–103.

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

Favero, C., Pagano, M. & 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. & Semmler, W. (2007). Testing Wage and Price Phillips Curves for the United States, *Metroeconomica* **58**(4): 550–581.

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

Geyer, A., Kossmeier, S. & 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. & 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. & 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. & 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. & 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. & Opsomer, J. (2011). Data-Driven Selection of the Spline Dimension in Penalized Spline Regression, *Biometrika* **98**(1): 225–230.

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

Krivobokova, T. & 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, *Journal of Economic Perspectives* **26**(3): 49–67.

Pástor, L. & Veronesi, P. (2011). Political Uncertainty and Risk Premia, *NBER Working Paper 17464*.

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

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

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

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric Regression*, Vol. 12 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge Univ Press.

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

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

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

Spiriti, S., Eubank, R., Smith, P. W. & 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*.

von Hagen, J., Schuknecht, L. & 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.

Appendix A: Tables and Figures

Table 3.1: 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 3.2: 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 3.4.1 (1); the baseline estimation, Subsections 3.4.2 to 3.4.6 (2); the estimation excluding Italian data, Subsection 3.5.2 (3); and the estimation with country dummies, Subsection 3.5.3 (4). An F-Test for the joint significance of the country dummies is reported for the last estimation.

Table 3.3: 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 3.4: 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.

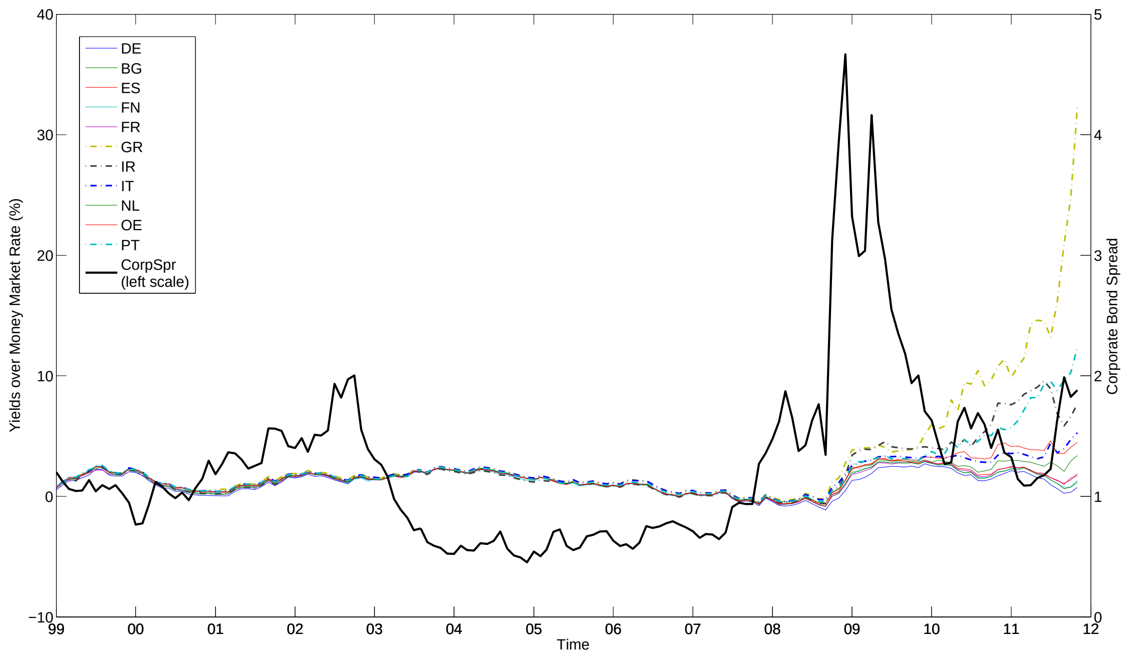


Figure 3.1: European government bond yields over money market rates and the corporate spread (right scale) between Januar 1999 and December 2012. Global uncertainty is measured by the corporate spread (the spread between corporate bond yields of AAA- and BBB-rated US-companies).

