Dissertation
an der
Fakultät für Mathematik, Informatik und Statistik
der
Ludwig-Maximilians-Universität München

Modelling Risk in Financial Economics

vorgelegt von

Alexander B. Matthies

München, den 13. März 2018

1. Gutachter: Prof. Dr. Stefan Mittnik
2. Gutachter: Prof. Dr. Markus Haas
Tag der Mündlichen Prüfung: 31.10.2018

Abstract

This work focuses on the modelling of two specific risk measures, namely the term structure of government bonds and firm credit ratings. We discuss estimation and forecasting issues relating to these risk measures and their relationship to macroeconomic trends.

Risk is an important component in any financial decision. Interest rates and credit ratings are two economic variables that reflect financial risk in different ways and are both interrelated. The term structure of interest rates — different rates for different maturities— represent a market perspective while credit ratings represent the view of a credit rating agency on the default probability of an economic entity.

Government bond interest rates determine the ability of a country to finance itself and are important factors in the fixture of other interest rates. Modelling the term structure helps in understanding this economic base variable and determine financial risk. We choose a data driven approach in which unseen dynamic factors of the term structure are estimated via principal components analysis in rolling time windows to produce yield curve forecasts. A statistical and economic evaluation is provided for different sets of predictors, estimation methods, and forecasting methods. The data consists of daily observations of government bond interest rates for Germany, Switzerland, the UK, and the US for the time period from 2000 to 2016. Implicitly this approach tests the basic assumptions of Nelson-Siegel type economic factor models of the term structure. Term structure forecasts are evaluated in terms of three complementary criteria or loss functions, namely the statistical mean squared forecast error criterion, and the two more economic criteria of directional accuracy and big hit ability. Factor analysis supports the idea that a level, slope, and curvature factor underly the yield curve. In a data set with all term structures we find evidence of a global level factor. A comparison to simple forecasts such as random walk and autoregressive forecasts shows that dynamic factor models can, in rare instances, improve on random walk forecasts and consistently outperform auto-regressive forecasts under both statistical and economic evaluation criteria. Statistical and economic criteria suggest that more than one factor should be employed for forecasting. Using additional term structures for factor extraction can improve forecasts for some countries depending on the forecast horizon. With regard to estimation methods the standard principal components method using ordinary least squares outperforms the alternative method using generalised least squares. The forecasting method employing autoregressive factors

outperforms the method exploiting the lagged correlation of factors and interest rates. These results support the concepts of of the Nelson-Siegel Model.

Corporate credit ratings are the traditional business of credit rating agencies. Credit ratings have important effects on financial markets and are a part of financial market regulation. Credit ratings agencies also provide assessments of the credit quality of other entities like countries and structured finance products. The financing of ratings, the competition on rating markets, and the power of debt issuers play an important role in the quality of credit ratings. There are numerous statistical methods to estimate corporate credit ratings. Here, we employ the ordered probit and an unordered logit approch, as well as an OLS approach developed here, that replaces ratings with their respective default rates. Methodologically this approach provides a way to integrate an estimate of the assumed continuous variable that underlies the probit and logit methods, which is unobservable in these models. Furthermore, this approach underscores the connection between credit ratings and default probability. This thesis uses ratings of selected US, UK, German, French, Japanese, Canadian, and Australian firms from 1990 - 2009 and their respective accounting data for corporate credit rating estimation. Here, previous findings are confirmed that show that credit rating agency standards have become more stringent over time given the same accounting data. Furthermore, the results shown here suggests that market pressure outside the US rating market can influence credit ratings agencies judgement.

Zusammenfassung

Risiko ist ein wichtiger Bestandteil der ökonomischen Entscheidungsfindung. Zinssätze und Kredit-Ratings sind zwei Variablen, die auf unterschiedliche Art und Weise finanzielle Risiken beschreiben und zugleich in vielerlei Hinsicht miteinander verbunden sind. Die Zinsstrukturkurve besteht aus den unterschiedlichen Zinssätzen für verschiedene Fälligkeiten und repräsentiert eine Marktperspektive der Ausfallwahrscheinlichkeit eines Wirtschaftssubjekts. Kredit-Ratings stellen hingegen die Sicht einer Ratingagentur auf die Ausfallwahrscheinlichkeit dar.

Die Zinssätze von Staatsanleihen beeinflussen die Fähigkeit eines Landes, sich selbst zu finanzieren und sind ein wichtiger Faktor zur Festlegung anderer Zinssätze. Ein aussagekräftiges Modell der Zinsstrukturkurve hilft zum Verständnis dieser ökonomischen Basisvariablen und bei der Bestimmung von ökonomischen Risiken. In dieser Arbeit wird ein datengesteuerter Ansatz benutzt um nicht beobachtbare dynamische Faktoren der Zinsstrukturkurve mittels Hauptkomponentenanalyse in einem rollenden Zeitfenster zu schätzen, um dadurch Prognosen der Zinsstrukturkurve zu erhalten. Eine statistische und ökonomische Bewertung wird für verschiedene Datensätze, Schätzmethoden und Prognosemethoden vorgenommen. Die Daten bestehen aus täglichen Beobachtungen von Staatsanleihen für Deutschland, die Schweiz, Großbritannien und die USA für den Zeitraum von 2000 bis 2016. Implizit testet dieser Ansatz die Grundannahmen des Nelson-Siegel Modells. Die Prognosen der Zinsstrukturkurve werden anhand drei verschiedener Evaluationskriterien untersucht, nämlich dem quadrierten mittleren Vorhersagefehler, der Richtungsgenauigkeit und der gewichteten Richtungsgenauigkeit. Die Faktorenanalyse unterstützt die Idee, dass der Zinsstrukturkurve drei Faktoren zugrundeliegen, ein Niveau- bzw. Level-, ein Steigungs- und ein Krümmungsfaktor. In einem Datensatz mit den Zinsstrukturkurven aller untersuchten Länder finden wir Hinweise auf einen globalen Levelfaktor. Vergleiche mit einfacheren Prognosenmethoden, wie dem Random Walk oder autoregressiven Ansätzen, zeigen, dass dynamische Faktormodelle in einigen Fällen Random-Walk-Prognosen verbessern können und zudem regelmäßig autoregressive Prognosen verbessern. Die Evaluation nach allen drei Kriterien deutet darauf hin, dass mehr als ein Faktor für die Prognosen eingesetzt werden sollte. Der Gebrauch von zusätzliche Zinsstrukturkurven für die Faktorextraktion kann die Prognosen für einige Länder je nach Prognosehorizont verbessern. Bezüglich der Schätzmethoden ist die Standard-Hauptkomponenten-Methode, die die Methode der gewöhnlichen kleinsten Fehler-Quadrate nutzt, einem alternativen Ansatz, der mit verallgemeinerten kleinsten Fehler-Quadraten arbeitet, vorzuziehen. Die Prognosemethode mit autoregressiven Faktoren übertrifft Methoden, die die verzögerte Korrelation von Faktoren und Zinssätzen ausnutzen. Diese Ergebnisse bestätigen die Annahmen

des Nelson-Siegel Modells.

Unternehmensratings sind das klassische Geschäft von Ratingagenturen. Kredit-Ratings haben wichtige Auswirkungen auf die Finanzmärkte und sind Teil der Finanzmarktregulierung. Ratingagenturen bewerten auch die Kreditqualität von Ländern und von strukturierten Finanzprodukten. Die Art der Ratingfinanzierung, der Wettbewerb auf den Ratingmärkten und die Marktmacht der Emittenten spielen eine wichtige Rolle für die Qualität der Ratings. Es gibt zahlreiche statistische Methoden, um Unternehmensratings zu schätzen. In dieser Arbeit werden geordnete Probit und ungeordnete Logit Ansätze benutzt sowie ein hier entwickelter Ansatz kleinster Fehler-Quadrate, bei dem Ratings durch die entsprechenden Ausfallwahrscheinlichkeiten ersetzt werden, um die Bonitätsbeurteilung ausgewählter amerikanischer, britischer, deutscher, französischer, japanischer, kanadischer und australischer Firmen aus den Jahren von 1990 bis 2009 mit ihren jeweiligen Buchhaltungsdaten zu reproduzieren. Der neu entwickelte methodische Ansatz bietet eine Möglichkeit, die nichtbeobachtbare kontinuierliche Variable zu schätzen, den die Probit- und Logit-Ansätzen unterliegt. Der Ansatz unterstreicht zudem die enge Verbindung zwischen Kredit-Ratings und Ausfallwahrscheinlichkeit. Diese Arbeit kann ältere Ergebnisse bestätigen, die zeigen, dass die Bewertungen von Ratingagenturen strenger geworden sind. Darüberhinaus zeigen die Ergebnisse, dass ausserhalb des US-Ratingmarktes Marktzwänge die Beurteilung von Ratingagenturen beeinflussen können.

Acknowledgements

I would like thank my family, my colleagues, and friends. This Research was supported by the German Research Foundation via the CRC 649 ‘Economic Risk’, Humboldt-University Berlin, and a stipend ‘Förderung des wissenschaftlichen und künstlerischen Nachwuchses des Landes Schleswig-Holstein’, which is gratefully acknowledged.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	—	Analysis of Variance
AR	—	Auto regressive
ARMSFE	—	Auto regressive mean squared forecast error
ARBHA	—	Auto regressive big hit ability
ARDA	—	Auto regressive directional accuracy
BHA	—	Big Hit Ability
BH	—	Blaskowitz and Herwartz
BLM	—	Blume, Lim, and MacKinley
CDS	—	Credit default swaps
CPI	—	Consumer price index
CRA	—	Credit rating agency
DA	—	Directional Accuracy
DFM	—	Dynamic factor models
EBIT	—	Earnings before interest and taxes
e.g.	—	exempli gratia—"for example"
Eq.	—	Equation
EU	—	European Union
Fig.	—	Figure
GDP	—	Gross domestic product
GER	—	Germany
GLS	—	Generalised least squares
IC	—	Interest coverage
i.e.	—	id est—"that is"
LTD	—	Long term debt
MDA	—	Multivariate discriminate analysis
ML	—	Maximum likelihood
MSFE	—	Mean Squared Forecast Error
NR	—	Not rated
NBER	—	National Bureau of Economic Research
NRSRO	—	Nationally Recognized Statistical Rating Organization
OI NS	—	Operating income to net sales
OLS	—	Ordinary least squares
PCA	—	Principal components analysis
PC-GLS	—	Principal components generalised least squares

PC-OLS	—	Principal components ordinary least squares
RE	—	Retained Earnings
RW	—	Random walk
ROA	—	Return on assets
RWMSFE	—	Random walk mean squared forecast error
SEC	—	U.S. Securities and Exchange Commission
S&P	—	Standard and Poor's
SW	—	Stock and Watson
SWI	—	Switzerland
TA	—	Total assets
Tab	—	Table
TD	—	Total debt
UK	—	United Kingdom
US	—	United States
VAR	—	Vector auto regressive
w.r.t.	—	with respect to

Chapter 1

Introduction

Risk is a central feature of life. There is no success without taking some risk. Indeed, high ambitions are often connected to taking larger risks. Understanding risk, being able to quantify and predict it, is an essential element of making informed decisions and understanding their consequences. Risk measures quantify the probability and potential size of financial loss (see for instance the value at risk measure). A financial loss may be the default of payment on debt or the reduction of value of any given asset. Interest rates and credit ratings (the two features of economic risk discussed here) quantify the probability of financial default. Modelling interest rates and measuring the probability of default are central features of investment decisions, portfolio management, as well as policy planning in finance, business, and governance.

Interest rates and credit ratings are interrelated. Interest rates represent a market point-in-time perspective. In contrast, credit ratings are an institutional through-the-cycle view on an entity's creditworthiness. Furthermore, a credit rating downgrade likely correlates with an interest rate increase and vice versa. A credit rating downgrade signals a higher likelihood of default which will increase interest rates. An increase in interest rates (possibly caused by a macroeconomic shift) increases interest payments and thereby decreases the likelihood of repayment. This might cause a rating downgrade.

Credit ratings reflect a long term perspective and focus on a time window of one year and more. Interest rate maturities range from over-night to 50 years. Interest rates and credit ratings can be estimated and forecasted. While interest rates change on a daily basis, credit ratings may remain constant over several years or even decades.

For interest rates, a further issue of interest is the relationship between different maturities (graphically displayed by the yield curve). The level, slope, and curvature of the term structure of interest rates are important features of financial risk over time and are correlated to financial and macroeconomic factors.

Default risk analysis (which is the essence of credit ratings) is an essential part of valuing corporate bonds and the management of bank loan portfolios, or any portfolio (Yue, 2010). For bonds, there are two related financial risks. The chance that a bond will default, and the possibility of loss of market value due to apparent increased probability that a bond will default (Zhou, 2001b).

Studying bond prices in relationship to corporate credit ratings, one finds that bonds are priced as if ratings capture real information (Elton et al., 2001). Nevertheless, credit risk and expected default as measured by credit ratings does not explain yield spreads completely (Elton et al., 2001, Huang & Huang, 2012). Yield curves may even have similar shapes within rating classes (He et al., 2000).

Interest rates across different economies may also be used to measure systemic risk. Instability in one economy may transfer to other economies via contagion (Recchioni & Tedeschi, 2017). The shape of the yield curve also correlates with macroeconomic variables and the business cycle (Dewachter & Lyrio, 2006). These macro variables are themselves important determinants of sovereign ratings (Afonso et al., 2011). The downgrades of sovereign ratings affect bank lending and corporate ratings (Adelino & Ferreira, 2016, Almeida et al., 2017). Corporate ratings can correlate with the bussiness cycle (Amato & Furfine, 2004) and appear to have become more stringent over time on a global level (Blume et al., 1998, Amato & Furfine, 2004, Jorion et al., 2009, Matthies, 2013c, Alp, 2013). Regulation, the set up, and market dynamics within the credit rating agency (CRA) industry may create or help facilitate market processes that may cause economic disruptions and shocks that cause changes in macroeconomic fundamentals, e.g. the 2008 financial crisis (Matthies, 2013b, Benmelech & Dlugozs, 2010). The relevant literature is now briefly summarised. More elaborate literature reviews are provided in the relevant chapters.

1.1 Previous Work

Research on term structure strongly related to the approach in this thesis focuses on the shape of term structure factors and their relationship to other financial and macroeconomic factors. The future development of interest rates forms the basis of most risk assessments and valuations of investments. Term structure forecasts are therefore an important element of financial planning. The study of Cochrane & Piazzesi (2005), Duffee (2011) and the empirical results of Blaskowitz & Herwartz (2009) suggest that factors that carry a negligible amount of the variation of the term structure can nevertheless improve forecast performance. Term structure factors are shown to be autocorrelated (Diebold & Li, 2006, Blaskowitz & Herwartz, 2009, 2011). The term structure is also highly correlated to macroeconomic factors so that these factors may improve forecast performance (Ang & Piazzesi, 2003, Abad & Novales, 2005, Diebold et al., 2005, Mönch, 2008).

For corporate credit ratings, the line of research that analyses to which degree publicly available information can reproduce ratings is most relevant for the topic of this thesis. Reproducing credit ratings reflects the determinants of corporate credit risk. There are three main distinct groups of variables that determine credit ratings and credit rating changes: financial data and ratios (Ederington, 1985, Blume et al., 1998, Altman & Rijken, 2004), corporate governance characteristics (Bhojraj & Sengupta, 2003, Ashbaugh-Skaife et al., 2006, Jorion et al., 2009), and macroeconomic factors (Nickel et al., 2000, Amato & Furfine, 2004). Financial ratios (such as leverage and earnings ratios) are the basic determinants of credit ratings. Corporate governance mechanisms influence the creditworthiness in basically two ways; firstly, with respect to the principal agent problem between stockholders and management (Bhojraj & Sengupta, 2003), and secondly, through wealth transfer effects from bondholders to shareholders (Ashbaugh-Skaife et al., 2006). Furthermore, Jorion et al. (2009) analyse the impact of changes in accounting quality on credit ratings. Macroeconomic factors can either be reflected in changing agency standards (Blume et al., 1998) or through a dependence on the business cycle (Nickel et al., 2000, Amato & Furfine, 2004).

1.2 The Contribution of this Thesis

This thesis expands on previous research by the author in term structure forecasts and corporate credit rating estimation. In the author's diploma thesis multiple different dynamic factor models were employed to forecast interest rates. Forecasts were performed for Libor and Euribor rates with maturities of 1, 6, and 12 months and German government bonds of 2, 5, and 10 years. The results focused on the best performing forecast strategies, and in an ANOVA evaluation selective results were discussed with respect to the influence of additional financial data, alternative estimation methods, alternative forecasting methods, the number of factors, the number of lags, and the size of a rolling estimation window. Following this, Matthies (2014) expands on these preliminary results. In Matthies (2014) the focus is on government bond interest rates, namely those of Germany, the US, the UK, and Switzerland. In extension to previous work, the employed number of factors is increased and a larger data set is used, a comparison to some basic forecasting methods is provided, and the comparison is via ANOVA evaluation, as this provides a better statistical and economic evaluation. Furthermore, linear combinations of the 2, 5, and 10 year maturities that may be interpreted as the level, slope, and curvature of the yield curve are forecasted. This line of research is further developed in this thesis. We show that the superiority of an estimation and forecasting method combination holds over a longer time period. Employing the entire term structure as predictors as compared to a single maturity is beneficial for forecasting. Furthermore, we show that additional term structures can, in certain cases, improve forecasts as financial data could in Matthies (2014).

With respect to credit ratings, there are three papers written by the author of the present work on this subject: Matthies (2013a), Matthies (2013c), and Matthies (2013b). Matthies (2013a) is a review on the current state of research and important older findings of empirical studies on corporate credit ratings and their relationship to ratings of other entities. Specifically, the focus is on the results of three lines of research: the correlation of credit ratings and corporate default, the influence of ratings on capital markets, and the determinants of credit ratings and rating changes. Results from each individual line are important and relevant for the construction and interpretation of studies in the other two fields, e.g. the choice of statistical methods. Moreover, the design and construction of credit ratings and the credit rating scale are essential to an understanding of empirical findings.

In Matthies (2013c) standard explanatory variables that determine credit ratings do not achieve significant effects in a sample of 100 US non-financial firms in an ordered "probit" panel estimation. Sample size and selection as well as the distribution of explanatory variables across rating classes may be the cause of this problem. Furthermore, we find evidence to suggest that variable coefficients vary over rating classes when analysed with an unordered "logit" model. The sample reproduces well-established macroeconomic effects of credit ratings and highlights the influence of the rating agencies' through-the-cycle approach on rating transitions.

In Matthies (2013b) a review, as in Matthies (2013a) is provided of the empirical literature with respect to results that are relevant to reputational capital and the effects of firm size. We conclude that in corporate and sovereign markets agencies seem to try to obtain reputational capital, in contrast to the market for structured finance products, in which they might have sold regulatory licences. A survey of the 100 largest non-financial public firms in 26 countries is included to determine which ones are rated by S&P. We find that in many developed and developing countries credit ratings are still rare. In our sample only all US corporations have a rating, making an international comparison difficult. With an ordered probit panel analysis for the US corporations of our sample from 1990 − 2009 the effects of standard firm specific variables and their comparison to results in the literature are evaluated. The empirical evidence can be interpreted as a concern for reputational capital on behalf of rating agencies. Standard explanatory variables do not achieve significant effects in our US-sample, which indicates that variable coefficients could depend, among others, on firm size. Alternatively to the probit estimator, Matthies (2013c) proposes an ordinary least squares (OLS) estimator that replaces credit ratings with ten year default rates of the respective rating category.

This thesis continues the author's previous research on corporate ratings by employing a larger data set of corporate credit ratings from six additional countries. Beyond the US firms used in Matthies (2013c), the thesis also includes credit ratings of German, Australian, UK, Japanese, Canadian, and French firms, thereby allowing estimation of country specific risk factors. It is shown that rating specific default rates can be used to replace credit ratings when modelling business default risk. We confirm the findings in Blume et al. (1998) and other studies that CRA standards have become more stringent. The results furthermore suggest that the position of CRAs can be compromised in markets outside the US.

Structure of the dissertation Chapter 2 deals with term structure forecasts from dynamic factor models. In Section 2.1 dynamic factor models are presented and alternative estimation, forecasting, and evaluation methods are discussed. In Section 2.2 the data is presented, factor analysis of the term structures is performed, and the forecasting strategies are discussed. Section 2.3 presents and discusses the results.

Chapter 3 addresses the estimation of corporate credit ratings. Section 3.1 discuses the history and regulation of credit rating agencies, and different markets for credit ratings. Section 3.2 presents an overview of statistical methods that may be employed to estimate credit ratings. Section 3.3 and Section 3.4 present and discuss the data and results respectively. Chapter 4 concludes.

Chapter 2

Term structure Forecasts

This chapter provides a statistical and economic evaluation of term structure forecasts via alternative variants of statistical dynamic factor models (DFM). Data-driven factor models are exploited for the purpose of ex-ante forecasting of four government term structures. Implicitly this approach tests the basic assumptions about the factors of Nelson-Siegel type term structure economic factor models. The evaluation of term structure forecasts provides an assessment of the possibility to model the daily variation of risk changes as represented by government bonds. Government bonds (in particular those of the US) are often regarded as "risk free" and are used as a benchmark for other rates.

Based on the concept from Chamberlain & Rothschild (1983) of —*the approximate factor model*— Stock & Watson (2002) develop an estimation and forecasting method based on classical principal components analysis (PCA). Under certain restrictions this method can be applied to dynamic factor models[1].

Factor models and PCA allows us to use a large number of variables in economic research. For the term structure forecasting in Matthies (2014), for example, no restrictive a priori decisions on which predictors to use are necessary. A large part of the variation of a set of observable correlated variables can be explained by unobservable factors. In a DFM these unobservable static factors are furthermore assumed to be the product of a linear combination of current and prior unobservable dynamic factors. A DFM can therefore be represented in its static form with static factors (Stock & Watson, 2002).

Bai & Ng (2002) introduce a method to determine the number of static factors in approximate factor models. For factor models of large dimensions Bai (2003) develops an inferential theory. Further approaches explicitly deal with dynamic factors. In Bernanke, Boivin & Eliasz (2005) and Stock & Watson (2005) dynamic factor meth-

[1]Factor analysis can be distinguished from PCA. We acknowledge that difference but use the terms principal components and factors interchangeably in this thesis.

ods are integrated into vector autoregression (VAR) analysis. Moreover, Breitung & Tenhofen (2011) expand on Stock & Watson (2005) and Doz, Giannone & Reichlin (2011) to propose a Gaussian pseudo maximum likelihood (ML)-estimator or principal components-generalised least squares (PC-GLS) estimator. This estimator can account for heteroscedastic and/or autocorrelated errors.

Nelson & Siegel (1987) develop a factor model with predetermined factor loadings for the yield curve. The model is parsimonious while using three factors. The model is able to reproduce numerous empirically observed shapes of the yield curve. Using a data driven approach for the factor loadings it is possible to use PCA to determine the underlying factors of interest rate term structures. Furthermore, it is possible to provide economic interpretation of the static factors. Due to the distribution of factor loadings along the yield curve, the first three factors are interpreted as level, slope, and curvature of the term structure (Diebold & Li, 2006, Blaskowitz & Herwartz, 2009).

