Erasmus School of Economics
Master Thesis Quantitative Finance

Contingent claims approach to measuring sovereign default risk in the Eurozone

Student:
Shahir Anwar

Student number:
532050

Supervisors:
prof. dr. Chen Zhou (Erasmus)
Hao Zhou (Deloitte)
Koen Nicolai (Deloitte)

Second assessor:
dr. Maria Grith

January 30, 2021

Abstract

We examine whether the contingent claims approach (CCA) to measure sovereign default risk for member states of the EMU can be improved by incorporating information from equity and bond markets. Market information is included by using equity and bond volatilities as a proxy for the asset volatility. Combining information from both markets significantly improves the CCA for core countries. For peripheral countries, enhancement of the CCA is only realised in non-crisis periods, as a result of market overreaction.

Keywords— Sovereign default, Structural model, Contingent claims approach, Credit risk, Economic and Monetary Union, Sovereign debt, Default probability

Contents

1 Introduction

Throughout history, sovereign default risk has primarily been perceived as an issue for developing countries. However, the global financial crisis of 2008 and specifically its transition into the European sovereign debt crisis have changed this view. Several member states of the Economic and Monetary Union of the European Union (EMU) were unable to meet their financial obligations and required assistance of third parties. Eventually, this led to defaults of Cyprus and Greece. While some member states are still recovering from this crisis and already have inflated debt levels, the novel COVID-19 virus has forced governments to further increase debt levels by providing stimulus packages. The upcoming crisis may have similar or even worse effects than the global financial crisis of 2008.

Although the previous events have changed the view on sovereign default risk, current policies and regulations do not incorporate this revised view. Under the current Basel regulations, exposures to highly rated sovereigns and to sovereigns denominated and funded in the domestic currency are perceived as risk-free. One of the issues with the current Basel regulations is that these depend on sovereign ratings, which are mainly provided by the "Big Three" global credit rating agencies - Standard & Poor's (S&P), Moody's and Fitch. These agencies have been accused of providing overly favourable credit ratings and being too slow in adjusting ratings. Highly rated sovereigns, above BBB-, are assigned a default probability of 0%. However, Credit Default Swap (CDS) spreads suggest that the default probability is higher than this. Therefore, there is a mismatch between the market's perception of sovereign default risk and ratings based methods. Besides the Basel regulations, most credit loss models also depend on credit ratings. As a consequence, no capital is held for exposures to highly rated sovereigns, which could have detrimental effects. Therefore, it is of utmost importance to model sovereign default risk correctly and close the gap between market and model implied sovereign default risk.

This research aims to determine whether the contingent claims approach (CCA) to measure sovereign default risk for member states of the Economic and Monetary Union of the European Union (EMU) can be improved by incorporating information from equity and bond markets.

The combination of two different strands of literature has led to our research and our motivation is best explained by discussing these. The first strand is about credit risk modeling and the second strand deals with the determinants of sovereign credit risk. Firstly, the available sovereign credit risk models which can be applied to member states of the EMU are very limited. Existing examples are CreditPortfolioView (Wilson, 1998) and CreditMetrics (Morgan, 1997), which depend on transition matrices supplied by credit rating agencies. Many member states of the EMU have high credit ratings, which correspond to a default probability of 0% according to these agencies. These models then do not provide any meaningful results. More recent machine learning techniques cannot be applied either due to insufficient data on defaults.

Our model is based on the sovereign contingent claims approach (CCA) as developed by Gray et al. (2007). They argue that the sovereign balance sheet shows implicit similarities with a corporate balance sheet, by treating local-currency liabilities as equity and foreign-currency liabilities as debt. Unobserved values for the sovereign assets and its volatility are then implied from balance sheet relations. Model implied Risk-Neutral Default Probabilities (RNDPs) are shown to closely follow CDS spreads for Brazil, Turkey, Korea, Mexico, South Africa and the Phillippines. All these countries have large amounts of debt issued in a "hard" currency, Dollars or Euros (foreign-currency liabilities), as well as in their local currencies (local-currency liabilities). Different from Gray et al. (2007), our study is concerned with member states of the EMU, which do not have large amounts of foreign-currency liabilities.

Although not explicitly stated in their research, it has been argued that the CCA only works for emerging countries with a large amount of foreign debt in a "hard" currency, the Dollar or Euro (Duyvesteyn and Martens, 2015). There are two reasons for this, which also give rise to our adaptations of the CCA. First of all, member states of the EMU mainly issue debt in Euros and have little (if any) debt in Dollars. This implies that the seniority of liabilities needs to be rearranged, such that we define a different priority structure. Second, member states of the "EMU have very limited control over the money supply of the European Central Bank (ECB)" (Duyvesteyn and Martens, 2015). The most important consequence of this is perhaps the loss of "information" from the forward exchange rate. The forward exchange rate plays a vital role in the CCA and is crucial to determine the RNDP. The RNDP mainly depends on the amount of debt, the asset value and the asset volatility. Especially the asset volatility is a key component, which is heavily influenced by the forward exchange rate. The forward exchange rate reflects the collective view of market participants on the future of a sovereign. However, the forward exchange rate of the Euro does not contain much "information" for member states of the EMU.

This leads to the other strand of literature which is about the determinants of sovereign credit risk. We argue that "information" can be collected from other sources, namely the equity and bond markets, as the sovereign asset value depends on these markets. In particular, we use the volatilities from these markets as a proxy for the asset volatility. Ang and Longstaff (2013) study the nature of systematic sovereign credit risk for the states in the U.S. and Eurozone countries. Their results show that "systemic sovereign risk is strongly related to financial market variables rather than to macroeconomic fundamentals" (Ang and Longstaff, 2013). Furthermore, Longstaff et al. (2011) study the nature of sovereign credit risk for 26 developed and emerging market countries, none of which is a member state of the EMU. They find "that the majority of sovereign credit risk can be linked to global factors" (Longstaff et al., 2011). In their research, CDS spreads are primarily driven by US equity, high-yield bonds and the VIX index. Country-specific stock returns and bond yields tend to show common patterns, but are not as significant as global factors. However, they conclude that country-specific factors may be more important in less correlated time periods.

We consider the volatility of (future) equity index returns, the VSTOXX index and the volatility of bond yields with a maturity of 1, 5 and 10 years, as proxy for the asset volatility. None of these volatilities will be a perfect proxy for the asset volatility. The bond market will result in a volatility that is too low, whereas the equity market will give a volatility that is too high. A low volatility results in a low RNDP, and vice versa. Therefore, we examine whether a combination of the information in these two markets explains sovereign default risk. Individual predictions of the RNDP, which differ in the asset volatility, are combined using forecast combinations. We have used longer forecast horizons to determine how often weights should be re-calibrated and the impact of these weights on the RNDP. Our estimates for the RNDP based on bond markets, equity markets and a combination of these are compared with the Market-Implied Default Probability that follows from Credit Default Swap (CDS) spreads. We use the Harvey, Leybourne and Newbold (HLN) test to determine whether our method results in a significantly higher accuracy than the benchmarks. As benchmarks, we use the RNDP obtained by applying the standard CCA method and also S&P rating-based default probabilities. The research is conducted for five member states of the EMU, Belgium (BE), France (FR), Germany (DE), Italy (IT) and Spain (ES) between the period of 2008 Q1 - 2020 Q1.

Our results show that the combined approach reduces the MSE for all countries when compared with the benchmarks, the standard CCA and a rating-based method. However, based on the Harvey, Leybourne and Newbold (HLN) test, the combined RNDP forecasts are only significantly more accurate, than both benchmarks and over the entire period, for core countries. For core countries, we obtain the best results for combinations including one equity RNDP and one bond RNDP, with new weights being determined more than once a year. The best combinations are: (C_1) RNDP_1 (equity index) & RNDP_4 (1y bond yield); (C_2) RNDP_1 (equity index) & RNDP_6 (10y bond yield); (C_3) RNDP_2 (future index) & RNDP_5 (5y bond yield). Furthermore, the HLN test shows that the combined RNDP forecasts are also significantly more accurate than both benchmarks for peripheral countries for the period of Q2 2015 - Q1 2020. For peripheral countries during this period, the best combinations are (C_1) RNDP_1 (equity index) & RNDP_4 (1y bond yield); (C_2) RNDP_1 (equity index) & RNDP_6 (10y bond yield). Our analysis shows that our approach is not optimal for peripheral countries in times of crisis due to market overreaction.

The structure of this paper is as follows: Section 2 contains the methodology of this research. In this section we describe the CCA and our adaptations, the incorporation of market information and the evaluation procedure. In Section 3 we describe the data that has been used. In Section 4 we show and analyse our results. Finally, a discussion and concluding remarks can be found in Section 5.

2 Methodology

In this section we describe the methodology of this research. We first describe the sovereign contingent claims approach (CCA) in Section 2.1. In Section 2.2 we show how the asset value and its volatility are computed for the standard CCA model, which we use as the benchmark model. We then describe how market information is incorporated into the CCA in Section 2.3. Finally, the evaluation procedure is described in Section 2.4.

2.1 Sovereign contingent claims approach (CCA)

A contingent claim is any financial asset for which the future payoff depends on the value of another asset. Gray et al. (2007) state that the CCA is based on three principles: (i) the values of liabilities are derived from assets; (ii) liabilities have different priority (i.e. senior and junior claims); and, (iii) assets follow a stochastic process. We describe the implementation of the CCA by discussing each principle separately.

2.1.1 The values of liabilities are derived from assets

Structural models are based on fundamental balance sheet information. Assets and liabilities of a sovereign depend on two entities, the government and the monetary authorities. Therefore, it is necessary to combine the balance sheets of these two entities and construct a consolidated balance sheet for the sovereign. We follow the methodology as described by Gray et al. (2007) and Gapen et al. (2008) to obtain such a balance sheet. The segregated balance sheets of the monetary authorities and the government are depicted in Tables 1a and 1b, respectively.

Certain items need to be rearranged, such that we obtain an economic sovereign balance sheet which only includes liabilities with observable quantities and market prices. The value of these liabilities should be derived from the assets. For this reason, open positions between the government and the monetary authorities are netted and guarantees provided to institutions of systemic importance are subtracted from the asset side. Finally, the balance sheet needs to be denominated in a common currency, which is the Euro in our case. The resulting economic balance sheet is given in Table 1c.

Table 1: Simplified representation of the balance sheets of the public sector.

(a) Balance sheet of the Monetary Authority.

Assets	**Liabilities**
Foreign Reserves	*Obligation to supply FX to Government to pay FX Debt*
Credit to Government	Base Money
Credit to other Sectors	

(b) Balance sheet of the Government.

Assets	**Liabilities**
Net Fiscal Asset	Guarantees (to too-important-to-fail entities)
Other Public Sector Assets	Foreign-currency Debt
Obligation from Monetary Authority to supply FX to Government to pay FX Debt	Local-currency Debt held outside of the Government and the Monetary Authorities
	Credit from Monetary Authorities

(c) Combined balance sheet for the entire public sector, the Sovereign.

Assets	**Liabilities**
Foreign Reserves	
Net Fiscal Assets	Foreign-currency Debt
Other Public Sector Assets	Local-currency Debt
- Guarantees	Base Money

Note. The assets and liabilities of the monetary authority and the government include items which cancel each other out when combining the two balance sheets. These items are stated in italic; The liabilities side includes equity-like and debt-like items. These items can be used to derive the market value of assets.

2.1.2 Liabilities have different priority

Another crucial part of the CCA is the definition of seniority within the sovereigns' capital structure. "Seniority of sovereign liabilities is not defined through legal status as in the corporate sector, but may be inferred from examining government behaviour" (Gray et al., 2007). Earlier literature on the debt restructuring of defaulted emerging countries suggests that foreign-currency liabilities are senior to local-currency liabilities (Ariyoshi et al. (2000), Kincaid and Collyns (2003)). Furthermore, Sims (1999) argues that domestic currency debt has similar characteristics when compared to equity issued by a company. Therefore, Gray et al. (2007) and Gapen et al. (2008) model the implicit capital structure as follows. Local-currency liabilities (LCL), the sum of the local-currency debt and the monetary base, represent the "equity" of a sovereign. Duyvesteyn and Martens (2015) explain the intuition behind this as follows: the public sector may create more domestic currency to repay their local-currency debt. However, printing more money may cause inflation such that the real value of payments to local-currency debt holders becomes lower. This effect is somewhat similar to the dilutive effect of issuing additional stocks. The equivalence between foreign-currency debt and a firm's risky debt follows from two reasons. Firstly, a sovereign cannot easily issue more foreign-currency as this will depreciate the foreign exchange rates. Second, most countries prefer to raise local-currency debt over defaulting on foreign-currency debt. As a result, foreign-currency debt is considered to be senior to local-currency debt.

There are two reasons why this priority structure does not work for member states of the EMU. First of all, developed countries can issue debt in their domestic currency due to liquid markets and therefore have a small amount of foreign-currency debt (Duyvesteyn and Martens, 2015). Second, individual member states of the EMU do not have much influence on the ECB's money supply. They do not have the possibility to increase their local-currency debt in order to pay off foreign-currency debt (De Grauwe (2012), Cochrane (2005) and Kopf (2011)).

Kahlert et al. (2017) define the priority structure by holder of the liabilities, instead of the currency of denomination, for member states of the EMU. Multilateral lenders such as the IMF, are then assumed to be the most senior credit holders. The logic behind this is that in future crisis, this lender of last resort may be needed again to borrow further (Steinkamp and Westermann, 2014). However, this assumption is contradicted by the fact that Greece defaulted on its IMF debt in 2015, while respecting other payments. Therefore, we propose a different priority structure. Senior claims include all debt, both local-currency and foreign-currency, held outside of the monetary authorities. Only the monetary base has a junior status.

2.1.3 Assets follow a stochastic process

The stochastic process for the asset value is given by,

$$\frac{dA_t}{A_t} = \mu dt + \sigma_A dW_t,$$

where A_t is the asset value of a sovereign at time t, μ is the expected growth rate of the assets, σ_A is the asset volatility and W_t is a Brownian motion $W_t \sim N[0, 1]$. The solution is given by,

$$A_t = A_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{t}W_t}, \tag{1}$$

which can be used to simulate the asset value of a sovereign.

The CCA is based on the Merton model (Merton, 1974). A sovereign defaults when the value of its assets is below a certain threshold, the distress barrier. The distress barrier, B_f, is based on the book value of senior liabilities and is equal to the short-term debt plus half of the long-term debt. On average, the sovereign will not default as the value of the assets will be above the distress barrier. The Risk-Neutral Default Probability (RNDP) is calculated by,

$$P(A_T \leq B_f) = \Phi\left(\frac{\log(B_f) - \log(A_0) - (r - \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}\right) =: \Phi(-d_2), \tag{2}$$

where r is the risk-free rate and $\Phi(\cdot)$ is the cumulative distribution of a standard Gaussian. We consider a time period of 1 year, such that $T = 1$. Hence, to calculate the RNDP we need the asset value, A_0, and its volatility, σ_A.

2.2 Standard CCA

Equations (1) and (2) of the Merton model require estimating the market value of the assets and its volatility, which are unobserved. Asset values are derived from the observable junior liabilities, the monetary base, and senior liabilities, all debt held outside of the monetary authorities. As the claim of senior liability holders has preference over the claim of junior liability holders, the monetary base is equal to the residual value of the assets after the debt payments have been made. The monetary base, MB, is modeled as an implicit call option on the market value of the sovereign assets, A, with the strike price equal to the distress barrier, B_f. The distress barrier is equal to short-term senior liabilities plus half of the long-term senior liabilities. We then obtain two equations and need to solve for two unknowns, the initial market value of the assets at the start of a year, A_0, and its volatility, σ_A.

$$MB_0 = A_0\Phi(d_1) - B_f e^{-rT}\Phi(d_2) \tag{3}$$

$$\sigma_{MB} = \frac{A_0}{MB_0}\sigma_A\Phi(d_1) \tag{4}$$

where,

$$d_1 = \frac{\log(A_0) - \log(B_f) + (r + \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}, \quad d_2 = d_1 - \sigma_A\sqrt{T}$$

and σ_{MB} is the volatility of the returns of MB. We further denote,

$$F(A_0, \sigma_A) := A_0\Phi(d_1) - B_f e^{-rT}\Phi(d_2) - MB_0$$
$$G(A_0, \sigma_A) := \frac{\sigma_A}{\sigma_{MB}}\Phi(d_1)A_0 - MB_0.$$

Theoretically, $F(A_0, \sigma_A) = G(A_0, \sigma_A) = 0$. Practically, we estimate A_0 and σ_A by minimizing

$$[F(A_0, \sigma_A)]^2 + [G(A_0, \sigma_A)]^2, \tag{5}$$

which is solved using an iterative process. This is the standard CCA model and is used as a benchmark. The only difference between the standard CCA and the method defined by Gray et al. (2007) is that we only consider the monetary base as the "equity" value. For emerging countries, all local-currency liabilities are considered to be "equity". The reason for not using the exact same model as a benchmark is that member states of the EMU have a negligible amount of foreign-currency liabilities (in Dollars). Moreover, rating agencies treat the Euro as both a local-currency as well as a foreign-currency liability for member states of the EMU.

2.3 Incorporating market information

The difference in the definition of the "equity" value has a large impact on the asset volatility and eventually the RNDP, as the RNDP measure depends heavily on the asset volatility. For emerging countries, the asset volatility is determined by the volatility of the local-currency liabilities. "If the exchange rate is floating the volatility comes largely from the exchange rate." (Gray et al., 2007).

Gray et al. (2007) use the forward exchange rate to determine the volatility of the local-currency liabilities (and hence the asset volatility). The exchange rate depends on the supply and demand of a currency, which in turn depends on various factors, including investment opportunities, political situation et cetera. Therefore, the forward exchange rate contains much "information" about the collective view of market participants and their expectations of the future development of a sovereign. Although the market is not always right, it is assumed to be the best available source of information for the future prospects of a sovereign (Gapen et al., 2008).

For developed countries with a "hard" currency, the Euro or Dollar, we cannot obtain such information from the exchange rate. Moreover, changes in the monetary base returns will only be reflected in the volatility when these are reported, which may be too late. We suggest to incorporate relevant market information from other sources. Longstaff et al. (2011) study the nature of sovereign credit risk for 26 countries. They find "that the majority of sovereign credit risk can be linked to global factors" (Longstaff et al., 2011). Likewise, Ang and Longstaff (2013) study the nature of systematic sovereign credit risk for states in the U.S. and Eurozone countries. Their results show that "systemic sovereign risk is strongly related to financial market variables rather than to macroeconomic fundamentals" (Ang and Longstaff, 2013). Therefore, we include information from equity and bond markets in our models. Individual RNDP forecasts are based on either equity or bond volatilities, whereas forecast combinations are based on both.