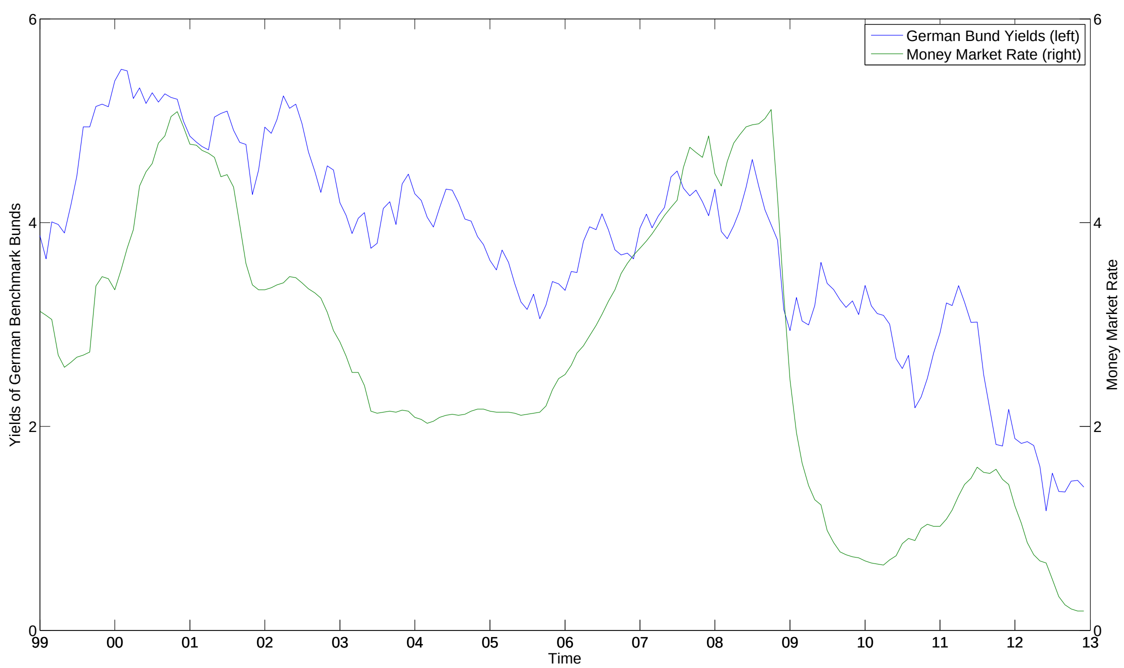


Figure 3.2: Yields of German benchmark Bunds (maturity ten years) (left axis) and three-months money market rate Euribor (right axis).

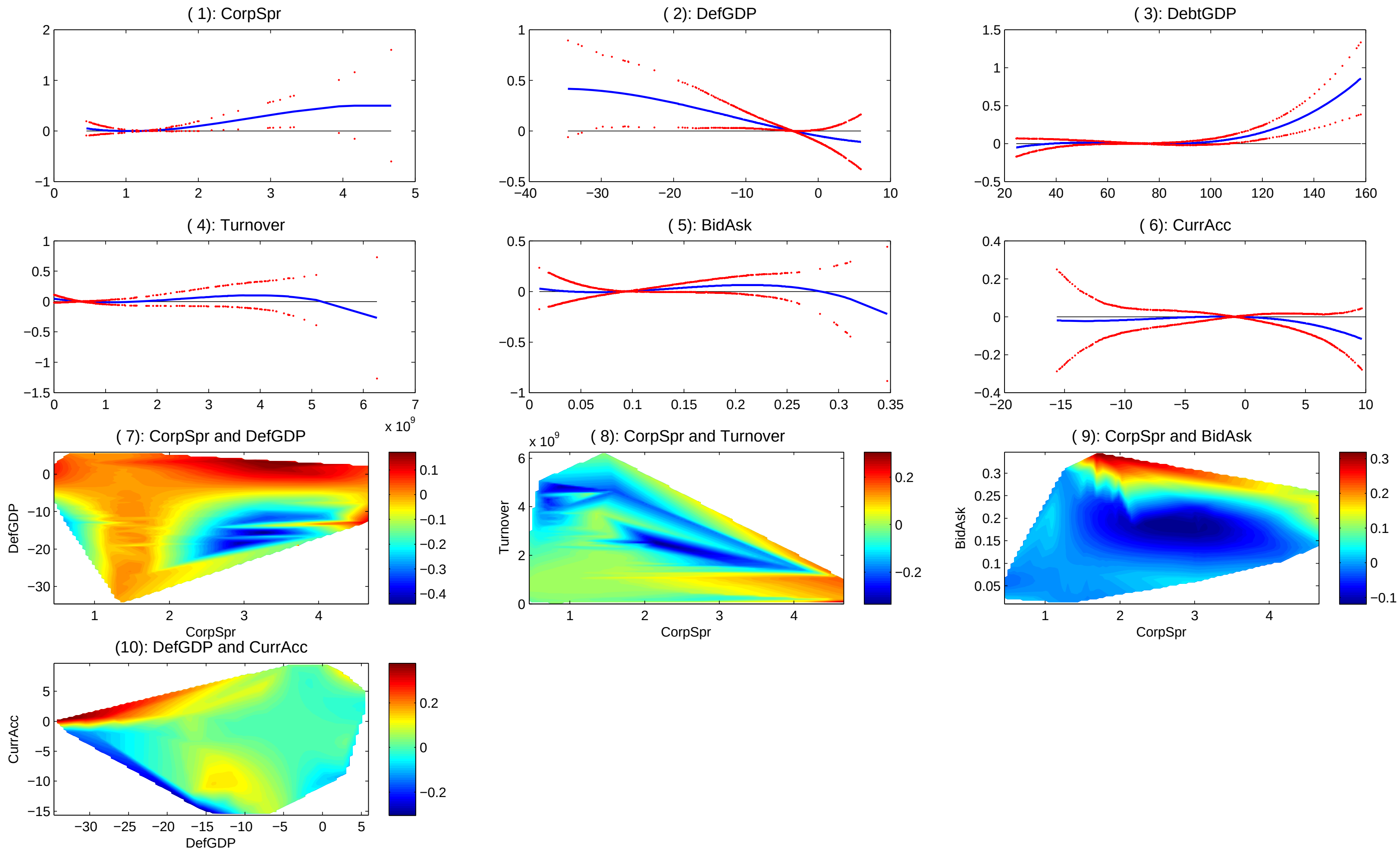


Figure 3.3: Individual and interaction effects for explanatory variables, estimation excluding splines.

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.