Here in Chapter 2 there are two distinct relevant groups of studies that use factors in term structure analysis. One group employs Nelson-Siegel type models (Diebold & Li, 2006, Yu et al., 2009, Yu & Zivot, 2011) while the other extracts factors on a data-driven basis via PCA (Litterman & Scheinkman, 1991, Blaskowitz & Herwartz, 2009, 2011, Matthies, 2014).

Using factor models in term structure forecasts it is important to note that a parsimonious model of the term structure does not automatically provide the best forecast performance. Empirical results (Blaskowitz & Herwartz, 2009, Matthies, 2014) and some studies (Cochrane & Piazzesi, 2005, Duffee, 2011) find that factors that carry a negligible amount of the variation of the term structure can nevertheless improve forecast performance. Furthermore, factors are found to be autocorrelated (Diebold & Li, 2006, Blaskowitz & Herwartz, 2009, 2011). Other studies find a correlation between the term structure and macroeconomic variables, and also that such variables can improve forecast performance (Ang & Piazzesi, 2003, Abad & Novales, 2005, Diebold et al., 2005, Mönch, 2008, Matthies, 2014).

Assuming locally homogeneous term structure dynamics, Blaskowitz & Herwartz (2009, 2011) motivate an adaptive procedure that employs dynamic model variation. Blaskowitz & Herwartz (2011) use data-driven selection algorithms to forecast the term structure of interest rates. Forecasts are improved to benchmark methods under economic criteria while using an adaptive approach. Matthies (2014) follows their approach, but expands their set of strategies in three ways. Strategies in Blaskowitz & Herwartz (2011) solely use the term structure as predictors, while Matthies (2014) compares strategies as those to strategies that include macro variables for the purpose of factor extraction and forecast performance. Similar to Ang & Piazzesi (2003), Matthies (2014) finds that macroeconomic data may improve forecasts with regard to economic evaluation criteria. Here, additional predictors are stock price indices, exchange rates, commodity prices, and stock market volatilities. The second aspect in which Matthies (2014) expands strategies is to test the empirical potential of an alternative estimation method. The alternative estimation methods are the common PC-OLS[2] approach (Stock & Watson, 2002) and the PC-GLS method (Breitung & Tenhofen, 2011). The PC-GLS method tries to exploit autocorrelated features within the factor model as observed in the literature (Diebold & Li, 2006, Blaskowitz & Herwartz, 2009). The third aspect addressed by Matthies (2014) is the accuracy of two competing forecast algorithms. Forecasts alternatively either come from the method of Stock & Watson (2002) (hereafter SW), or from the approach of Blaskowitz & Herwartz (2011) (hereafter BH), which is motivated by the Nelson-Siegel AR approach of Diebold & Li (2006). SW conditions the forecasted variable directly on the factors. In contast, BH proceeds with factor predictions and a subsequent forecasting of the variable of interest by means of estimated factor loadings. Matthies (2014) finds that a method combination of PC-OLS estimation and BH forecasting is preferable to others, thereby confirming the assumptions that underly the Nelson-Siegel AR approach of Diebold & Li (2006).

Closely related to the factor analysis of term structures is the co-integration analysis of interest rates (Engle & Granger, 1987, Hall et al., 1992, Zhang, 1993, Carstensen, 2003, Giese et al., 2008). Co-integration analysis specifically tests if a linear combination of two nonstationary interest rate times series is stationary. If such a linear combination exists, the interest rates are regarded as co-integrated. Term structures are found to be co-integrated (Engle & Granger, 1987, Hall et al., 1992, Giese et al., 2008). Common non-stationary trends are interpreted similarly to common factors (Zhang, 1993).

This chapter expands the research in Matthies (2014). It tests the predictive power of the estimation and forecasting methods on a longer time scale, hereby testing in different term structure settings, such as negative rates. Furthermore, it reduces the variables in the alternative larger data set to a set of select other government term structures.

Term structure forecasts are evaluated as in Matthies (2014) in terms of three complementary criteria or loss functions, namely the statistical mean squared forecast error criterion, and the two more economic criteria of directional accuracy and big hit ability.

Results are that factor models have to be carefully designed if they are to perform better than simple forecasting methods, i.e. an AR(p) processes[3] or a random walk

[2] This is the principal components ordinary least squares method, i.e. the standard PCA method as discussed in Section 2.1

[3] This is a autoregressive process of order p.

model. The inclusion of further term structures may improve forecasts. The PC-OLS estimation method and the BH forecasting method outperform their respective alternative counterparts, the PC-GLS and SW method. The BH method success confirms the assumptions that underly the Nelson-Siegel AR approach of Diebold & Li (2006).

2.1 Methods

The classical factor model poses restrictions on the error terms that are not easily fulfilled. The approximate factor model can be estimated with PCA. For that reason we make no distinctions between factors and principal components.

We first provide a representation of a dynamic factor model. We transform the model into its static representation and its corresponding static form. Then two alternative estimation methods for the static form are presented, the PC-OLS and PC-GLS methods. PC-OLS is the standard PCA method and equivalent to the estimation of the static model. PC-GLS is a two step method that uses PC-OLS as a first step. In the second step dynamic features are exploited.

We then present two alternative forecasting methods. One is the standard approach by Stock & Watson (2002) which uses simple OLS. The alternative is the method by Blaskowitz & Herwartz (2009) which uses autoregressive features in the factors.

Finally, we present and discus three alternative evaluation criteria for the term structure forecasts. These are the statistical mean squared forecast error MSFE criterion and the economic directional accuracy DA and big hit ability BHA criteria.

2.1.1 Dynamic Factor Model

The motivation for using factor analysis and PCA is the idea that a large set of N variables can be explained by a smaller set of R factors or principal components. We now present a dynamic factor model to highlight the underlying assumptions for our model. It is then transformed into its static representation and the corresponding static form for simpler estimation.

Following the design in the literature (Breitung & Eickmeier, 2006, Stock & Watson, 2011) a dynamic factor model has the form

$$\tilde{x}_{n,t} = \theta_n(L)'\boldsymbol{g}_t + e_{n,t}.$$

Here $\tilde{x}_{n,t}$ is the n'th appropriately modified variable observed in period t (for $n = (1, \ldots, N)$ and $t = (1, \ldots, \tau)$), $\boldsymbol{g}_t$ is a $(k \times 1)$ vector of dynamic factors, $\theta_n(L) = \theta_{0,n} + \theta_{1,n}L + \ldots + \theta_{u,n}L^u$ is a $(k \times 1)$ polynomial of factor loadings, L is the lag operator, and $e_{n,t}$ is an error term. It is assumed that $E[\tilde{x}_{n,t}] = E[e_{n,t}] = 0$. If one allows $\Theta(L) = \Theta_0 + \Theta_1 L + \ldots + \Theta_u L^u$ and $\Theta_j = [\theta_{j,1}, \ldots, \theta_{j,N}]'$ [4], and defines $G_t = [\boldsymbol{g}'_t, \ldots, \boldsymbol{g}'_{t-u}]'$, the static representation of the dynamic factor model is

$$\tilde{\boldsymbol{x}}_{\bullet,t} = \Theta G_t + \boldsymbol{e}_{\bullet,t},$$

where $\tilde{\boldsymbol{x}}_{\bullet,t} = [x_{1,t}, \ldots, x_{N,t}]'$, $\Theta = [\Theta_0, \ldots, \Theta_u]$, and $\boldsymbol{e}_{\bullet,t} = [e_{1,t}, \ldots, e_{N,t}]'$. It is not necessary for Θ to be of full column rank. Assume that the rank of Θ is then R, where $R \leq (u+1)k$. Then a $N \times R$ matrix A exists so that $\Theta G_t = A\boldsymbol{f}_{\bullet,t}$, where $\boldsymbol{f}_{\bullet,t} = DG_t$, where D is a non-singular $R \times (u+1)k$ matrix and $\boldsymbol{f}_{\bullet,t}$ is a vector of static factors.

With R and D we can transform the dynamic factor model into its corresponding static factor model. In full matrix notation the corresponding static factor model of the dynamic factor model is

$$\widetilde{X}_s = F_s A'_s + E_s, \qquad s = T_0, \ldots, T-h, \tag{2.1}$$

Here $\widetilde{X}_s = (\tilde{\boldsymbol{x}}'_{\bullet,1}, \ldots, \tilde{\boldsymbol{x}}'_{\bullet,\tau})' = (\tilde{\boldsymbol{x}}_{1,\bullet}, \ldots, \tilde{\boldsymbol{x}}_{N,\bullet})$ is a $\tau \times N$ matrix of N appropriately transformed variables. X_s is observed in a time window of length τ. Moreover, R unobservable factors are collected in $F_s = (\boldsymbol{f}'_{\bullet,1}, \ldots, \boldsymbol{f}'_{\bullet,\tau})' = (\boldsymbol{f}_{1,\bullet}, \ldots, \boldsymbol{f}_{R,\bullet})$ that is a $\tau \times R$ matrix, A_s is a $N \times R$ matrix of loadings, and E_s is a $\tau \times N$ matrix of disturbances. We denote the n'th row and the r'th column of A_s as $\boldsymbol{a}_{n\bullet}$ and $\boldsymbol{a}_{\bullet r}$. Here, time indices are omitted for the sake of brevity to allow notations to be readable. Furthermore, h is a forecast horizon as explained below in Section 2.1.3, and the size of the overall sample and the first forecast origin are T and T_0 respectively.

The estimation in a moving time window is motivated by possible temporal instability in the available sample information. This approach follows Blaskowitz & Herwartz (2009) and estimates Eq. (2.1) from a rolling time window of size τ sequentially at each time instance s. The rolling time window estimation of the static model is the basis from which we proceed to the alternative estimation and forecasting methods. These are now discussed respectively.

2.1.2 Estimation Methods

The alternative estimation methods that try to extract R factors from a set of N variables, i.e. PC-OLS and PC-GLS, are described. The PC-OLS approach is the standard PCA (Stock & Watson, 2002). The PC-GLS method builds on the PC-OLS estimation and seeks to exploit dynamic features in the factor errors to improve factor and factor loading estimation (Breitung & Tenhofen, 2011).

[4] $j = 0, \ldots, u$.

PC-OLS

Estimation via this method is analogous to the estimation of the static model in Eq. (2.1). The nonlinear objective function conditioned on a agiven time window of sample information is

$$V_{OLS}(\tilde{F}_s, \tilde{A}_s) = \sum_{n=1}^{N} \sum_{t=s-(\tau-1)}^{s} (\tilde{x}_{n,t} - \tilde{\boldsymbol{f}}_{\bullet,t}\tilde{\boldsymbol{a}}'_{n\bullet})^2. \tag{2.2}$$

This function is then minimised with respect to the possible values of the loadings and factors $\tilde{A}_s$ and $\tilde{F}_s$ respectively. The minimising arguments for V_{OLS}, denoted by $\hat{A}_s$ and $\hat{F}_s$, are the eigenvectors $\hat{\boldsymbol{a}}_{\bullet,r}$ corresponding to the largest eigenvalues λ_r, $r = 1, \ldots, R$, of the matrix $(\tilde{X}_s'\tilde{X}_s)$ and $\hat{F}_s = \tilde{X}_s\hat{A}_s$.

PC-GLS

Proceeding from the PC-OLS estimator we now discuss an extension that addresses the statistical features of the error term E_s. If the idiosyncratic errors are heteroskedastic or autocorrelated, the PC-OLS estimator is not efficient. Breitung & Tenhofen (2011) propose a GLS estimator that accounts for heteroskedastic and autocorrelated idiosyncratic components.

The approach follows the Stock & Watson (2005) representation of the idiosyncratic component which is stationary autoregressive

$$e_{n,t} = \rho_{1,n}e_{n,t-1} + \ldots + \rho_{p,n}e_{n,t-p} + \epsilon_{n,t},$$

so that $\tilde{\rho}_n(L)\hat{e}_{n,t} = \epsilon_{n,t}$. Here, $\epsilon_{n,t}$ are idiosyncratic error terms, and $\tilde{\rho}_n(L)$ is a lag polynomial. For each variable $\boldsymbol{x}_{n\bullet}$, for $n = 1, \ldots, N$, there exists a Toeplitz matrix $R(\boldsymbol{\rho}^{(n)})$ of size $(\tau - p) \times \tau$. Denoting $\boldsymbol{\rho}^{(n)} = (\rho_{n,1}, \ldots, \rho_{n,p})$, this matrix is given as

$$R(\boldsymbol{\rho}^{(n)}) = \begin{bmatrix} -\rho_{n,p} & -\rho_{n,p-1} & \ldots & -\rho_{n,1} & 1 & 0 & 0 & \ldots \\ 0 & -\rho_{n,p} & \ldots & -\rho_{n,2} & -\rho_{n,1} & 1 & 0 & \ldots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & \ldots \\ 0 & \ldots & 0 & -\rho_{n,p} & \ldots & -\rho_{n,2} & -\rho_{n,1} & 1 \end{bmatrix}.$$

Then, with $\Sigma = \operatorname{diag} \sigma_n^2$, the PC-GLS objective function to be minimised is

$$\begin{aligned} V_{GLS}(\tilde{F}_s, \tilde{A}_s, \rho, \Sigma) = & -\frac{\tau}{2} \sum_{n=1}^{N} \ln [\sigma_n^2] \\ & - \sum_{n=1}^{N} \sum_{t=s-(\tau-p-1)}^{s} \frac{(\hat{e}_{n,t} - \rho_{n,1}\hat{e}_{n,t-1} - \ldots - \rho_{n,p}\hat{e}_{n,t-p})^2}{2\sigma_n^2}. \end{aligned} \quad (2.3)$$

If the $\tilde{x}_{n,t}$ are normally distributed and $N \to \infty$, the PC-GLS estimator is asymptotically equivalent to the maximum likelihood (ML)-estimator, as shown by Breitung & Tenhofen (2011).

A necessary condition for a minimum of the objective function is the vanishing of the gradient. In order to minimise V_{GLS} in Eq. (2.3) we obtain the following set of gradients for all $\boldsymbol{a}_{n,\bullet}$, $\boldsymbol{f}_{\bullet,t}$, $\rho_{k,n}$ and σ_n^2:

$$\begin{aligned} g_{\boldsymbol{a}_{n,\bullet}}(\cdot) = \frac{\partial V_{GLS}(\cdot)}{\partial \boldsymbol{a}_{n,\bullet}} &= \frac{1}{\sigma_n^2} \left(\sum_{t=p_n+1}^{\tau} \epsilon_{n,t} [\rho_n(L) \boldsymbol{f}_{\bullet,t}] \right) \\ g_{\boldsymbol{f}_{\bullet,t}}(\cdot) = \frac{\partial V_{GLS}(\cdot)}{\partial \boldsymbol{f}_{\bullet,t}} &= \sum_{n=1}^{N} \frac{1}{\sigma_n^2} (\epsilon_{n,t}\boldsymbol{a}_{n,\bullet} - \rho_{1,n}\epsilon_{n,t+1}\boldsymbol{a}_{n,\bullet} - \cdots - \rho_{p,n}\epsilon_{n,t+p_n}\boldsymbol{a}_{n,\bullet}) \\ &= \sum_{n=1}^{N} \frac{1}{\sigma_n^2} \rho_n(L^{-1}) \epsilon_{n,t} \boldsymbol{a}_{n,\bullet} \\ g_{\rho_{k,n}}(\cdot) = \frac{\partial V_{GLS}(\cdot)}{\partial \rho_{k,n}} &= \frac{1}{\sigma_n^2} \sum_{t=p_n+1}^{\tau} \epsilon_{n,t} (\tilde{x}_{n,t-k} - \boldsymbol{a}_{n,\bullet} \boldsymbol{f}_{\bullet,t-k}) \\ g_{\sigma_n^2}(\cdot) = \frac{\partial V_{GLS}(\cdot)}{\partial \sigma_n^2} &= \frac{1}{2\sigma_n^4} \sum_{t=p_n+1}^{\tau} \epsilon_{n,t}^2 - \frac{\tau}{2\sigma_n^2} \end{aligned}$$

The PC-GLS estimator is obtained by setting all gradients equal to zero and solving the resulting system iteratively. The problem of solving this system of $2NR + N + \sum p_n$ equations via the Newton method would require the inverse Hessian matrix. And although there are methods which do not require the inverse Hessian, e.g. Quasi Newton methods, they are still fairly complicated (Press et al., 2007, p.521 ff).

Therefore, Breitung & Tenhofen (2011) suggest a simple two-step estimator that is asymptotically equivalent to the PC-GLS estimator. A feasible two-step estimator of $\boldsymbol{f}_{\bullet,t}^{\star}$ $t \in \{1, \ldots, \tau\}$ and $\boldsymbol{a}_{n,\bullet}^{\star}$ $n \in \{1, \ldots, N\}$ is obtained by solving the set of equations

$$\begin{aligned} g_{\boldsymbol{f}_{\bullet,t}}(\hat{A}_s, \tilde{\boldsymbol{f}}_{\bullet,t}, \hat{\rho}, \hat{\Sigma}_e) &\overset{!}{=} 0, \\ g_{\boldsymbol{a}_{n,\bullet}}(\tilde{\boldsymbol{a}}_{n,\bullet}, \hat{\boldsymbol{f}}_{\bullet,t}, \hat{\rho}, \hat{\Sigma}_e) &\overset{!}{=} 0. \end{aligned}$$

Here, $\hat{F}_s$ and $\hat{A}_s$ are the PC-OLS estimators of F_s and A_s and $\hat{\rho}^{(n)}$ are the OLS estimators of the coefficients from $\hat{e}_{n,t} = \rho_{1,n}\hat{e}_{n,t-1} + \ldots + \rho_{p_n,n}\hat{e}_{n,t-p_n} + \epsilon_{n,t}$, where $\hat{e}_{n,t} = \tilde{x}_{n,t} - \hat{\boldsymbol{a}}'_{n,\bullet}\hat{\boldsymbol{f}}_{\bullet,t}$, $\hat{\Sigma} = \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2)$, and $\hat{\sigma}_n^2$ is estimated from the residuals $\tilde{\epsilon}_{n,t}$. The two-step estimator of $\boldsymbol{a}_{n,\bullet}$ is then equivalent to the least squares estimator of $\boldsymbol{a}_{n,\bullet}$ in the regression:

$$\hat{\rho}_n(L)\tilde{x}_{n,t} = \hat{\rho}_n(L)\left(\hat{\boldsymbol{a}}_{n,\bullet}\hat{\boldsymbol{f}}_{\bullet,t}\right) - \tilde{\epsilon}_{i,s} \qquad \forall t \in \{1, \ldots, \tau\},$$

where $\hat{\rho}_n(L) = 1 - \hat{\rho}_{1,n}L - \cdots - \hat{\rho}_{p_n,n}L^{p_n}$.

The two-step estimator of $\boldsymbol{f}_{\bullet,t}$ is similarly obtained as the least squares estimator of the regression:

$$\frac{1}{\hat{\omega}_n}\tilde{x}_{n,t} = \frac{1}{\hat{\omega}_n}\hat{\boldsymbol{a}}_{n,\bullet}\boldsymbol{f}_{\bullet,t} + \tilde{e}_{n,t} \qquad \forall n \in \{1, \ldots, N\} \ \wedge \ \forall t \in \{1, \ldots, \tau\}.$$

where $\hat{\omega}_n^2 = \frac{1}{\tau}\sum_{t=1}^{\tau}(\tilde{x}_{n,t} - \hat{\boldsymbol{a}}'_{n,\bullet}\hat{\boldsymbol{f}}_{\bullet,t})^2$ is a consistent estimator of the variance $\hat{\omega} = \mathbf{E}[e_{n,t}^2]$.

$F_s^\star$ and $A_s^\star$, are obtained in two steps. First, PC-OLS estimates $\hat{F}_s$ and $\hat{A}_s$ are determined and, second, the rows $\boldsymbol{a}_{n,\bullet}^\star$ and $F^\star$ are estimated by means of $N+1$ regressions for $n = 1, \ldots, N$,

$$\begin{aligned} \boldsymbol{a}_{n,\bullet}^{\star\prime} &= \left[\left(R(\hat{\boldsymbol{\rho}}^{(n)})\hat{F}_s\right)'\left(R(\hat{\boldsymbol{\rho}}^{(n)})\hat{F}_s\right)\right]^{-1}\left(R(\hat{\boldsymbol{\rho}}^{(n)})\hat{F}_s\right)'\left(R(\hat{\boldsymbol{\rho}}^{(n)})\tilde{\boldsymbol{x}}_{n,\bullet}\right), && (2.4) \\ \text{and } F_s^{\star\prime} &= \left[(\hat{\Omega}^{-1}\hat{A}_s)'(\hat{\Omega}^{-1}\hat{A}_s)\right]^{-1}(\hat{\Omega}^{-1}\hat{A}_s)'(\hat{\Omega}^{-1}\tilde{X}_s'). && (2.5) \end{aligned}$$

Note that if the error terms are homoscedastic, $\sigma_n^2 = \sigma^2$, and $\hat{\rho}_n(L) = 1$ for all n, the PC-GLS and the PC-OLS estimators coincide.

2.1.3 Forecasting Methods

Two forecasting methods are used alternatively in this study. We first sketch the method suggested by Stock & Watson (2002), and then consider the approach in Blaskowitz & Herwartz (2011). Following Blaskowitz & Herwartz (2009) and Blaskowitz & Herwartz (2011) we consider a M-dimensional vector of interest rates $\boldsymbol{y}_{s+h} = (y_{1,s+h}, \ldots, y_{m,s+h}, \ldots, y_{M,s+h})'$ to be forecasted. These elements $y_{m,s}, m = 1, \ldots, M$, are a subset of the $x_{n,s}, n = 1, \ldots, N$. Ex-ante prediction is performed on the deviations from unconditional in-sample means, $\tilde{y}_{m,t} = y_{m,t} - \bar{y}_{m,s}$, where $\bar{y}_{m,s} = \frac{1}{\tau}\sum_{t=s-(\tau-1)}^{s} y_{m,t}$. Forecasts of vectors $\tilde{\boldsymbol{y}}_{s+h}$ are throughout determined by means of sample information available at time s, $\Xi_{s,\tau} = \{\tilde{\boldsymbol{x}}_t | t = s - (\tau - 1), \ldots, s\}$. Thus these are pseudo ex-ante forecasting exercises. These forecasts $\boldsymbol{y}_{s+h|s} = \mathbf{E}[\tilde{\boldsymbol{y}}_{s+h}|\Xi_{s,\tau}] + \bar{\boldsymbol{y}}_s$ can then also be used to forecast level, slope and curvature of the term structure.