2.3.1 Individual forecasts

The minimization procedure given in Equation (5) depends on both the asset value, A_0, and its volatility, σ_A. Hence, there are many combinations of A_0 and σ_A which satisfy Equations (3) and (4). We incorporate market information by setting σ_A equal to the volatility obtained from equity or bond markets. Let:

$$\sigma_A = \sigma_j \quad \text{for } j = \{1, \ldots, 6\},$$

where we define (1) the volatility of the equity index returns, σ_1; (2) the volatility of the future equity index returns, σ_2; (3) the level of VSTOXX (the implied volatility of the Euro STOXX 50), σ_3; (4) the volatility of the 1-year bond yields, σ_4; (5) the volatility of the 5-year bond yields, σ_5; (6) the volatility of the 10-year bond yields σ_6. All these volatilities are annualized over the daily returns or yields using a rolling window of 60 business days.

We then determine the initial asset values, $A_{0,j}$ for $j = \{1, \ldots, 6\}$. In theory, we obtain the asset values by setting $F(A_{0,j}, \sigma_A) = 0$. Practically, we estimate $A_{0,j}$ by minimizing $[F(A_{0,j}, \sigma_A)]^2$. Hence, we obtain 6 sets of asset values and volatilities. These are used to calculate the individual forecasts RNDP_j, $j = \{1, \ldots, 6\}$. RNDP_1, RNDP_2 and RNDP_3 are based on the volatility from equity markets, which we call equity RNDPs. Likewise, RNDP_4, RNDP_5 and RNDP_6 are named bond RNDPs as these are based on the volatility from bond markets.

We do not expect that individual forecasts will have a higher accuracy than the standard CCA. However, we do think that a combination of these forecasts may improve the standard approach significantly. The reason for this is that the RNDP measure depends heavily on the asset volatility and the two are positively related. A higher asset volatility implies a higher RNDP whereas a lower asset volatility results in a lower RNDP, *ceteris paribus*. The volatility of bond yields will probably be too low, whereas the volatility of the equity index returns may be too high. The actual asset volatility is probably between the two. Eventually, we want to determine whether there is a linear combination of forecasts that performs significantly better than the standard approach. Ideally, we would only want to include two forecasts in the linear combination, one based on bond yield volatility and the other based on the volatility of equity returns.

Forecast selection

To determine which of the individual RNDPs are included in the linear combination, we regress the Market-Implied Default Probability on the individual RNDPs. Under the assumption that CDS spreads represent pure default risk (Longstaff et al., 2005), the Market-Implied Default Probability is given by,

$$PD_{M,t} = \frac{1 - e^{-S_{CDS,t}}}{1 - RR}, \tag{6}$$

where $S_{CDS,t}$ is the CDS spread at time t and RR is the recovery rate which is set equal to 50%. The linear regression is then given by, $PD_M = \alpha + RNDP'\beta + \epsilon$. The selection of individual RNDPs to include in the linear combinations is based on the R^2 and the Akaike Information Criterion (Akaike, 1998). The Akaike Information Criterion is given by, AIC $= 2k - 2\ln(\hat{L})$, where k is the number of estimated parameters and $\hat{L}$ is the maximum value of the likelihood function for the model. The likelihood is given by: $L(\beta, \sigma^2; PD_M, RNDP) = (\frac{1}{\sqrt{2\pi\sigma^2}})^T e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(PD_{M,t} - RNDP_t'\beta)^2}$, where T is the number of time periods (observations). The models with the highest R^2 and/or the lowest AIC are optimal.

2.3.2 Combined forecasts

For each country at time t, the individual forecasts are combined by,

$$RNDP_{C_i,t} = \sum_{j \in C_i} w_{j,t} * RNDP_{j,t},$$

where $i = \{1, \ldots, 8\}$ and $w_{j,t}$ is the weight for the RNDP$_j$ at time t. Hence, we define 8 different combinations of the individual forecasts. The ultimate goal is to obtain a simple combination of forecasts, without compromising too much on the accuracy. The simple combination should include as less individual forecasts as possible. Ideally, we would only want to include two individual RNDPs in the linear combination, one based on bond yield volatility and the other based on the volatility of equity returns. Therefore, C_1, C_2, C_3 and C_4 only include 2 individual forecasts; C_5 and C_6 combine 3 individual forecasts. For comparison, we also define C_7 which includes 5 RNDPs and C_8 which includes all 6 RNDPs. The 8 combinations that we consider include following individual RNDPs: (C_1) 1, 4; (C_2) 1, 6; (C_3) 2, 5; (C_4) 3, 4; (C_5) 1, 4, 6; (C_6) 2, 4, 5; (C_7) 1, 2, 3, 5, 6; (C_8) 1, 2, 3, 4, 5, 6.

The weight for each forecast j at time t, is given by,

$$w_{j,t} = \frac{(\sum_{\tau=t-v}^{t-1} |\varepsilon_{j,\tau}|)^{-1}}{\sum_{k \in j}(\sum_{\tau=t-v}^{t-1} |\varepsilon_{k,\tau}|)^{-1}}, \tag{7}$$

where $j \in C_i$ represents the different combinations of asset values and volatilities. Which RNDP$_j$ is included, depends on the combination, C_i. For example, for C_8 we would have $j = \{1, \ldots, 6\}$. For each country, the forecast error at time t is given by $\varepsilon_{j,t} = PD_{M,t} - RNDP_{j,t}$. Hence, a larger weight will be assigned to predictions with a smaller forecast error and vice versa. We set $v = 3$, such that the weights are based on a rolling window which includes the three most recent forecast errors. Predictions and forecast errors are determined on a quarterly basis. This gives different weights for each country and each time period.

In an ideal world, the weights that the individual forecasts receive would be constant across the entire time period and hold for all countries. We do not expect this to be the case for two reasons. Firstly, we expect the weights to differ in times of crisis. In times of crisis, CDS spreads are higher and financial markets tend to show increasing correlations (Ang and Bekaert, 2002). Hence, we would expect that the models based on equity volatility would receive a higher weight in such periods. Whereas, in periods of economic stability, the forecasts based on bond yield volatility would receive a higher weight. Second, we expect a difference between core and peripheral countries. Core countries within the Eurozone are Austria, Belgium, Finland, France, Germany and the Netherlands. Peripheral countries are Cyprus, Greece, Ireland, Portugal and Spain. Peripheral countries are perceived to be riskier and hence we would expect that the models based on equity volatility would receive a higher weight for these countries than for the core countries. We determine the impact of the weights by forecasting for longer horizons.

2.3.3 Forecasting for longer horizons

We examine the performance of the combinations when forecasting for longer horizons. There are two reasons for this. Firstly, we want to know whether the standard CCA can be improved when forecasting for longer horizons. Second and most importantly, if the combined forecasts result in a significantly higher accuracy than the standard CCA, we want to determine what the main driver of that gain is. Whether the gain is truly achieved by incorporating market information or whether it is merely a result of the way in which the individual forecasts are combined. The combinations depend on the weights of the individual forecasts, which are based on past CDS spreads. The standard CCA model does not include this, hence the combined forecasts have an "unfair advantage" over the standard CCA model. A gain in predictive performance that is mainly a result of the weighting structure is not what we aim to achieve.

For the forecast combinations in Section 2.3.2, we calculate new weights at each time period t. To examine the performance for longer horizons, we fix the weights for a longer time period. More specifically, after calculating $w_{j,t}$ with Equation (7). Let:

$$w_{j,t} = w_{j,t+1} = \cdots = w_{j,t+m},$$

such that we fix the weight for RNDP_j for m periods after t. We use $m \in \{2, 4, 8, 16, 23, 100\}$, which we call the holding period. As we determine forecasts for each quarter, this entails that we set the weights equal for a period of 0.5, 1, 2, 4, 5.75, 25 years. We use 5.75 years as this is exactly half our total time frame. This means that the weights are determined twice. For $m = 100$, the weights are only determined once.

2.4 Evaluation

The eventual goal of this research is to determine whether we can improve the CCA for member states of the EMU by incorporating market information. More specifically, we test whether a model with bond and equity information is more accurate than the standard CCA method in describing sovereign default risk. The model can either be based on volatilities from the bond market, the equity market or a combination of the two. Our standard CCA model is almost identical to the model used by Gray et al. (2007). The only difference is the definition of the "equity" value. Using the exact same model would not provide any meaningful results such that slight adaptations were needed. As our benchmark model is different, we also include a second benchmark based on credit ratings.

The second benchmark is a rating-based method. Credit risk models and Basel regulations depend on these ratings. Therefore, these rating-based methods form the basis for many applications in credit risk. The rating-based method depends on credit ratings and historical transition matrices. Credit ratings are mainly provided by the "Big Three" global credit rating agencies - Standard & Poor's (S&P), Moody's and Fitch. These agencies use various metrics, ranging from financial to political, to assess the credit quality of a sovereign. A sovereign is then assigned a credit rating. We use S&P ratings, where AAA is the highest rating and CC the lowest. Based on historical observations, these credit ratings translate to historical default probabilities, PD_H. S&P provide an annual report in which the sovereign default probabilities are determined for each credit rating. These reports are based on data starting in 1975 until the most recent year. Unfortunately, we could not find this report for every year. Appendix A.2 shows the historical default probabilities corresponding to S&P ratings. We use the 2010 report for the period up to and including 2010. For the remaining years, we use the most recent report.

We use Credit Default Swap (CDS) spreads to evaluate our RNDP forecasts. CDS spreads are considered to be the most direct measure of credit risk. The final evaluation is based on the out-of-sample mean squared error. To determine the MSE, we use the difference between the RNDP forecasts and the Market-Implied Default Probability given in Equation (6). As we use a 3-quarter rolling window to determine the weights in our combinations, only the first three quarters will not be included in the test set. To test whether the various RNDP forecasts have a significantly higher accuracy than the benchmarks, we use the Harvey, Leybourne and Newbold (HLN) test.

2.4.1 Harvey, Leybourne and Newbold (HLN) Test

We test whether our RNDP forecasts have a significantly higher accuracy in forecasting the Market-Implied Default Probability than the two benchmarks, the standard CCA method and the rating-based method. Therefore, we define the actual series $\{PD_{M,t};\ t=1,\ldots,T\}$, the prediction from the standard CCA $\{RNDP_{S,t};\ t=1,\ldots,T\}$, the historical default probability $\{PD_{H,t};\ t=1,\ldots,T\}$, and our forecasts $\{RNDP_{h,t};\ h=1,\ldots,6,C_1,\ldots,C_8;\ t=1,\ldots,T\}$. The forecast error for the standard CCA is given by $e_{S,t}=RNDP_{S,t}-PD_{M,t}$; for the rating-based method this is given by $e_{H,t}=PD_{H,t}-PD_{M,t}$; and for our RNDP forecasts these are given by $e_{h,t}=RNDP_{h,t}-PD_{M,t}$. We then use the squared error loss function $g(e_{i,t})=(e_{i,t})^2$, to define the loss differential between one of our forecasts and the benchmark forecasts as $d_t=g(e_{h,t})-g(e_{B,t})$, where $B\in\{S,H\}$. Under the assumption that the loss differential is covariance stationary, we say that the two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t. The alternative hypothesis is that our forecast is more accurate than the benchmark, such that the loss differential has an expectation larger than zero for all t. Therefore, we test the following:

$$H_0: E(d_t)=0 \quad vs \quad H_1: E(d_t)>0 \ \ \forall t.$$

The HLN test (Harvey et al., 1997) is a modified version of the Diebold-Mariano (DM) test (Diebold and Mariano, 2002). The DM test tends to reject the null hypothesis too often for small samples. The HLN test obtains improved small-sample properties by including a bias correction to the DM test statistic and comparing the statistic with a Student-t distribution with $T-1$ degrees of freedom. Therefore, the HLN test statistic is given by,

$$HLN=\sqrt{\frac{T+1-2h+h(h-1)}{T}}DM \sim t(T-1),$$

where,

$$DM=\frac{\bar{d}}{\sqrt{\frac{\hat{\gamma}_d(0)+2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)}{T}}},$$

$$\hat{\gamma}_d(k)=\frac{1}{T}\sum_{t=|k|+1}^{T}(d_t-\bar{d})(d_{t-|k|}-\bar{d}),$$

hence, the autocovariance at lag k is denoted by $\hat{\gamma}_d(k)$ and we define the mean of the loss differential as $\bar{d}=\frac{1}{T}\sum_{t=1}^{T}d_t$. Lastly, for h-steps ahead forecasts, it is assumed that all autocorrelations of d_t of order h or higher are zero. We determine h with the Tiao-Box procedure (Tiao and Box, 1981) for each country and loss differential series. With this procedure, we search for the lag length beyond which the autocorrelations are between $-2T^{-0.5}$ and $2T^{-0.5}$. We assume that correlations within this range imply that the autocorrelation is essentially zero.

3 Data

In this section we describe the data that is used for this research. In total, 5 countries are considered: Belgium, France, Germany, Italy and Spain. Other member states of the EMU, Austria, Cyprus, Finland, Greece, Ireland, the Netherlands and Portugal, could not be included due to insufficient data. This study is conducted over the period between 2008 Q1 - 2020 Q1, where results are given for the beginning of each quarter. Credit Default Swap (CDS) spreads were only available after 2008, and this time frame includes periods of crisis as well as periods of economic stability.

The balance sheet items that are relevant for the CCA are given in Table 2. Base money is retrieved on a monthly basis from Datastream. Quarterly debt data is found on Eurostat's Quarterly Public Sector Debt database. The total debt includes the following debt items for the General Government: Currency and deposits; Debt securities; Loans; Insurance, pensions, and standardized guarantee schemes; Special Drawing Rights; Other accounts payable. Short-term debt includes all items of the total debt with payments due in one year or less. Long-term debt includes all items of the total debt with payments due in more than one year. Finally, the share of holdings by the public sector have been collected from the Bruegel database of sovereign bond holdings developed in Merler and Pisani-Ferry (2012). Credit ratings have been derived from Bloomberg, where we use the long-term foreign currency debt ratings, supplied by Standard & Poor's. The ratings for each country are given in Appendix A.

Table 2: Descriptive statistics for the balance sheet items included in the Contingent Claim Approach.

Country	Base Money	Debt		Share
		Short-term	Long-term	public
Belgium	$6.7 * 10^{10}$	$6.24 * 10^{10}$	$3.63 * 10^{11}$	0.04
	$(3.61*10^{10})$	$(7.44 * 10^{9})$	$(5.47*10^{10})$	(0.04)
France	$3.94 * 10^{11}$	$4.36 * 10^{11}$	$1.69 * 10^{12}$	0.08
	$(1.99*10^{11})$	$(3.85*10^{10})$	$(3.25*10^{11})$	(0.07)
Germany	$5.62 * 10^{11}$	$3.13 * 10^{11}$	$1.73 * 10^{12}$	0.05
	$(2.74*10^{11})$	$(1.38*10^{11})$	$(1.28*10^{11})$	(0.06)
Italy	$2.03 * 10^{11}$	$3.45 * 10^{11}$	$1.54 * 10^{12}$	0.09
	$(6.35*10^{10})$	$(2.11*10^{10})$	$(2.19*10^{11})$	(0.06)
Spain	$1.63 * 10^{11}$	$1.47 * 10^{11}$	$7.95 * 10^{11}$	0.15
	$(4.83*10^{10})$	$(1.76*10^{10})$	$(2.62*10^{11})$	(0.04)

Note. Data is collected for the period between 28-12-2007 and 30-12-2019. Base money is found on a monthly basis $N = 144$, while other items are stated on a quarterly basis $N = 48$. For each country, the first row represents the mean and the second row (in parentheses) the standard deviation of a certain item. Base money, short-term debt and long-term debt are reported in Euros. Share public shows what share of the total outstanding debt securities are held by the public sector.

To incorporate market information, equity indexes, a global volatility index and bond yields have been included. The equity indexes that are used in this research are the largest and most liquid indexes per country. The following equity indexes are included, BEL20 for Belgium, CAC for France, DAX for Germany, FTSEMIB for Italy and IBEX for Spain. The future indexes are, BE1 for Belgium, CF1 for France, GX1 for Germany, ST1 for Italy and IB1 for Spain. VSTOXX is used as a proxy for the global implied volatility index. VSTOXX is the implied volatility based on options on the Euro STOXX 50 index. This index tracks the 50 largest companies in Europe. Furthermore, sovereign bond yields with maturities of 1, 5 and 10 years are considered and the 12 month LIBOR is used as a proxy for the risk-free rate, r. Daily data on all indexes and bond yields has been collected from Bloomberg. Tables 3 and 4 show descriptive statistics for the index data and bond data, respectively. Lastly, we have used CDSs with a maturity of 1 year. These CDSs are denominated in Euros, rather than Dollars. Although Dollar-denominated CDSs are more liquid, these also include exchange rate risk and therefore do not truly reflect sovereign default risk. Descriptive statistics on the CDS spreads are given in Table 5.

Table 3: Descriptive statistics for the equity indexes.

Country	Index			Future		
	Mean	SD	N	Mean	SD	N
Belgium	3056.06	651.22	3071	2521.50	757.97	3076
France	4317.83	780.38	3071	3527.24	998.33	3076
Germany	8873.07	2649.66	3042	8938.47	2549.67	3058
Italy	20418.29	3988.09	3044	16061.91	3217.47	3044
Spain	9752.57	1387.48	3064	7364.09	1323.19	3064
VSTOXX	23.27	9.33	3052			

Note. Data is collected from Bloomberg for the period between 28-12-2007 and 30-12-2019 on a daily basis. All data is shown in Euros, except for VSTOXX which is in percentages. SD represents the standard deviation and N the number of observations.

Table 4: Descriptive statistics for the bond yields.

Country	1-year			5-year			10-year		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Belgium	-0.15	0.57	2198	1.35	1.58	3130	2.18	1.54	3130
France	0.34	1.18	3117	1.13	1.33	3130	1.98	1.36	3130
Germany	0.26	1.20	3123	0.82	1.31	3128	1.54	1.34	3127
Italy	1.02	1.35	3126	2.39	1.58	3130	3.39	1.46	3130
Spain	0.51	1.22	2184	2.18	1.79	3129	3.17	1.75	3128
LIBOR 12M	0.94	1.41	3053						

Note. We consider bonds with maturities of 1, 5 and 10 years. Daily yields are collected from Bloomberg for the period between 28-12-2007 and 30-12-2019. The 12 month LIBOR is used as a proxy for the risk-free rate. All data is shown in percentages. SD represents the standard deviation and N the number of observations.

Table 5: Descriptive statistics for the Credit Default Swap (CDS) spreads with a maturity of 1 year and denominated in Euros.

Country	Mean	SD	Minimum	Median	Maximum	N
Belgium	26.70	44.27	0.88	7.00	308.97	3131
France	15.16	20.14	0.72	6.29	128.78	3131
Germany	7.07	9.37	0.28	2.70	50.49	3131
Italy	75.51	89.25	2.50	43.10	550.85	3131
Spain	71.08	87.11	2.25	27.28	426.63	3131

Note. Daily spreads are collected from Datastream for the period between 28-12-2007 and 30-12-2019. All spreads are shown in basis point. SD represents the standard deviation and N the number of observations.

4 Results

In this chapter we describe the results of this research. We first show the results of the individual forecasts and the selection of these in Section 4.1. In Section 4.2 we describe the results of the combined forecasts and examine the forecast combination weights. We discuss the impacts of considering longer forecast horizons in Section 4.3. Lastly, we show the effects of market overreaction for peripheral countries in Section 4.4.