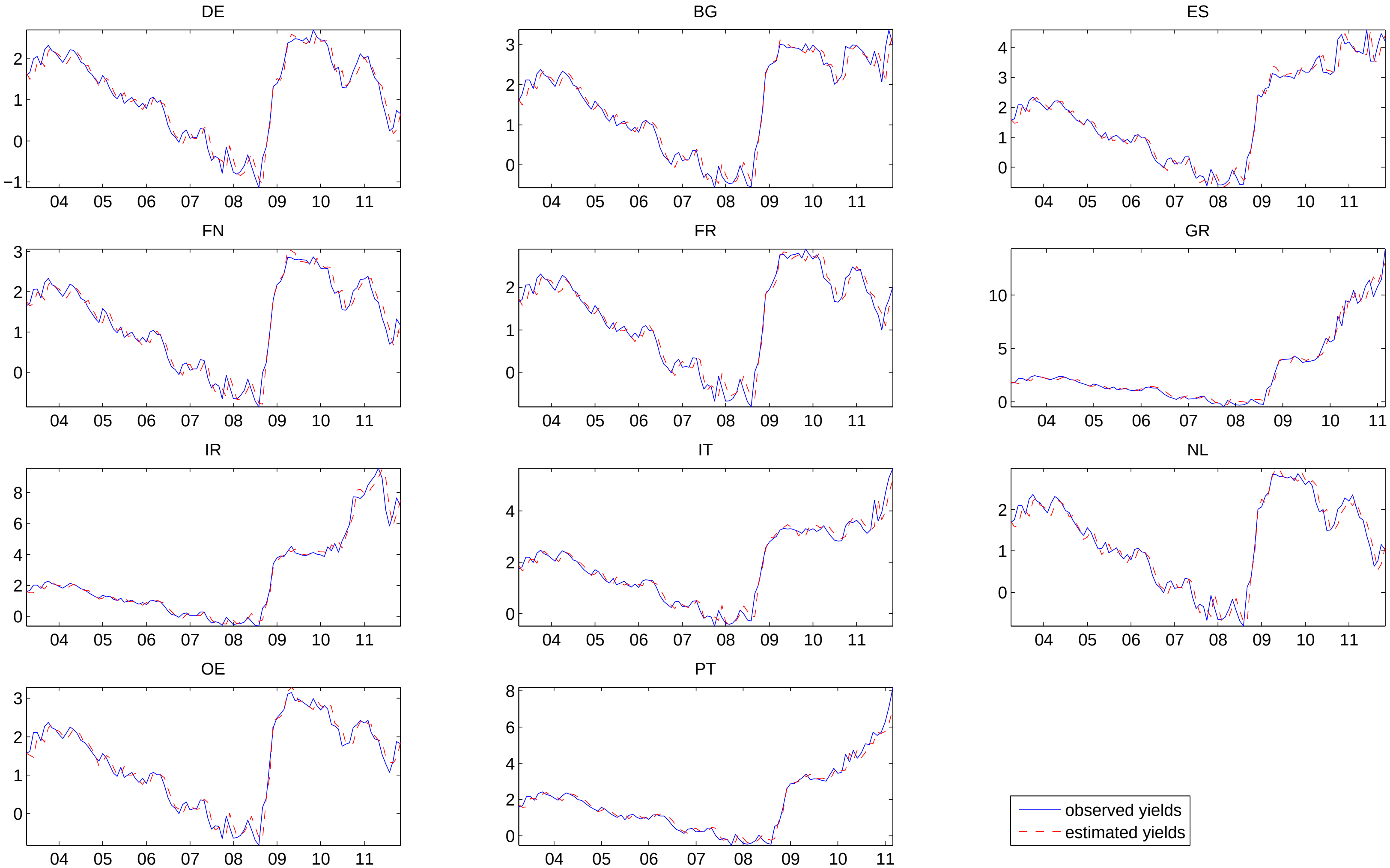


Figure 3.4: Observed and estimated yield spreads (blue/red) for the Euro-12 countries without Luxembourg, baseline estimation.