Using the factors to forecast The method is specifically developed to forecast one series using a large set of predictor series. After extracting the factors, the relationship between the factors and the predictors is estimated via linear regression. To obtain ex–ante predictions Stock & Watson (2002) employ the model

$$\tilde{y}_{m,s+h} = \alpha_m + \boldsymbol{\beta}'_m \boldsymbol{f}_{\bullet,s} + \boldsymbol{\gamma}'_m \boldsymbol{w}^m_s + \varepsilon_{m,s+h}, \qquad m = 1, \ldots, M, \tag{2.6}$$

where the vector of observable predictors $\boldsymbol{w}^m_s$ contains $\tilde{y}_{m,s}$ and lagged values of it, α_m, $\boldsymbol{\beta}_m$, and $\boldsymbol{\gamma}_m$ are respective parameter vectors, and $\varepsilon_{m,s+h}$ is a disturbance term. In this application we only forecast single maturities by means of M independent regressions. The factor estimates $\hat{f}_{r,t}$ are determined by means of the PC-OLS or the PC-GLS approach. The parameters $\hat{\alpha}_m$, $\hat{\boldsymbol{\beta}}_m = (\hat{\beta}_{m,1}, \ldots, \hat{\beta}_{m,R})'$, and $\hat{\boldsymbol{\gamma}}_m = (\hat{\gamma}_{m,1}, \ldots, \hat{\gamma}_{m,q})'$ are estimated via an OLS regressions within the time-window $[s-(\tau-q-1), \ldots, s-h]$. Conditional on $\Xi_{s,\tau}$ the h–step ahead forecast is then

$$\hat{y}_{m,s+h|s} = \hat{\alpha}_m + \hat{\beta}_{m,1}\hat{f}_{1,s} + \ldots + \hat{\beta}_{m,R}\hat{f}_{R,s} + \hat{\gamma}_{m,1}\tilde{y}_{m,s} + \ldots + \hat{\gamma}_{m,q}\tilde{y}_{m,s-(q-1)} + \bar{y}_{m,s}.$$

This method employs the factors like any other variable. The relationship of factors and variables known through the factor model is not exploited.

Forecasting the factors Using factor models in forecasting exercises, it is intuitive to exploit the relationship between factors and variables given by the factor loadings. In the approach from Blaskowitz & Herwartz (2011), the factors $\check{\boldsymbol{f}}_{s+h|s}$ are first predicted, and then the factor model Eq. (2.1) is used to determine forecasts $\hat{y}_{m,s+h|s}$. The basic assumption is that factors follow a VAR process in first differences,

$$\Delta\hat{\boldsymbol{f}}_{\bullet,t} = \boldsymbol{\nu} + \Phi_1 \Delta\hat{\boldsymbol{f}}_{\bullet,t-1} + \cdots + \Phi_q \Delta\hat{\boldsymbol{f}}_{\bullet,t-q} + \boldsymbol{\eta}_t,$$

where $\Delta\hat{\boldsymbol{f}}_{\bullet,t} := \hat{\boldsymbol{f}}_{\bullet,t} - \hat{\boldsymbol{f}}_{\bullet,t-1}$, the constant vector $\boldsymbol{\nu}$ is $R \times 1$, the Φ_l are $R \times R$ matrices for $l = 1, \ldots, q$, and $\boldsymbol{\eta}_t$ is a $R \times 1$ vector of serially uncorrelated disturbances. Due to the fact that factors are estimated with PCA, they are orthogonal. Thus the factors $\hat{f}_{r,t}$ are treated as cross-sectionally uncorrelated with each other[5]. The matrices Φ_l are accordingly presumed to be diagonal. For each factor a univariate autoregression is then employed for the purpose of forecasting

$$\Delta\check{f}_{r,s+t|s} = \hat{\nu}^{(r)} + \hat{\phi}^{(r)}_1 \Delta\check{f}_{r,s+t-1|s} + \ldots + \hat{\phi}^{(r)}_q \Delta\check{f}_{r,s+t-q|s}. \tag{2.7}$$

[5] This approach follows Blaskowitz & Herwartz (2011). Blaskowitz & Herwartz (2009) nevertheless point out that the time series are not serially correlated does not automatically follow from orthogonality. AR estimation is however far more convenient than VAR estimation. Moreover, AR Nelson-Siegel models can outperform their VAR counterparts (Yu & Zivot, 2011).

The OLS estimates in Eq. (2.7) $\hat{\nu}^{(r)}$, $\hat{\phi}_1^{(r)}, \ldots, \hat{\phi}_q^{(r)}$ are conditioned on $\Xi_{s,\tau}$, and $\Delta\check{f}_{r,s+t|s} = \Delta\hat{f}_{r,s+t}$ if $t \leq 0$. Iteratively for $t = 1, \ldots, h$ the prediction obtained from Eq. (2.7) is performed until we obtain $\Delta\check{f}_{r,s+h|s}$. The level factor forecast is then

$$\check{f}_{r,s+h|s} = \hat{f}_{r,s} + \sum_{t=1}^{h} \Delta\check{f}_{r,s+t|s}.$$

The h-step forecast of $\boldsymbol{y}_{s+h}$ is thus

$$\hat{\boldsymbol{y}}_{s+h|s} = \mathbf{E}[\tilde{\boldsymbol{y}}_{s+h}|\Xi_{s,\tau}] + \bar{\boldsymbol{y}}_s = \hat{A}_s^{(Y)}\check{\boldsymbol{f}}_{s+h|s} + \bar{\boldsymbol{y}}_s. \tag{2.8}$$

The loading matrix $\hat{A}_s^{(Y)}$ here is a $M \times R$ sub-matrix of $\hat{A}_s$ or A_s^*. The matrix contains the factor loadings of the elements in $\tilde{\boldsymbol{y}}_s$ that are estimated as described in Section 2.1.2. The forecasting method relies on the assumption that the factor loadings are stable out-of-sample.

It is noteworthy here to point out that both the BH forecasting method and the PC-GLS estimation method exploit dynamic features within factor models. They differ in that in the PC-GLS method the factor errors can incorporate an autoregressive structure while in the BH method it is assumed in the factors themselves. As $R \leq N$, it follows that in the BH method fewer parameters are to be estimated than for PC-GLS. The BH method therefore has a practical advantage in its applications.

2.1.4 Forecast Evaluation

To evaluate forecasts in statistical and economic terms three criteria are employed. Diebold & Mariano (2002) point out that statistical criteria might not necessarily conform with economic measures. For instance, Leitch & Tanner (1991) find that, compared to the mean squared forecast error (MSFE), the directional accuracy (DA) of forecasts is highly correlated with profits in a term structure analysis. We therefore consider two economic criteria in addition to the standard MSFE criterion. As in Blaskowitz & Herwartz (2011), evaluation criteria are the MSFE, DA, and big hit ability (BHA). The overall performance of competing forecasting methods is evaluated by means of an analysis of variance (ANOVA), in order to determine the average impact of any particular method on forecast efficiency.

Mean squared forecast error The most-used and intuitive criterion to evaluate forecast strategies is the MSFE, i.e.

$$MSFE^{h,m} = \frac{1}{T - (T_0 + \bar{h} - 1)} \sum_{s=T_0+\bar{h}}^{T} \hat{\varepsilon}^2_{m,s+h|s}, \tag{2.9}$$

where $T_0 = 252$ and $\bar{h} = 15$ is the largest forecast horizon applied in this study. The realised forecast error is $\hat{\varepsilon}_{m,s+h|s} = \hat{y}_{m,s+h} - y_{m,s+h|s}$. With an increasing forecast horizon h, the $MSFE$ values should logically increase.

Directional accuracy In order to have a measure of directional accuracy (DA) "loss", we use the function

$$\overline{DA}^{h,m} = \frac{1}{T - (T_0 + \bar{h} - 1)} \sum_{s=T_0+\bar{h}}^{T} I\left[(\hat{y}_{m,s+h|s} - y_{m,s})(y_{m,s+h} - y_{m,s}) > 0\right], \qquad (2.10)$$

where $I[\bullet]$ is an indicator function. Perfect directional accuracy is indicated by a DA value equal to one.

Big hit ability The "Big Hit Ability" (BHA) is a measure for forecast accuracy proposed by Hartzmark (1991). The functions weighs the correct prediction of large changes by their size. The BHA "loss" function is

$$\overline{BHA}^{h,m} = \frac{1}{T - (T_0 + \bar{h} - 1)} \sum_{s=T_0+\bar{h}}^{T} BHA^{h,m}_{s+h|s}, \qquad (2.11)$$

with

$$BHA^{h,m}_{s+h|s} = \begin{cases} |y_{m,s+h} - y_{m,s}| & \text{if } (\hat{y}_{m,s+h|s} - y_{m,s})(y_{m,s+h} - y_{m,s}) > 0 \\ -|y_{m,s+h} - y_{m,s}| & \text{if } (\hat{y}_{m,s+h|s} - y_{m,s})(y_{m,s+h} - y_{m,s}) < 0. \end{cases}$$

Here, a positive $\overline{BHA}$ can be obtained from a few directionally accurate forecasts that generate large profits while most other forecasts are inaccurate but have small losses. In contrast to $\overline{MSFE}$, the best models according to the $\overline{DA}$ and $\overline{BHA}$ criteria are those that maximise the corresponding loss function.

These three criteria or loss functions address three different aspects that are associated with the risk of forecasting financial time series. MSFE reflects how close a forecast is to the realised value. The DA loss function represents the ability to predict the direction a financial time series will take in the future. The BHA loss function gives a weight to this ability. It thereby reflects the size of possible losses and therefore, together with the other two loss functions, quantifies the inherent risk of financial forecasts.

2.2 Variables, Eigenvalues and Factors, Forecasts

Government bonds are an important part in many portfolios. The interest rates for these bonds reflect a measure for the probability that the given government will not be able to pay the interest rates or repay its debt. The interest rates and the respective size of the bonds in the portfolio therefore reflect their risk in the portfolio. We now present the data and some term structure specific features, perform eigenvalue analysis, describe the factor loadings, and describe the forecasting strategies and their ANOVA evaluation.

2.2.1 Data

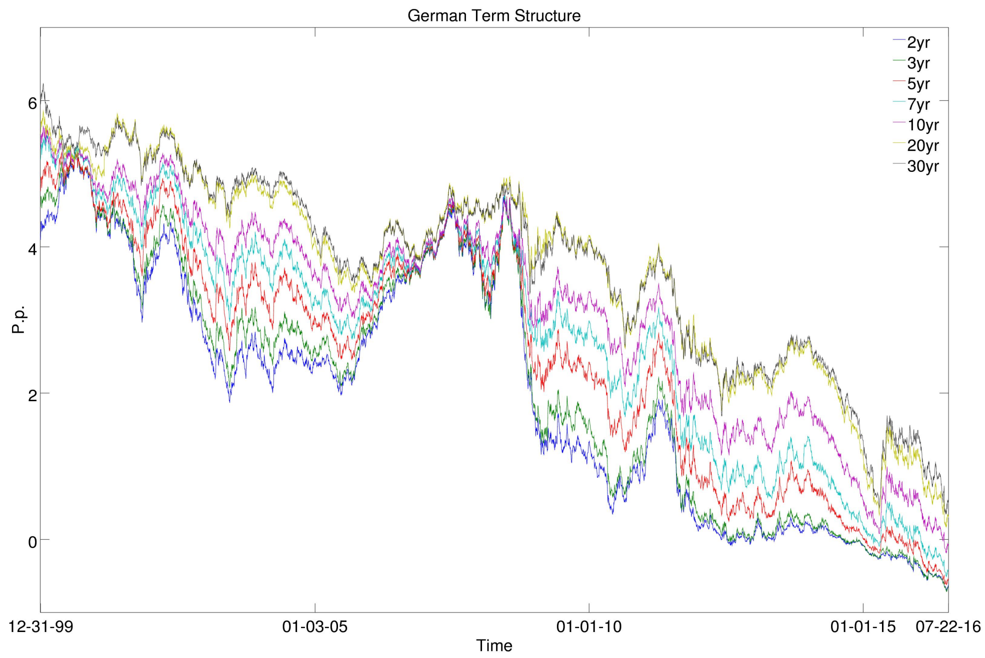


Figure 2.1: Evolution of the term structure of German government bonds. The vertical axis is in percentage points and the horizontal axis is time.

To evaluate forecasts for the term structure of interest rates, we select yields of bonds issued by the federal government of Germany (GER), Swiss confederation bond yields (SWI), UK government liability nominal spot rates (UK), and US treasury benchmark bonds (US).

Forecasting exercises are performed for seven GER bond interest rates with maturities, collected in $\boldsymbol{y}_s^{GER} = (y_{1,s}, \ldots, y_{7,s})'$, as well as five SWI, eight UK, and six US interest rates collected in analogous vectors, $\boldsymbol{y}_s^{SWI}$, $\boldsymbol{y}_s^{UK}$, and $\boldsymbol{y}_s^{US}$. The respective

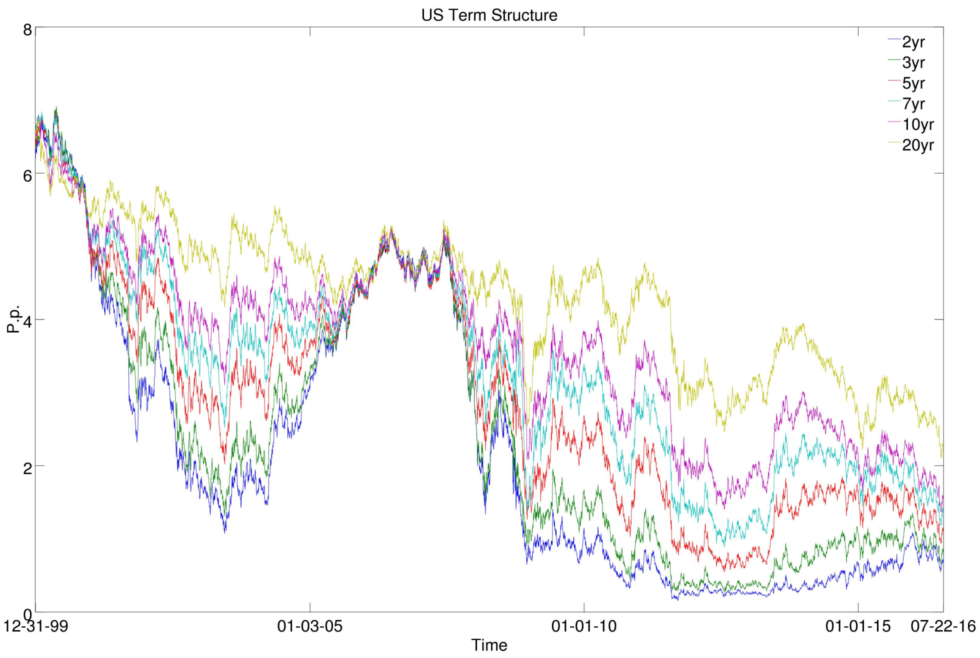


Figure 2.2: Evolution of the term structure of US government bonds. The vertical axis is in percentage points and the horizontal axis is time.

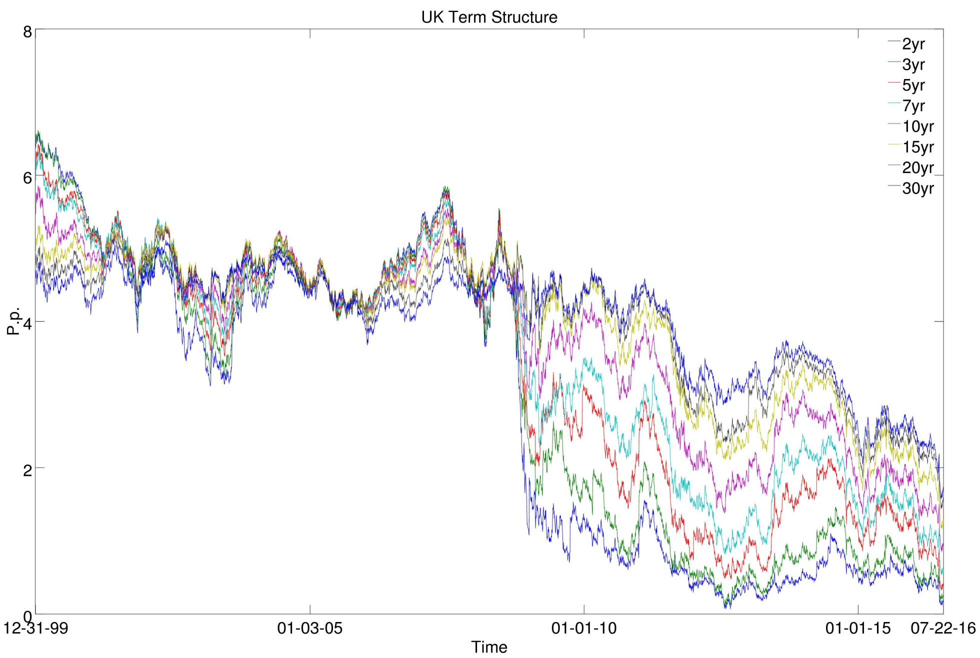


Figure 2.3: Evolution of the term structure of UK government bonds. The vertical axis is in percentage points and the horizontal axis is time.

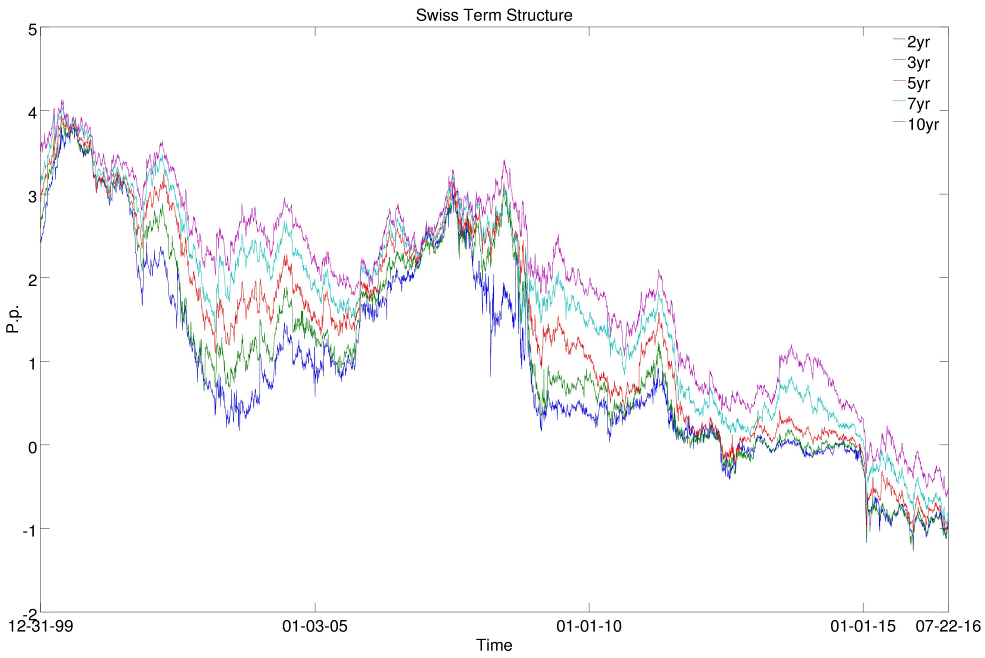


Figure 2.4: Evolution of the term structure of Swiss government bonds. The vertical axis is in percentage points and the horizontal axis is time.

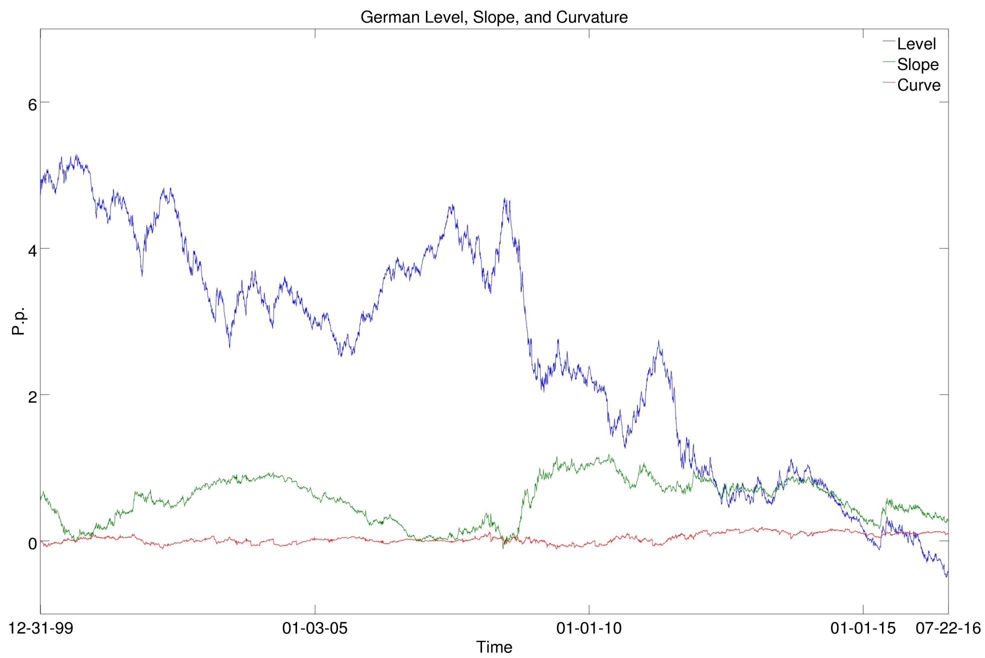


Figure 2.5: Evolution of the German level, slope, and curvature. The vertical axis is in percentage points and the horizontal axis is time.

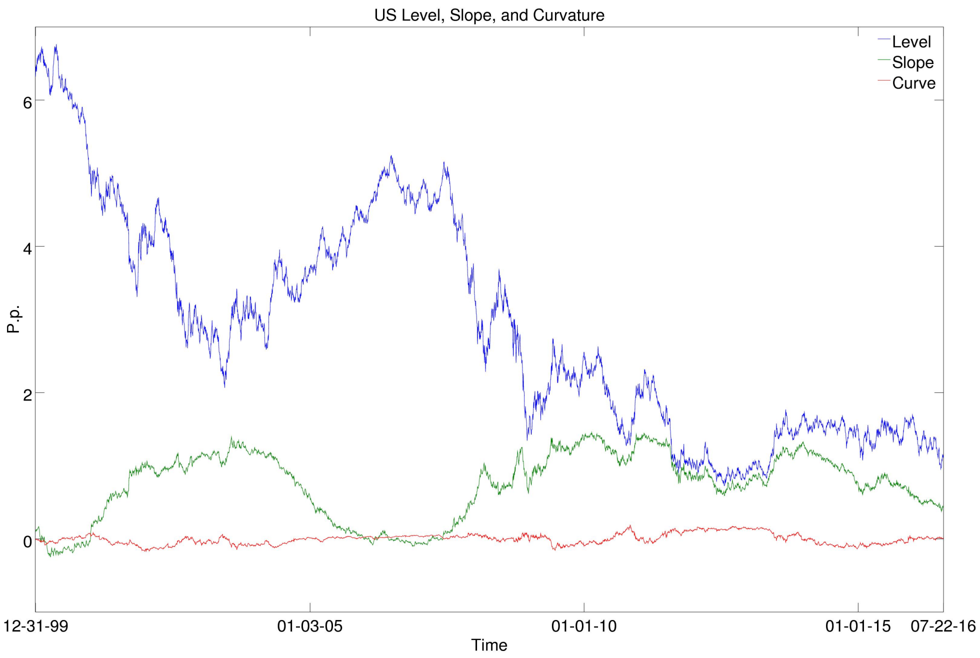


Figure 2.6: Evolution of the US level, slope, and curvature. The vertical axis is in percentage points and the horizontal axis is time.