4.1 Individual forecasts

The Market-Implied Default Probability (PD_M), the two benchmarks ($RNDP_S$, PD_H) and the individual RNDP forecasts are shown in Figures 1, 2 and 3. These Figures show the actual PD_M relative to the benchmarks, the equity RNDPs and the bond RNDPs, respectively. When examining the benchmark forecasts in Figure 1, the problem immediately becomes clear. Firstly, the rating-based method only results in a default probability of 0%. For our entire time horizon, all countries had a rating of at least BBB-, which translates to $PD_H = 0$. The lowest rating, BBB-, was assigned to Spain and Italy during the sovereign debt crisis. Second, the standard CCA method does not perform well either. RNDP$_S$ is often equal to 0 as well, when this is not the case, it is either higher or lower than PD_M. Although, there is some co-movement between RNDP$_S$ and PD_M, the relation does not generalize well and no clear patterns appear. It can be observed that RNDP$_S$ occasionally moves up or down during the same period as PD_M. For example, RNDP$_S$ is relatively higher during the financial and sovereign debt crisis than in other periods. However, this does not hold for all countries. For Germany this is only the case during the financial crisis and for Belgium this does not hold at all. Moreover, PD_M is higher during the sovereign debt crisis than during the financial crisis for all countries. RNDP$_S$ shows the opposite for all countries except Spain, for which the difference is almost negligible. Lastly, an up or down movement of RNDP$_S$ is mostly a quarter after a similar movement by PD_M. As balance sheet items are only reported at the end of a quarter, the change in RNDP$_S$ is only reflected afterwards, while the change may be reflected earlier in PD_M.

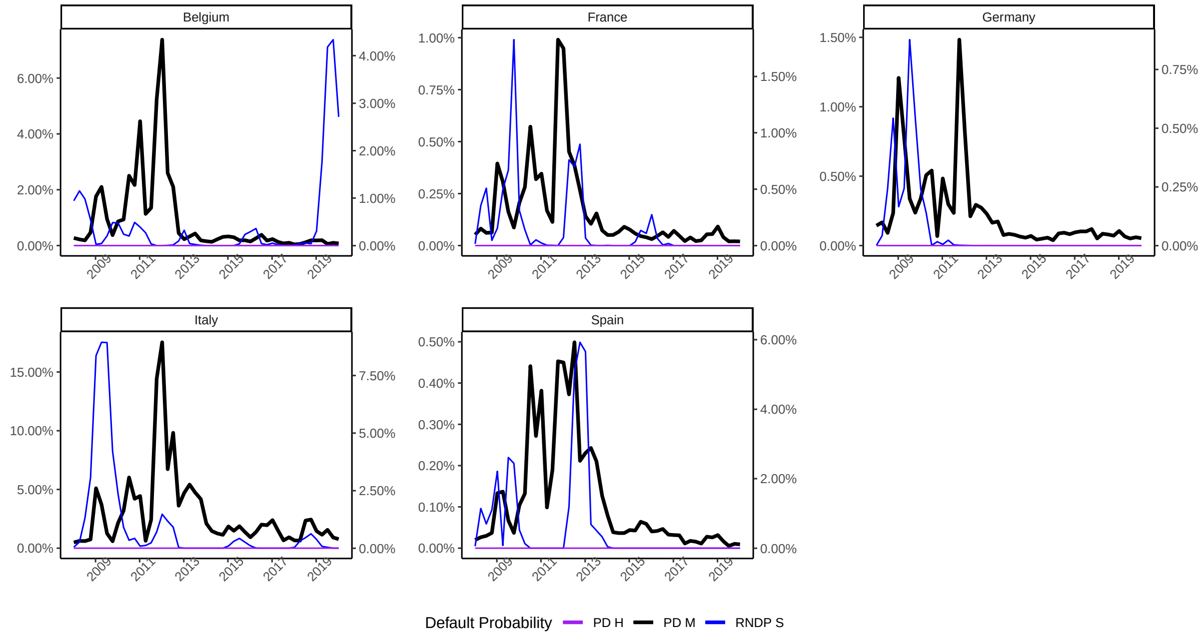


Figure 1: Benchmarks (left scale) vs Market-Implied Default Probability (right scale)

Figures 2 and 3 show that equity RNDPs are too high, whereas bond RNDPs are often too low to serve as a proxy for PD_M. For equity RNDPs, we make two observations. First of all, equity RNDPs drop significantly after the end of 2015. We attribute this to two different causes, lower equity volatilities and lower debt-to-asset ratios. Equity markets were less volatile from the end of 2015 onward. The height of the sovereign debt crisis lowered around the end of 2012, but full recovery was not achieved before 2015. The level of debt relative to the asset value also lowered after the sovereign debt crisis. Improved debt-to-asset ratios determine the speed of the decrease of equity RNDPs. RNDP_3, which has the same volatility for all countries, decreases slower for Italy and Spain. Peripheral countries, Italy and Spain, were affected more by the sovereign debt crisis than core countries such as France and Germany. Although Belgium is also a core country, it was hit harder by the sovereign debt crisis than France and Germany, but not as hard as Italy and Spain. As a result, these peripheral countries had higher debt-to-asset ratios and took longer to recover. Our second observation is that equity RNDPs are more volatile than PD_M during the financial and sovereign debt crisis. Comparing the movements of equity RNDPs with PD_M, we observe that these do not show the same pattern. As a result of these two observations, scaling the equity RNDPs by a constant would not provide adequate results to serve as a proxy for PD_M. Hence, equity RNDPs alone are not sufficient to describe sovereign default risk.

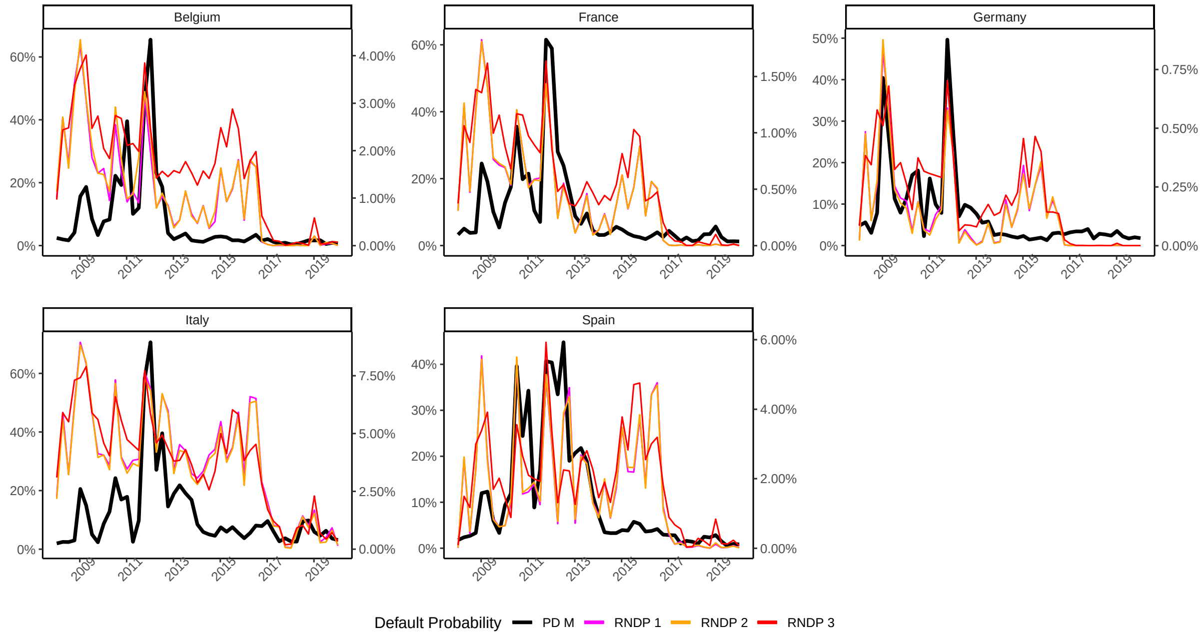


Figure 2: Equity RNDPs (left scale) vs Market-Implied Default Probability (right scale)

Scaling the bond RNDPs by a constant would not be a solid proxy for the PD_M either. Bond RNDPs are (almost) always equal to zero. For France and Germany, bond RNDPs are so low that it appears as if these are always equal to zero. Comparing the movements of the bond RNDPs with PD_M, we observe that bond RNDPs do not show enough volatility. However, for Italy, Spain and Belgium the bond RNDPs display peaks at the same time as PD_M. For these countries, the co-movements between bond RNDPs and PD_M already appear to be better than the co-movements between the benchmarks and PD_M, but are still far from optimal. It should be noted that, due to data limitations, we do not have RNDP$_4$ prior to Q4 2011 for Belgium and Spain.

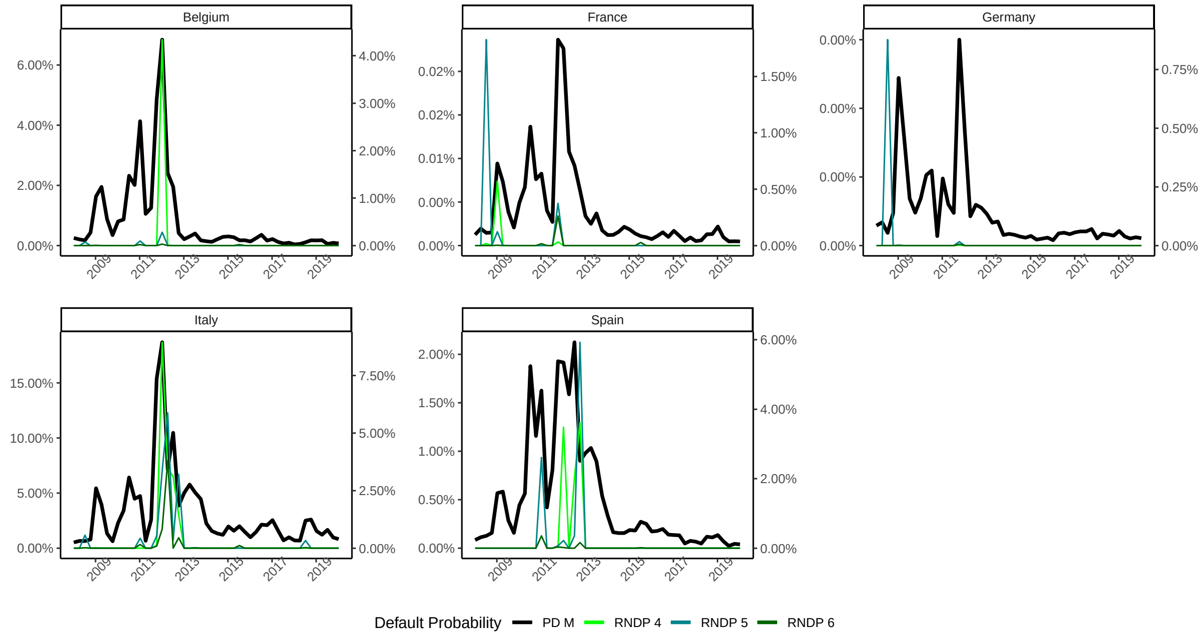


Figure 3: Bond RNDPs (left scale) vs Market-Implied Default Probability (right scale)

Table 6 shows the MSE for RNDP_S, PD_H and the individual RNDPs. It also displays which individual RNDPs are significantly more accurate than the benchmarks based on the Harvey, Leybourne and Newbold test using a significance level of $\alpha = 5\%$. The MSE is notably higher for equity RNDPs than for RNDP_S, PD_H and the bond RNDPs. Which was expected as equity RNDPs are much higher than PD_M. Furthermore, bond RNDPs mostly result in a lower MSE than RNDP_S. For Belgium, RNDP_4, RNDP_5 and RNDP_6 are significantly more accurate than RNDP_S. However, this is not because the bond RNDPs are very accurate, but rather due to the fact that RNDP_S is a weak proxy for PD_M. As PD_M is low, the MSE will be small for forecasts that are close to zero, which is the case for bond RNDPs. This is illustrated by the fact that these RNDPs are not significantly more accurate than PD_H, which is always equal to zero. Only RNDP_5 for Spain shows a significantly higher accuracy than both benchmarks. However, it only gives a slight reduction of the MSE and the result does not hold for other countries. Considering a different time horizon may have yielded different results and we attribute the significance to chance. Hence, individual RNDP forecasts based on either equity or bond volatilities are not sufficient to outperform both benchmarks.

Table 6: MSE and Harvey, Leybourne and Newbold (HLN) test results for the individual forecasts.

Country	Benchmark		Risk-Neutral Default Probability					
	S	H	1	2	3	4	5	6
Belgium	3.74 (*4.88*)	0.98 (*0.96*)	462.71	503.51	823.12	0.59*	0.89*	0.97*
France	0.21	0.24	443.50	439.61	596.92	0.24	0.24	0.24
Germany	0.09	0.05	162.01	160.84	239.45	0.05	0.05	0.05
Italy	19.69	5.10	1116.57	1057.87	1072.25	4.87	4.78	4.55
Spain	4.30 (*4.05*)	4.55 (*4.35*)	261.56	263.04	265.19	3.62	**4.22**	4.52

Note. MSE is in percentages; (*italic*) is the MSE for the period of Q4 2011 - Q1 2020; * significantly more accurate than $RNPD_S$, ** significantly more accurate than PD_H, **bold** significantly more accurate than both $RNPD_S$ and PD_H ($\alpha = 0.05$).

Upon closer inspection of Figures 2 and 3, we notice that a combination of the two different groups of RNDPs may show good results. For Italy, Spain and Belgium, equity RNDPs show peaks of roughly the same magnitude during the financial and sovereign debt crisis. The peaks of PD_M are notably higher during the sovereign debt crisis than during the financial crisis for these countries. Bond RNDPs show much higher peaks during the sovereign debt crisis than during the financial crisis. Hence, a combination of equity and bond RNDPs may result in forecasts that closely follow PD_M.

Forecast selection

The selection of individual forecasts is based on the R^2 and AIC. Before moving to regression models with multiple individual RNDPs, we first discuss models that only include one individual RNDP. Table 7 shows the results of regressing PD_M on each of the individual RNDPs separately. For all countries, at least one of the regressions based on individual RNDPs has a higher R^2 than the regression based on RNDP_S. There are two important remarks to be noticed. First, the coefficients for equity and bond RNDPs show different results. For equity RNDPs, the coefficients are similar for RNDP_1, RNDP_2 and RNDP_3 for each country. Moreover, the coefficients are not too different across countries as these are between 0.01 and 0.07. This is not the case for bond RNDPs. For all countries except Italy, the coefficients for RNDP_4, RNDP_5 and RNDP_6 are different. Also across countries, the coefficients vary widely. For example, the coefficient for RNDP_5 is negative for Germany, which is not in line with the other coefficients or countries. Furthermore, when performing the regressions for different time periods, the coefficients also differ.

Second, there is no single RNDP which results in a higher R^2 than the other RNDPs. Neither do the results show that there is a clear "winner" between equity and bond RNDPs. For Belgium, Italy and Spain, RNDP_4 gives the highest R^2. While RNDP_1 and RNDP_2 achieve this for France and Germany. This is an interesting result as we did not expect this. However, we do understand what causes this. For France and Germany, the bond yield volatility is extremely low for most periods, which resulted in the RNDP being (almost) equal to zero. This is reflected in the coefficients of RNDP_4 and RNDP_6, which are large for Germany and France. These two remarks again point out that individual RNDPs are not sufficient and a combination of both equity and bond RNDPs is needed for better results.

Table 7: Results for OLS regressions of each individual RNDP on the Market-Implied Default Probability.

Country	Value	Risk-Neutral Default Probability						
		1	2	3	4	5	6	S
Belgium	Intercept	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.01
	Coefficient	0.05	0.05	0.03	0.59	9.25	57.31	-0.07
	R^2	0.39	0.45	0.25	0.59	0.59	0.38	0.02
France	Intercept	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Coefficient	0.01	0.01	0.01	70.59	6.55	458.03	0.24
	R^2	0.34	0.34	0.29	0.04	0.00	0.32	0.01
Germany	Intercept	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Coefficient	0.01	0.01	0.01	$1.83 * 10^6$	-452.00	$8.01 * 10^5$	0.11
	R^2	0.52	0.52	0.39	0.37	0.00	0.39	0.03
Italy	Intercept	0.00	0.00	0.00	0.01	0.01	0.01	0.01
	Coefficient	0.05	0.05	0.04	0.42	0.38	0.46	0.03
	R^2	0.24	0.24	0.17	0.53	0.23	0.10	0.01
Spain	Intercept	0.00	0.00	0.00	0.01	0.01	0.01	0.01
	Coefficient	0.07	0.07	0.07	2.93	0.84	45.25	6.37
	R^2	0.25	0.26	0.20	0.32	0.03	0.07	0.26

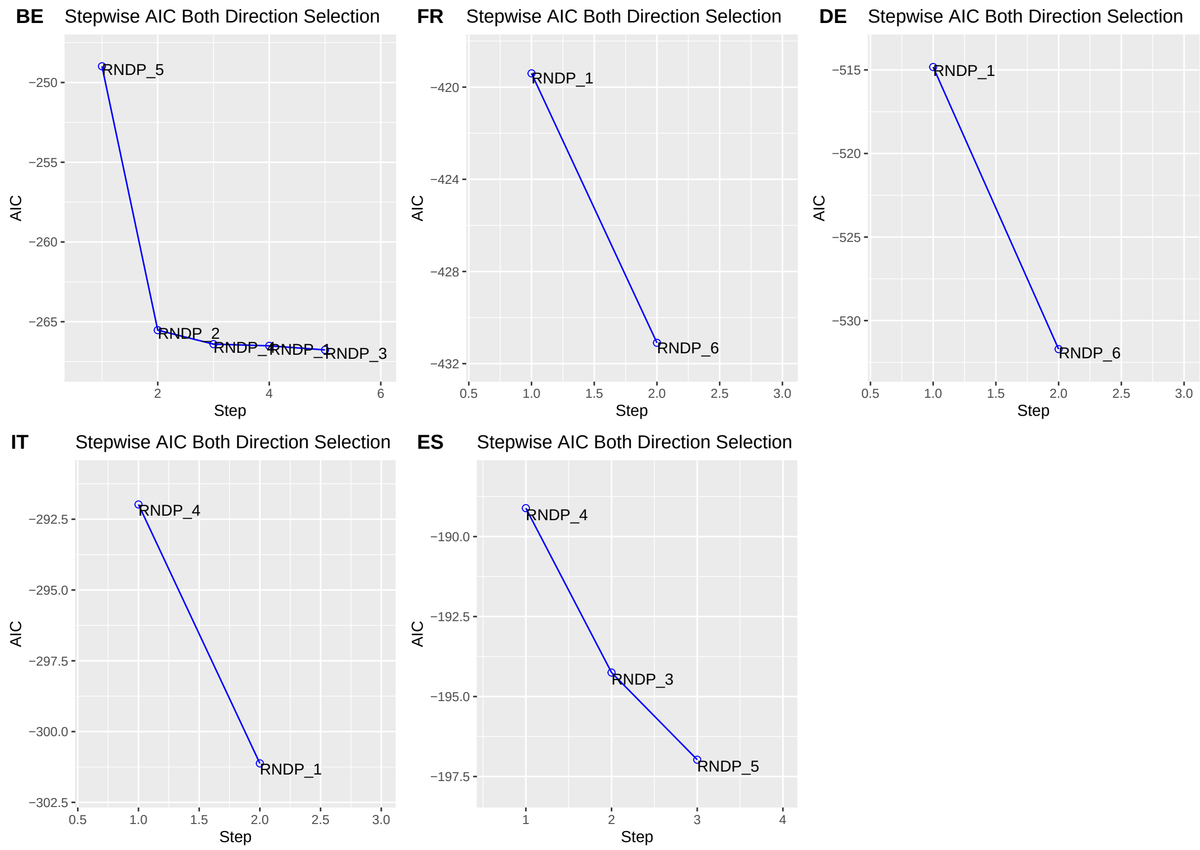


Figure 4: Stepwise AIC selection

Figure 4 shows the individual RNDPs that lower the AIC for each country. The results are not similar across all countries. Only for France and Germany the same RNDPs appear to be selected, namely RNDP_1 and RNDP_6. However, there is a clear pattern that both equity and bond RNDPs need to be combined. For France, Germany and Italy, we observe that the AIC reaches a minimum when including one equity RNDP, RNDP_1, and one bond RNDP, either RNDP_4 or RNDP_6. For Belgium and Spain, adding more forecasts lowers the AIC, but the improvement is also made on an alternating basis. That is to say, that the AIC is lowered by including a RNDP based on the other market than the last RNDP. So for Belgium, the first predictor is RNDP_5 (bond), the second is RNDP_2 (equity), the third is RNDP_4 (bond), et cetera.