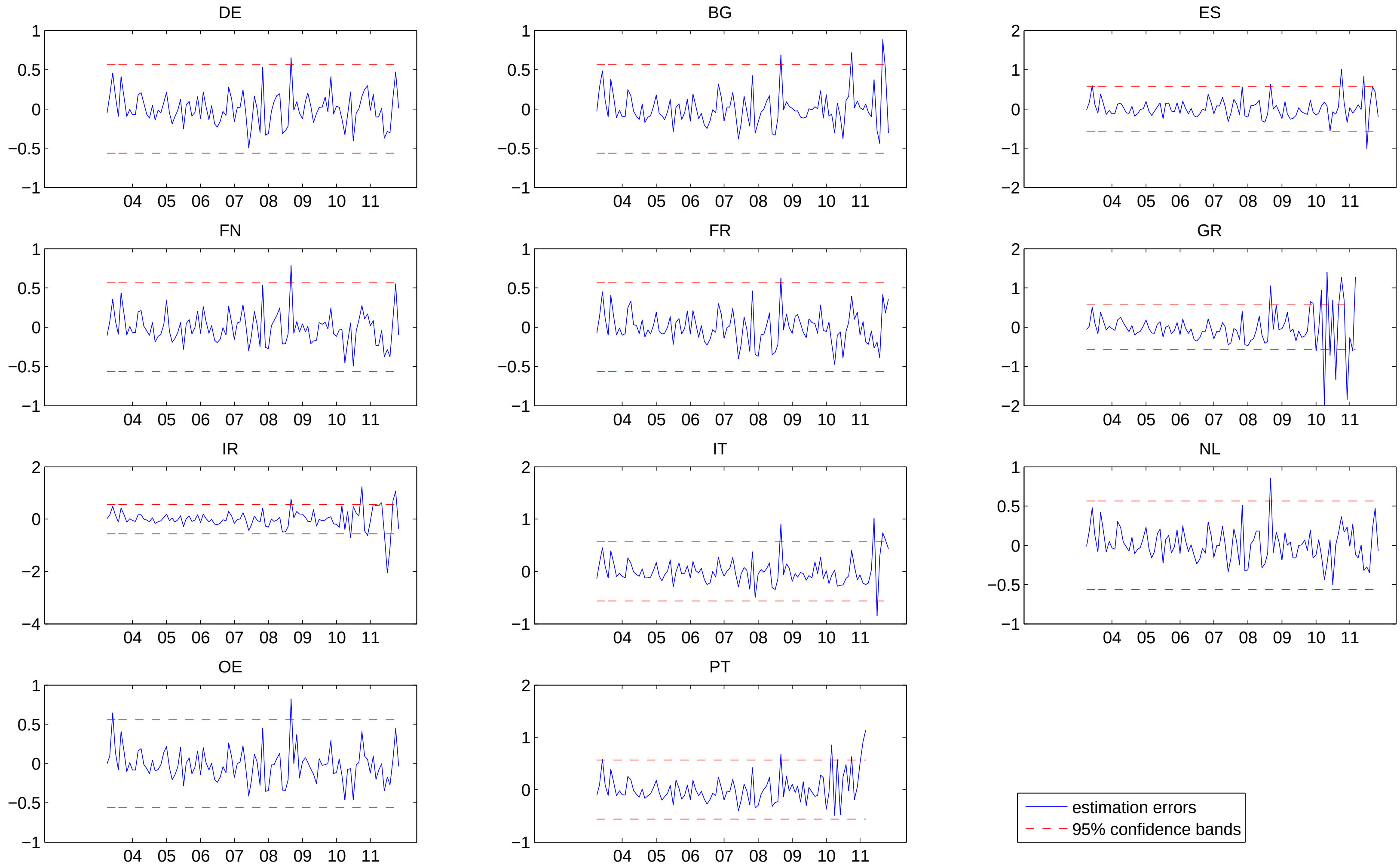


Figure 3.5: Residuals and 95% confidence bands (not adapted to sample size) of the baseline estimation.

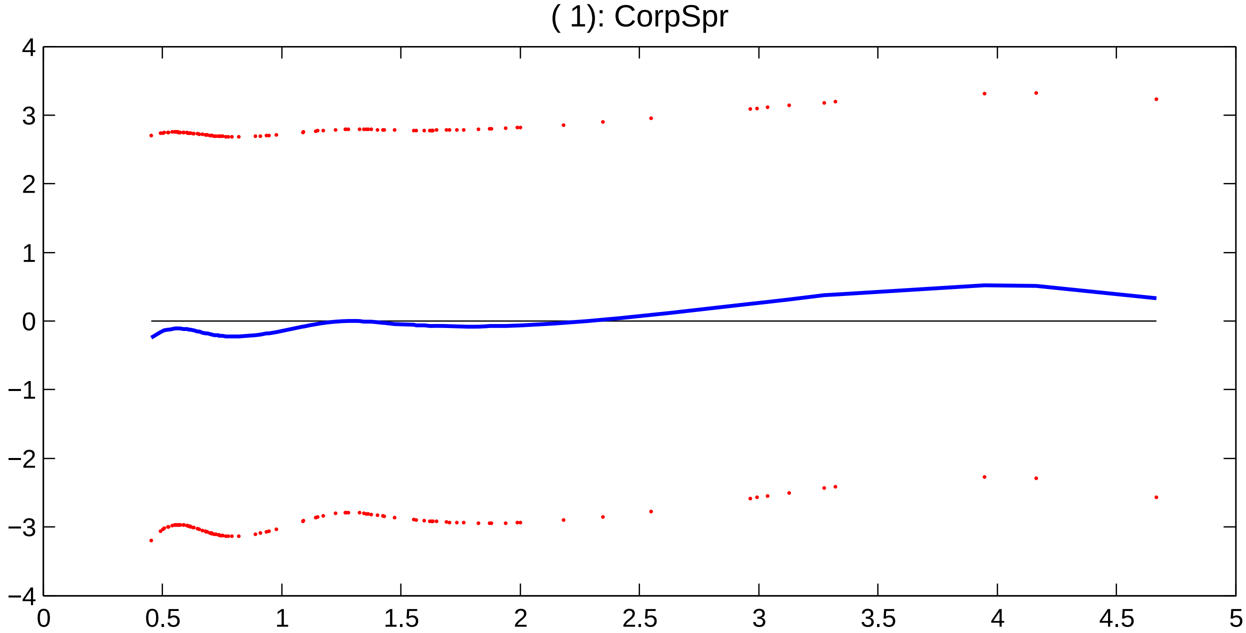


Figure 3.6: Individual effect of global risk, baseline estimation.

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.