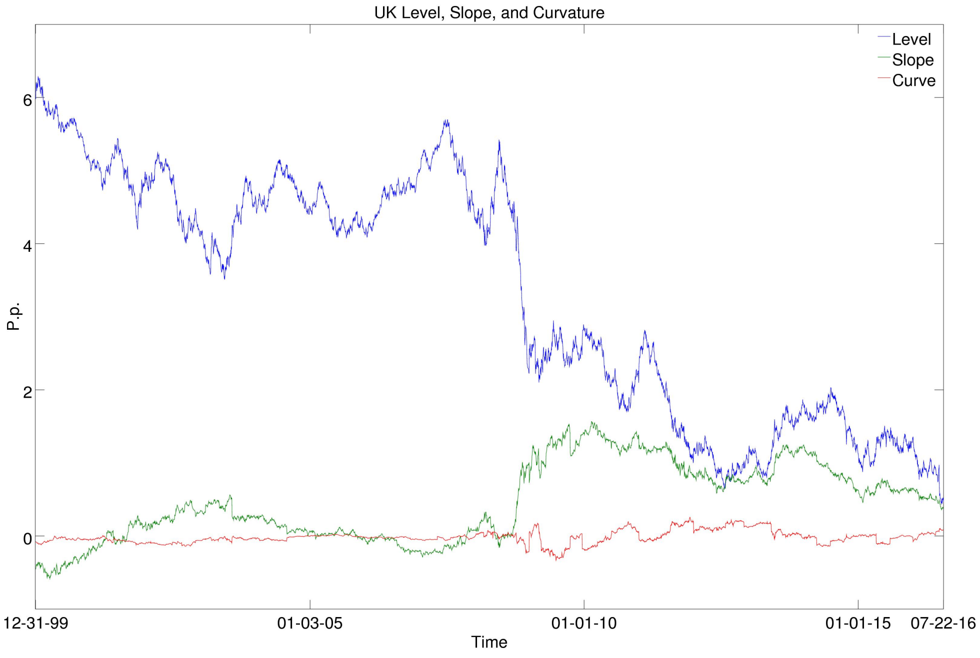


Figure 2.7: Evolution of the UK level, slope, and curvature. The vertical axis is in percentage points and the horizontal axis is time.

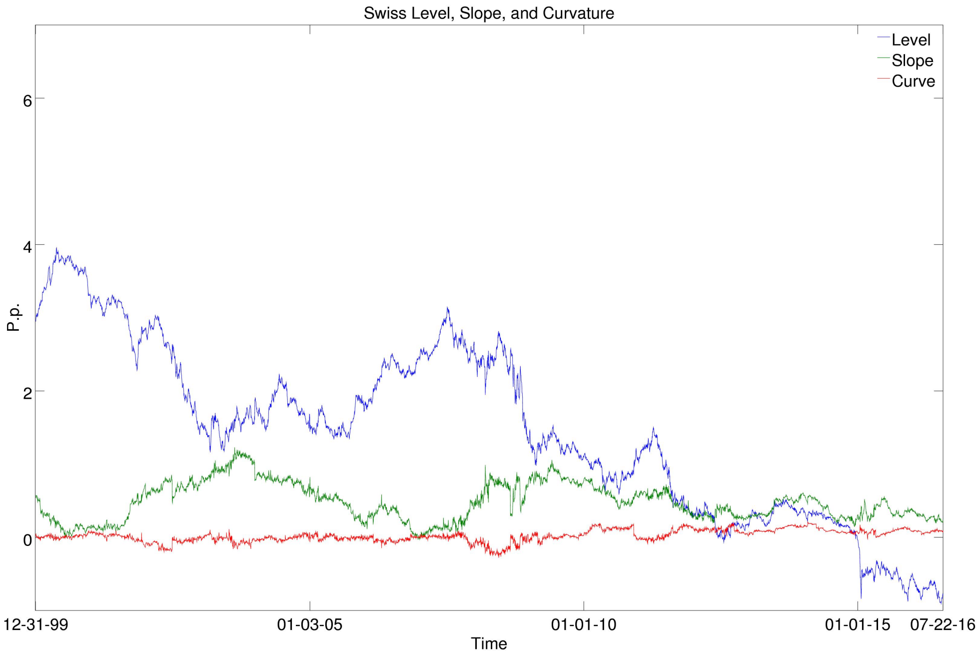


Figure 2.8: Evolution of the Swiss level, slope, and curvature. The vertical axis is in percentage points and the horizontal axis is time.

yr	1	2	3	5	7	10	15	20	30	N
GER		×	×	×	×	×		×	×	7
SWI		×	×	×	×	×				5
UK		×	×	×	×	×	×	×	×	8
US		×	×	×	×	×		×		6

Table 2.1: Maturities used for GER, SWI, the UK, and the US.

maturities are listed in Table 2.1. In addition, the potential of strategies to forecast the level, slope and curvature of the term structure is investigated. These are obtained from rates with maturities of 2, 5, and 10 years. The market characteristics are determined as in Blaskowitz & Herwartz (2011) as: $lev_s = (2yr_s + 5yr_s + 10yr_s)/3$ (level), $slo_s = (10yr_s - 2yr_s)/2$ (slope), and $cur_s = 2yr_s/4 - 5yr_s/2 + 10yr_s/4$ (curvature).

In our first data set, similar to the one used in Blaskowitz & Herwartz (2011), only interest rate content is exploited with $N = 7$ (GER), $N = 5$ (SWI), $N = 8$ (UK), and $N = 6$ (US) dimensioned types of distinct maturities, Table 2.1 gives an overview of the applied maturities. These are used for factor extraction and prediction ($X_s = Y_s$). In a second forecasting exercise, the data set uses all of these interest rates totalling $N = 26$ predictors. All series are collected at the daily frequency for the period from 1999-12-31 to 2016-07-22, comprising a total of 4321 observations. The series are drawn from Thomson Datastream (`http://www.datastream.com`). As the largest size of rolling estimation windows is $\bar{\tau} = 252$, the highest forecast horizon is $\bar{h} = 15$, we arrive at a total of 4055 forecasts starting at 2001-01-09 which are compared with actual realisations.

Adding more data as in the second data set might supply additional information to improve ex–ante forecasts. Including additional interest rates might help to estimate the underlying factors governing the term structure of interest rates. Further predictors may proxy macroeconomic effects. Ang & Piazzesi (2003) use an inflation and an economic growth factor in term structure forecasts. The factors are extracted from variables that can either be considered as inflation or real activity variables. Furthermore, Diebold et al. (2005) find that inflation is correlated with the level factor, the slope factor with real activity, while the curvature factor does not appear to be correlated with any main macroeconomic variable. Mönch (2008) uses 160 macroeconomic monthly time series (which is similar in size to the data set in Matthies (2014)) in a no-arbitrage model of interest rate forecasts.

However, some of the variables from which for instance Ang & Piazzesi (2003) extract these two factors are not available on a daily basis, and thus cannot be used here. Furthermore, in contrast to Mönch (2008), other interest rates are not excluded, as this is not a no-arbitrage application.

Figures 2.1, 2.2, 2.3, and 2.4 present the evolution of the GER, US, UK, and Swiss term structure from 1999-12-31 to 2016-07-22 respectively. Furthermore, figures 2.5, 2.6, 2.7, and 2.8 present the evolution of the GER, US, UK, and Swiss level, slope, and curvature from 1999-12-31 to 2016-07-22 respectively.

Figures 2.1, 2.2, 2.3, and 2.4 show there is a downward trend in all term structures and a larger spread between short and long term rates after 2008. The downward trend is stronger for GER and SWI. Moreover, some rates for GER and SWI become negative from mid 2014 and May 2012 respectively. There are no negative rates for the US and the UK in this time period.

In figures 2.5, 2.6, 2.7, and 2.8, it can be seen that level, slope, and curvature of the term structure are most likely time varying. Figure 2.5 specifically shows the development of the level, slope, and curvature of the German term structure. The level is occasionally negative between February and May 2015 and from June 2016 onwards. A negative level means that the average of the 2yr, 5yr ,and 10yr rates are negative. On four occasions the slope changes its sign and becomes negative; in August 2000, from October till December 2006, in March 2007, and in June 2016. A negative slope might be interpreted as an indicator of a future recession (see for example Estrella & Hardouvelis (1991)). In Figure 2.6 it can be seen that the US had falling level after 2008. The slope is frequently negative for half a year from February 2000 until August 2000 and for one and a half years from December 2005 until June 2007. For the UK we can see a falling level in Figure 2.7. The slope has frequent phases of negativity until 09-04-2001 and from September 2005 until September 2008, covering a period of three years. In Figure 2.8 it is observable that SWI experienced a negative level between June and September 2012 and from December 2014 onwards. Furthermore, the slope is negative from December 2006 till January 2007.

We find no negative rates for the UK and the US but longer periods with a negative slope. Analysing the level, slope, and curvature provides further motivation for using a data-driven rolling time window approach.

2.2.2 Eigenvalues and Factorloadings

The term structures are assumed to be governed by common factors (Nelson & Siegel, 1987, Litterman & Scheinkman, 1991, Diebold & Li, 2006). Empirically these factors can be for example extracted via PCA as in Blaskowitz & Herwartz (2011) or given predetermined loadings as in Diebold & Li (2006) estimated via a Kalman filter (Christensen et al., 2011). Here, we use PCA to extract the factors. We now discuss the eigenvalues and eigenvectors which are interpreted as factor loadings.

Eigenvalues Figures 2.9 and 2.10, 2.11 and 2.12, 2.13 and 2.14, 2.15 and 2.16, and 2.17 and 2.18 show the development of eigenvalues of the term structure set of GER, the US, the UK, SWI, and for all data respectively for two time windows of the size $\tau = 42$ and $\tau = 252$.

The eigenvalues are normalised. They are divided by the total sum of all eigenvalues. These (normalised) eigenvalues show what share of the variance the corresponding factor explains. The first eigenvalues tend to be of higher magnitude in larger time

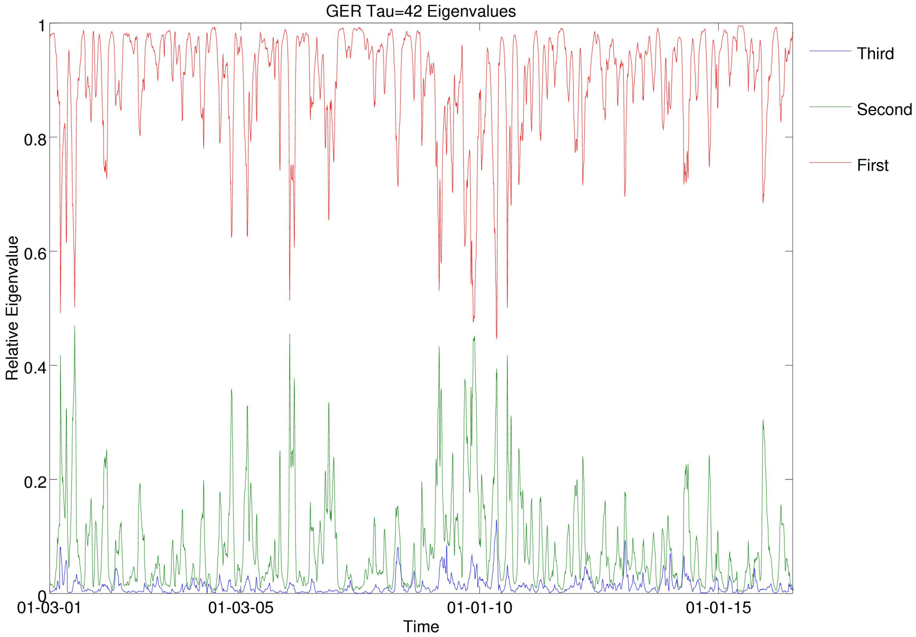


Figure 2.9: Evolution of the German eigenvalues for $\tau = 42$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

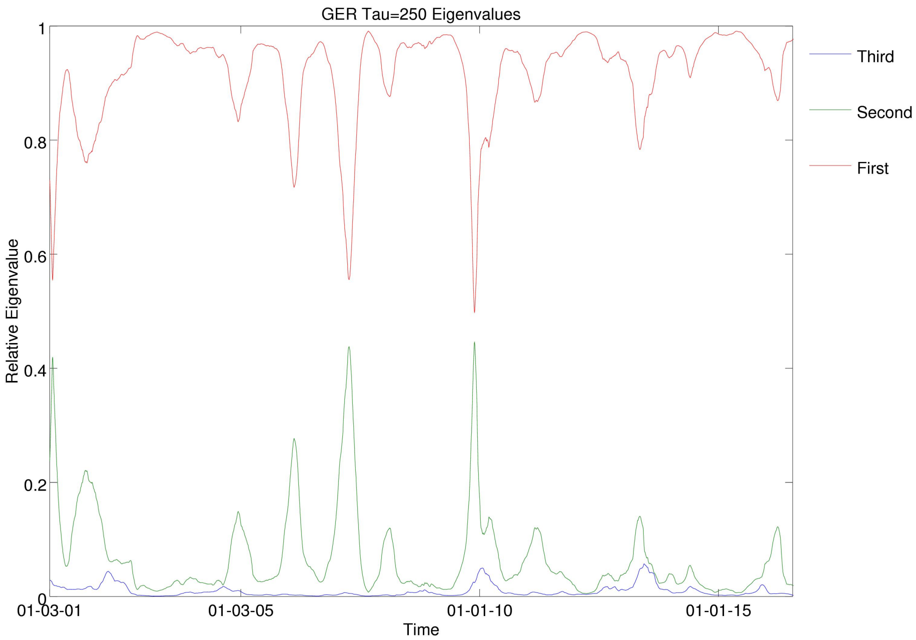


Figure 2.10: Evolution of the German eigenvalues for $\tau = 250$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

windows. Therefore, the first factor explains more of the variance in larger time windows than in smaller ones. The first eigenvalue is usually interpreted as the level of the term structure.

We can observe for the GER $\tau = 42$ time window in Fig. 2.9 that the first factor often explains 90% and upwards of the variance. At multiple instances, the second factor explains more than 20% and sometimes more than 40%. For $\tau = 250$ in Fig. 2.10 the eigenvalue processes are smoother yet the main features remain. The first factor explains most of the variance. The increases of the second eigenvalue correspond to those in the $\tau = 42$ time window, yet they are now increases over a continuous period instead of erratically clustered spikes.

For the US data set we see for the $\tau = 42$ time window in Fig. 2.11 that the first two factors behave similar to the GER term structure. In contrast, the second eigenvalue has less spikes. This feature is more profound for the $\tau = 250$ time window in Fig. 2.12. The third eigenvalue explains less in comparison to the GER data set.

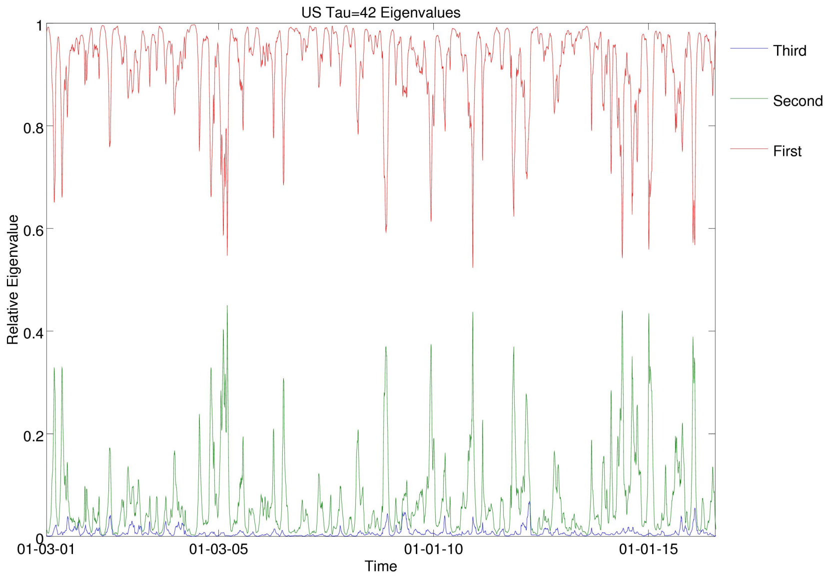


Figure 2.11: Evolution of the US eigenvalues for $\tau = 42$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

Examining the evolution of the UK eigenvalues for the $\tau = 42$ time window in Fig. 2.13 we find that in contrast to GER and the US the second factor spikes more frequently. In particular, this occurs after the 2008 crisis. There are periods when the third factor explains up to 20% of the variance. The $\tau = 250$ time window in Fig. 2.14 further highlights the third eigenvalue contribution.

In the SWI data set the second eigenvalue spikes are most frequent as can be seen in

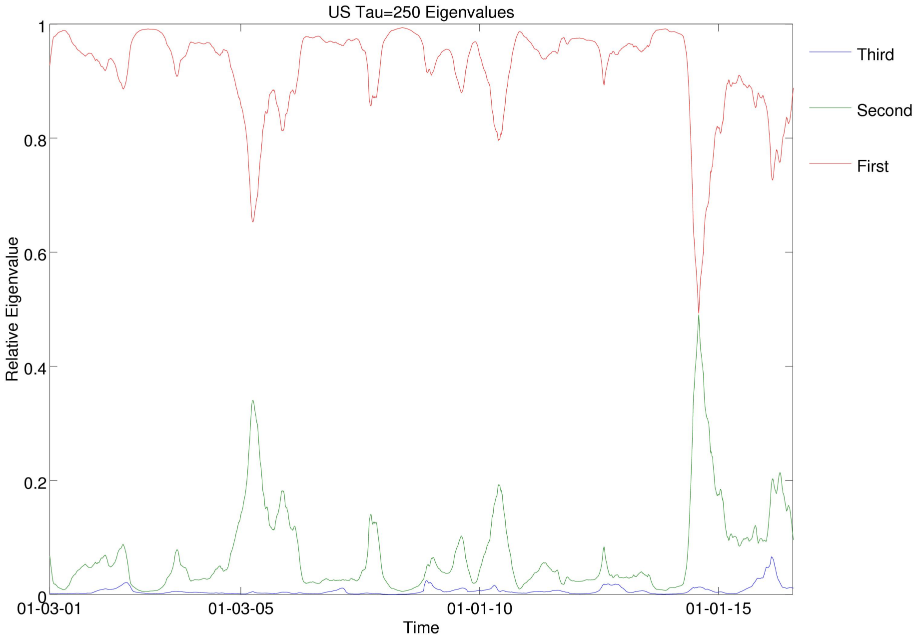


Figure 2.12: Evolution of the US eigenvalues for $\tau = 250$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

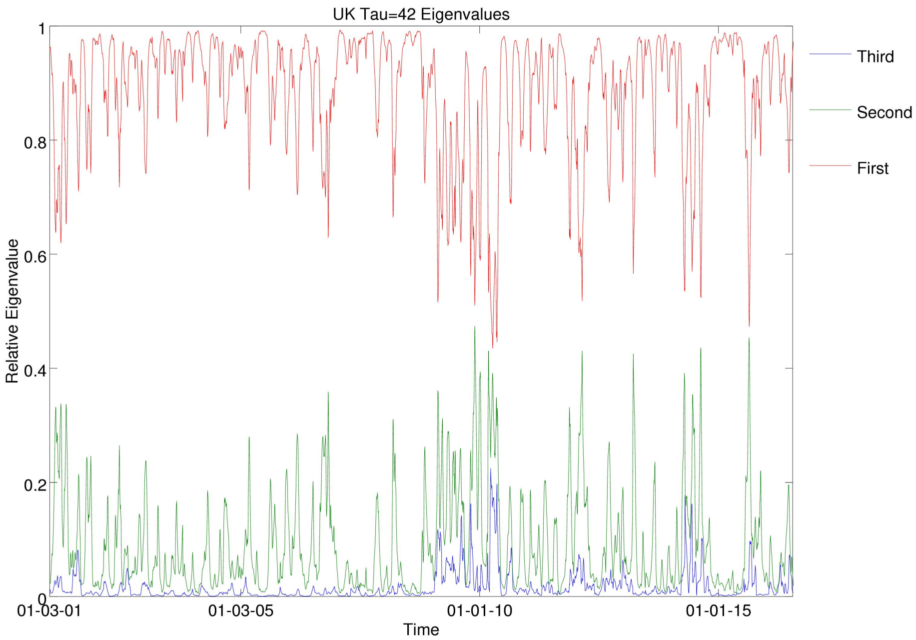


Figure 2.13: Evolution of the UK eigenvalues for $\tau = 42$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

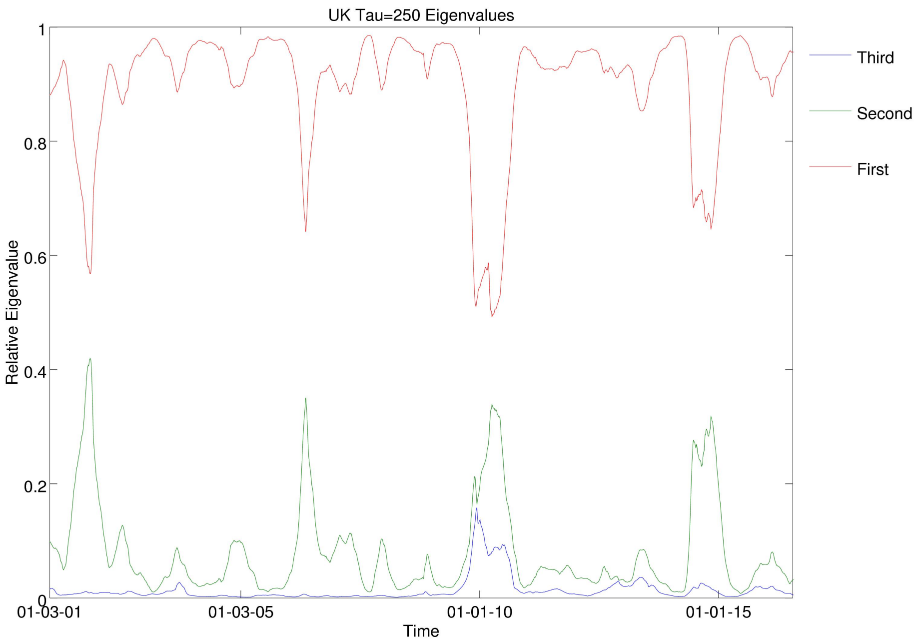


Figure 2.14: Evolution of the UK eigenvalues for $\tau = 250$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