The R^2 that results from regressions shows similar results. When including two individual RNDPs, the highest R^2 is achieved when one of these is an equity RNDP and the other a bond RNDP. The combinations resulting in the highest R^2 (given in parentheses) are the following: RNDP_2 and RNDP_5 for Belgium (0.77); RNDP_1 and RNDP_6 for France (0.50); RNDP_1 and RNDP_6 for Germany (0.68); RNDP_1 and RNDP_4 for Italy (0.63); RNDP_3 and RNDP_4 for Spain (0.45). Adding a third individual RNDP only improves the R^2 for Belgium and Spain. The combinations including three RNDPs resulting in the highest R^2 are the following: RNDP_2, RNDP_4 and RNDP_5 for Belgium (0.78); RNDP_3, RNDP_4 and RNDP_5 for Spain (0.52). Note that for both Belgium and Spain the analysis is made after Q4 2011, as RNDP_4 is not available prior to this. This may be the cause of the difference that we see in the results.

4.2 Combined forecasts

We include 4 combinations with 2 RNDPS, one equity and one bond, and define 2 combinations with 3 RNDPs, one equity and two bond. For comparison we also define a combination that includes all RNDPs except for RNDP_4, as we do not have data on RNDP_4 prior to Q4 2011 for Belgium and Spain. Lastly, we consider a combination that includes all RNDPs. The 8 combinations that we consider are the following: (C_1) 1, 4; (C_2) 1, 6; (C_3) 2, 5; (C_4) 3, 4; (C_5) 1, 4, 6; (C_6) 2, 4, 5; (C_7) 1, 2, 3, 5, 6; (C_8) 1, 2, 3, 4, 5, 6.

The combined RNDP forecasts are shown in Figures 5, 6 and 7. It immediately becomes clear that the combined forecasts improve upon the individual forecast shown in Figures 1, 2 and 3. The combined RNDPs appear to describe sovereign default risk quite well. Although the combined RNDPs do not perfectly follow PD_M, these follow similar patterns. For both, the combined RNDPs and PD_M, highs and lows appear at the same time. This is a major improvement when compared to the benchmarks and individual RNDPs. However, the magnitude of the highs and lows of the combined RNDPs is not always exactly the same as for PD_M. This is mainly due to the construction of the weights.

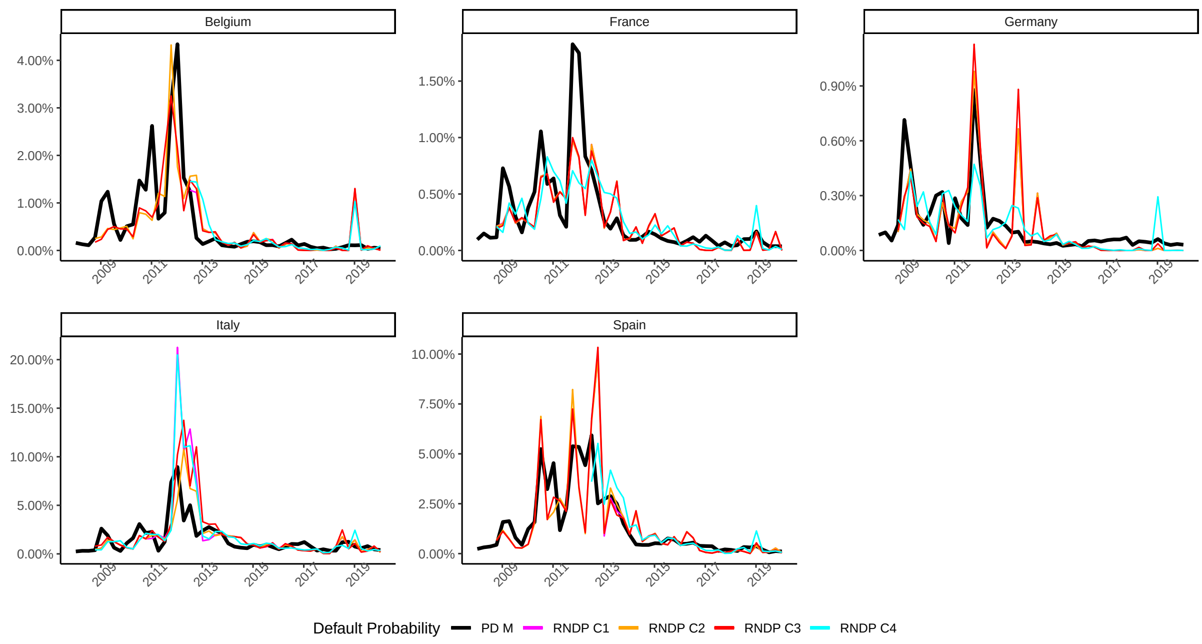


Figure 5: Forecasts based on combinations of 2 individual RNDPs vs PD_M

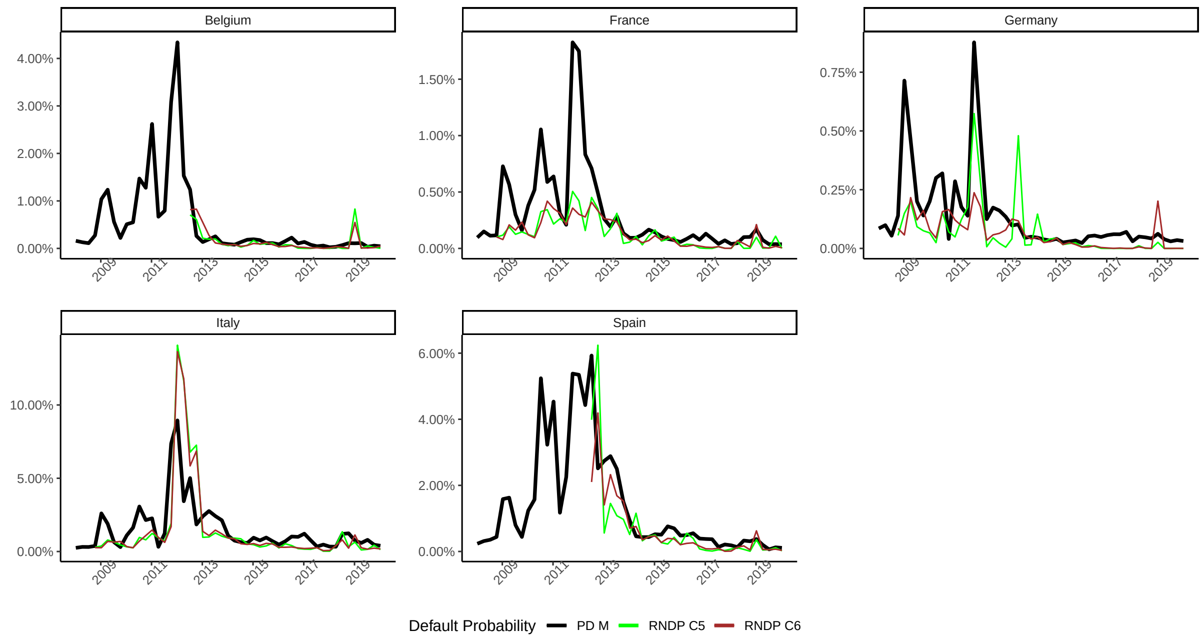


Figure 6: Forecasts based on combinations of 3 individual RNDPs vs PD_M

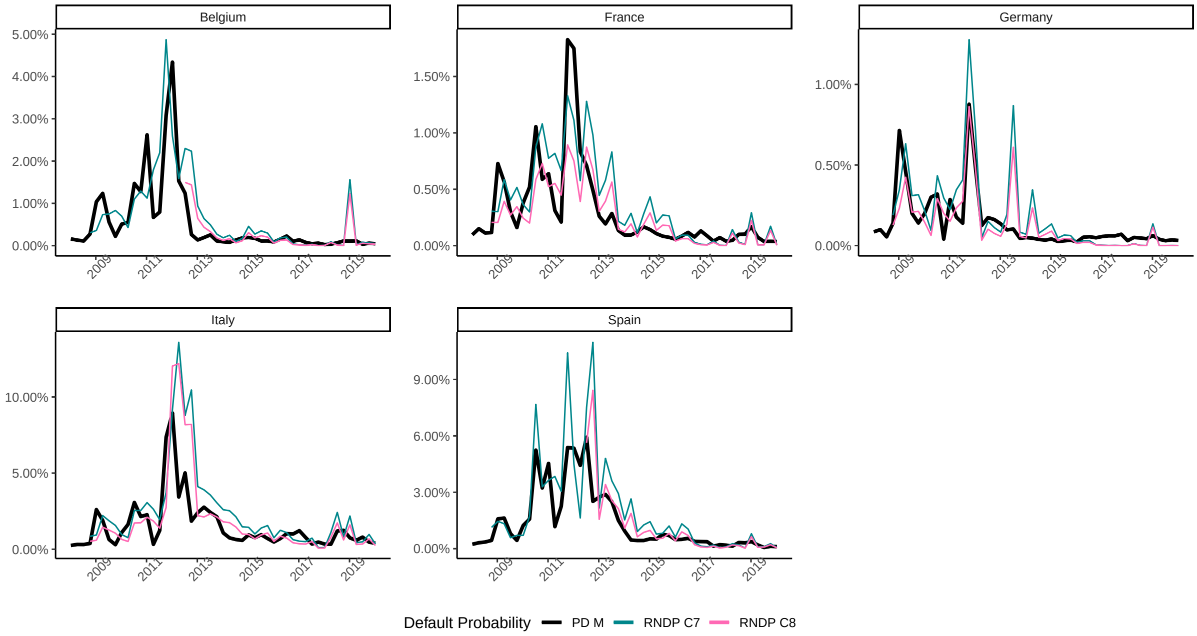


Figure 7: Forecasts based on combinations of 5 or 6 individual RNDPs vs PD_M

The weights are based on historical errors, which may include "noise". This becomes most apparent when comparing Figures 5 and 7. Figure 5 shows the RNDPs for C_1, C_2, C_3 and C_4, which only include 2 individual RNDPs. Figure 7 shows the RNDPs for C_7 and C_8, which include 5 and 6 individual RNDPs, respectively. For all countries, the pattern of both Figures is (almost) the same. However, for most countries and quarters, the highs and lows are more volatile for the combinations based on two individual RNDPs. The highs and lows in Figure 7 are closer to PD_M. For example, consider the peaks of Italy and France during 2012. Figure 5 shows that the combined forecasts are too high for Italy, and too low for France. Although Figure 7 shows this as well, the difference between the forecasts and PD_M is smaller. The reason for this is model averaging. As the combinations in Figure 7 are based on more RNDPs, the "noise" in the historical errors is averaged out. While the difference between the highs and lows for C_7, C_8 and PD_M is indeed smaller than for the other combinations, the relative gain is little. Table 8 shows the MSE and HLN results for the combined RNDPs. For each country, there is at least one other combination that leads to a lower MSE than C_7 and C_8.

Furthermore, almost all combined RNDPs lead to a lower MSE than both benchmarks. When comparing the MSE for each of the 8 combined forecasts, we find contradicting results. For core countries, the results are similar to what we derived from the AIC and R^2. The lowest MSE is achieved by C_5 and C_6 for Belgium. For France and Germany, C_1 and C_2 (among others), result in the lowest MSE. Our results for the individual RNDPs gave us that for France and Germany we only needed to include two RNDPs. Adding more RNDPs did not lead to better results. For Belgium, including a third RNDP led to both, a higher R^2 as well as a lower AIC. Contrarily, this does not hold for peripheral countries. Based on the individual RNDP results, we expected that the optimal combinations for Italy and Spain would be C_1 and C_6, respectively. In contrast, C_1 actually results in the highest MSE, while C_2 gives the lowest for Italy. For Spain, C_4 leads to the lowest MSE, but C_6 does not differ too much.

Although a lower MSE is desirable, the ultimate goal is to find forecast combinations that are significantly more accurate than both benchmarks. When comparing the significance of the combined forecasts, we make two observations. Firstly, we observe that a lower MSE does not necessarily lead to a higher power of significance. In fact, Table 8 shows that some RNDPs that result in a higher MSE are actually significant, while RNDPs with a lower MSE are not. For example, C_7 leads to a higher MSE than C_3 for Spain. However, RNDP_{C_7} is significantly more accurate than PD_H, while RNDP_{C_3} is not. The reason for this is that the HLN test depends on the autocorrelation of the loss differentials. The Tiao-Box procedure resulted in $h = 0$, for forecasts that show such results. This means that the autocorrelation is essentially zero for all lags, which makes it nearly impossible to use for forecasting.

Secondly, the HLN test results show a clear distinction between core and peripheral countries. For core countries, almost all combined forecasts are significantly more accurate than both benchmarks. When compared with RNDP_S, only C_3 and C_7 for Germany do not result in a significantly higher accuracy. When compared with PD_H, the combinations that do not result in a significantly higher accuracy are: C_1, C_4 C_5, C_6 and C_8 for Belgium; C_3 and C_7 for Germany. Note that for Belgium, all these combinations have in common that they include RNDP_4 and therefore forecasts are only made after Q3 2012. From this quarter onward, PD_M is low, such that PD_H results in a low MSE. Also for Belgium, all combined RNDPs that are based on the entire time horizon, result in a significantly higher accuracy than both benchmarks. For peripheral countries on the other hand, there is no single combination that leads to a significant outperformance of both benchmarks. For Italy, none of the 8 combinations is significantly more accurate than either RNDP_S or PD_H. For Spain, only C_2, C_6 and C_7 show a significantly higher accuracy than the PD_H.

Table 8: MSE and Harvey, Leybourne and Newbold (HLN) test results for the combined forecasts.

Country	Benchmark		Risk-Neutral Default Probability							
	S	H	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
Belgium	3.81	1.05		**0.40**	**0.34**				**0.46**	
	4.37	*0.07*	0.08*			0.11*	0.03*	0.03*		0.10*
France	0.22	0.25	**0.06**	**0.06**	**0.06**	**0.09**	**0.12**	**0.14**	**0.07**	**0.07**
Germany	0.10	0.05	**0.02**	**0.02**	0.03	**0.02**	**0.02**	**0.03**	0.03	**0.02**
Italy	20.86	5.43	7.42	2.84	5.06	6.56	4.03	3.88	5.24	3.82
Spain	4.58	4.84		2.19**	2.17				3.03**	
	1.95	*2.28*	1.99			0.67	0.92	0.69**		1.30

Note. MSE in percentages; *italic* is the MSE for the period of Q3 2012 - Q1 2020; * significantly more accurate than RNPD$_S$, ** significantly more accurate than PD_H, **bold** significantly more accurate than both RNPD$_S$ and PD_H ($\alpha = 0.05$).

Our results up until now suggest that our method may be better suited for core countries. For peripheral countries, the combined forecasts are not significantly more accurate than both benchmarks. Eventually, we want to determine the economic reason for this difference and evaluate whether our method works for peripheral countries when taking this reason into account. However, the combined RNDP forecasts are made up of various layers, such that there could be several explanations for this difference. Therefore, we first give mechanical explanations for the difference and afterwards move to an economic explanation in Section 4.4. We found that for core countries the optimal combinations were the same as the results from the AIC and R^2 in Section 4.1. We would expect that this would hold for all countries, but this was not the case for peripheral countries. Hence, we examine the forecast combination weights to find an explanation for this.

Forecast combination weights

Figure 8 shows the weights for all 8 combinations. Overall, bond RNDPs are assigned the largest weights. Before 2017, bond RNDPs account for a total weight of more than 90%. This can clearly be seen for C_1, C_2, C_3 and C_4, which only include one bond RNDP. For combinations including more bond RNDPs, the total weight of all bond RNDPs is similar to the weight of the individual bond RNDP in C_1, C_2, C_3 and C_4. For all countries this starts to change at the end of 2016. Lower equity RNDPs cause equity weights to increase and bond weights to decrease. The equity RNDPs drop sharpest for core countries. Both equity and bond weights converge quickly and are equal to around 50% for Q2 2017. Thereafter, the composition of the combined forecasts is more balanced in terms of equity and bond RNDPs for France and Germany. The equity and bond weights for Belgium diverge again in 2019. The convergence is slower for peripheral countries. For Spain, equity and bond weights are close to 50% in Q3 2018, more than a year later. For Italy, the weights never move close to 50%. Furthermore, we notice that C_1, C_2 and C_3, result in similar equity and bond weights. For each country, the weight pattern is about the same, while it is not the case for C_4.

Lower equity RNDPs, and thus higher equity weights, are caused by both lower equity volatilities and lower debt-to-asset ratios. Our observations can be explained by examining the weights of C_4 after 2017. C_4 is a combination of RNDP_3, based on VSTOXX, and RNDP_4, based on the 1 year bond yield volatility. For Belgium, France, Germany and Spain, VSTOXX is higher than their local equity volatilities. Germany is the only country for which the debt-to-asset ratio becomes sufficiently low, such that even for C_4, the weights converge. For Belgium, France and Spain, the debt-to-asset ratio also lowers, but not as much. The relatively higher VSTOXX causes RNDP_3 to remain large compared to bond RNDPs and hence the weights do not converge for C_4. However, the debt-to-asset ratios are low enough such that in combination with even lower local equity volatilities, RNDP_1 and RNDP_2 become sufficiently low and the weights do converge for C_1, C_2 and C_3. Lastly, for Italy, the weights do not converge for any of the combinations. Although, both the equity volatility and the debt-to-asset ratio decrease, these remain relatively high. Also, the local equity volatilities are not lower than VSTOXX. Therefore, all equity RNDPs remain relatively high for Italy and the weights do not converge.