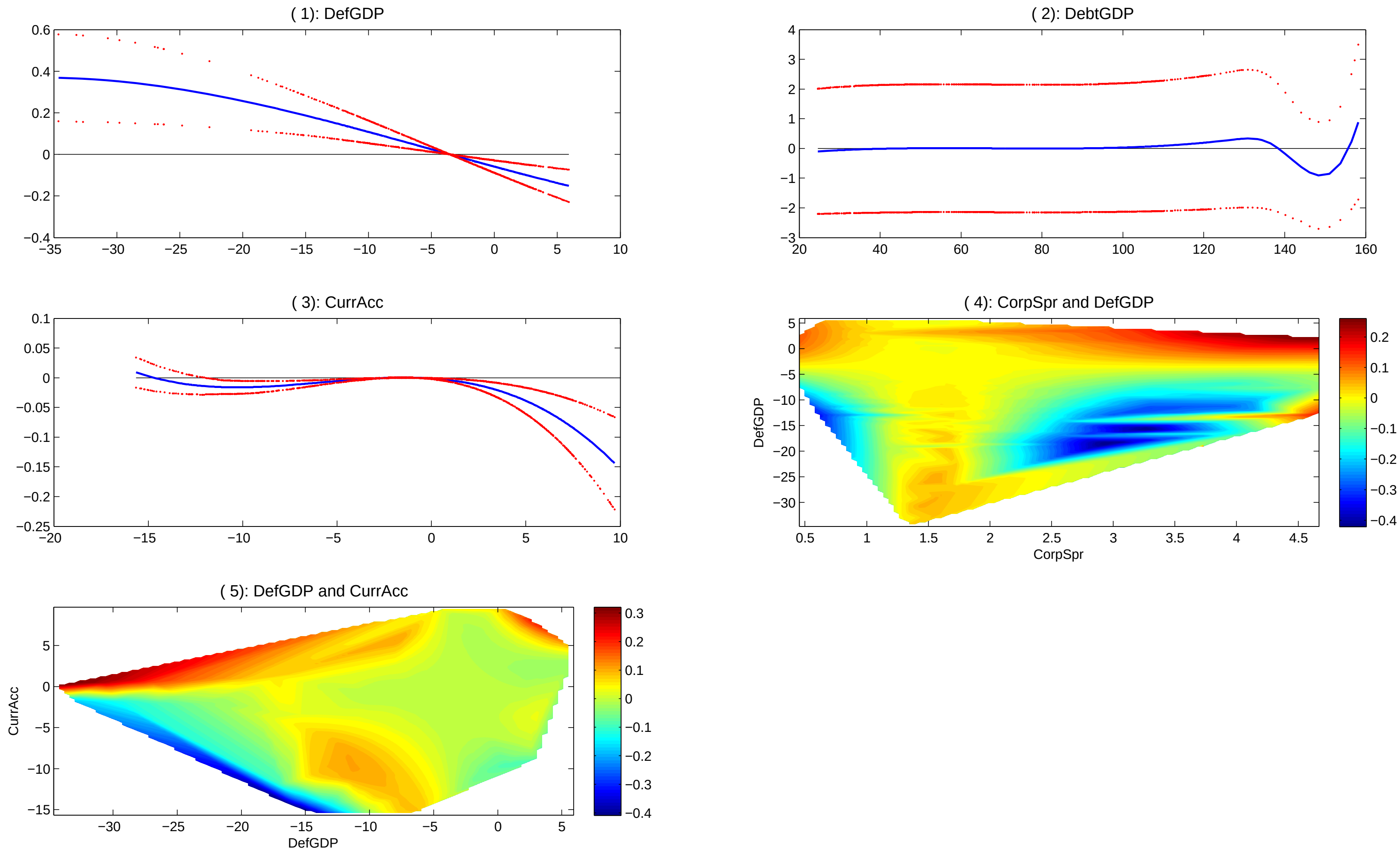


Figure 3.7: Individual and interaction effects for the credit risk variables *GovDef*, *GovDebt* and *CurrAcc*, baseline estimation.

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.

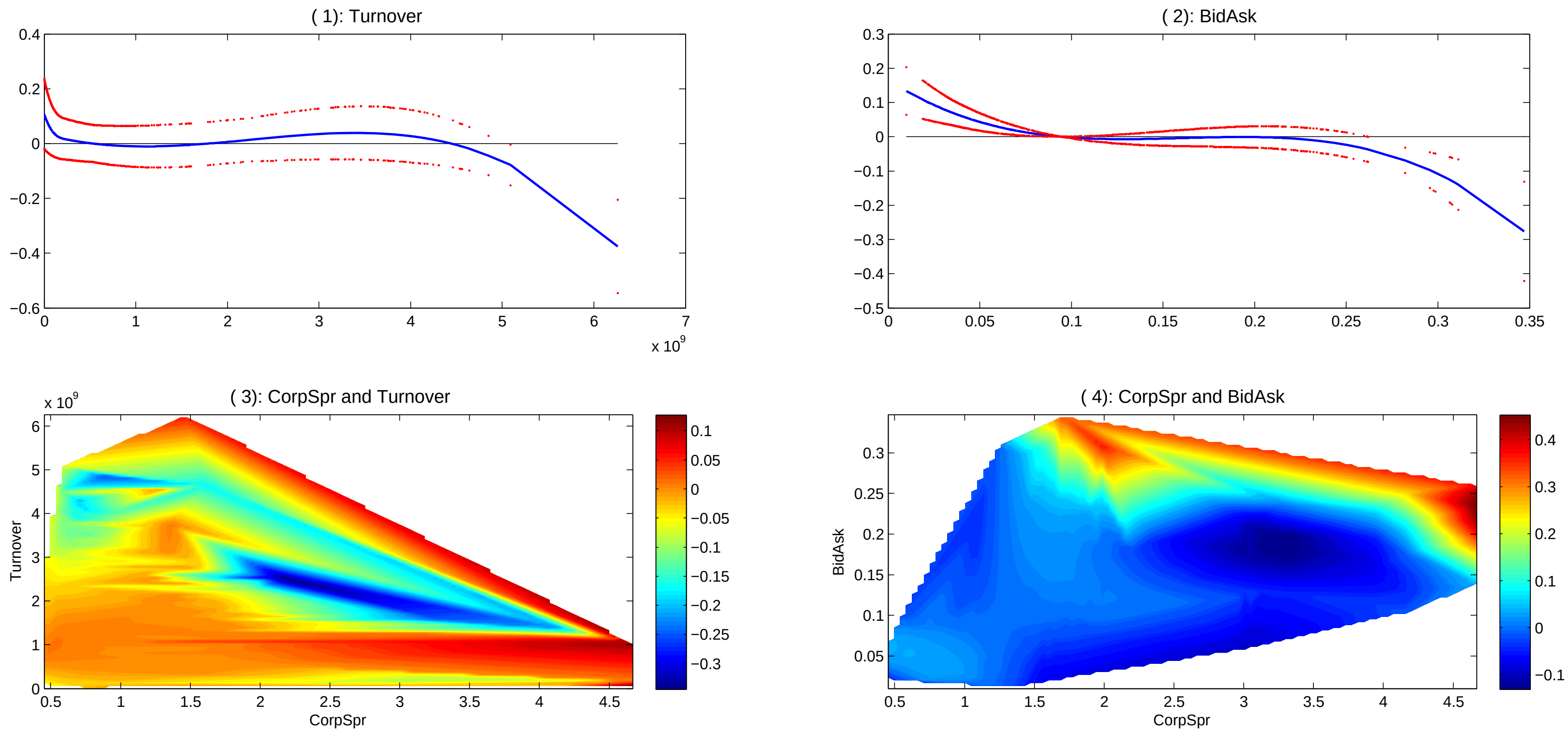


Figure 3.8: Individual and interaction effects for the liquidity risk variables *Turnover* and *BidAsk*, baseline estimation.