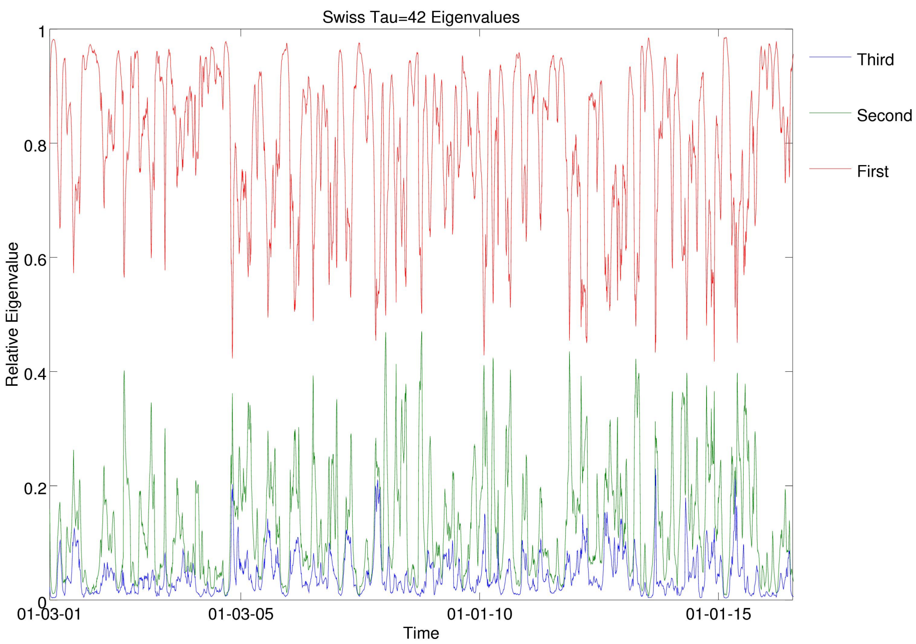


Figure 2.15: Evolution of the Swiss eigenvalues for $\tau = 42$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

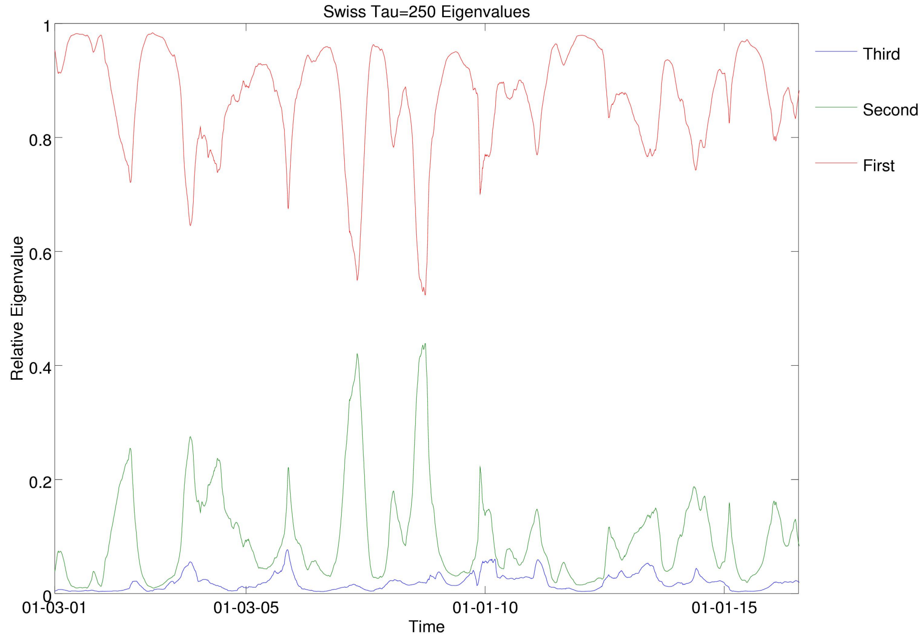


Figure 2.16: Evolution of the Swiss eigenvalues for $\tau = 250$. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

Figures 2.15 and 2.16. The variance explained by the third eigenvalue is also higher than in the other data sets. For the single term structure data sets two factors are sufficient to represent 90% of the variance, which could be seen as a criterion to determine the number of factors. This issue will be discussed in more detail in Section 2.3.2.

For the data set containing all term structures in the $\tau = 42$ time window the first eigenvalue frequently explains less than 90% of the variance as can be seen in Figure 2.17. The second eigenvalue exhibits frequent spikes above 20% throughout the entire time span of the data set as the first factor may often fall below 50% and the third factor is above 10%. For the $\tau = 250$ time window seen in Figure 2.18 the influence of the third eigenvalue is somewhat reduced. The features for the first and second eigenvalues remain, except for the smoother process in larger time windows observed for all data sets.

Factors The factor loadings for the first factors play an important role in understanding the term structure. Based on the idea found in Blaskowitz & Herwartz (2011) for intuitive factor loadings for the $2yr$, $5yr$, and $10yr$ maturities we develop the factor loadings along the yield curve as seen in Figure 2.19. We can then compare this theoretical construct with the empirical observations from the rolling time windows PCAs.

For the alternative data sets we perform rolling PCAs and document some res-

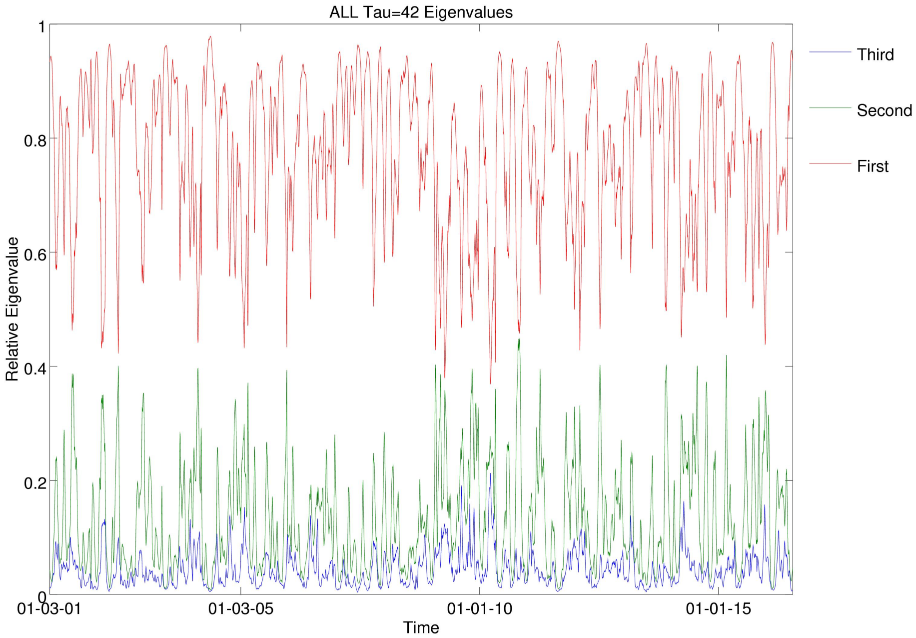


Figure 2.17: Evolution of the eigenvalues for $\tau = 42$ for all. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

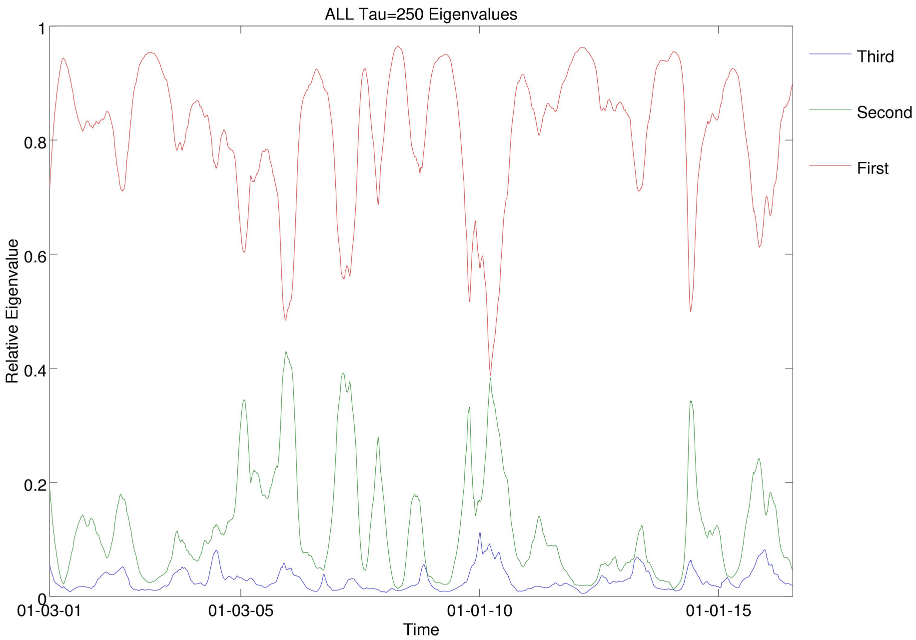


Figure 2.18: Evolution of the eigenvalues for $\tau = 250$ for all. The vertical axis is the standardised eigenvalue size and the horizontal axis is time.

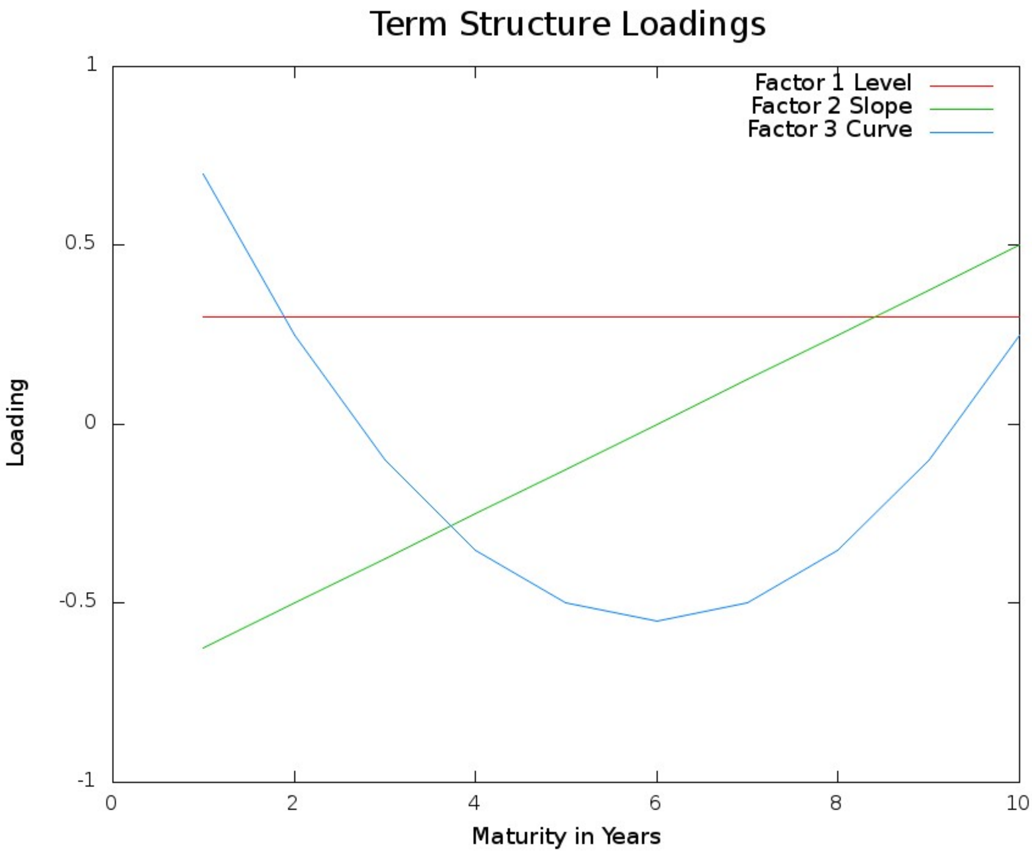


Figure 2.19: Factor loadings for level, slope, and curvature.

ults for the eigenvectors (factor loadings) $\boldsymbol{a}_{\bullet r}$. We discuss the properties and possible interpretations of the factors for the individual term structures (GER, US, UK, and SWI) and all term structures collected in a single data set. We first discuss the factor loadings in the individual data sets and then in the larger data set.

The first three factors are often interpreted as the level, slope, and curvature of the term structure. For GER we can see in Figure 2.20 for the $\tau = 250$ time window that the loadings are all positive and of similar size. For the second factor we find that negative (positive) loadings on the $2yr$ and $3yr$ maturities correspond to positive (negative) loadings for the $10yr$, $20yr$, and $30yr$ maturities, as seen in Figure 2.21. For the third factor we observe in Figure 2.22 positive (negative) loadings on the $2yr$, $3yr$, $10yr$, $20yr$, and $30yr$ maturities while the corresponding $5yr$ and $7yr$ maturities have negative (positive) loadings.

In contrast to GER, the US data set has no $30yr$ maturity. Nevertheless, we find a similar picture for the first three factors of the US term structure as seen in Figures 2.23, 2.24, and 2.25. In comparison to the GER term structure, the UK data set additionally has a $15yr$ maturity as seen in Figures 2.26, 2.27, and 2.28 and the SWI has no maturities longer than $10yr$ as seen in Figures 2.29, 2.30, and 2.31. The structure of the loadings of the first three eigenvectors exhibits the same features. These results support the idea that the first three factors of the term structure represent the level, slope, and curvature of the term structure.

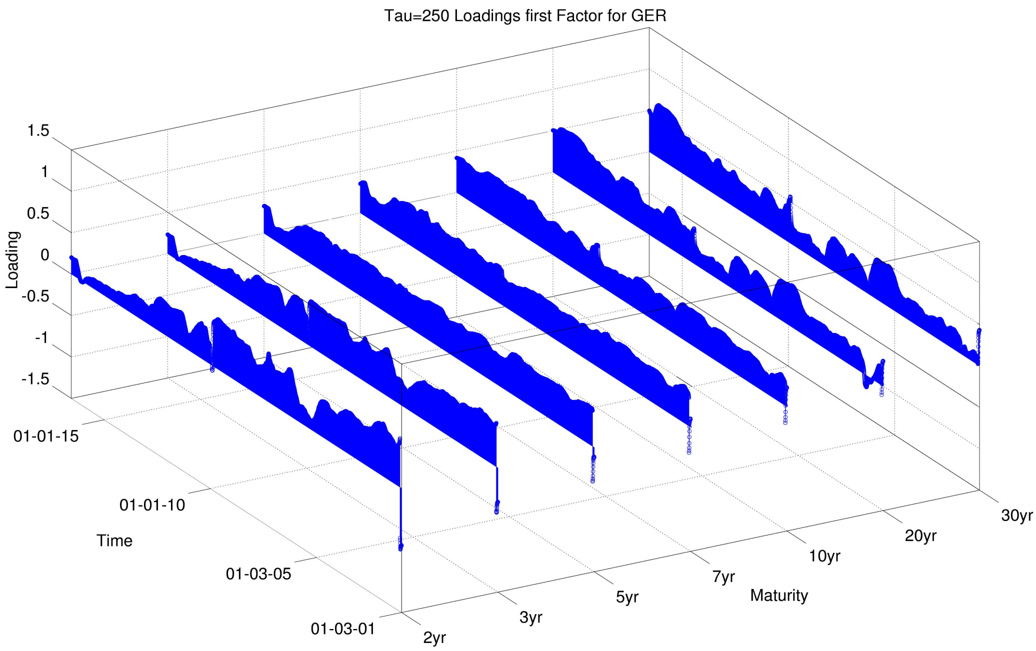


Figure 2.20: Evolution of the first factor loadings for GER with $\tau = 250$.

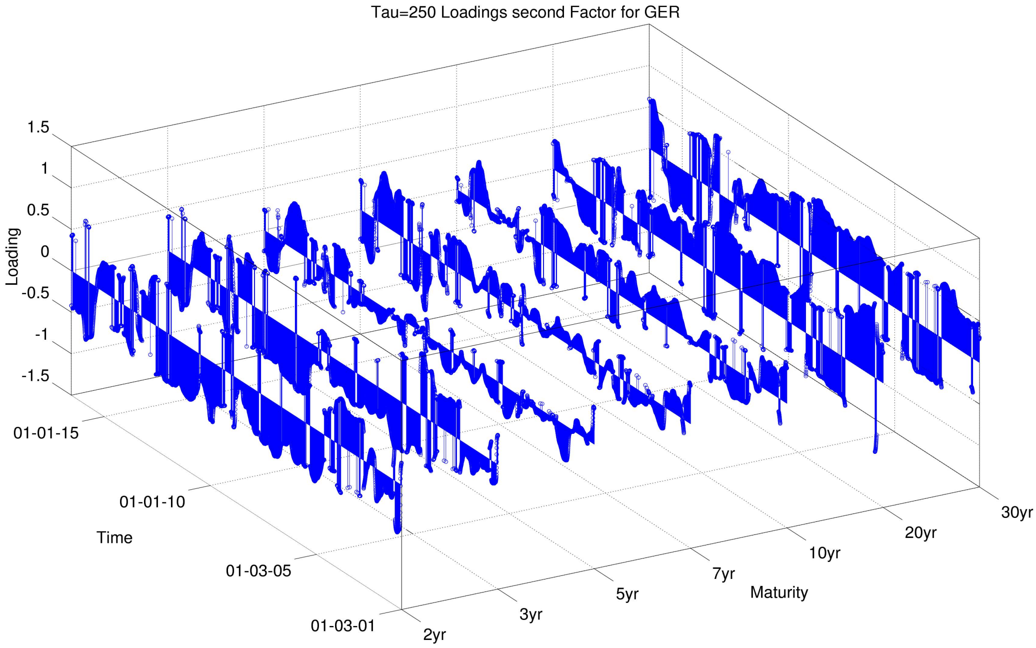


Figure 2.21: Evolution of the second factor loadings for GER with $\tau = 250$.

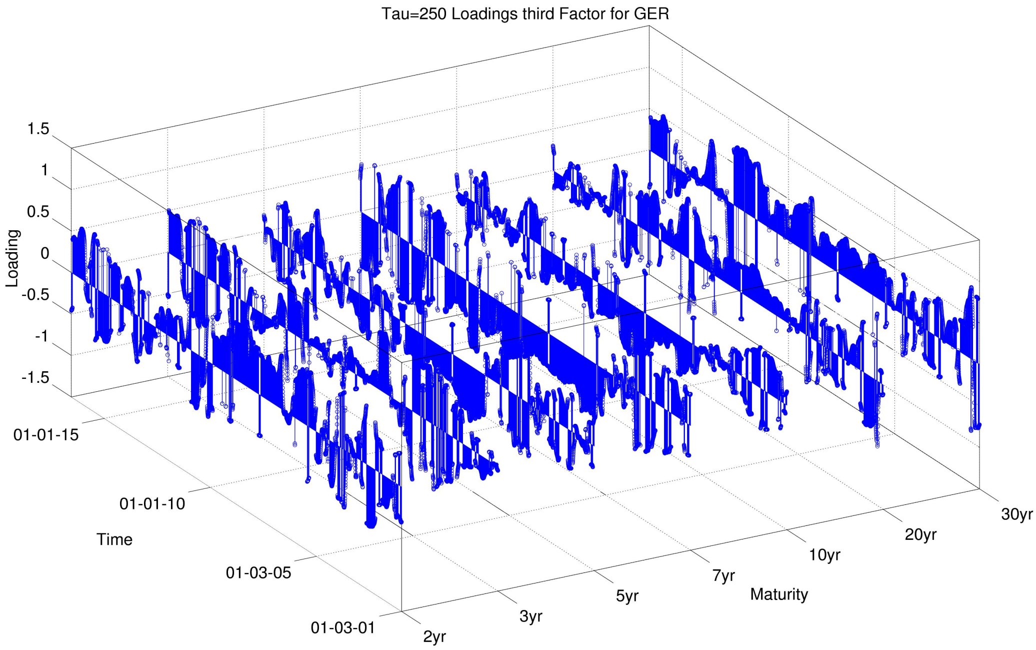


Figure 2.22: Evolution of the third factor loadings for GER with $\tau = 250$.

The factor loadings of the three largest factors in the larger data set describe the correlation between and within the term structures. In the larger data set that contains

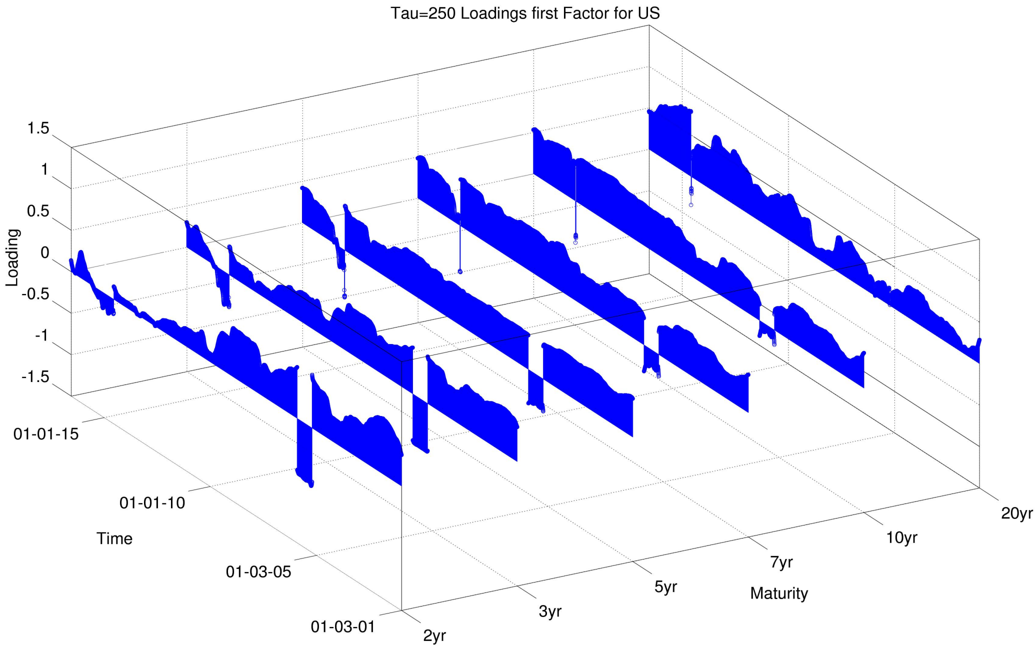


Figure 2.23: Evolution of the factor loadings for US of the first factor for $\tau = 250$.