The difference in performance between core and peripheral countries is a result of the variability in the weights. For core countries, we observe that weights are relatively stable over time. Especially before convergence, equity and bond weights almost follow a straight line. For peripheral countries, the variance in the weights is higher. Furthermore, when including more RNDPs, the equity and bond weights are almost equally divided between the number of equity and bond RNDPs for core countries. For example, C_5 and C_6 for Germany and France almost only show 2 lines, while 3 RNDPs are included. The two bond RNDPs receive the exact same weight. For Belgium and Spain, this is also almost the case. While for Italy, which shows the worst results, we do see clear differences in the bond weights.

The risk of a higher variance is that more "noise" is captured in the weights. As these are based on historical errors, the "noise" of the previous periods is reflected in the weights of the next periods. The weights then become unreliable and are not optimal. This is also the reason why we found different optimal combinations in Section 4.1 and Section 4.2 for peripheral countries.

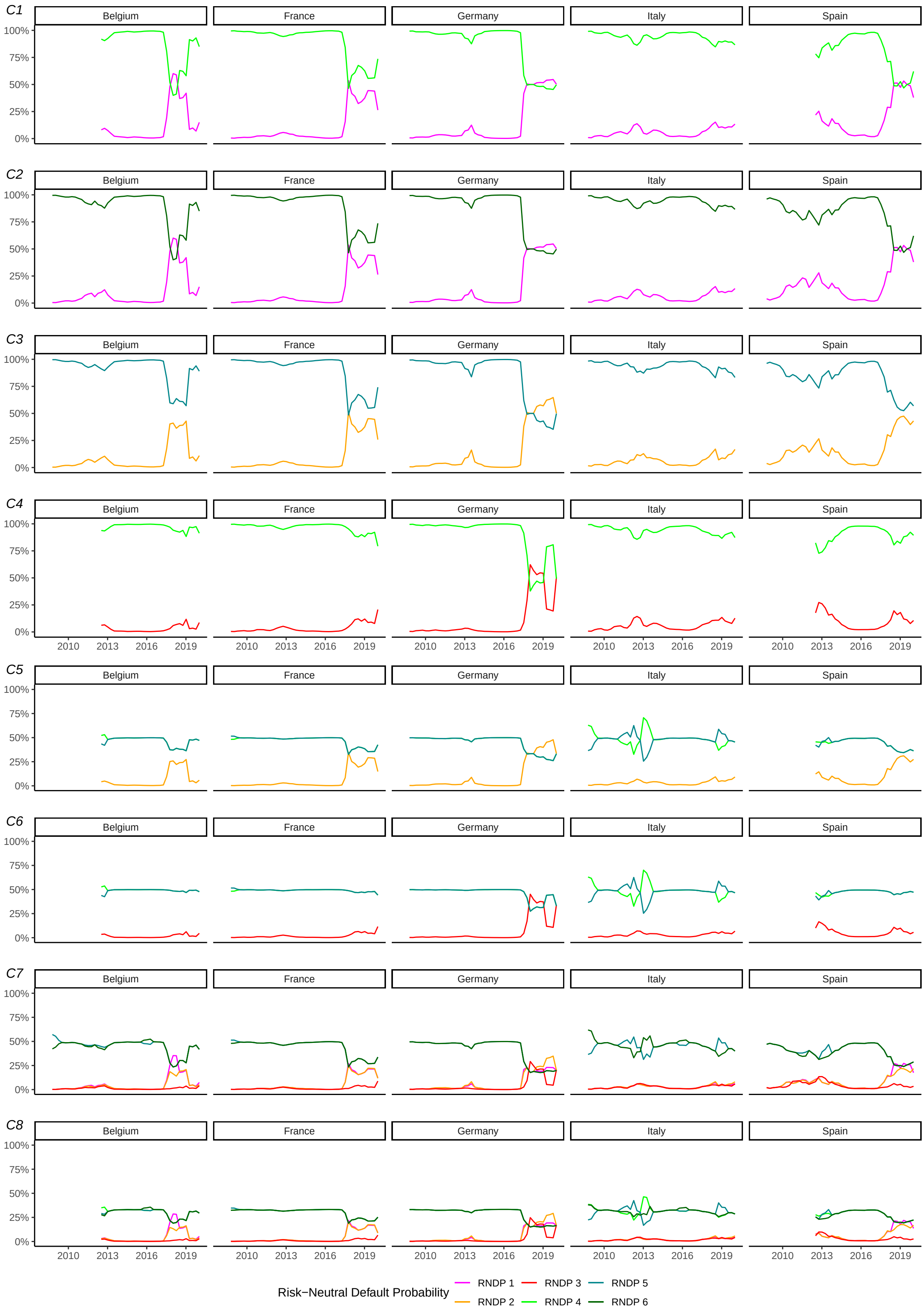


Figure 8: Individual RNDP weights in each combined forecast

4.3 Forecasting for longer horizons

To examine the performance of the forecast combinations for longer horizons, we set weights equal for a specific time period, the holding period. This means that we do not determine new weights every quarter, but "hold" the weights for m quarters. We consider 2 quarters, 1 year, 2 years, half time and all time as the time frame for which weights are equal. Hence, this means that weights are only determined twice a year, once a year, once every two years, twice in total and once in total, respectively. Note that the weights are determined in the same way as for the combined forecasts and are still based on a 3-quarter rolling window. Therefore, the new weights are also present in Figure 8. The only difference is that the weights are set equal for a longer period.

The results of the combined RNDP forecasts for longer horizons are shown in Figures 9, 10, 11, 12, 13, 14, 15 and 16. For comparison, we also include the combined RNDP forecasts with a holding period of 1 quarter. Within each combination, all forecasts are the same at the beginning, as the weights are determined in exactly the same way. The forecasts will only differ once new weights have been determined. The shorter the holding period, the quicker the forecasts will differ from the others. So for example, forecasts based on a 2 quarter holding period show different results than the others after 2 quarters. Whereas, combined RNDPs based on a half time or all time holding period will give the exact same forecasts for the first 23 quarters. For Belgium and Spain, combinations that include RNDP$_4$ (C_1,C_4,C_5,C_6 and C_8), only include forecasts from Q3 2012 onward. From here on, we will call this group 2. All other countries and combinations belong to group 1. Hence, group 1 includes all combinations for France, Germany and Italy and all combinations that exclude RNDP$_4$ (C_2,C_3 and C_7) for Belgium and Spain.

Figures 9, 10, 11, 12, 13, 14, 15 and 16 show us that the time when weights are determined is crucial for the performance of the forecasts. This becomes clear when looking at the forecasts for combinations with longer holding periods, 4 years, half time and all time. We first discuss group 1 and then move to group 2. For group 1, we observe that a holding period of 4 years does not perform well. After the initial setting, the weights are adapted for the first time in Q4 2012, which was during the sovereign debt crisis. As a consequence, almost all RNDP forecasts are below PD_M before Q4 2012. This can also be seen for holding periods of half time and all time, which result in the same forecasts as a holding period of 4 years before Q4 2012. Moreover, the all time holding period actually always results in low RNDP forecasts with respect to PD_M. The only exception is the peak during the sovereign debt crisis for Italy. Figure

8 shows that for most countries, bond RNDP weights are highest and equity RNDP weights are lowest in the initial period. The weights come close to those of the initial period during the peak of the sovereign debt crisis for Italy. For the other countries, this only happens once economic stability is returned. For a 4-year holding period, the weights are adapted in Q4 2012 for the first time. This quarter is at the end of the peak of the sovereign debt crisis, such that the weights are based on this crisis period and we obtain "crisis weights". Equity RNDPs have a relatively larger weight and the bond RNDP weight is smaller than in other periods. The effect of this can be seen in the 4 years following Q4 2012, RNDP forecasts based on a 4-year holding period are at their highest levels for group 1. Small increases in PD_M correspond to large increases in RNDP forecasts, as too much weight is placed on the (high) equity RNDPs.

For group 2, we observe that forecasts based on holding periods of 4 years, half time and all time, are all high when compared with PD_M. The reason for this is that the initial period for group 2 is Q3 2012, which also was during the sovereign debt crisis. The elevated forecasts are then caused by the same reasons as explained for the forecasts after Q4 2012 based on a 4-year holding period for group 1. For group 2, it also becomes clear that the all-time holding period is worst. The shorter the holding period, the quicker the forecasts drop and become closer to PD_M.