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.

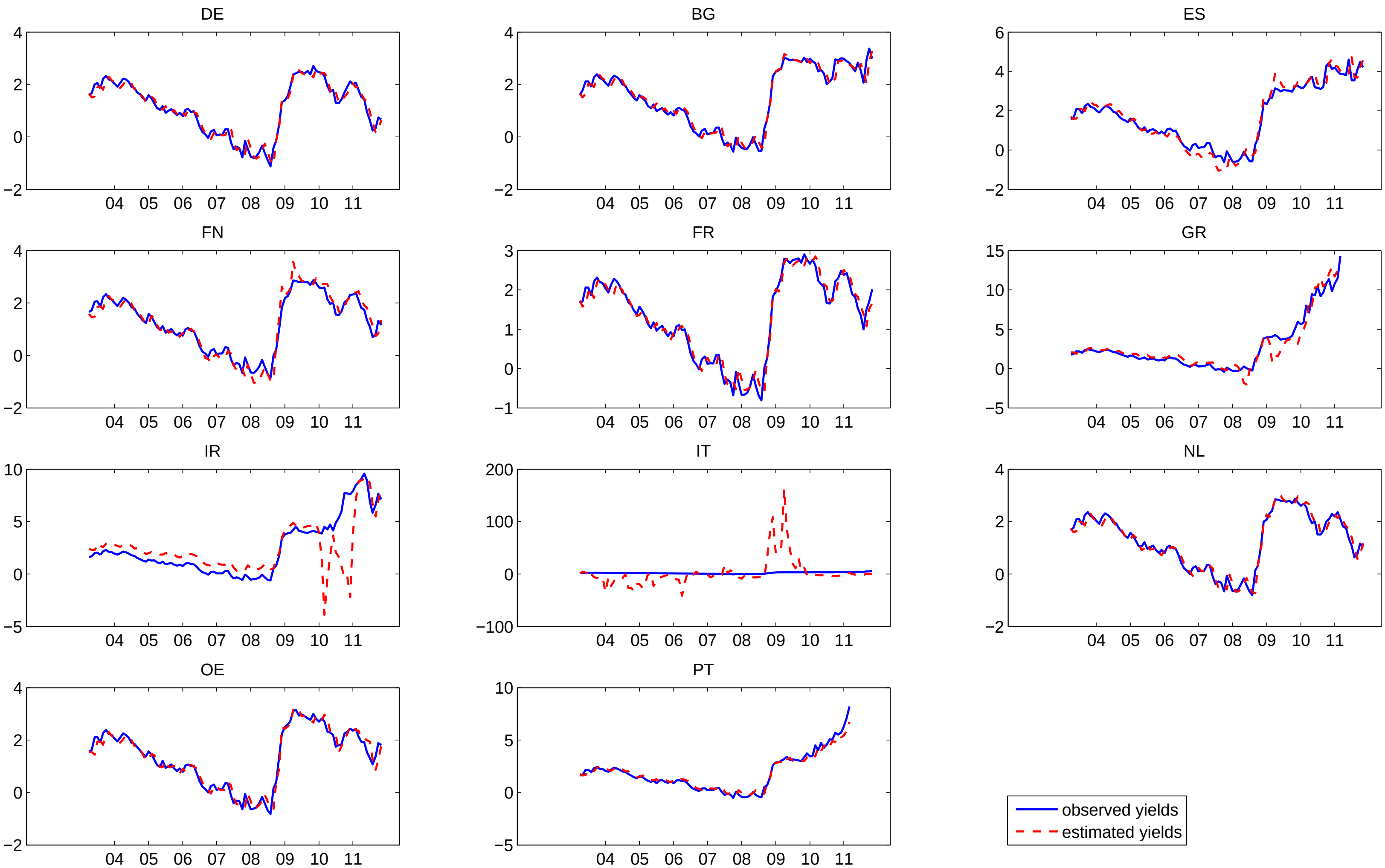


Figure 3.9: Yield spreads in the panel-out-of-sample estimation.

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.