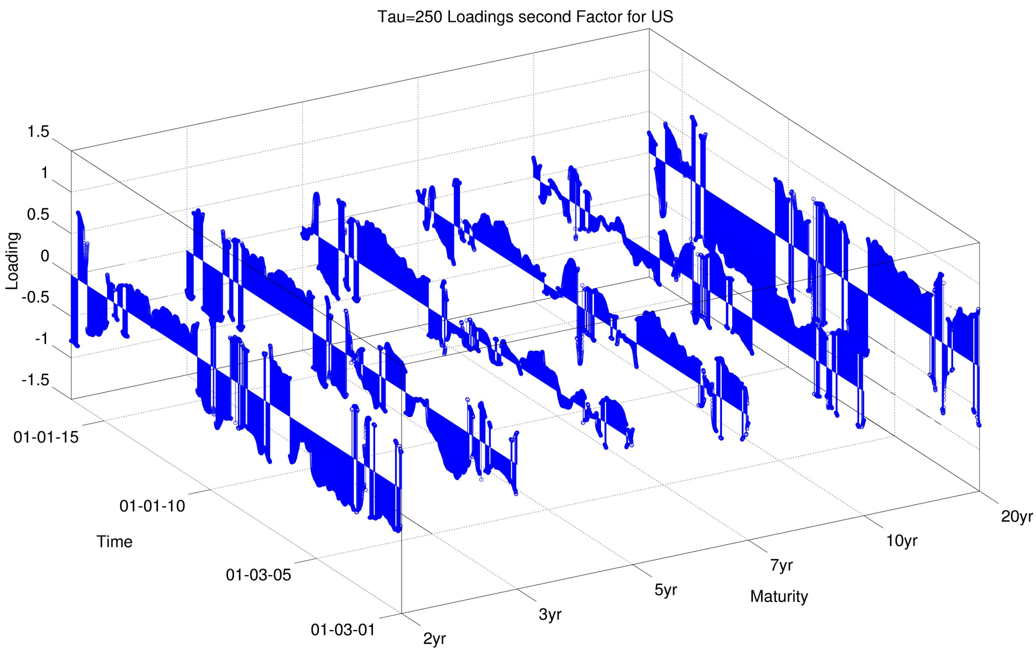


Figure 2.24: Evolution of the second factor loadings of the US for $\tau = 250$.

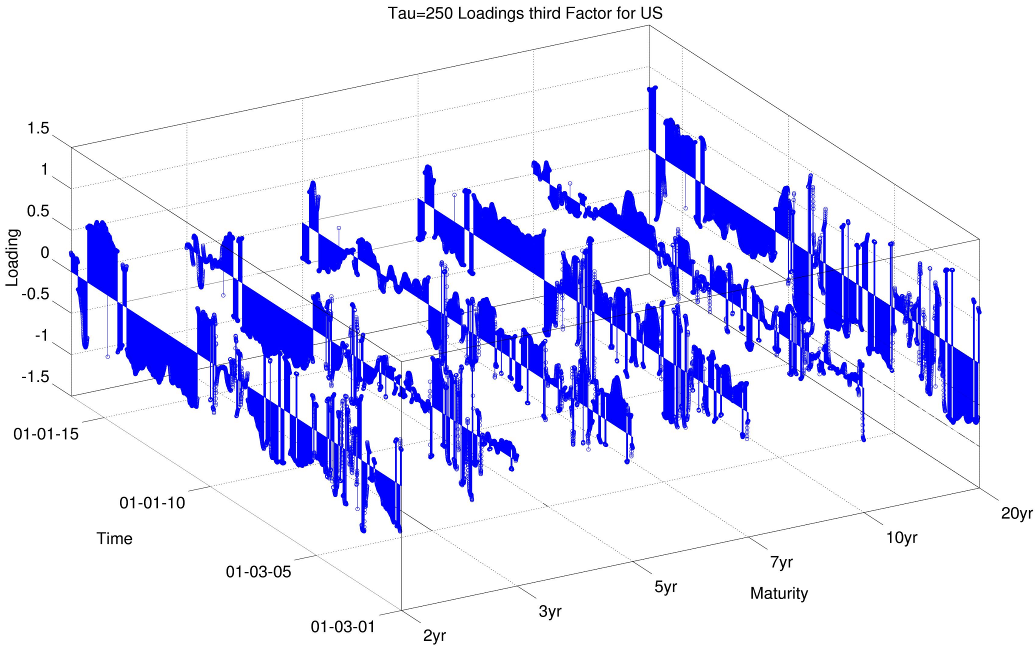


Figure 2.25: Evolution of the third factor loadings of the US for $\tau = 250$.

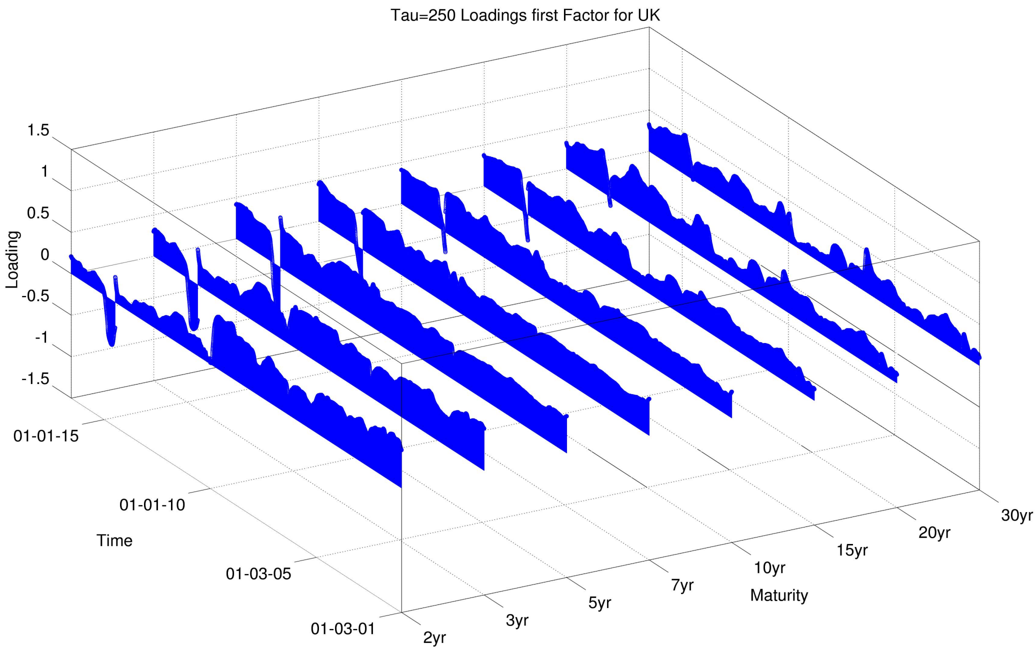


Figure 2.26: Evolution of the first factor loadings of the UK for $\tau = 250$.

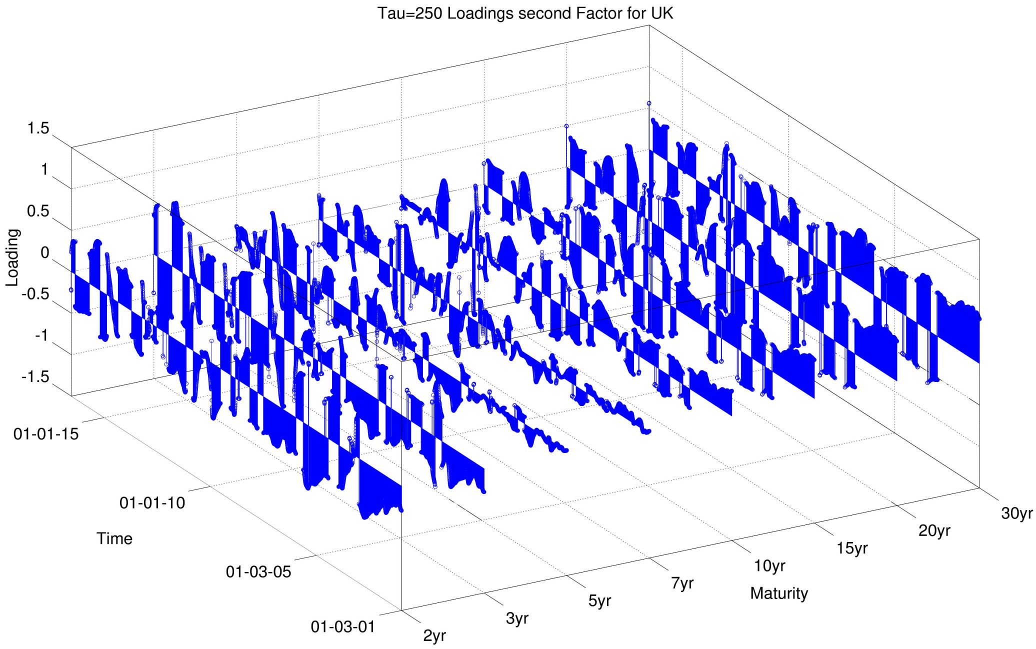


Figure 2.27: Evolution of the second factor loadings of the UK for $\tau = 250$.

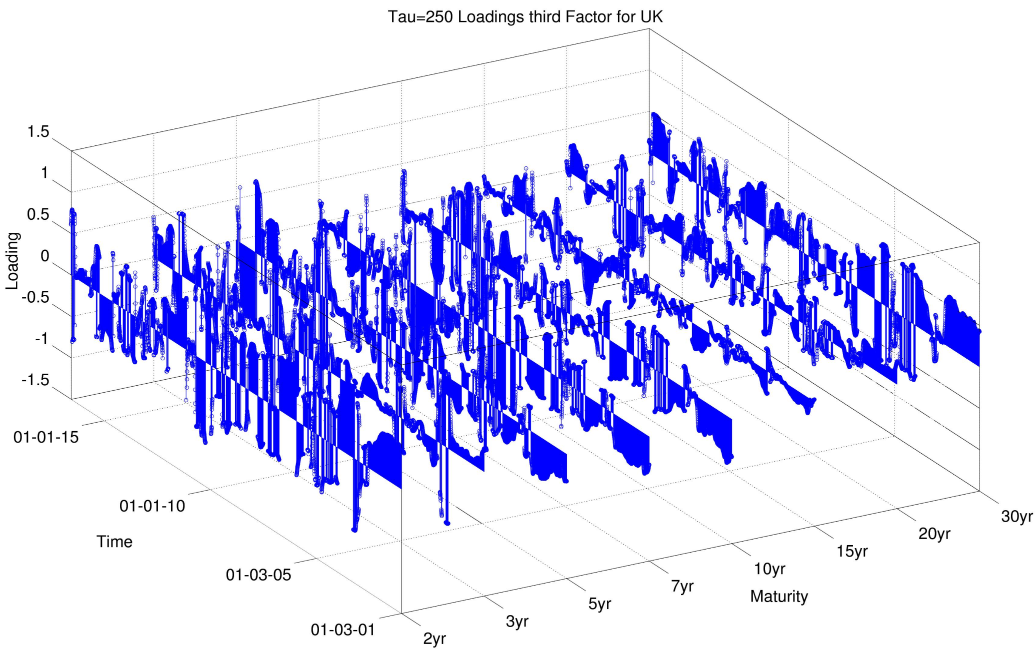


Figure 2.28: Evolution of the third factor loadings of the UK for $\tau = 250$.

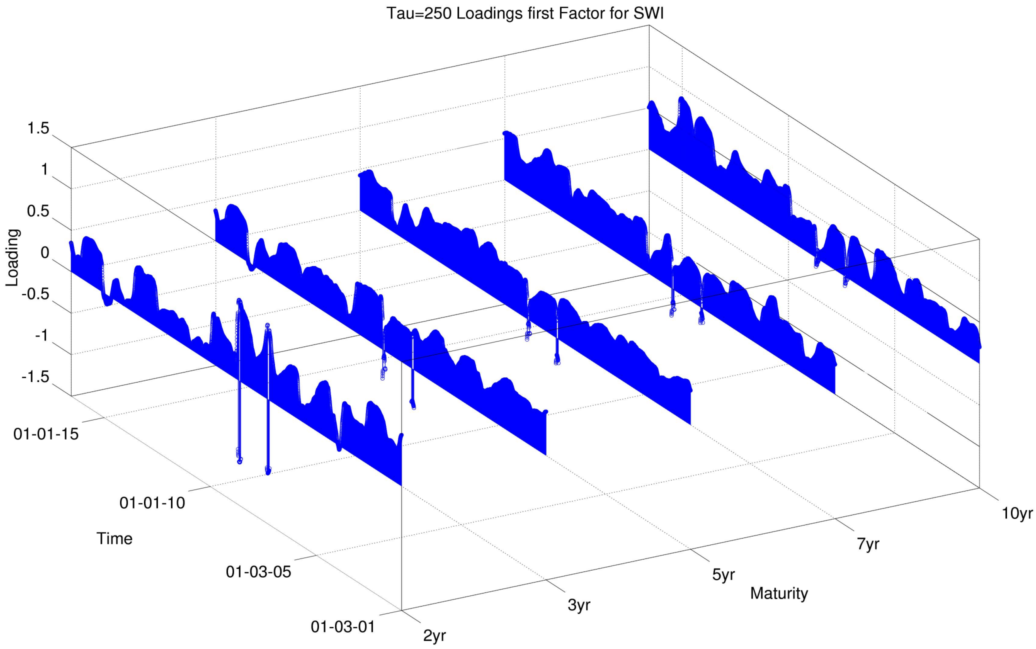


Figure 2.29: Evolution of the first factor loadings of SWI for $\tau = 250$.

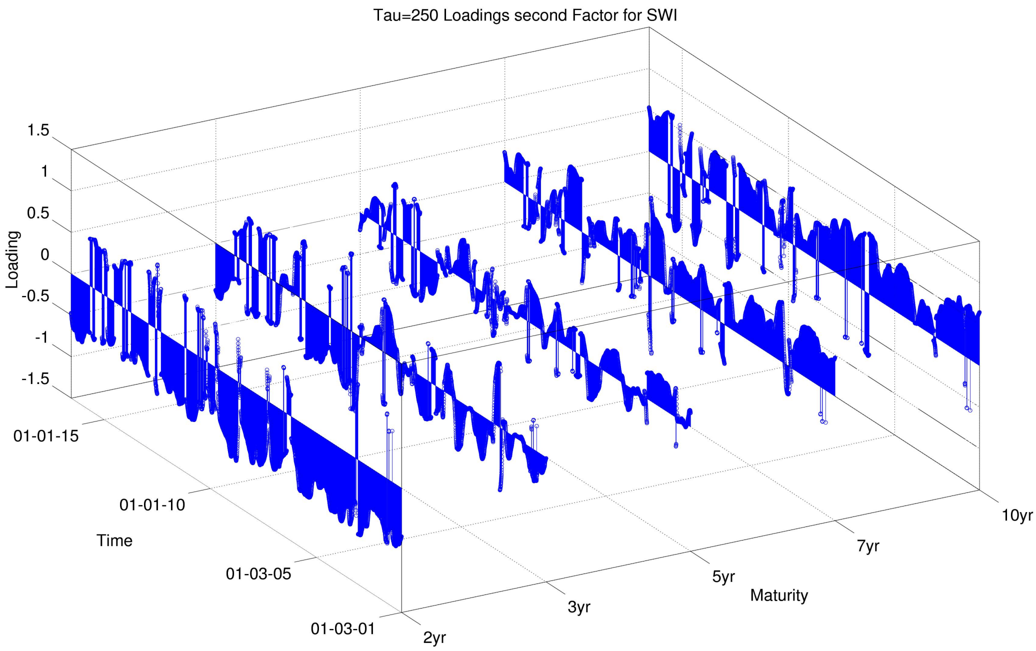


Figure 2.30: Evolution of the second factor loadings of SWI for $\tau = 250$.

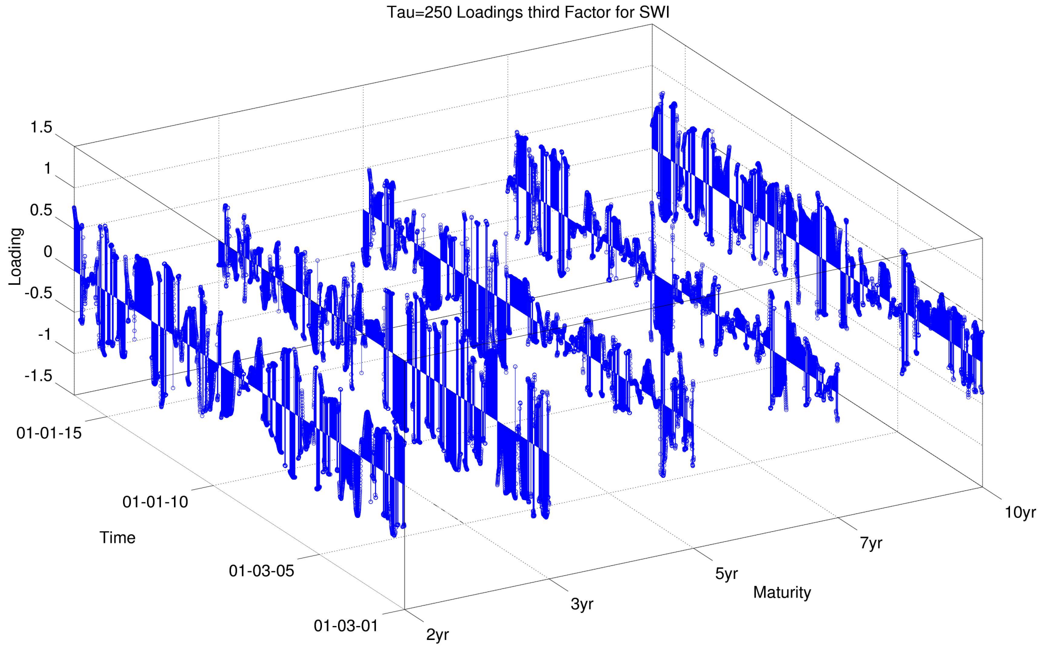


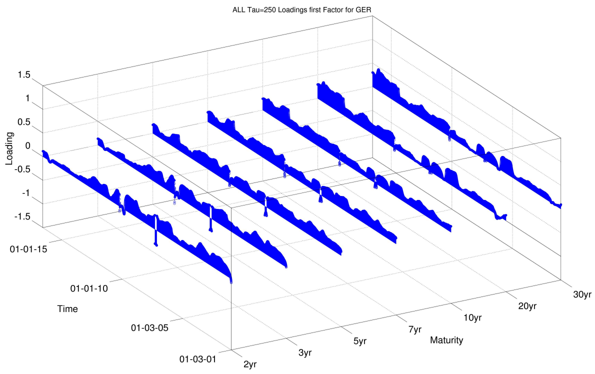
Figure 2.31: Evolution of the third factor loadings of SWI for $\tau = 250$.

all term structures we observe the loadings for each country separately in the individual panels of Figures 2.32, 2.33, and 2.34. All loadings for GER are positive and of similar value in the first factor as seen in Panel (a) of Figure 2.32. This holds for the US in Panel (b), the UK in Panel (c), and the SWI maturities in Panel (d). Therefore, the first factor can be interpreted as an overall level factor. The second and third factor displayed in Figures 2.33 and 2.34 respectively at different periods for different countries have country specific loadings that appear like slope or curvature loadings. Yet, these features are not consistent.

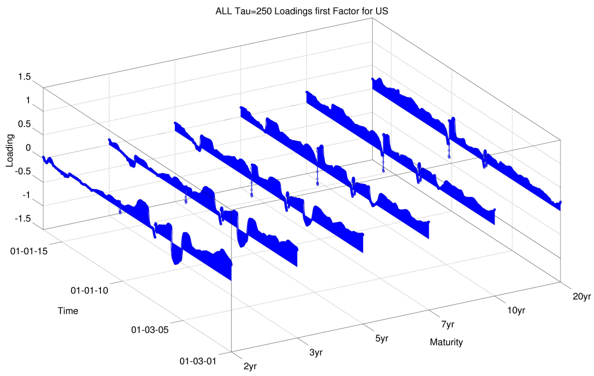
2.2.3 Forecasts

Accounting for alternative data sets, estimation methods, forecasting methods, time windows, number of factors, and number of lags allows for a large number of distinct forecast strategies. The relative impact on forecast performance can be determined by an ANOVA for the three alternative evaluation criteria. Dynamic factor model forecasts need to be compared with benchmark strategies, providing a comparison with a larger set of forecasting methods and a broader perspective.

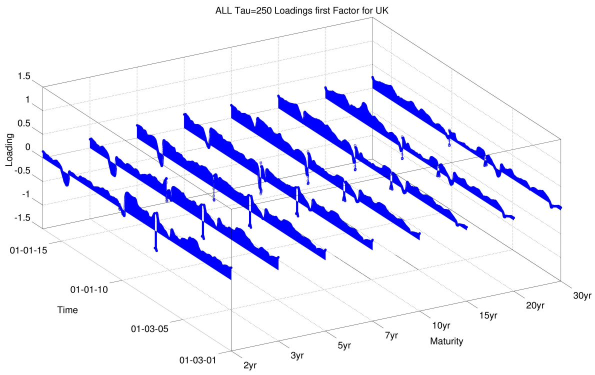
Blaskowitz & Herwartz (2011) tested 100 different PCA/AR strategy combinations which represent a much broader approach than the single strategy used by Diebold



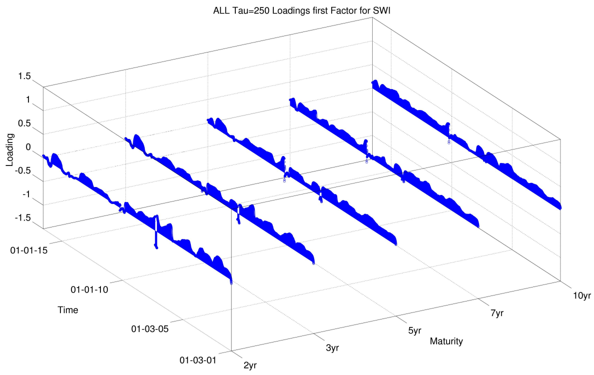
(a) GER



(b) US

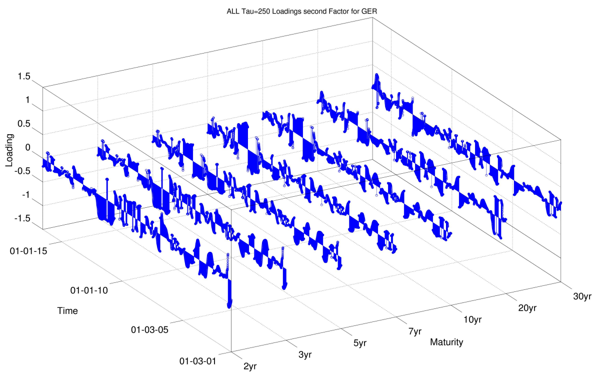


(c) UK

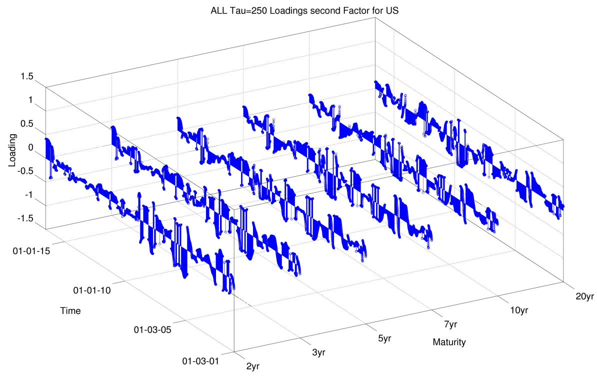


(d) SWI

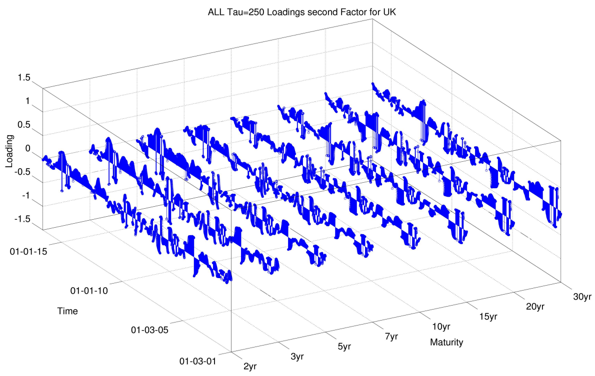
Figure 2.32: Evolution of the first factor loadings of all term structures for $\tau = 250$. Panel (a) displays the factor loadings for the German maturities, Panel (b) displays those for the US maturities, Panel (c) those for the UK maturities, and Panel (d) for the Swiss maturities.