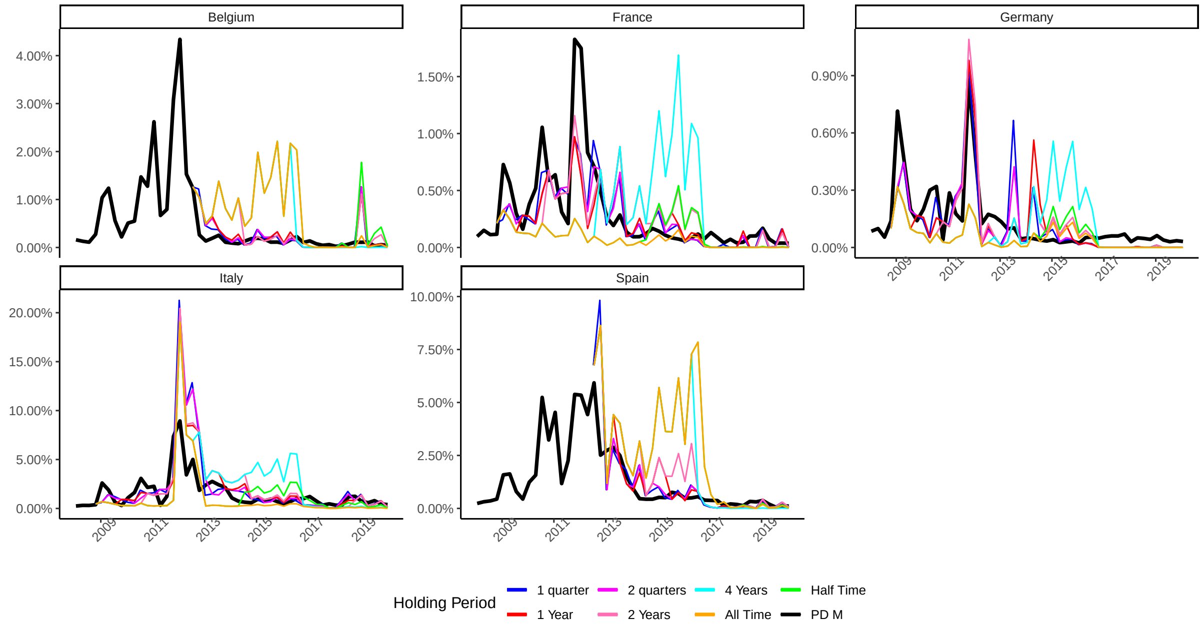


Figure 9: Longer horizon RNDP forecasts for C_1

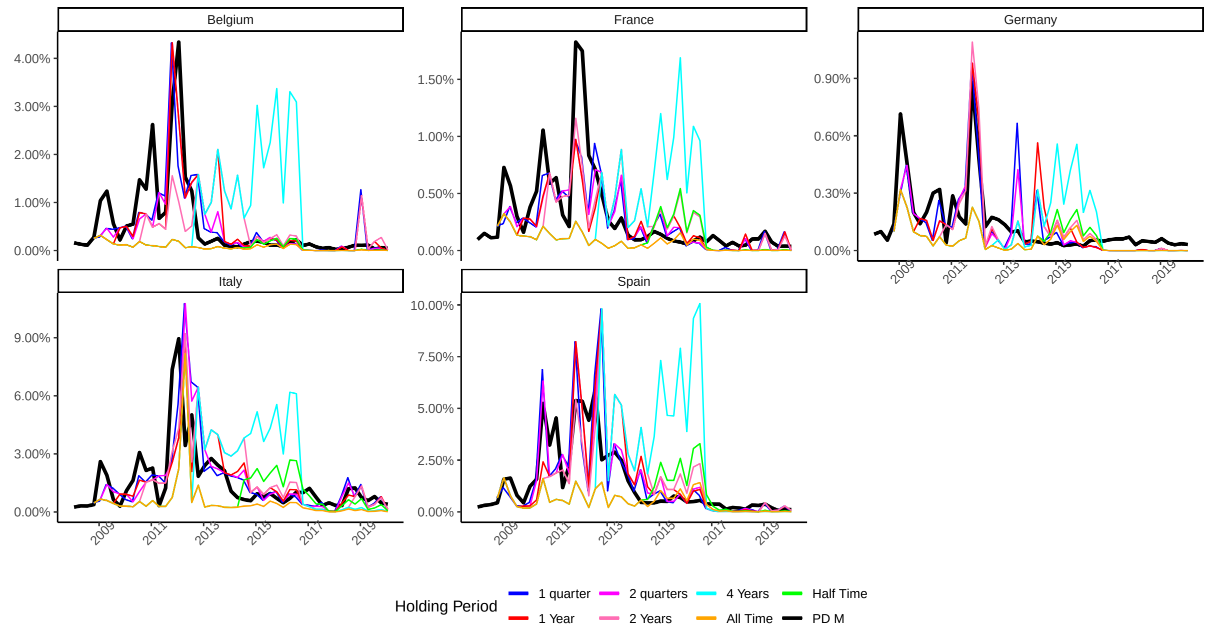


Figure 10: Longer horizon RNDP forecasts for C_2

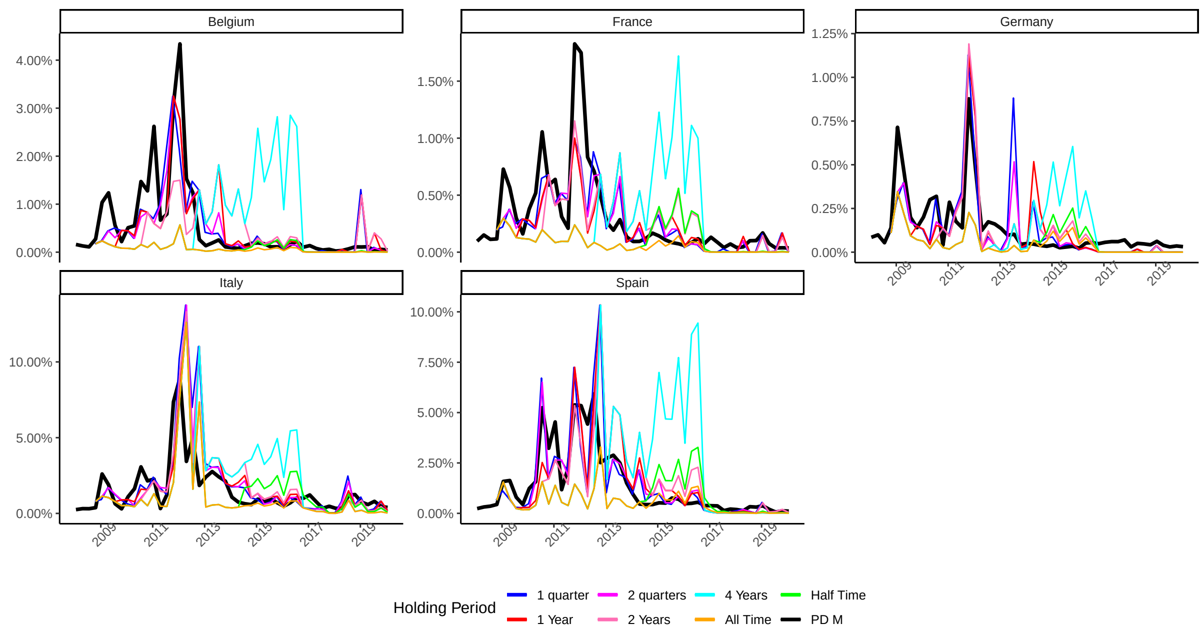


Figure 11: Longer horizon RNDP forecasts for C_3

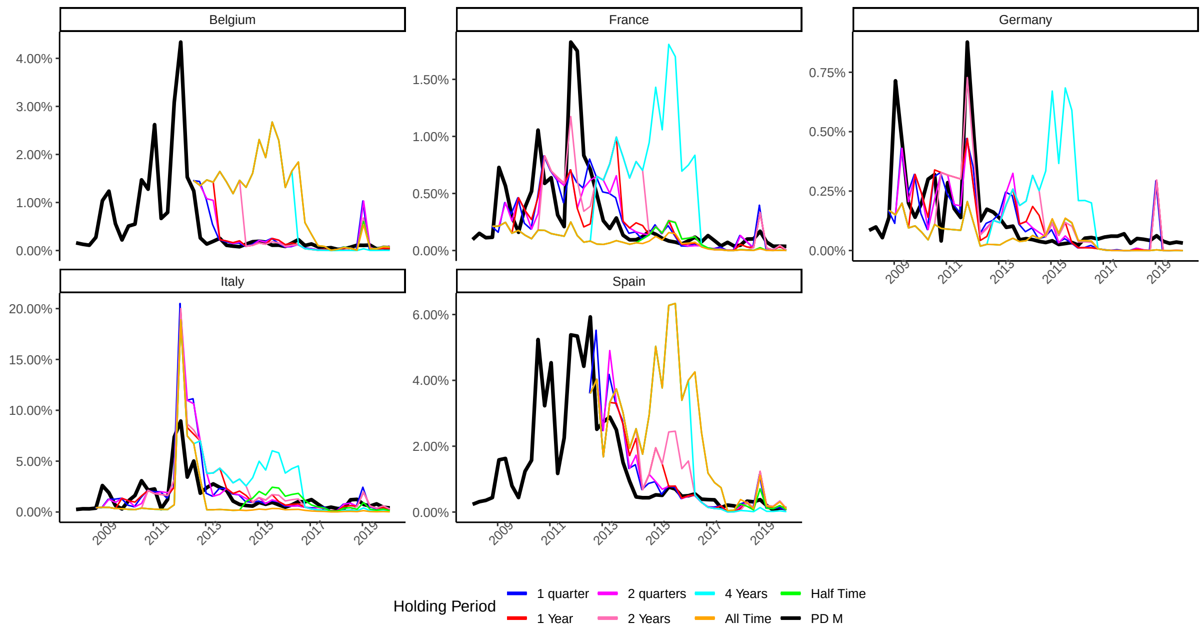


Figure 12: Longer horizon RNDP forecasts for C_4

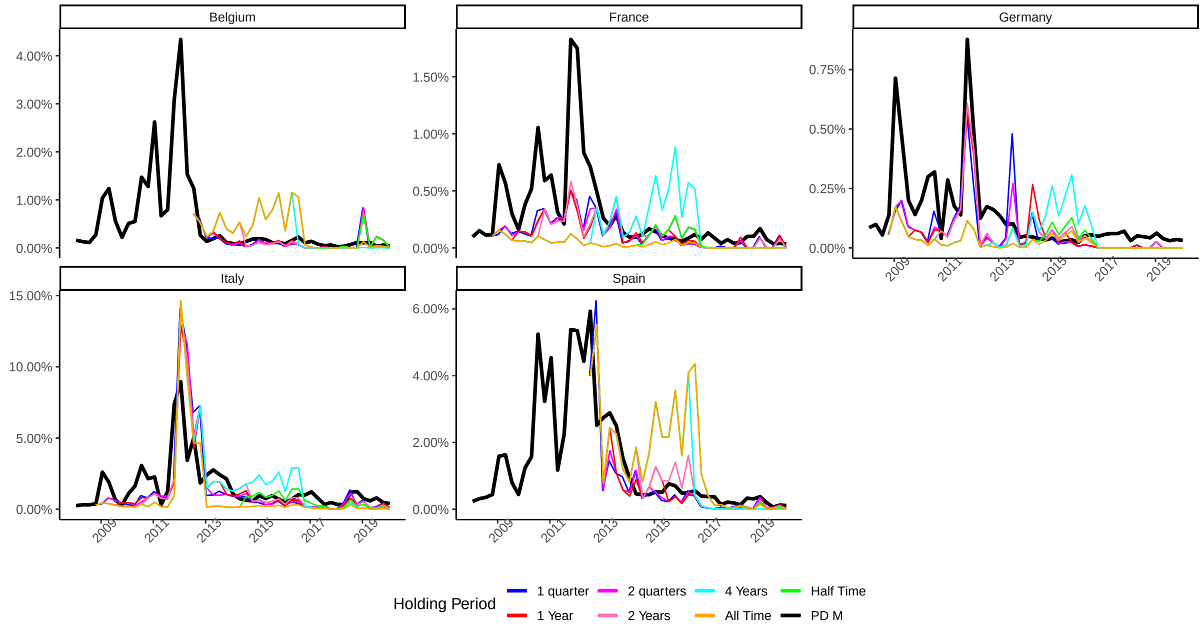


Figure 13: Longer horizon RNDP forecasts for C_5

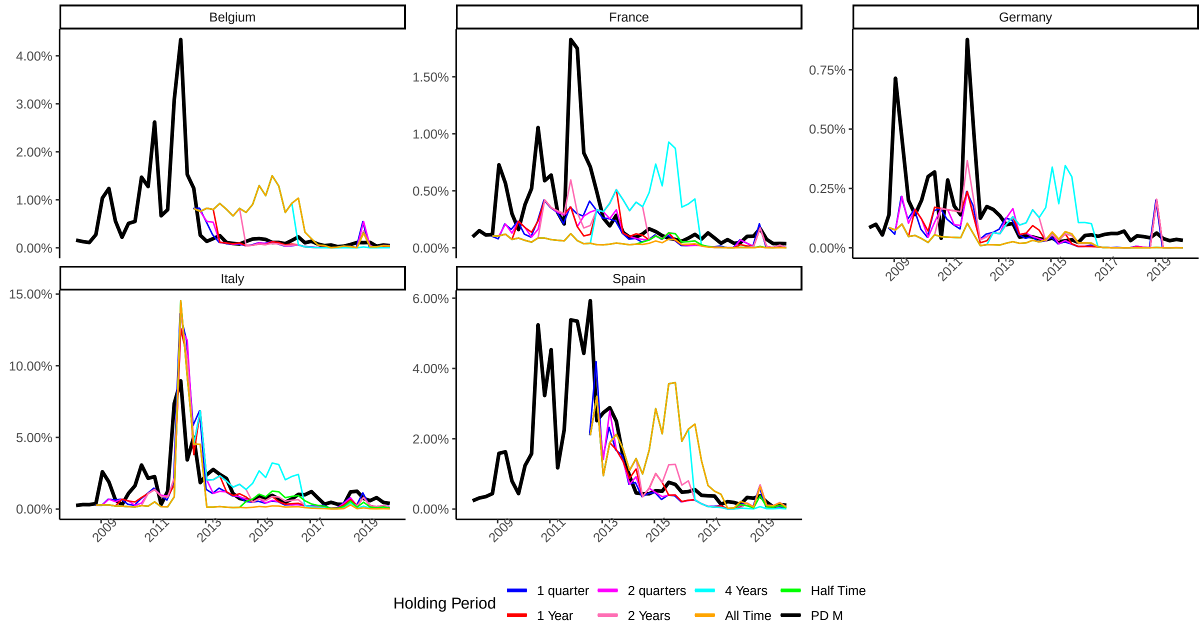


Figure 14: Longer horizon RNDP forecasts for C_6

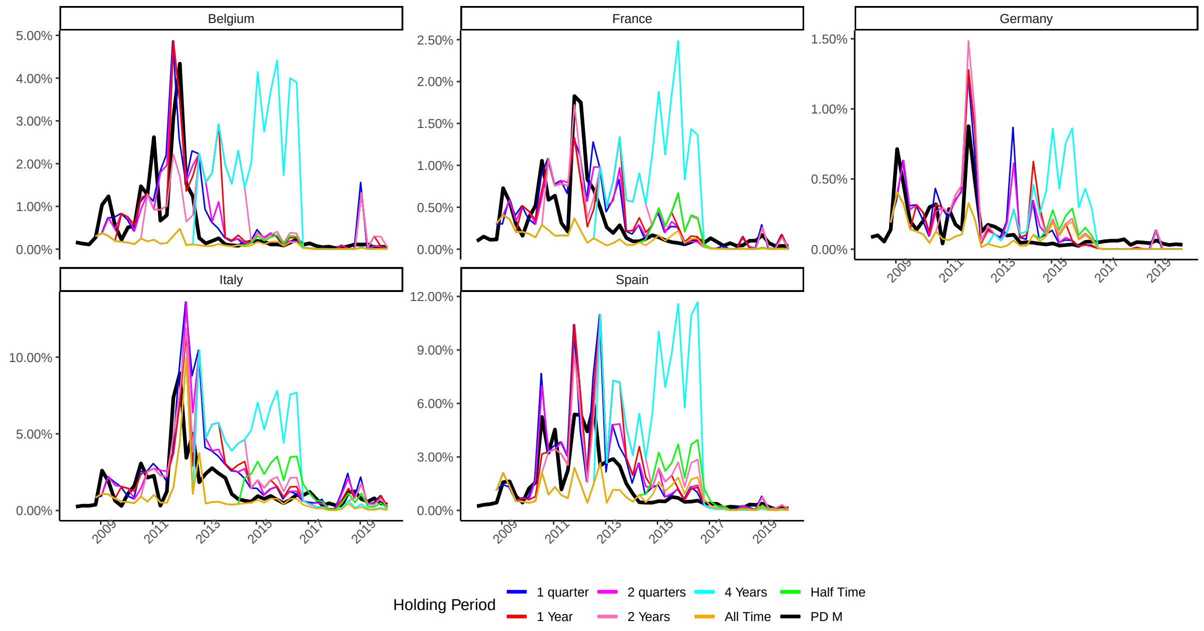


Figure 15: Longer horizon RNDP forecasts for C_7

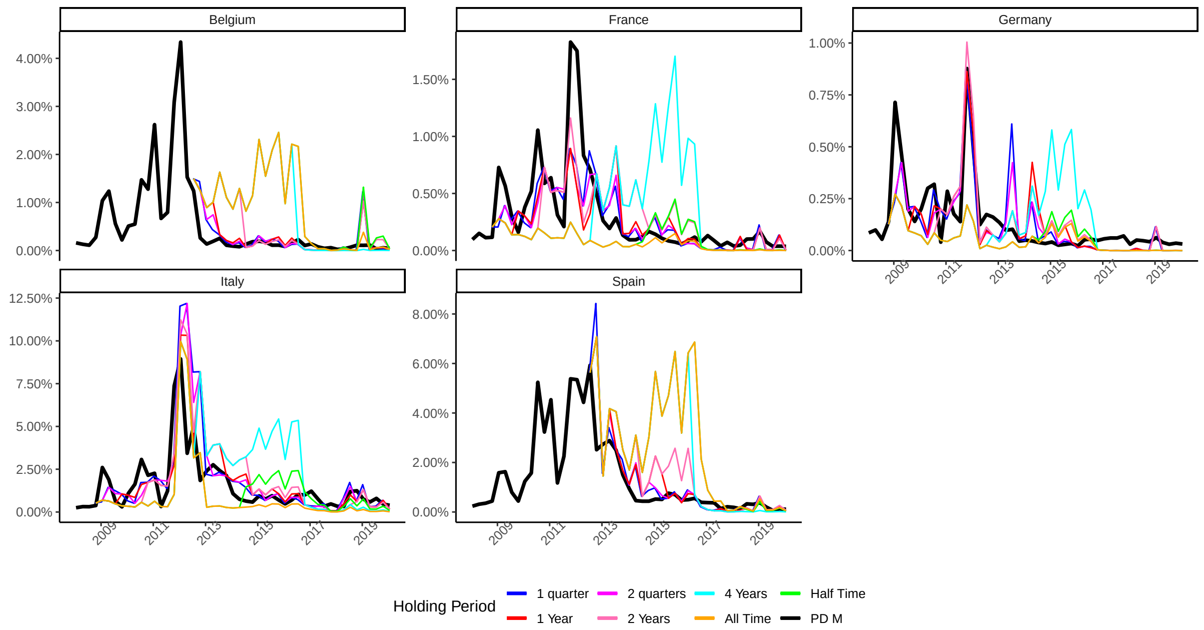


Figure 16: Longer horizon RNDP forecasts for C_8

Table 9 shows the HLN test results of the combined forecasts for longer horizons. We draw three conclusions from the results. First of all, the results from Section 4.2 appear to hold here as well. We find significant improvements of both benchmarks for core countries, while for peripheral countries the results remain poor. For core countries, the combinations that resulted in a significant improvement of both benchmarks, also show significant improvements for longer holding periods. For Italy, none of the forecasts results in a significant improvement of both benchmarks. For Spain, there are some specific forecast combinations and holding periods which now give a significant improvement of both benchmarks. These are the following C_4 with a holding period of 2 quarters; C_6 with a holding period of 2 years; C_2, C_7 with a holding period of half time; C_2, C_3, C_7 with a holding period of all time.

Second, weights play an important role in the combined RNDP forecasts, but only to a certain extent. The combined RNDP forecasts show significant improvements even if the weights are not "perfect". It is more important to use different weights for different times in the economic cycle than it is to frequently re-balance. The results in Table 9 confirm that a 4-year holding period performs worst for group 1 and long-term holding periods, half time and all time, result in the worst forecasts for group 2. For group 1, long-term holding periods do not perform that bad. Equity and bond weights do not dramatically change from quarter to quarter. However, equity and bond weights are significantly different in times of crisis and in times of economic stability.

Lastly, we notice that there are different optimal holding periods for core countries and peripheral countries. For core countries, shorter holding periods, 1 quarter to 1 year, give better results than longer holding periods. For France, most forecasts do not significantly outperform the two benchmarks for longer holding periods. For Belgium and Germany, forecasts with longer holding periods still significantly outperform the benchmarks, however the MSE deteriorates. For peripheral countries on the other hand, the opposite appears to hold. Although none of the forecast combinations is significant for Italy, holding periods of 2 years, half time and all time result in the lowest MSE for most RNDP forecasts. For Spain, 5 out of 7 significant combined RNDP forecasts are based on holding periods of half time and all time. The combinations for which this is not the case are C_4 and C_6. However, as mentioned before, half time and all time holding periods perform worst for group 2 due to the timing of the weights. C_1, C_4, C_5, C_6 and C_8 all belong to group 2. Hence, we confirm our beliefs that there is too much "noise" included in the weights for peripheral countries.

Table 9: MSE and Harvey, Leybourne and Newbold (HLN) test results of the combined forecasts for longer horizons.

Country	Holding period	Risk-Neutral Default Probability C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
Belgium	1 quarter	0.08*	**0.40**	**0.34**	0.11*	0.03*	0.03*	**0.46**	0.10*
	2 quarters	0.08*	**0.33**	**0.29**	0.12*	0.03*	0.03*	0.44*	0.10*
	1 year	0.07*	0.41*	**0.34**	0.16*	0.03*	0.05*	0.58*	0.11*
	2 years	0.16*	0.78*	0.64*	0.38	0.05*	0.11*	0.97*	0.27*
	4 years	0.61	2.11	1.70	1.21	0.15*	0.34	3.32	0.95
	Half time	0.81	**0.90**	**0.84**	1.31	0.18*	0.36	**0.81**	1.12
	All time	0.72	**0.90**	**0.84**	1.31	0.17*	0.36	**0.80**	1.07
France	1 quarter	**0.06**	**0.06**	**0.06**	**0.09**	**0.12**	**0.14**	**0.07**	**0.07**
	2 quarters	**0.08**	**0.08**	**0.08**	**0.11**	**0.13**	**0.15**	0.09**	**0.08**
	1 year	**0.09**	**0.09**	**0.09**	0.13**	**0.14**	**0.16**	0.11	**0.10**
	2 years	0.10**	0.10**	0.10**	0.14	**0.14**	0.15**	0.14	0.11**
	4 years	0.34	0.34	0.35	0.45	0.24	0.26	0.60	0.37
	Half time	0.19**	0.19**	0.19	0.18**	0.22**	0.21**	0.17	0.19**
	All time	0.18**	0.18**	0.18**	0.18**	0.21**	0.22**	**0.15**	0.18**
Germany	1 quarter	**0.02**	**0.02**	0.03	**0.02**	**0.02**	**0.03**	0.03	**0.02**
	2 quarters	**0.01**	**0.01**	**0.02**	**0.02**	**0.02**	**0.03**	**0.02**	**0.01**
	1 year	**0.02**	**0.02**	**0.02**	**0.02**	**0.02**	**0.03**	**0.03**	**0.01**
	2 years	**0.02**	**0.02**	**0.02**	0.02*	**0.02**	**0.03**	0.03	**0.02**
	4 years	0.05	0.05	0.05	0.06	0.04	0.04	0.08	0.05
	Half time	**0.03**	**0.03**	**0.03**	**0.03**	0.04**	0.04**	**0.02**	**0.03**
	All time	**0.03**	**0.03**	**0.03**	**0.03**	0.04**	0.04**	**0.02**	**0.03**
Italy	1 quarter	7.42	2.84	5.06	6.56	4.03	3.88	5.24	3.82
	2 quarters	6.77	3.19	4.99	6.10	3.76	3.69	5.21	3.55
	1 year	5.32	2.95	4.79	4.88	3.32	3.18	4.77	2.96
	2 years	5.80	3.21	4.98	5.47	3.47	3.32	5.41	3.29
	4 years	7.69	7.41	7.49	7.92	4.16**	4.19**	11.80	6.03
	Half time	4.94	4.16	4.45	4.86	3.65	3.67	4.12	2.90
	All time	4.74	3.94**	4.15	4.78	3.75	3.80	3.29	2.71
Spain	1 quarter	1.99	2.19**	2.17	0.67	0.92	0.69**	3.03**	1.30
	2 quarters	1.50	2.10**	2.09	**0.56**	0.76	0.61**	3.12	0.89
	1 year	1.63	2.59	2.52	0.62	0.72	0.70**	3.95	1.02
	2 years	2.27	3.18	3.14	0.98	0.72	**0.71**	4.74	1.62
	4 years	6.07	11.51	10.80	4.54	1.81	1.70	18.46	5.60
	Half time	7.85	**3.51**	3.37*	5.15	2.27	1.83	**3.29**	6.98
	All time	7.85	**3.03**	**2.87**	5.16	2.27	1.83	**2.37**	6.98

Note. MSE is in percentages; * significantly more accurate than $RNPD_S$, ** significantly more accurate than PD_H, **bold** significantly more accurate than both $RNPD_S$ and PD_H ($\alpha = 0.05$).

4.4 Market overreaction

In the previous sections we have mainly provided mechanical reasons why our method may be better suited for core countries than for peripheral countries. We argued that there is too much “noise” included in the weights for peripheral countries. However, there actually is an explanation for this “noise” phenomenon which has its roots in behavioral economics, known as market overreaction.

Financial markets tend to overreact in times of crisis. Rather than acting rationally, investors are driven by emotions and hence the information obtained from financial markets may be biased. Moreover, riskier assets are affected more than assets that are considered to be less risky. Peripheral countries are regarded as riskier than core countries. As a result, both equity and bond markets of peripheral countries are affected more by market overreaction than the financial markets of core countries. The forecast combination weights in Figure 8 clearly show this as the weights for core countries are “smooth” even in crisis periods, while for peripheral countries these are only “smooth” after the crisis.

Based on the entire testing period, we concluded for peripheral countries that none of the 8 combinations led to a significant higher accuracy than both benchmarks, Table 8. This generally holds for all other testing periods that include observations prior to Q2 2015. Note that RNDP forecasts prior to Q2 2015 were affected by the crisis. The only period for which the method shows improvements for peripheral countries is from Q2 2015 onward.

Table 10 shows the HLN test results for peripheral countries over the period of Q2 2015 - Q1 2020. For Spain, only C_7 does not show a significant higher accuracy than both benchmarks. All the 7 other combinations lead to a significantly higher accuracy. For Italy, C_1, C_2 and C_8 result in a significant higher accuracy than both benchmarks. While all combinations lead to a significant outperformance of PD_H.

Table 10: MSE and Harvey, Leybourne and Newbold (HLN) test results for the combined forecasts of peripheral countries based on Q2 2015 - Q1 2020.

Country	Benchmark		Risk-Neutral Default Probability							
	S	H	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
Italy	0.33	0.63	**0.18**	**0.17**	0.24**	0.26**	0.25**	0.25**	0.35**	**0.18**
Spain	0.16	0.16	**0.06**	**0.06**	**0.06**	**0.05**	**0.05**	**0.05**	0.12	**0.04**

Note. MSE in percentages; * significantly more accurate than RNPD_S, ** significantly more accurate than PD_H, **bold** significantly more accurate than both RNPD_S and PD_H ($\alpha = 0.05$).

Note that it is not a coincidence that our method only works from Q2 2015 onward for peripheral countries. In March 2015, the ECB started buying assets from commercial banks to support economic growth across the euro area and control inflation. This asset purchase program is known as quantitative easing (QE). Over almost four years, the ECB has spent 2.6 trillion euros buying up mostly government bonds. Consequently, bond prices increased and yields dropped to record lows. At the same time, QE has caused bond markets to become less volatile and more predictable. There is less overreaction in bond markets. Moreover, the economy was relatively stable in this period and equity markets displayed less volatility as there was no cause to overreact. Therefore, we conclude that our method does not perform well for peripheral countries due to market overreaction. For non-crisis periods, we improve the CCA by incorporating information from both equity and bond markets for peripheral countries.