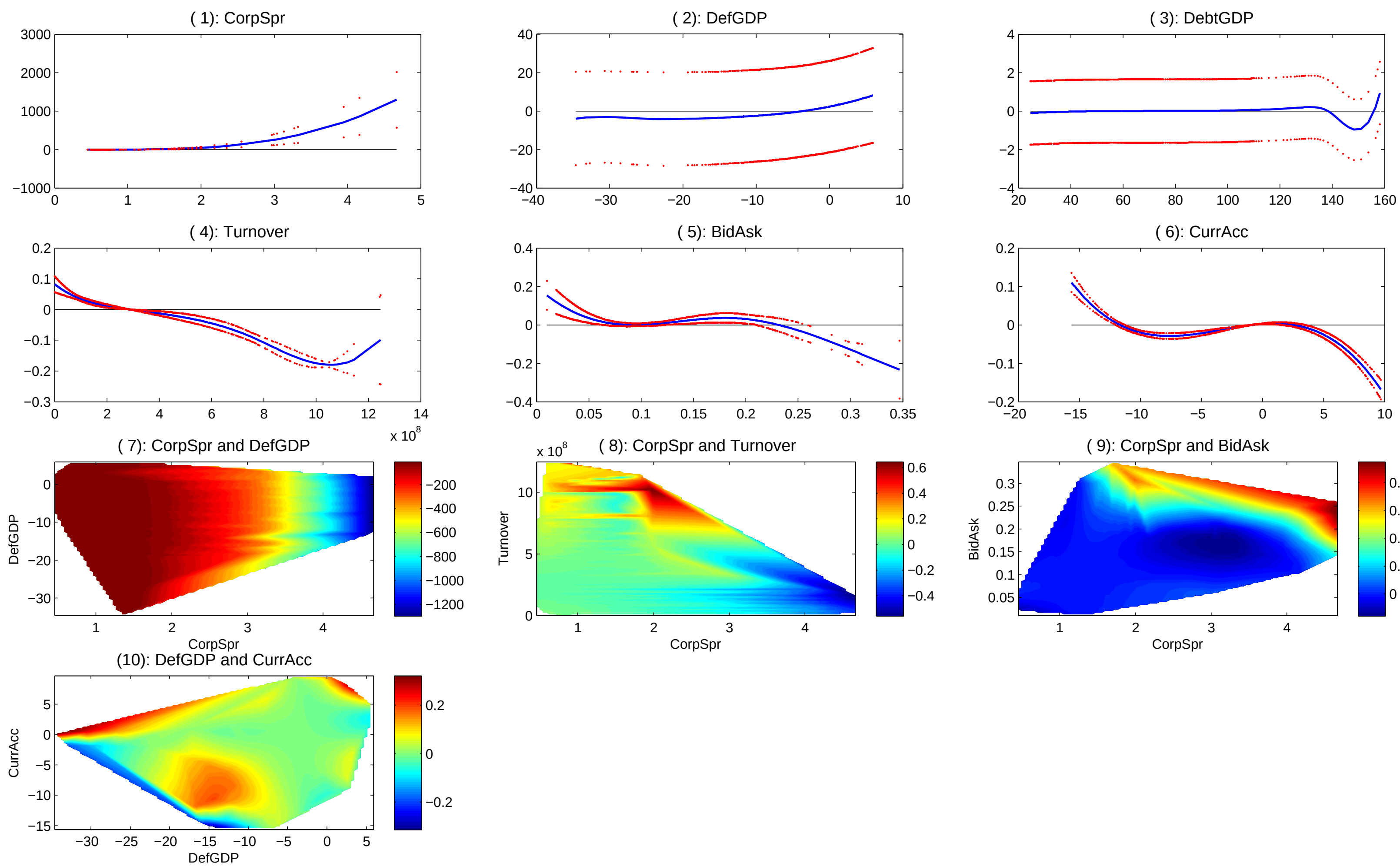


Figure 3.10: Individual and interaction functions for the estimation excluding Italy.

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.