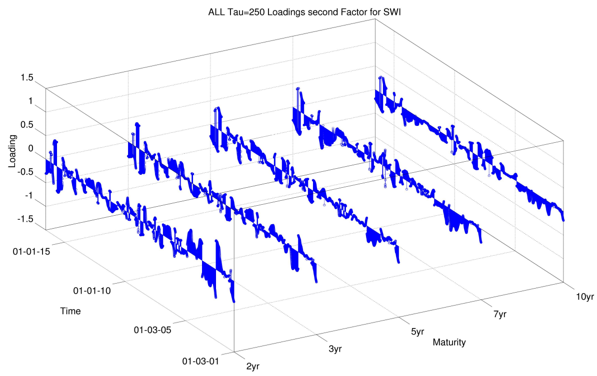
(a) GER



(b) US

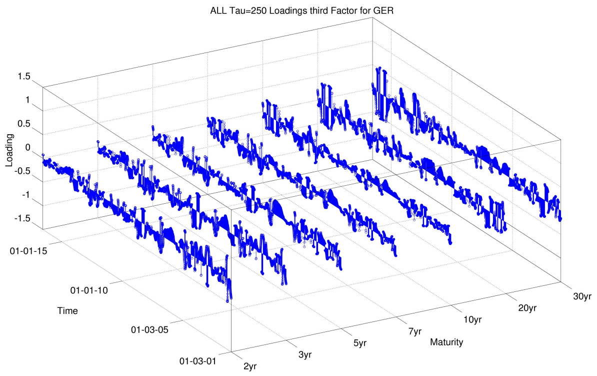


(c) UK

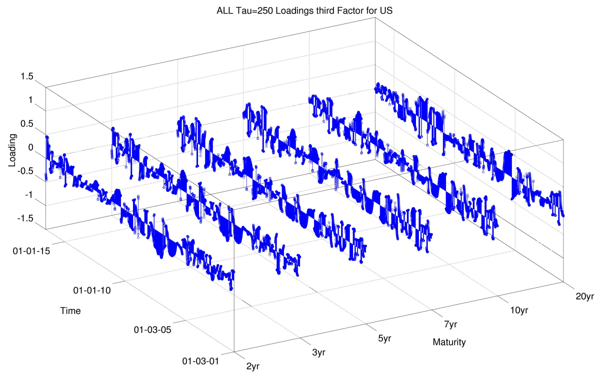


(d) SWI

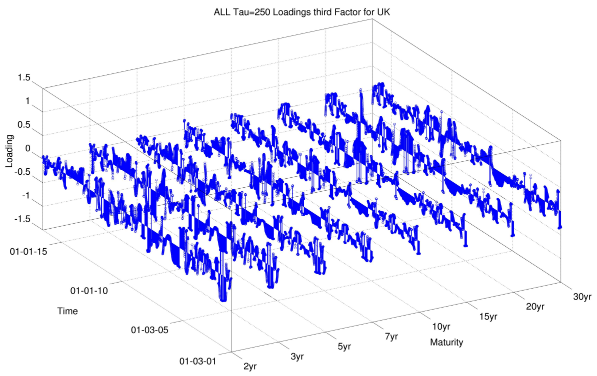
Figure 2.33: Evolution of the second factor loadings of all term structures for $\tau = 250$. For an explanation of the panels, see the notes under Fig 2.33.



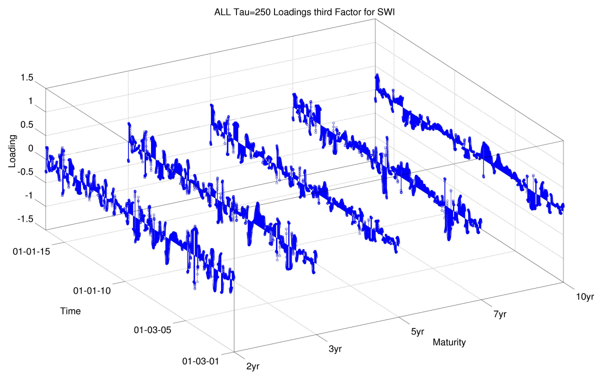
(a) GER



(b) US



(c) UK



(d) SWI

Figure 2.34: Evolution of the third factor loadings of all term structures for $\tau = 250$. For an explanation of the panels, see the notes under Fig 2.33.

& Li (2006)[6]. Blaskowitz & Herwartz (2011) show that an adaptive model selection strategy of PCA/AR can outperform the Nelson-Siegel AR approach, but adaptive strategies are beyond the scope of this thesis. With alternative predictor sets, estimation methods, and forecasting methods Matthies (2014) expands the set of strategies compared to Blaskowitz & Herwartz (2011) to 640. We focus mainly on these strategies, as Blaskowitz & Herwartz (2011), Diebold & Li (2006), and Yu & Zivot (2011) among others already include more comprehensive method comparisons. This chapter focuses on the evaluation of dynamic PCA approaches in term structure forecasts and employs varying numbers of forecasting strategies based on the specific research question.

Benchmark strategies Simple forecasting methods are easily implemented and intuitive. They establish benchmarks that more sophisticated methods need to outperform. For this purpose, interest rates are forecasted with the random walk method (RW) $\hat{y}_{m,s+h|s} = y_{m,s}$. Furthermore each individual interest rate is forecasted with multiple $AR(p_{ar})$ processes $\hat{y}_{m,s+1} = \hat{\varphi}_{0,m} + \hat{\varphi}_{1,m}\tilde{y}_{m,s} + \ldots + \hat{\varphi}_{p_{ar},m}\tilde{y}_{m,s-p_{ar}}$, where $\hat{\varphi}_{0,m}$ is an intercept term and $\hat{\varphi}_{1,m}, \ldots, \hat{\varphi}_{p_{ar},m}$ are autoregressive parameters estimated from five rolling time windows of $\tau = 42, 63, 126, 189, 252$ and four different lags $p_{ar} = 0, 1, 2, 3$. Specifically, the random walk yields a benchmark statistic for MSFE. As the random walk predicts no change, it by definition has no score for DA and BHA. Therefore, the DA and BHA results obtained from dynamic factor specifications are compared with the respective results for the respective best $AR(p_{ar})$ strategy as the benchmark.

Forecast strategies based on dynamic factor models Forecasting strategies Ψ_i^h are defined as: $\Psi_i^h = \Psi^h(macro, GLS, SW, \tau, r, q)$, where $h \in \{1, 5, 10, 15\}$ is the forecast horizon and $i = 1, \ldots, 640$. $macro$, GLS, and SW are binary variables. The binary variable $macro$ indicates if a strategy only exploits a single term structure of respective interest rates — i.e. the first data set for $macro = 0$ — or uses all considered term structures in the second data set ($macro = 1$). GLS determines whether factors and factor loadings are estimated by means of the PC-GLS approach ($GLS = 1$), or whether the forecasts rely on the standard PC-OLS estimator ($GLS = 0$). The PC-GLS lag order is $p = 1$ throughout. The third binary variable SW selects the forecasting method. In the case of $SW = 0$ forecasts are implemented according to the BH method in Eq. (2.8), otherwise ($SW = 1$) forecasts are determined by means of the SW method Eq. (2.6). Furthermore, forecast strategies are characterised by alternative time window sizes $\tau \in \{42, 63, 126, 189, 252\}$, the number of factors $r \in \{1, 2, 3, 4\}$,

[6]Using only one PCA/AR strategy, which was outperformed by the Nelson-Siegel AR Diebold & Li (2006) approach under only one forecast criterion, Diebold & Li (2006) conclude that a Nelson-Siegel approach has an empirical advantage against a PCA/AR approach.

and the lag order $q \in \{0, 1, 2, 3\}$. Altogether the implementations can yield up to $(2 \times 2 \times 2 \times 5 \times 4 \times 4) = 640$ alternative strategies for each forecast horizon h.

In order to determine the average effect of the strategy defining characteristics, ANOVAs with 7 dummy variables are implemented[7]. The benchmark strategy in each case is $(0, 0, 0, 42, 1, 0)$. In the ANOVA a dummy variable is attached for each of the three binary characteristics and for the four time window sizes $\tau = 63, 126, 189, 252$.

In contrast, the average effect of the factors, time windows, and lags are estimated via ANOVAs with 11 dummy variables. The benchmark strategy in this case is $(42, 1, 0)$. In these ANOVAs a dummy variable is attached for each of the four time window sizes $\tau = 63, 126, 189, 252$, each of the four factors $R = 2, 3, 4, 5$, and for each of the three lags $q = 1, 2, 3$.

In order to keep forecast statistics comparable, we eliminate unreliable forecasts. In particular, these might occur at higher forecast horizons when using small windows of sample information. If a model forecast of any interest rate in time $s + h$ differs more than 10 percentage points from the value in period s[8], the forecast is ignored and the random walk forecast is taken instead $y_{m,s+h|s} = y_{m,s}$.

2.3 Results

Evaluation of the forecast results is done for the four term structures and with respect to the three distinct evaluation criteria. First the DFM forecast performance is compared to the benchmark strategies. Then different estimation time window sizes, factor numbers, and lag selections are compared for a specific estimation-forecasting method combination. Finally, the alternative data-sets, estimation methods, and forecasting methods are compared.

2.3.1 Benchmark Strategies

In order to put forecasts of the term structure by factor-based methods into a larger perspective, we compare their results with those obtained from simpler forecasting methods. For this purpose a comparison of factor-based methods with a RW and the best performing AR model estimated from moving time windows is done. The results show the fraction of factor-based forecasting strategies that can outperform the RW and the respective best AR method under the MSFE, DA, and BHA criteria. We now discuss the results for GER, the US, the UK, and SWI successively.

[7] In each ANOVA there is also a universal model constant.

[8] This occurs very rarely (less than 10 times for any given strategy), and mainly using the BH forecasting method if $\hat{\phi} > 1$, sometimes leading to forecasts of $y_{m,s+h} > 1000$. Inclusion of these obviously unbelievable forecasts would highly skew the results.

GER										
h	2	3	5	7	10	20	30	Lev	Slo	Cur
					RWMSFE					
1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
					ARMSFE					
1	0	0	0	0	0	0	0	0	0	0
5	0.05	0.0328	0	0.0266	0.0266	0.0078	0.0672	0.0375	0.0016	0
10	0.0719	0.0969	0.0328	0.0516	0.0859	0.0547	0.1297	0.0594	0.0625	0
15	0.0766	0.1141	0.0688	0.0563	0.1219	0.125	0.1969	0.0672	0.075	0.0063
					ARDA					
1	0.05	0.1031	0.3781	0.3031	0.2734	0.0125	0.0922	0.3375	0	0.7031
5	0.2078	0.3531	0.6172	0.5672	0.5875	0.2922	0.275	0.4078	0.0437	0.3563
10	0.3016	0.425	0.4953	0.4859	0.5094	0.3953	0.4344	0.425	0.0625	0.3328
15	0.3641	0.4156	0.5047	0.4219	0.5484	0.325	0.5172	0.4031	0.2562	0.3
					ARBHA					
1	0.1437	0.1125	0.1766	0.3438	0.3016	0.0266	0.0406	0.3141	0.0219	0.4109
5	0.3516	0.3828	0.6172	0.3531	0.3219	0.3156	0.4766	0.2719	0.0719	0.2484
10	0.3422	0.4	0.5188	0.3703	0.3906	0.4063	0.4969	0.2656	0.2	0.2078
15	0.3703	0.3812	0.4922	0.4297	0.4641	0.4047	0.5922	0.2656	0.2188	0.2062

Table 2.2: Fraction of forecast strategies based on factor analysis outperforming naïve forecast strategies under the MSFE, DA, and BHA criteria for GER. Rows 2–5: forecast horizon $h = 1, 5, 10, 15$ comparison to RW w.r.t. MSFE (RWMSFE). Rows 6–9: forecast horizon $h = 1, 5, 10, 15$ comparison to the respective best AR-model w.r.t. MSFE (ARMSFE). Rows 10–13: forecast horizon $h = 1, 5, 10, 15$ comparison to the respective best AR-model w.r.t. DA (ARDA). Rows 14–17: forecast horizon $h = 1, 5, 10, 15$ comparison to the respective best AR-model w.r.t. BHA (ARBHA).

Germany The results in Tab 2.2 show that no DFM-strategy outperforms RW under MSFE. Neither does any strategy improve on the best AR-model under MSFE at the $h = 1$ forecast horizon. At longer horizons often 5% and more of the DFM-strategies improve on the respective best AR-model. That percentage increases with with the length of the forecast horizon. Curvature is an exception. Here, only at the $h = 15$ horizon is the only instance where any DFM-strategies improve on the best AR-model under the MSFE criterion. These are 4 of 160 or 0.625%. Under the DA criterion for the $h = 1$ forecast horizon, slope is the only case where there are no improvements to the respective best AR-model. Furthermore, for $2yr$, $20yr$, and $30yr$ less than 10% of the strategies improve. For $5yr$, $7yr$, and level there are more than 30% and for curvature there are more than 70%. At the longer forecast horizons mostly over 20% of the strategies improve. Under BHA there are at least some strategies that outperform the respective best AR-model for each maturity or linear combination at every forecast horizon. For the maturities from $h = 5$ onwards mostly 30% plus strategies outperform the best AR-model. In contrast, for the three linear combinations level, slope, and curvature more than 20% outperform the best AR-model. At $h = 1$ the percentages are higher for level and curvature.

					US				
h	2	3	5	7	10	20	Lev	Slo	Cur
					RWMSFE				
1	0	0	0	0	0	0	0	0	0
5	0.0187	0	0	0	0	0	0	0.0094	0
10	0.0187	0	0	0	0	0	0	0.025	0
15	0.025	0	0	0	0	0	0.0047	0.0234	0
					ARMSFE				
1	0.0109	0	0	0	0	0.0047	0.0016	0	0
5	0.0813	0.0453	0.0125	0.0219	0.0813	0.0719	0.0797	0.1156	0.0047
10	0.1125	0.075	0.0688	0.0453	0.1125	0.0875	0.1125	0.1797	0.0969
15	0.1625	0.1125	0.0984	0.0516	0.0938	0.1016	0.1359	0.1594	0.2047
					ARDA				
1	0	0.2578	0.1031	0.1281	0.3391	0.6516	0.0313	0.0844	0.0422
5	0.0375	0.0844	0.2578	0.0344	0.1812	0.0016	0.1109	0.1844	0.0141
10	0.1219	0.1891	0.2562	0.0578	0.2422	0.0125	0.175	0.2437	0.0187
15	0.0766	0.2313	0.1141	0.0453	0.0609	0.0594	0.1047	0.3937	0
					ARBHA				
1	0.1391	0.0563	0.225	0.0406	0.2734	0.1812	0.1391	0.0891	0.0141
5	0.2047	0.5656	0.4625	0.0906	0.2188	0.0266	0.4313	0.2687	0.0078
10	0.1688	0.3594	0.3453	0.1	0.1703	0.0281	0.4281	0.3234	0.0125
15	0.1234	0.25	0.2938	0.1	0.0922	0.0766	0.3047	0.4406	0.0266

Table 2.3: Fraction of forecast strategies based on factor analysis outperforming naïve forecast strategies under the MSFE, DA, and BHA criteria for the US. For an explanation see Tab 2.2

US The results for the US are displayed in Tab 2.3. There are no strategies that can improve on RW under the MSFE criterion at the $h = 1$ forecast horizon. Neither for the $3yr$, $5yr$, $2yr$, $10yr$, and $20yr$ maturities and curvature at $h = 5$. For the $2yr$ there are 12 (1.87%) strategies that improve on RW at the $h = 5$ and $h = 10$ forecast horizons. This increases to 16 (2.5%) at $h = 15$. For slope, 6 (0.94%), 16 (2.5%), and 15 (2.3%) improve at the $h = 5$, $h = 10$, and $h = 15$ forecast horizon respectively. The best AR-models perform well at the $h = 1$ horizon under MSFE. Only for $2yr$, $20yr$, and level are there some strategies that improve on the simple AR-models. For the $h = 5$, $h = 10$, and $h = 15$ forecast horizon mostly around 10% of the DFM-strategies outperform the respective best AR-models. With respect to DA, only for $2yr$ and curvature at $h = 15$ is AR not outperformed. In every other case there are at least some strategies that improve on the best AR-models. For BHA, there is no case in which there are no strategies that do not improve on the respective AR-model.

					UK						
h	2	3	5	7	10	15	20	30	Lev	Slo	Cur
					RWMSFE						
1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
					ARMSFE						
1	0	0	0	0	0	0	0	0	0	0	0
5	0.0141	0.0031	0	0.0391	0	0	0	0	0	0	0
10	0.0453	0.0547	0	0.1281	0.0016	0.0859	0.0187	0.0063	0.0063	0.0297	0
15	0.0625	0.0594	0.0078	0.2031	0.1	0.1109	0.0453	0.0063	0.0234	0.075	0.0125
					ARDA						
1	0.0063	0.1453	0.3672	0.5484	0.4172	0.3688	0.0641	0.0125	0.2094	0.0016	0
5	0.0781	0.1359	0.3375	0.6172	0.3766	0.4984	0.4375	0.0109	0.1844	0.0172	0.0891
10	0.0703	0.3187	0.4844	0.5344	0.2281	0.3688	0.2547	0	0.2609	0	0.0969
15	0.2234	0.3047	0.4063	0.5672	0.3453	0.2281	0.0609	0.0016	0.2297	0.0469	0.1187
					ARBHA						
1	0.0047	0.2281	0.2469	0.4562	0.3172	0.1922	0.0234	0.0094	0.2891	0.0016	0.0063
5	0.0953	0.3969	0.525	0.6719	0.1844	0.1641	0.0563	0	0.1812	0.0031	0.0344
10	0.2672	0.5031	0.4891	0.6547	0.2875	0.2031	0.0453	0	0.3266	0	0.0422
15	0.4703	0.5016	0.4969	0.6344	0.3578	0.3547	0.0156	0	0.3172	0.0422	0.0375

Table 2.4: Fraction of forecast strategies based on factor analysis outperforming naïve forecast strategies under the MSFE, DA, and BHA criteria for the UK. For an explanation see Tab 2.2

UK The results for the UK are displayed in Tab 2.4. Here, we see that no DFM-strategies improve on RW under the MSFE criterion. Furthermore, in comparison to the respective best AR-models all DFM-strategies are outperformed at $h = 1$ under MSFE. At the $h = 5$ forecast horizon for $2yr$, $3yr$, and $7yr$ maturities only some strategies improve on the respective best AR-model. In contrast, for $h = 10$ only $5yr$ and curvature are not improved. For $h = 15$ there are at least some strategies for every maturity and linear combination that improve on the respective best AR-forecast model. Across all maturities and linear combinations over all forecast horizons under DA there are only three cases where no DFM-strategy improves on the respective best AR-model. These are $30yr$ and slope at $h = 10$ and curvature at $h = 1$. For $2yr$, $30yr$, slope, and curvature the percentage of DFM-strategies that improve stays under 10% but can increase at longer horizons. For $3yr$, $5yr$, $7yr$, $10yr$, $15yr$, and $20yr$ 20% and more of the DFM-strategies improve on the best respective AR-model in most cases. Under BHA for the $30yr$ maturity at $h = 5$, $h = 10$, and $h = 15$ and slope at $h = 10$ there are no DFM-strategies that improve on the respective best AR-model. Furthermore, for $30yr$ at $h = 1$ only 1%, i.e. 6, strategies improve. For the $2yr$ to $15yr$ maturities and level double digit percentages of strategies improve on the best respective AR-model. The number often increases at larger forecast horizons. For $20yr$, $30yr$, slope, and curvature the numbers are lower.

				SWI				
h	2	3	5	7	10	Lev	Slo	Cur
				RWMSFE				
1	0.0828	0.0359	0	0	0	0.0016	0.0484	0.0141
5	0.0437	0	0	0	0	0	0.0391	0.0266
10	0	0	0	0	0	0	0	0.0141
15	0	0	0	0	0	0	0	0.0094
				ARMSFE				
1	0.0516	0.0359	0.0047	0.0031	0	0.0016	0.0266	0.0047
5	0.1563	0.1078	0.0563	0.0734	0.0313	0.0828	0.1313	0.0766
10	0.1703	0.1734	0.1625	0.1078	0.0766	0.2344	0.2625	0.1328
15	0.3141	0.3422	0.3969	0.1609	0.1016	0.3859	0.2437	0.0656
				ARDA				
1	0.1563	0.5219	0.5719	0.5938	0.5641	0.2859	0.4484	0.6906
5	0.3516	0.225	0.5734	0.9484	0.5984	0.5891	0.1047	0.4984
10	0.3484	0.0266	0.4078	0.7047	0.8422	0.5359	0.0813	0.1828
15	0.3047	0.0141	0.3609	0.5719	0.7141	0.5109	0.0547	0.1078
				ARBHA				
1	0.3859	0.8328	0.9703	0.7031	0.5469	0.2687	0.4875	0.7734
5	0.2625	0.5844	0.6578	0.7031	0.5719	0.5813	0.0422	0.4813
10	0.2125	0.3875	0.4	0.4766	0.7844	0.5078	0.0063	0.1016
15	0.1953	0.2141	0.3609	0.3937	0.7391	0.5281	0.0187	0.0781

Table 2.5: Fraction of forecast strategies based on factor analysis outperforming naïve forecast strategies under the MSFE, DA, and BHA criteria for SWI. For an explanation see Tab 2.2

Switzerland In Tab 2.5 we find the results for Switzerland. In contrast to GER and UK yield curves, there are some cases in which DFM-strategies improve on RW under MSFE. For $2yr$, $3yr$, level, slope, and curvature we find multiple instances where this is the case. In comparison to the respective best AR-model under MSFE, DA, and BHA there is only one case in which no DFM-strategies can improve, namely for the $10yr$ maturity at $h = 1$ under MSFE. Otherwise, there is a mostly consistent picture under MSFE. There are some DFM-strategies that improve at $h = 1$, from below 1% to 5%. Then, the number of improving strategies increases at longer forecast horizons. Under DA there are many cases in which more than 50% of DFM-strategies outperform the respective best AR-model. This also holds true for BHA.

Most importantly, a significant number of strategies can outperform the AR method, justifying the use of factor models under economic criteria. For GER no strategies are better than the simple random walk method under MSFE, as all values in the first three columns are 0. It is usual not to outperform the RW-assumption, in particular at small forecast horizons. For instance in Diebold & Li (2006), RW is strongly performing at the 1 month forecast horizon, which is the smallest in that study. Here, our largest forecast horizon is three weeks. Hence, the fact that none of our methods outperforms RW under the MSFE is neither surprising, unusual, or worrying.