5 Discussion and Conclusion

This study examines whether the contingent claims approach (CCA) to measure sovereign default risk for member states of the Economic and Monetary Union of the European Union (EMU) can be improved by incorporating information from equity and bond markets. More specifically, we have used the volatility of equity indices, future equity indices, STOXX 50 and the volatility of bond yields with a maturity of 1, 5 and 10 years to determine the Risk-Neutral Default Probability (RNDP). The individual RNDPs that result from these volatilities have been combined using forecast combinations. Both individual RNDP forecasts as well as combined forecasts are evaluated using the Market-Implied Default Probability that follows from CDS spreads. We have used longer forecast horizons to determine how often weights should be re-calibrated and the impact of these weights on the RNDP. Our analysis is conducted over the period of 2008 Q1 - 2020 Q1 and includes three core countries, Belgium (BE), France (FR) and

Germany (DE), and two peripheral countries, Italy (IT) and Spain (ES).

Our results show that the combined approach reduces the MSE for all countries when compared with the benchmarks, the standard CCA and a rating-based method. However, based on the Harvey, Leybourne and Newbold (HLN) test, the combined RNDP forecasts are only significantly more accurate, than both benchmarks and over the entire period, for core countries. For core countries, we obtain the best results for combinations including one equity RNDP and one bond RNDP, with new weights being determined more than once a year. The best combinations are: (C_1) RNDP_1 (equity index) & RNDP_4 (1y bond yield); (C_2) RNDP_1 (equity index) & RNDP_6 (10y bond yield); (C_3) RNDP_2 (future index) & RNDP_5 (5y bond yield). Furthermore, the HLN test shows that the combined RNDP forecasts are also significantly more accurate than both benchmarks for peripheral countries for the period of Q2 2015 - Q1 2020. For peripheral countries during this period, the best combinations are (C_1) RNDP_1 (equity index) & RNDP_4 (1y bond yield); (C_2) RNDP_1 (equity index) & RNDP_6 (10y bond yield). Our analysis shows that our approach is not optimal for peripheral countries in times of crisis due to market overreaction.

The first combined forecast C_1 does not show a significantly higher accuracy than both benchmarks for Belgium. This brings us to the first caveat, we did not have data on 1-year bond yields prior to 12-07-2011 for Belgium and prior to 05-08-2011 for Spain. Hence, for both countries we could only include RNDP_4 from Q4 2011 onward. As a consequence, we could only determine the combined RNDP forecasts including RNDP_4 from Q3 2012 onward. Note that CDS spreads started to recover since this period. The heightened Market-Implied Default Probabilities, observed during the financial and sovereign debt crisis, started to drop. Hence, we could not analyse the performance of C_1, C_4, C_5, C_6 and C_8 during crisis periods. Especially for Belgium, we would expect that these combinations would have shown promising results.

A second caveat is the number of observations and number of countries that we have included in the research. Our test results include 49 and 46 observations for the individual and combined RNDPs, respectively. For Belgium and Spain, this is equal to 34 for RNDP_4 and 31 for combinations that include RNDP_4. The reason for only including this many observations is that sovereign balance sheet items were only available on a quarterly basis. Hence, we took the maximum number of quarters for which data was available. In total we only include 5 countries, 3 core and 2 peripheral. Also this was a consequence of data limitations. For other member states of the EMU, either balance sheet items, bond yields, equity returns or CDS spreads were not available. Even in our current set of countries, we still lack data on 1 year bond yields for Belgium and Spain. However, there was enough data such that these countries could be included. For other countries, this was not possible.

A third caveat, which severely limits our results, are the benchmarks. Although our benchmarks are based on literature and are commonly used, they are weak at best. The standard CCA method is actually intended for emerging countries with a weak currency. Our standard CCA benchmark is also an adapted version of the method by Gray et al. (2007). Using the exact same method would have given the same results as our other benchmark, the rating-based method. The rating-based method is based on "real-world" (historical) default probabilities, rather than risk-neutral default probabilities. During the entire time horizon, the historical default probability is equal to 0 for all countries included. Therefore, these benchmarks are not necessarily hard to "beat". However, as we were aware of this, we take this into consideration in our analysis. We do not merely conclude that a forecast is "good" because it significantly outperforms both benchmarks. Forecasts are only labelled as "good", if they significantly outperform both benchmarks, result in a sufficiently low MSE and describe the pattern of the Market-Implied Default Probability well for several countries. Unfortunately, these weak benchmarks are a direct result of the limited literature on sovereign default risk for developed countries. Which leaves room for further research on this topic.

Future recommendations on our work are as follows. First and foremost, our research can be improved by addressing the first two caveats. This can be achieved by including more member states of the EMU, both core and peripheral, and including more observations. More observations can either follow from considering a longer time horizon, or by using a weekly or monthly frequency rather than quarterly. It would be interesting to see whether our results would also hold for other countries and periods. Another approach to solve these caveats is by pooling data across countries. As for some countries, there were "gaps" in the data, these "gaps" could be filled with estimates. By regressing on the available data and then estimating the missing data. This may give adequate results, but also gives an added layer of uncertainty due to estimation.

Second, the current research can also be extended by considering different forecast combinations or by varying the way in which the volatility or weights are determined. Different forecasts could be included in the combination, outside of the 8 that we consider. Apart from VSTOXX, the other volatilities are determined using a 60-day rolling window. The weights are determined using historical errors of the past 3 quarters. Using a longer or shorter window could be beneficial. Thirdly, future research could be performed on the construction of the sovereign balance sheet. The potential effect of what is included in the equity or liabilities could be researched. Similarly, determining what impact different percentages of long-term liabilities included in the distress barrier may have.

Furthermore, our work could also be extended by using a linear combination of the volatilities rather than combining the different RNDPs. In fact, this was our initial approach, where our aim was to find a linear combination of the volatilities which could be used as a proxy for the asset volatility. To determine the combination, we calculated the implied volatility from the Market-Implied Default Probability and regressed this on the six different equity and bond volatilities. However, the problem with this was that the eventual RNDP is very sensitive to the volatility. Small changes in the volatility had relatively large effects on the RNDP. For example, if the implied volatility was 0.21043, then a volatility below 0.21040 would only give 0's while a volatility above 0.21045 would result in a RNDP that was significantly higher than the Market-Implied Default Probability. The reason for this, among others, is that the RNDPs are very low (in basis points, and close to 0) and hence small shifts in the volatilities would cause the RNDPs to either be too low or too high. Moreover, for all countries the RNDP was equal to 0 after the sovereign debt crisis, when volatilities were relatively lower. Hence, it is crucial to find am extremely accurate estimate of the asset volatility to use this approach. Our regressions did not provide such accurate estimates. Therefore, we dropped this idea and moved to forecasts combinations.

That being said, we must note that the regressions we used at the time were performed on each country separately. These regressions included at least 20 observations (quarters), which is about half the entire time period. Due to this, we did not use a regression which only included observations after the sovereign debt crisis for example. As there were not enough observations in this period. It may be possible that if the coefficients from the regression would only be based on the period after the sovereign debt crisis, that the resulting RNDP would not only give 0's. With the knowledge we have now, we realise that the time frame is more important than considering countries separately. Hence, using regressions on shorter time periods, it may have been possible to find better results with the initial approach. We realise now that this could be done by pooling countries together, such that the regressions can be performed on a shorter time period, as more observations would be included in the regression. However, this still is not that simple as the RNDP is sensitive to the volatility. Quite some research would need to be performed to find regression results which would give such accurate volatilities. Although we do think that this approach would give better results than what we initially used, we doubt that it would beat the forecast combination approach.

Lastly, although not a direct extension of our work, we urge that more research is performed on this topic. We think that the main limitation for this is the amount of available balance sheet data. Once more data becomes available, other techniques, such as machine learning, could also be deployed for research. In the meanwhile, researchers should aim to develop other models, such that more promising benchmarks become available. As a final note, we would like to conclude that, irrespective of the model decided to use, sovereign default risk can only be described adequately by incorporating information from both the sovereign balance sheet and financial markets.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Ang, A. and Longstaff, F. A. (2013). Systemic sovereign credit risk: Lessons from the us and europe. *Journal of Monetary Economics*, 60(5):493–510.

Ariyoshi, A., Kirilenko, A., Ötker, I., Laurens, B., Kriljenko, J. I. C., and Habermeier, K. F. (2000). *Capital controls: Country experiences with their use and liberalization*. Number 190. International Monetary Fund.

Cochrane, J. H. (2005). Money as stock. *Journal of Monetary Economics*, 52(3):501–528.

De Grauwe, P. (2012). The governance of a fragile eurozone. *Australian Economic Review*, 45(3):255–268.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Duyvesteyn, J. and Martens, M. (2015). Forecasting sovereign default risk with merton's model. *Journal of Fixed Income*, 25(2):58–71.

Gapen, M., Gray, D., Lim, C. H., and Xiao, Y. (2008). Measuring and analyzing sovereign risk with contingent claims. *IMF Staff Papers*, 55(1):109–148.

Gray, D. F., Merton, R. C., and Bodie, Z. (2007). Contingent claims approach to measuring and managing sovereign credit risk. *Journal of Investment Management*, 5(4):5.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Kahlert, D., Wagner, N., and Weipert, L. (2017). Contingent claims analysis of sovereign default risk in the eurozone. *Available at SSRN 2957385*.

Kincaid, G. R. and Collyns, C. (2003). *Managing financial crises: Recent experience and lessons for Latin America*. Number 217. International Monetary Fund.

Kopf, C. (2011). Restoring financial stability in the euro area. No. 237, CEPS Policy Briefs.

Longstaff, F. A., Mithal, S., and Neis, E. (2005). Corporate yield spreads: Default risk or liquidity? new evidence from the credit default swap market. *Journal of Finance*, 60(5):2213–2253.

Longstaff, F. A., Pan, J., Pedersen, L. H., and Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

Merler, S. and Pisani-Ferry, J. (2012). Who's afraid of sovereign bonds? Technical report, Bruegel Policy Contribution.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2):449–470.

Morgan, J. (1997). Creditmetrics-technical document. *JP Morgan, New York*.

Sims, C. A. (1999). Domestic currency denominated government debt as equity in the primary surplus. In *Latin American meetings of the Econometric Society, Cancun, Mexico. August*, volume 19.

Steinkamp, S. and Westermann, F. (2014). The role of creditor seniority in europe's sovereign debt crisis. *Economic Policy*, 29(79):495–552.

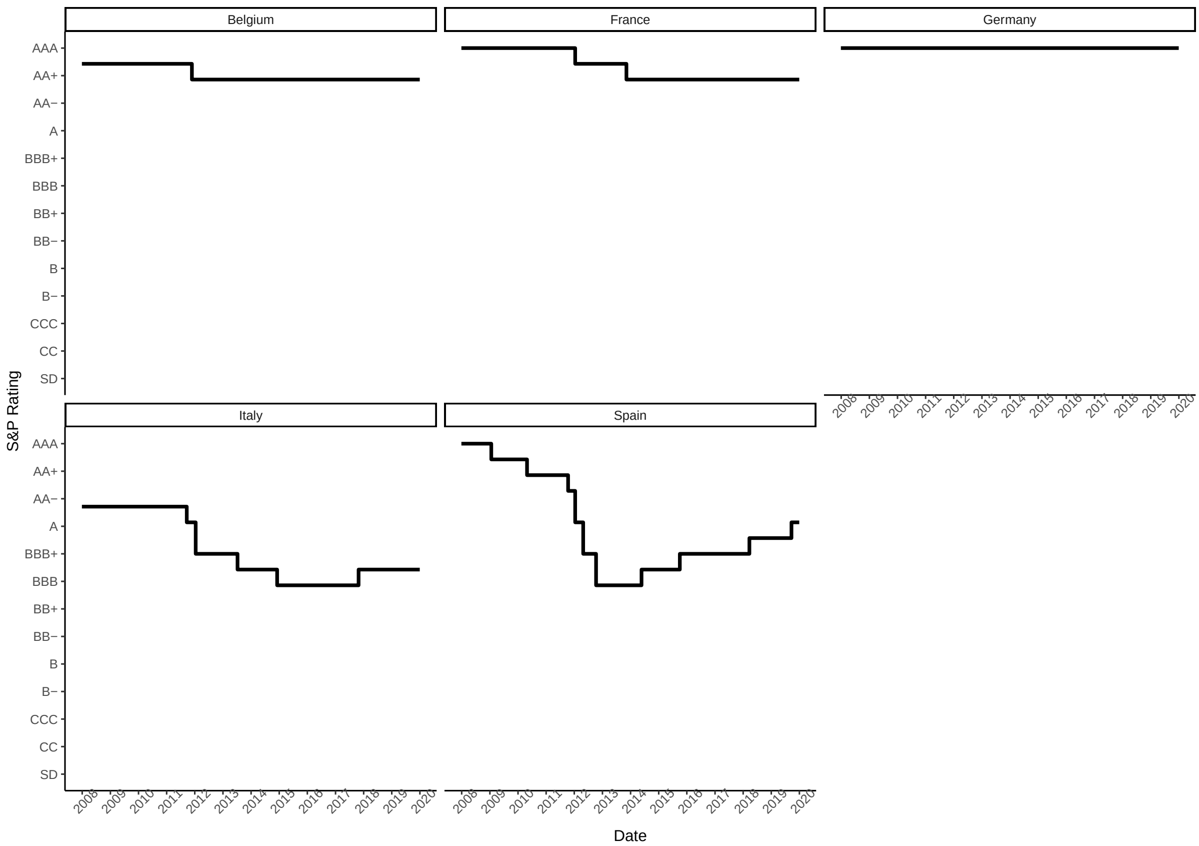
Tiao, G. C. and Box, G. E. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816.

Wilson, T. C. (1998). Portfolio credit risk. *Economic Policy Review*, 4(3).

Appendices

A S&P ratings

A.1 Ratings



This figure shows the historical credit ratings for all countries that are considered in this research. The ratings in this graph are the ratings for the long-term foreign currency debt and are determined by Standard & Poor's. Note that NR means Not Rated and SD means Selective Default.

A.2 Historical one-year default probabilities

	1975 until					
Initial rating	2010	2011	2014	2016	2017	2018
AAA	0.00	0.00	0.00	0.00	0.00	0.00
AA+	0.00	0.00	0.00	0.00	0.00	0.00
AA	0.00	0.00	0.00	0.00	0.00	0.00
AA-	0.00	0.00	0.00	0.00	0.00	0.00
A+	0.00	0.00	0.00	0.00	0.00	0.00
A	0.00	0.00	0.00	0.00	0.00	0.00
A-	0.00	0.00	0.00	0.00	0.00	0.00
BBB+	0.00	0.00	0.00	0.00	0.00	0.00
BBB	0.00	0.00	0.00	0.00	0.00	0.00
BBB-	0.00	0.00	0.00	0.00	0.00	0.00
BB+	0.00	0.00	0.20	0.20	0.20	0.19
BB	0.00	0.00	0.10	0.10	0.11	0.10
BB-	2.10	2.00	1.40	1.20	1.12	1.05
B+	0.00	0.00	0.50	0.50	0.69	0.64
B	2.00	1.80	2.20	2.30	2.32	2.10
B-	5.60	5.10	8.20	7.00	7.91	7.34
CCC+	15.40	15.40	23.10	23.40	23.41	19.21
CCC	40.00	40.00	42.90	35.10	36.84	37.50
CCC-	100.00	100.00	77.80	78.80	78.79	78.95
CC	100.00	100.00	100.00	100.00	100.00	100.00

This table shows the historical one-year default probabilities corresponding to credit ratings at the start of the year. The default probabilities are based on annual sovereign default reports supplied by Standard & Poor's. We use the sovereign foreign currency average one-year transition rates with rating modifiers to derive the PD. For each report, the historical PD is based data starting in 1975 until the year in which the report is published.

B Harvey, Leybourne and Newbold (HLN) test results

B.1 Individual forecasts

Country	Value	Benchmark S	Benchmark H	Risk Neutral Default Probability 1	2	3	4	5	6
Belgium	MSE	3.74	0.98	462.71	503.51	823.12	0.59	0.89	0.97
Belgium	Statistic			2.80	2.94	3.71	-1.96	-1.86	-1.80
Belgium	p-value			1.00	1.00	1.00	0.03	0.03	0.04
Belgium	Statistic_H			2.83	2.96	3.73	-1.05	-1.23	-1.67
Belgium	p-value_H			1.00	1.00	1.00	0.15	0.11	0.05
France	MSE	0.21	0.24	443.50	439.61	596.92	0.24	0.24	0.24
France	Statistic			4.19	2.68	2.95	1.88	1.88	1.89
France	p-value			1.00	1.00	1.00	0.97	0.97	0.97
France	Statistic_H			4.19	2.68	2.95	-1.16	-1.38	-1.05
France	p-value_H			1.00	1.00	1.00	0.12	0.09	0.15
Germany	MSE	0.09	0.05	162.01	160.84	239.45	0.05	0.05	0.05
Germany	Statistic			3.08	2.83	3.44	-1.25	-1.25	-1.25
Germany	p-value			1.00	1.00	1.00	0.11	0.11	0.11
Germany	Statistic_H			3.08	2.83	3.44	-1.03	-1.31	-1.03
Germany	p-value_H			1.00	1.00	1.00	0.15	0.10	0.15
Italy	MSE	19.69	5.10	1116.57	1057.87	1072.25	4.87	4.78	4.55
Italy	Statistic			7.29	7.08	3.78	-1.10	-1.11	-1.13
Italy	p-value			1.00	1.00	1.00	0.14	0.14	0.13
Italy	Statistic_H			5.06	4.87	3.70	-0.42	-0.15	-0.92
Italy	p-value_H			1.00	1.00	1.00	0.34	0.44	0.18
Spain	MSE	4.30	4.55	261.56	263.04	265.19	3.62	4.22	4.52
Spain	Statistic			4.56	4.65	5.30	-1.11	-2.99	1.29
Spain	p-value			1.00	1.00	1.00	0.14	0.00	0.90
Spain	Statistic_H			4.55	4.65	5.30	-1.54	-2.13	-1.43
Spain	p-value_H			1.00	1.00	1.00	0.07	0.02	0.08