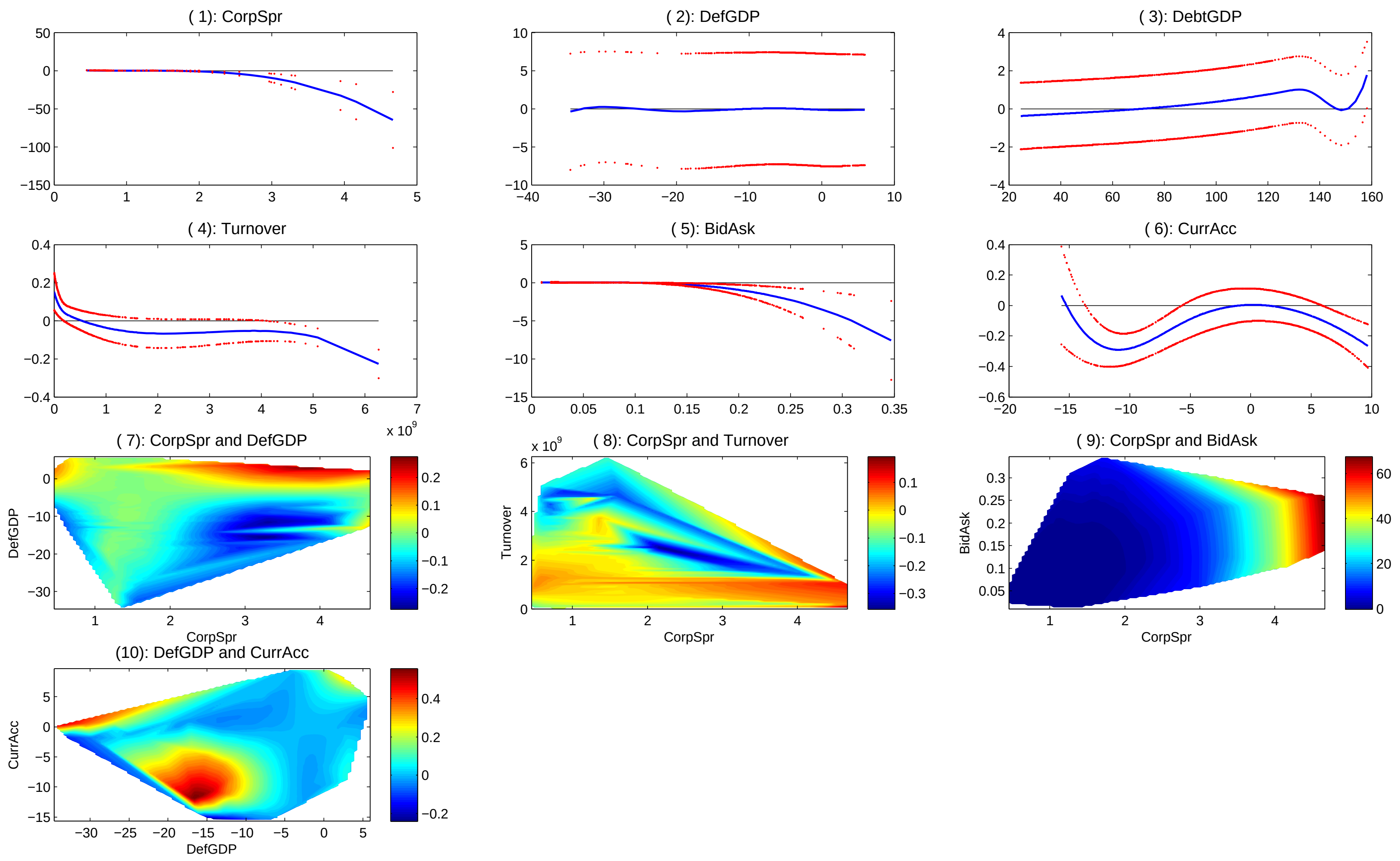


Figure 3.11: Marginal results for an estimation with country fixed effects.

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.

Annex B: Penalized Splines

Penalized Splines are used to estimate an arbitrary, unknown function f with

$$y = f(x_1, \ldots, x_N) + \epsilon.$$

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 & Carroll 1997).[24] Both univariate and bivariate components can be approximated by a third-order Taylor expansion (around x_0):[25]

$$f_i(x_i) = f(x_0) + \frac{\partial f}{\partial x_i}(x_i - x_0) + \frac{\partial f}{2!\partial^2 x_i}(x_i - x_0)^2 + \frac{\partial f}{3!\partial^3 x_i}(x_i - x_0)^3 + \mathcal{O}(x_i^4) \quad (3.3)$$

$$= \sum_{l=0}^{3} \beta_l x_i^l + \varepsilon_i, \quad (3.4)$$

where the equation (3.3) is the normal Taylor expansion and (3.4) 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

$$f_i(x_i) = \sum_{l=0}^{3} \beta_l x_i^l + \sum_{k=0}^{K_i} b_k (x_i - \kappa_{i,k})_+^3 + \varepsilon_i, \quad (3.5)$$

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

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

[25] 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.

$f_{i,j}$ by multiplication of all polynomial and all spline terms (without cross-multiplication). That is, $f_{i,j}$ is given by

$$f_{i,j}(x_i, x_j) = \sum_{l_i=0}^{3}\sum_{l_j=0}^{3}\beta_{l_i,l_j}x_i^{l_i}x_j^{l_j} + \sum_{k_i=0}^{K_i}\sum_{k_j=0}^{K_j} b_{k_i,k_j}(x_i - \kappa_{i,k_i})_+^3(x_j - \kappa_{j,k_j})_+^3 + \varepsilon_{i,j}.^{26} \tag{3.6}$$

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

$$y = X\beta + Zb + \varepsilon, \tag{3.7}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ are iid and $b \sim \mathcal{N}(0, \sigma_\varepsilon^2(\Lambda\Lambda'))$.[27] The objective function to be minimized is

$$\min_{\beta,b,\Lambda} \|y - X\beta - Zb\|^2 + b'(\Lambda\Lambda')^{-1}b, \tag{3.8}$$

where $b'(\Lambda\Lambda')^{-1}b$ is the penalty term giving the method its name.[28] 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 (3.7). For example, dummies or fixed effects only enter as constants. Similarly,

[26] 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.

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

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

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 (3.8), 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 (3.7) is equivalent to a Linear Mixed Model (Ruppert et al. 2003).[29] The linear transformation $b = \Lambda u$, where u is independent of Λ, transforms the objective function to

$$\min_{\beta,u,\Lambda} \|y - X\beta - Z\Lambda u\|^2 + \|u\|^2. \tag{3.9}$$

With this transformation, Bates (2012) achieves to integrate both β and u out of the likelihood function corresponding to the objective function (3.9), 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 (also known as *concentrated* maximum likelihood, Davidson & MacKinnon 2004) 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 (3.9) being only minimized over regression parameters β and u:

$$\tilde{d}(y|\Lambda) = \min_{\beta,u} \|y - X\beta - Z\Lambda u\|^2 + \|u\|^2 = \left\| \begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda & X \\ I_q & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2. \tag{3.10}$$

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:

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

$$L(\Lambda)L(\Lambda)' = \begin{bmatrix} \Lambda & 0 \\ 0 & I_p \end{bmatrix} A \begin{bmatrix} \Lambda & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.11}$$

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):

$$-2l_R(\Lambda|y) = 2\log(|L(\Lambda)|) + (n-p)\left(1 + \log\left(\frac{2\pi\tilde{d}(y|\Lambda)}{n-p}\right)\right), \tag{3.12}$$

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 is rather efficient. The dimension of the parameter space in the likelihood function (3.12) is reduced to the number of different penalty terms. The estimation of $p+q$ regression parameters in the second step is straightforward.

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

$$\sigma_\varepsilon^2 = \frac{\tilde{d}(y|\Lambda)}{n-p}. \tag{3.13}$$

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 & 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 (3.4) and the density of the datapoints x_i (Agarwal & 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 & 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 (3.12).[30] 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

[30] 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.

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 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 (SAGA, Hinterding, Michalewicz & Peachey 1996) with five subpopulations. However, more refined genetic algorithms failed to reproduce stable results and mostly arrived in a local optimum only.