Note. MSE in percentages

B.2 Combined forecasts

Country	Value	Benchmark S	H	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
				Risk Neutral Default Probability							
Belgium	MSE	3.81	1.05	0.08	0.40	0.34	0.11	0.03	0.03	0.46	0.10
Belgium	Statistic			-1.79	-2.08	-2.10	-1.78	-1.81	-1.82	-2.02	-1.78
Belgium	p-value			0.04	0.02	0.02	0.04	0.04	0.04	0.02	0.04
Belgium	Statistic_H			0.24	-1.99	-1.88	0.58	-0.75	-0.76	-1.45	0.47
Belgium	p-value_H			0.59	0.03	0.03	0.72	0.23	0.23	0.08	0.68
France	MSE	0.22	0.25	0.06	0.06	0.06	0.09	0.12	0.14	0.07	0.07
France	Statistic			-2.15	-2.15	-2.14	-2.27	-2.21	-2.43	-1.72	-2.21
France	p-value			0.02	0.02	0.02	0.01	0.02	0.01	0.05	0.02
France	Statistic_H			-2.50	-2.50	-2.49	-2.65	-2.82	-1.99	-2.01	-2.57
France	p-value_H			0.01	0.01	0.01	0.01	0.00	0.03	0.03	0.01
Germany	MSE	0.10	0.05	0.02	0.02	0.03	0.02	0.02	0.03	0.03	0.02
Germany	Statistic			-1.74	-1.74	-1.65	-1.94	-1.85	-1.76	-1.57	-1.82
Germany	p-value			0.04	0.04	0.05	0.03	0.04	0.04	0.06	0.04
Germany	Statistic_H			-2.13	-2.13	-1.11	-2.19	-1.86	-2.62	-1.58	-1.83
Germany	p-value_H			0.02	0.02	0.14	0.02	0.03	0.01	0.06	0.04
Italy	MSE	20.86	5.43	7.42	2.84	5.06	6.56	4.03	3.88	5.24	3.82
Italy	Statistic			-0.89	-1.27	-1.09	-0.96	-1.18	-1.19	-1.07	-1.19
Italy	p-value			0.19	0.11	0.14	0.17	0.12	0.12	0.14	0.12
Italy	Statistic_H			0.91	-1.33	-0.11	0.64	-0.72	-0.77	-0.06	-0.67
Italy	p-value_H			0.82	0.10	0.46	0.74	0.24	0.22	0.48	0.25
Spain	MSE	4.58	4.84	1.99	2.19	2.17	0.67	0.92	0.69	3.03	1.30
Spain	Statistic			0.02	-1.55	-1.40	-1.53	-1.13	-1.54	-0.84	-0.45
Spain	p-value			0.51	0.06	0.08	0.07	0.13	0.07	0.20	0.33
Spain	Statistic_H			-0.15	-1.71	-1.55	-1.64	-1.32	-2.26	-1.71	-0.64
Spain	p-value_H			0.44	0.05	0.06	0.06	0.10	0.02	0.05	0.26

Note. MSE in percentages

B.3 Time period after QE - Peripheral countries

Country	Value	Benchmark S	Benchmark H	Risk Neutral Default Probability C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
Italy	MSE	0.33	0.63	0.18	0.17	0.24	0.26	0.25	0.25	0.35	0.18
Italy	Statistic			-2.02	-2.13	-0.81	-0.47	-1.30	-1.21	0.11	-1.82
Italy	p-value			0.03	0.02	0.21	0.32	0.10	0.12	0.54	0.04
Italy	Statistic_H			-5.25	-5.41	-5.13	-2.45	-5.53	-6.16	-2.15	-5.38
Italy	p-value_H			0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00
Spain	MSE	0.16	0.16	0.06	0.06	0.06	0.05	0.05	0.05	0.12	0.04
Spain	Statistic			-2.86	-2.87	-3.02	-2.51	-2.11	-2.47	-0.77	-3.46
Spain	p-value			0.00	0.00	0.00	0.01	0.02	0.01	0.23	0.00
Spain	Statistic_H			-2.86	-2.87	-3.02	-2.51	-2.11	-2.47	-0.77	-3.46
Spain	p-value_H			0.00	0.00	0.00	0.01	0.02	0.01	0.23	0.00

Note. MSE in percentages

C Harvey, Leybourne and Newbold (HLN) test results for longer horizon forecasting

C.1 Belgium

Qs	Value	Benchmark		Risk Neutral Default Probability							
		S	H	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
2	MSE	3.81	1.05	0.08	0.33	0.29	0.12	0.03	0.03	0.44	0.10
2	Statistic			-1.79	-2.10	-2.13	-1.77	-1.82	-1.82	-2.02	-1.79
2	p-value			0.04	0.02	0.02	0.04	0.04	0.04	0.02	0.04
2	Statistic_H			0.18	-1.80	-1.84	0.77	-0.78	-0.74	-1.38	0.45
2	p-value_H			0.57	0.04	0.04	0.78	0.22	0.23	0.09	0.67
4	MSE	3.81	1.05	0.07	0.41	0.34	0.16	0.03	0.05	0.58	0.11
4	Statistic			-1.80	-2.05	-2.10	-1.76	-1.82	-1.81	-1.92	-1.78
4	p-value			0.04	0.02	0.02	0.04	0.04	0.04	0.03	0.04
4	Statistic_H			0.08	-1.55	-1.70	1.01	-0.90	-0.40	-0.99	0.61
4	p-value_H			0.53	0.06	0.05	0.84	0.19	0.35	0.16	0.73
8	MSE	3.81	1.05	0.16	0.78	0.64	0.38	0.05	0.11	0.97	0.27
8	Statistic			-1.76	-1.84	-1.93	-1.65	-1.81	-1.78	-1.69	-1.71
8	p-value			0.04	0.04	0.03	0.05	0.04	0.04	0.05	0.05
8	Statistic_H			1.25	-1.02	-1.36	1.59	-0.40	0.72	-0.14	1.99
8	p-value_H			0.89	0.16	0.09	0.94	0.35	0.76	0.44	0.97
16	MSE	3.81	1.05	0.61	2.11	1.70	1.21	0.15	0.34	3.32	0.95
16	Statistic			-1.54	-0.99	-1.25	-1.28	-1.76	-1.67	-0.17	-1.39
16	p-value			0.07	0.16	0.11	0.11	0.04	0.05	0.43	0.09
16	Statistic_H			2.25	1.66	7.66	2.11	6.45	4.61	38.26	2.01
16	p-value_H			0.98	0.95	1.00	0.98	1.00	1.00	1.00	0.97
23	MSE	3.81	1.05	0.81	0.90	0.84	1.31	0.18	0.36	0.81	1.12
23	Statistic			-1.46	-1.78	-1.82	-1.23	-1.75	-1.66	-1.84	-1.32
23	p-value			0.08	0.04	0.04	0.11	0.05	0.05	0.04	0.10
23	Statistic_H			2.45	-2.97	-2.00	4.71	1.71	4.95	-2.49	2.42
23	p-value_H			0.99	0.00	0.03	1.00	0.95	1.00	0.01	0.99
100	MSE	3.81	1.05	0.72	0.90	0.84	1.31	0.17	0.36	0.80	1.07
100	Statistic			-1.50	-1.78	-1.82	-1.23	-1.75	-1.66	-1.85	-1.34
100	p-value			0.07	0.04	0.04	0.11	0.05	0.05	0.04	0.10
100	Statistic_H			1.98	-2.97	-1.99	4.68	6.78	4.95	-2.51	2.23
100	p-value_H			0.97	0.00	0.03	1.00	1.00	1.00	0.01	0.98

Note. MSE in percentages

C.2 France

Qs	Value	Benchmark		Risk Neutral Default Probability							
		S	H	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
2	MSE	0.22	0.25	0.08	0.08	0.08	0.11	0.13	0.15	0.09	0.08
2	Statistic			-2.14	-2.14	-2.12	-2.16	-2.15	-2.40	-1.66	-2.18
2	p-value			0.02	0.02	0.02	0.02	0.02	0.01	0.05	0.02
2	Statistic_H			-2.52	-2.52	-2.49	-2.61	-2.85	-2.10	-1.96	-2.58
2	p-value_H			0.01	0.01	0.01	0.01	0.00	0.02	0.03	0.01
4	MSE	0.22	0.25	0.09	0.09	0.09	0.13	0.14	0.16	0.11	0.10
4	Statistic			-1.91	-1.91	-1.90	-1.66	-1.93	-1.95	-1.30	-1.88
4	p-value			0.03	0.03	0.03	0.05	0.03	0.03	0.10	0.03
4	Statistic_H			-2.32	-2.32	-2.30	-2.14	-2.70	-1.97	-1.61	-2.31
4	p-value_H			0.01	0.01	0.01	0.02	0.00	0.03	0.06	0.01
8	MSE	0.22	0.25	0.10	0.10	0.10	0.14	0.14	0.15	0.14	0.11
8	Statistic			-1.63	-1.63	-1.60	-1.08	-1.76	-1.68	-0.89	-1.55
8	p-value			0.05	0.06	0.06	0.14	0.04	0.05	0.19	0.06
8	Statistic_H			-1.99	-1.99	-1.97	-1.44	-2.42	-2.34	-1.17	-1.91
8	p-value_H			0.03	0.03	0.03	0.08	0.01	0.01	0.12	0.03
16	MSE	0.22	0.25	0.34	0.34	0.35	0.45	0.24	0.26	0.60	0.37
16	Statistic			1.03	1.03	1.10	4.10	0.64	1.24	1.37	1.14
16	p-value			0.85	0.84	0.86	1.00	0.74	0.89	0.91	0.87
16	Statistic_H			0.73	0.73	0.80	2.77	-0.32	0.46	1.24	1.92
16	p-value_H			0.76	0.76	0.79	1.00	0.37	0.68	0.89	0.97
23	MSE	0.22	0.25	0.19	0.19	0.19	0.18	0.22	0.21	0.17	0.19
23	Statistic			-1.21	-1.21	-1.05	-1.55	-0.61	-0.69	-1.42	-1.32
23	p-value			0.12	0.12	0.15	0.06	0.27	0.25	0.08	0.10
23	Statistic_H			-1.71	-1.71	-1.64	-2.09	-2.16	-2.24	-1.50	-1.87
23	p-value_H			0.05	0.05	0.05	0.02	0.02	0.01	0.07	0.03
100	MSE	0.22	0.25	0.18	0.18	0.18	0.18	0.21	0.22	0.15	0.18
100	Statistic			-1.56	-1.56	-1.44	-1.58	-0.63	-0.65	-1.98	-1.53
100	p-value			0.06	0.06	0.08	0.06	0.27	0.26	0.03	0.07
100	Statistic_H			-2.15	-2.14	-2.15	-2.13	-2.20	-2.19	-2.08	-2.15
100	p-value_H			0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Note. MSE in percentages

C.3 Germany

Qs	Value	Benchmark S	Benchmark H	Risk Neutral Default Probability C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
2	MSE	0.10	0.05	0.01	0.01	0.02	0.02	0.02	0.03	0.02	0.01
2	Statistic			-1.82	-1.82	-1.75	-1.91	-1.93	-1.76	-1.77	-1.88
2	p-value			0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03
2	Statistic_H			-2.32	-2.32	-2.20	-2.11	-2.06	-2.66	-1.91	-2.40
2	p-value_H			0.01	0.01	0.02	0.02	0.02	0.01	0.03	0.01
4	MSE	0.10	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.01
4	Statistic			-1.72	-1.72	-1.70	-1.87	-1.92	-1.74	-1.70	-1.83
4	p-value			0.05	0.05	0.05	0.03	0.03	0.04	0.05	0.04
4	Statistic_H			-2.16	-2.16	-2.15	-2.00	-1.95	-2.62	-1.68	-1.93
4	p-value_H			0.02	0.02	0.02	0.03	0.03	0.01	0.05	0.03
8	MSE	0.10	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.02
8	Statistic			-1.84	-1.84	-1.92	-1.83	-1.97	-1.75	-1.64	-1.88
8	p-value			0.04	0.04	0.03	0.04	0.03	0.04	0.05	0.03
8	Statistic_H			-2.43	-2.43	-2.38	-1.67	-1.94	-2.08	-1.55	-2.47
8	p-value_H			0.01	0.01	0.01	0.05	0.03	0.02	0.06	0.01
16	MSE	0.10	0.05	0.05	0.05	0.05	0.06	0.04	0.04	0.08	0.05
16	Statistic			-1.27	-1.27	-1.24	-0.83	-1.47	-1.34	-0.32	-1.15
16	p-value			0.11	0.11	0.11	0.21	0.07	0.09	0.38	0.13
16	Statistic_H			-0.34	-0.34	-0.26	0.79	-1.40	-1.06	0.52	-0.04
16	p-value_H			0.37	0.37	0.40	0.78	0.08	0.15	0.70	0.49
23	MSE	0.10	0.05	0.03	0.03	0.03	0.03	0.04	0.04	0.02	0.03
23	Statistic			-1.80	-1.80	-1.79	-1.76	-1.58	-1.52	-1.91	-1.80
23	p-value			0.04	0.04	0.04	0.04	0.06	0.07	0.03	0.04
23	Statistic_H			-2.09	-2.09	-2.01	-2.29	-2.31	-2.39	-1.91	-2.15
23	p-value_H			0.02	0.02	0.03	0.01	0.01	0.01	0.03	0.02
100	MSE	0.10	0.05	0.03	0.03	0.03	0.03	0.04	0.04	0.02	0.03
100	Statistic			-1.85	-1.85	-1.85	-1.76	-1.59	-1.52	-1.98	-1.83
100	p-value			0.04	0.04	0.04	0.04	0.06	0.07	0.03	0.04
100	Statistic_H			-2.32	-2.32	-2.33	-2.28	-2.40	-2.39	-2.21	-2.32
100	p-value_H			0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01

Note. MSE in percentages

C.4 Italy

Qs	Value	Benchmark S	Benchmark H	Risk Neutral Default Probability C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
2	MSE	20.86	5.43	6.77	3.19	4.99	6.10	3.76	3.69	5.21	3.55
2	Statistic			-0.95	-1.24	-1.09	-1.00	-1.20	-1.21	-1.08	-1.21
2	p-value			0.17	0.11	0.14	0.16	0.12	0.12	0.14	0.12
2	Statistic_H			0.75	-1.25	-0.13	0.45	-0.79	-0.82	-0.07	-0.74
2	p-value_H			0.77	0.11	0.45	0.67	0.22	0.21	0.47	0.23
4	MSE	20.86	5.43	5.32	2.95	4.79	4.88	3.32	3.18	4.77	2.96
4	Statistic			-1.07	-1.26	-1.11	-1.11	-1.24	-1.25	-1.11	-1.26
4	p-value			0.15	0.11	0.14	0.14	0.11	0.11	0.14	0.11
4	Statistic_H			-0.07	-1.67	-0.22		-1.11	-1.19	-0.24	-1.13
4	p-value_H			0.47	0.05	0.41		0.14	0.12	0.41	0.13
8	MSE	20.86	5.43	5.80	3.21	4.98	5.47	3.47	3.32	5.41	3.29
8	Statistic			-1.03	-1.24	-1.09	-1.06	-1.23	-1.24	-1.06	-1.24
8	p-value			0.15	0.11	0.14	0.15	0.11	0.11	0.15	0.11
8	Statistic_H			0.26	-1.33	-0.14		-1.11	-1.19	-0.01	-0.97
8	p-value_H			0.60	0.09	0.45		0.14	0.12	0.50	0.17
16	MSE	20.86	5.43	7.69	7.41	7.49	7.92	4.16	4.19	11.80	6.03
16	Statistic			-0.90	-0.92	-0.91	-0.88	-1.17	-1.17	-0.59	-1.02
16	p-value			0.19	0.18	0.18	0.19	0.12	0.12	0.28	0.16
16	Statistic_H			1.67	4.02	0.67	1.44	-1.92	-1.75	9.68	0.28
16	p-value_H			0.95	1.00	0.75	0.92	0.03	0.04	1.00	0.61
23	MSE	20.86	5.43	4.94	4.16	4.45	4.86	3.65	3.67	4.12	2.90
23	Statistic			-1.10	-1.17	-1.14	-1.11	-1.21	-1.21	-1.17	-1.27
23	p-value			0.14	0.12	0.13	0.14	0.12	0.12	0.12	0.11
23	Statistic_H			-0.81	-1.53	-0.39	-1.00	-1.37	-1.35	-0.84	-1.40
23	p-value_H			0.21	0.07	0.35	0.16	0.09	0.09	0.20	0.08
100	MSE	20.86	5.43	4.74	3.94	4.15	4.78	3.75	3.80	3.29	2.71
100	Statistic			-1.12	-1.19	-1.16	-1.12	-1.21	-1.20	-1.23	-1.28
100	p-value			0.13	0.12	0.13	0.13	0.12	0.12	0.11	0.10
100	Statistic_H			-1.23	-1.84	-0.51	-1.17	-1.30	-1.25	-1.42	-1.51
100	p-value_H			0.11	0.04	0.31	0.12	0.10	0.11	0.08	0.07

Note. MSE in percentages

C.5 Spain

Qs	Value	Benchmark S	Benchmark H	Risk Neutral Default Probability C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
2	MSE	4.58	4.84	1.50	2.10	2.09	0.56	0.76	0.61	3.12	0.89
2	Statistic			-0.30	-1.56	-1.43	-1.74	-1.37	-1.50	-0.78	-0.91
2	p-value			0.38	0.06	0.08	0.05	0.09	0.07	0.22	0.18
2	Statistic_H			-0.50	-1.72	-1.58	-1.79	-1.51	-2.26	-0.92	-1.12
2	p-value_H			0.31	0.05	0.06	0.04	0.07	0.02	0.18	0.13
4	MSE	4.58	4.84	1.63	2.59	2.52	0.62	0.72	0.70	3.95	1.02
4	Statistic			-0.21	-1.28	-1.21	-1.60	-1.41	-1.50	-0.33	-0.79
4	p-value			0.42	0.10	0.12	0.06	0.08	0.07	0.37	0.22
4	Statistic_H			-0.41	-1.44	-1.36	-1.68	-1.54	-2.18	-0.46	-1.01
4	p-value_H			0.34	0.08	0.09	0.05	0.07	0.02	0.32	0.16
8	MSE	4.58	4.84	2.27	3.18	3.14	0.98	0.72	0.71	4.74	1.62
8	Statistic			0.21	-0.90	-0.85	-1.14	-1.38	-2.24	0.08	-0.28
8	p-value			0.58	0.19	0.20	0.13	0.09	0.02	0.53	0.39
8	Statistic_H			-0.00	-1.07	-1.01	-1.28	-1.52	-2.14	-0.05	-0.52
8	p-value_H			0.50	0.15	0.16	0.10	0.07	0.02	0.48	0.30
16	MSE	4.58	4.84	6.07	11.51	10.80	4.54	1.81	1.70	18.46	5.60
16	Statistic			1.76	1.49	1.46	6.73	-0.14	-0.21	1.65	1.54
16	p-value			0.96	0.93	0.92	1.00	0.45	0.42	0.95	0.93
16	Statistic_H			1.58	1.43	1.39	4.07	-0.40	-0.40	1.62	1.35
16	p-value_H			0.94	0.92	0.91	1.00	0.35	0.35	0.94	0.91
23	MSE	4.58	4.84	7.85	3.51	3.37	5.15	2.27	1.83	3.29	6.98
23	Statistic			2.12	-2.27	-2.35		0.28	-0.10	-1.82	
23	p-value			0.98	0.01	0.01		0.61	0.46	0.04	
23	Statistic_H			1.96	-2.61	-1.26		-0.01	-0.31	-2.08	
23	p-value_H			0.97	0.01	0.11		0.50	0.38	0.02	
100	MSE	4.58	4.84	7.85	3.03	2.87	5.16	2.27	1.83	2.37	6.98
100	Statistic			2.12	-2.05	-2.01		0.28	-0.10	-1.82	
100	p-value			0.98	0.02	0.03		0.61	0.46	0.04	
100	Statistic_H			1.96	-2.07	-2.06		-0.01	-0.31	-1.86	
100	p-value_H			0.97	0.02	0.02		0.50	0.38	0.03	

Note. MSE in percentages