Given that MSFE does not necessarily correlate with the economic criteria DA and BHA, we put a stronger focus on the latter. On average 10 to 20 percent of factor based methods can outperform AR under MSFE, with values increasing at larger forecast horizons. With regard to DA and BHA more factor based strategies outperform the AR model at lower forecast horizons. The results for SWI, the UK, and the US are similar.

In summary, factor methods have to be carefully selected if they are to perform better than naïve benchmark approaches. This highlights the importance of the ANOVA results below, as only specific dynamic factor model features are worth applying in forecasting exercises. Furthermore, strategies based on dynamic factor models are generally not able to improve on RW under MSFE.

2.3.2 Factors, Time Windows, and Lags

A main interest in designing factor models with moving time windows using AR processes to forecast interest rates is the average forecast performance of strategies employing specific time windows, and the number of factors and autoregressive lags.

Previously, Diebold & Li (2006) had assumed that three factors (level, slope, and curvature) govern the term structure of interest rates. The magnitude of eigenvalues can be used as a simple measure for the number of factors governing the term structure.

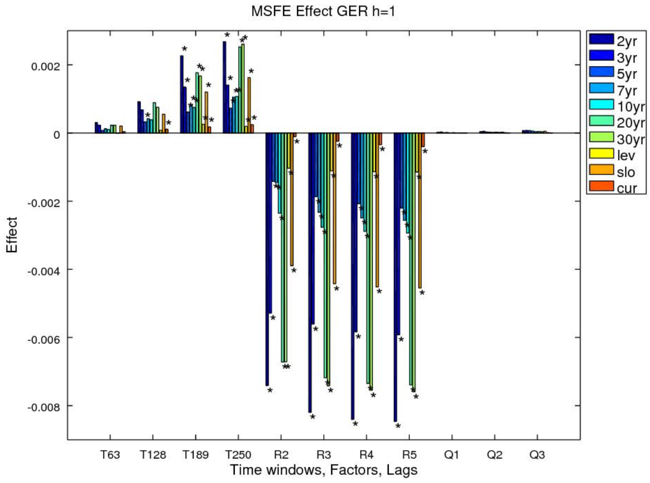
The Figures 2.9 and 2.10, 2.11 and 2.12, 2.13 and 2.14, and 2.15 and 2.16 in Section 2.2.2 show the development of eigenvalues for the term structure set of GER, the US, the UK, and SWI for two time windows of the size $\tau = 42$ and $\tau = 252$ respectively. The first eigenvalues tend to be of higher magnitude in larger time windows. In all cases two factors are sufficient to represent 90% of the variance, which could be seen as a criterion to determine the number of factors. Employing three penalty functions developed by Bai & Ng (2002) suggests that there are mostly one or two factors in the $\tau = 42$ time windows. The number of estimated factors increases with larger time windows. In the case of $\tau = 250$ the penalty functions often suggest the maximum number factors, thereby including factors that provide little additional information. This would imply that no meaningful reduction of dimensionality is possible in larger time windows. Bai & Ng (2002) point out that their criteria are not suitable for dynamic factor models. This finding further motivates the estimation of dynamic factors. As the method of Bai & Ng (2002) can yield results which are inconsistent with what can be observed with the naked eye in Figures 2.9 to 2.16 forecast results gain in importance to determine the number of factors governing the term structure of interest rates.

For this purpose we conduct an analysis with strategies that are only based on the single term structure data, PC-OLS estimation, BH forecasting combination. Moreover, 5 different time windows $\{42, 63, 125, 189, 250\}$, 5 distinct factor numbers $\{1, 2, 3, 4, 5\}$, and 4 lags $\{0, 1, 2, 3\}$ are employed. The base strategy in the ANOVA is $\tau = 42$, $r = 1$, and $q = 0$, i.e. $\{42, 1, 0\}$. We now discuss the results with respect to MSFE, DA, and BHA for each country successively.

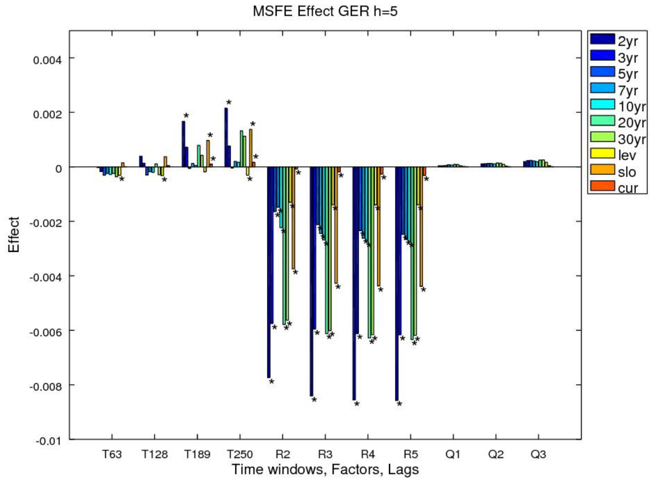
MSFE For GER on the $h = 1$ forecast horizon seen in Panel (a) of Fig. 2.35, we find that more factors improve forecasts, while strategies that use larger time windows are outperformed, and additional lags produce no significant results based on the MSFE criterion. The results for $h = 5$ in Panel (b) are similar. For $h = 10$ in (c) and $h = 15$ in (d) larger time windows improve forecasts and models with additional lags are significantly outperformed.

Forecasting the US yield curve at $h = 1$, as seen in Panel (a) of Fig. 2.36 we note that the use of smaller time windows is preferred. Employing more factors improves forecasts while additional lags produce no results. At $h = 5$ forecast horizon in (b) results are mixed for larger time windows while more factors improve and additional lags create no clear results. At the $h = 10$ horizon in (c) larger time windows and additional factors both improve and strategies with $q = 3$ lags are outperformed. For $h = 15$ in (d) larger time windows and more factors improve. Yet in contrast to $h = 10$ the effects for larger time windows are now stronger.

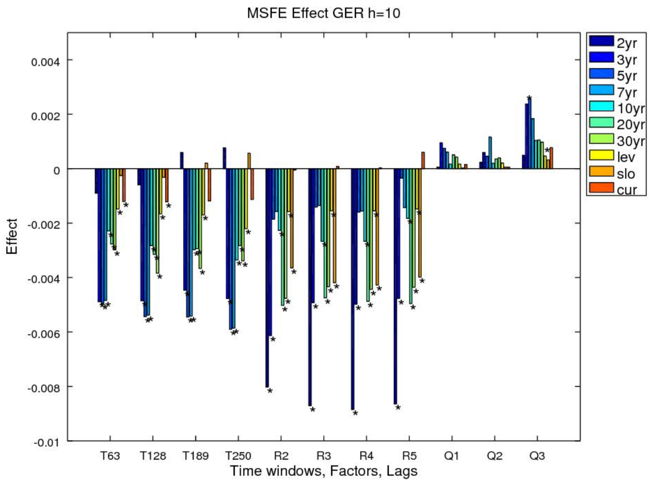
In the case of the UK term structure seen in Fig. 2.37 at a $h = 1$ forecast horizon



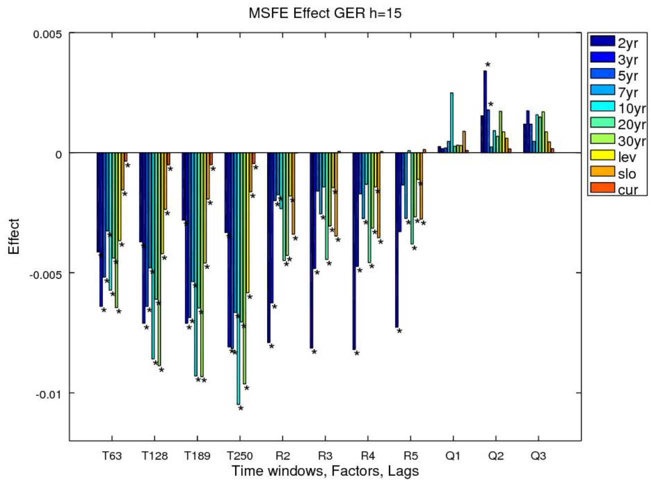
(a) $h = 1$



(b) $h = 5$

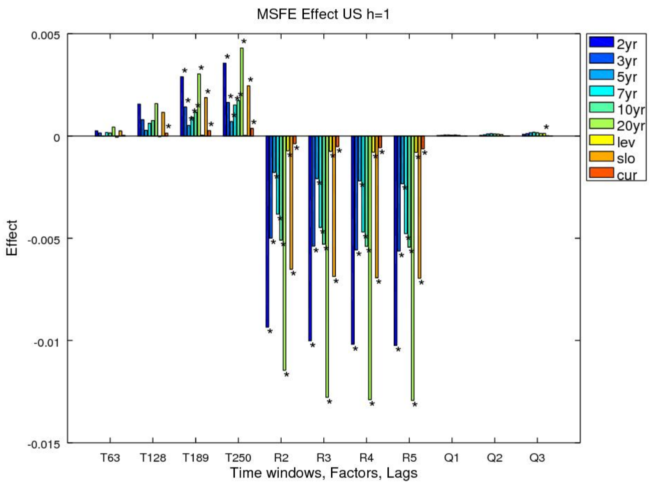


(c) $h = 10$

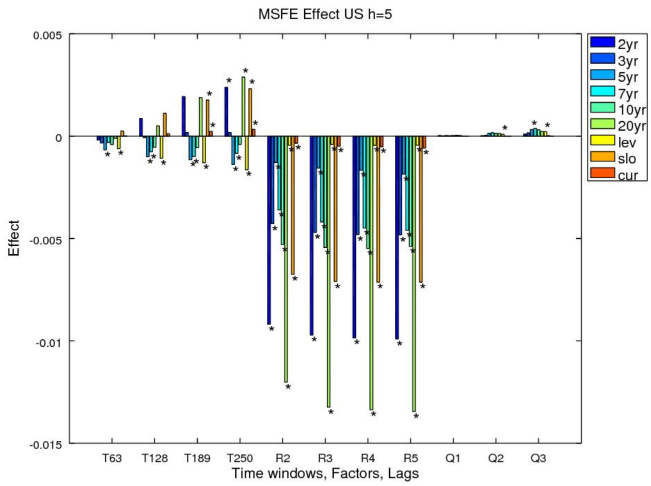


(d) $h = 15$

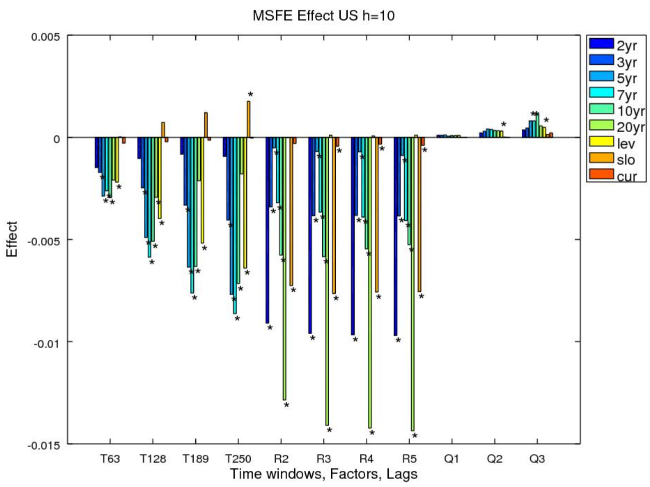
Figure 2.35: MSFE effects for Germany. Panel (a) displays the results for $h = 1$, Panel (b) for $h = 5$, Panel (c) for $h = 10$, and Panel (d) for $h = 15$.



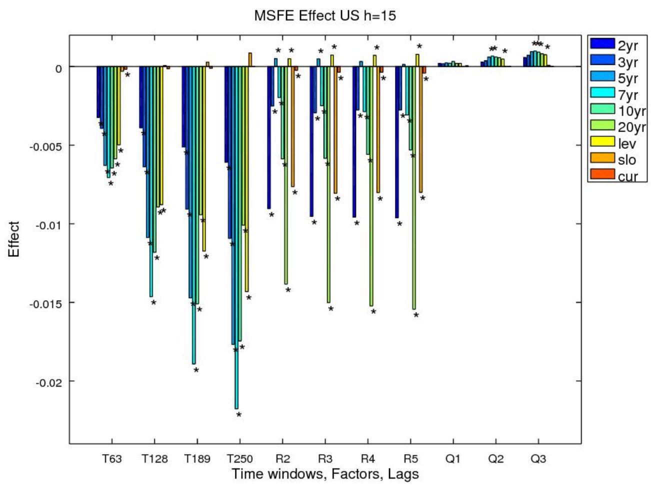
(a) $h = 1$



(b) $h = 5$

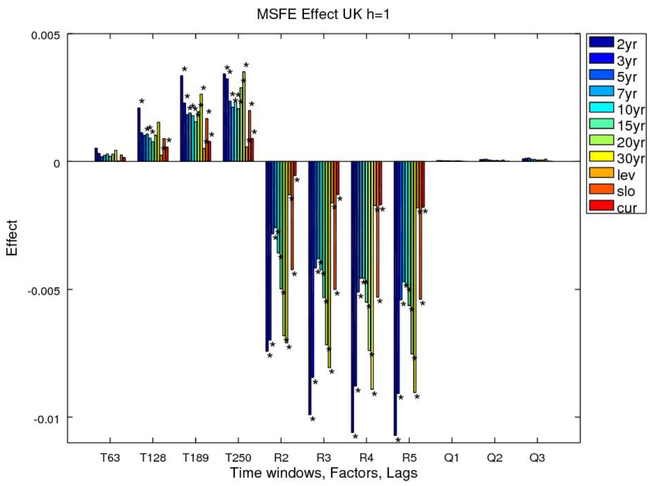


(c) $h = 10$

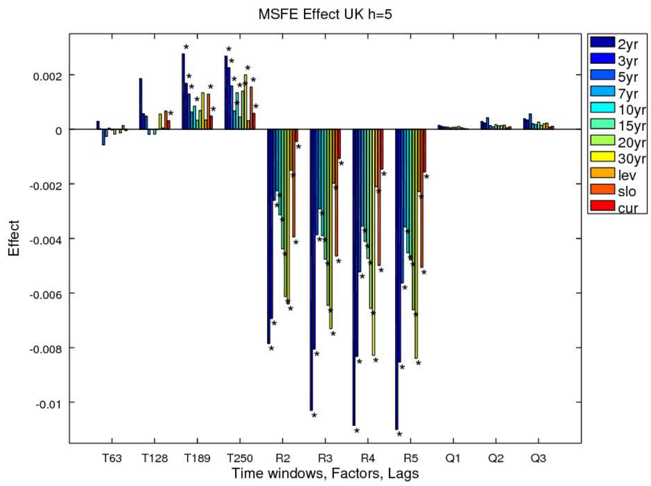


(d) $h = 15$

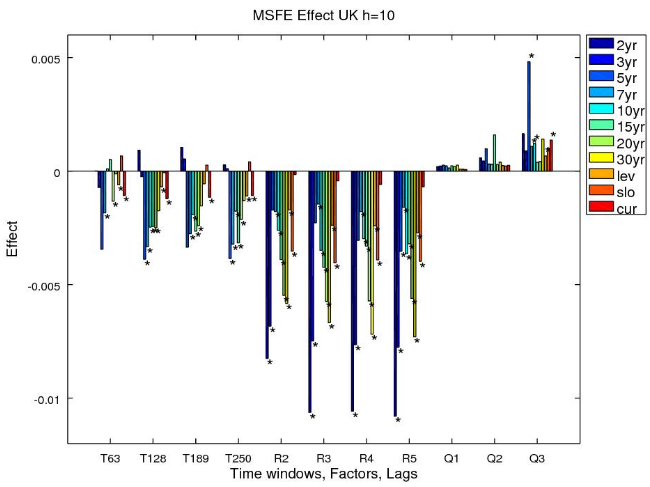
Figure 2.36: MSFE effects for the US. For an explanation of the panels, see the notes under Fig. 2.35.



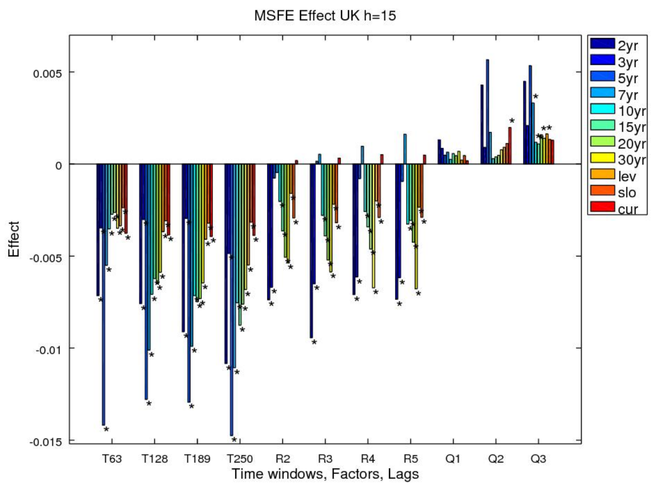
(a) $h = 1$



(b) $h = 5$

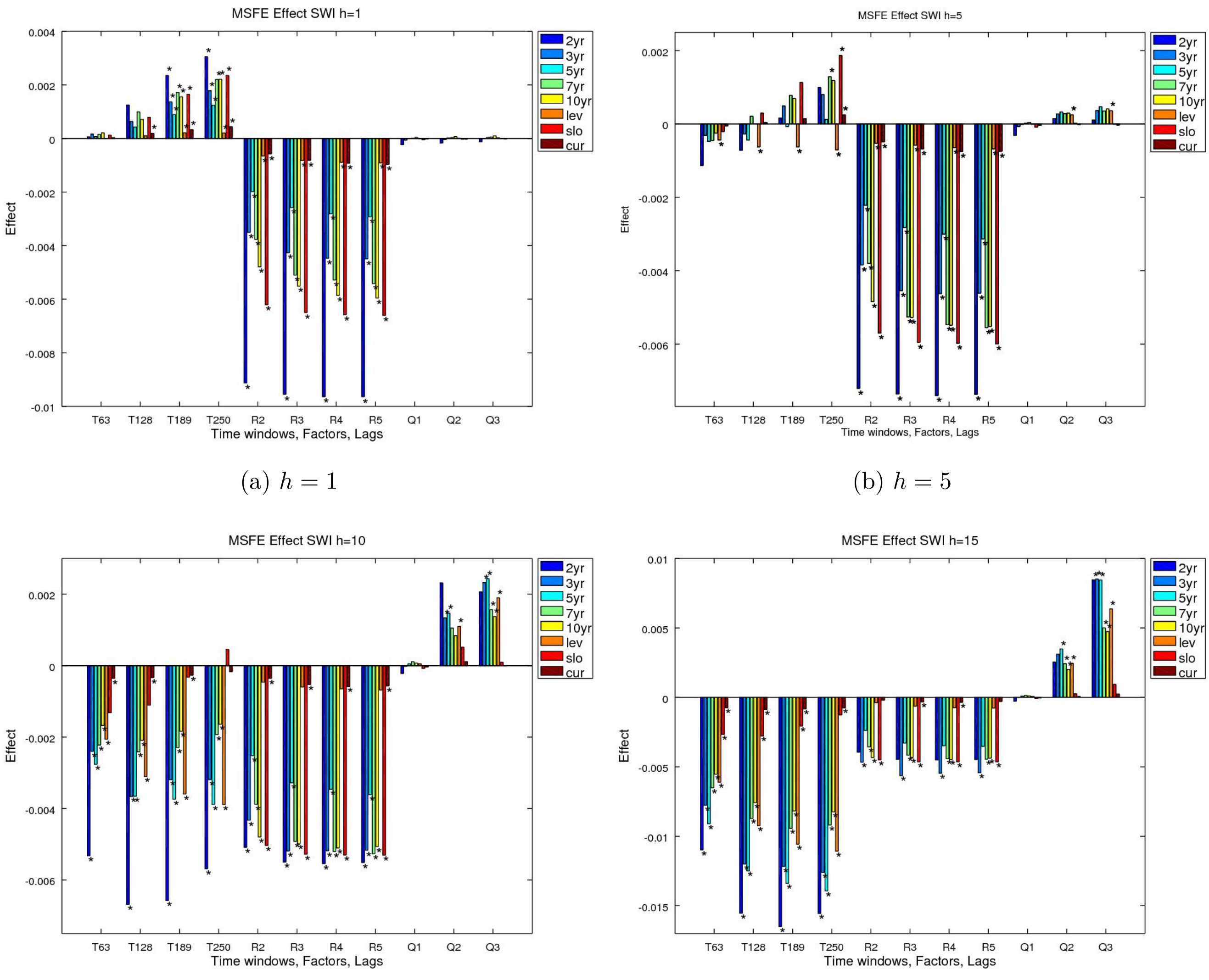


(c) $h = 10$



(d) $h = 15$

Figure 2.37: MSFE effects for the UK. For an explanation of the panels, see the notes under Fig. 2.35.



(a) $h = 1$

(b) $h = 5$

(c) $h = 10$

(d) $h = 15$

Figure 2.38: MSFE effects for Switzerland. For an explanation of the panels, see the notes under Fig. 2.35.

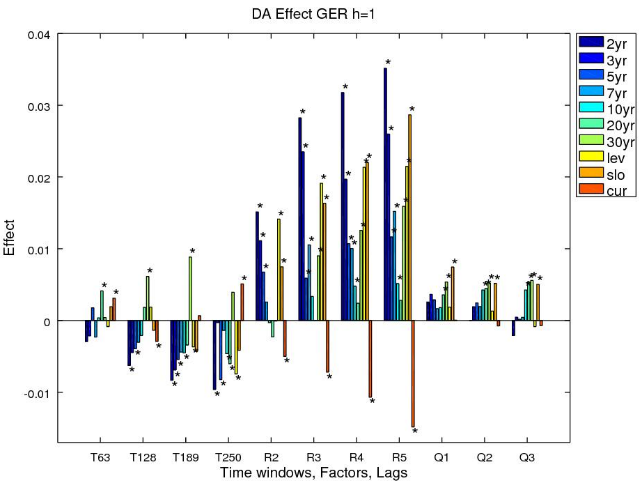
strategies with time windows larger than $\tau = 63$ are statistically outperformed. Using more factors improves forecasts while additional lags yield no significant results. At $h = 5$ strategies with time windows larger than $\tau = 125$ are significantly outperformed. Results for more factors and lags are like those for $h = 1$. At $h = 10$ results for time windows change. Time windows larger than $\tau = 63$ significantly improve forecasts. Yet the effects are not as strong as those of additional factors. This changes at the $h = 15$ forecast horizon. Using a time window larger than $\tau = 42$ improves forecasts. And the effects are stronger than those of additional factors. The coefficients of additional factors remain similar in size over all forecast horizons. Strategies with more than 2 lags are significantly outperformed for the $h = 10$ and $h = 15$ forecast horizons.

For SWI in Fig. 2.38 we find a similar picture as for the other three countries. Strategies with larger time windows are outperformed at $h = 1$, produce mixed results for $h = 5$, and improve forecasts for $h = 10$ and $h = 15$. More factors improve forecasts across all forecast horizons. Additional lags are outperformed at longer forecast horizons.

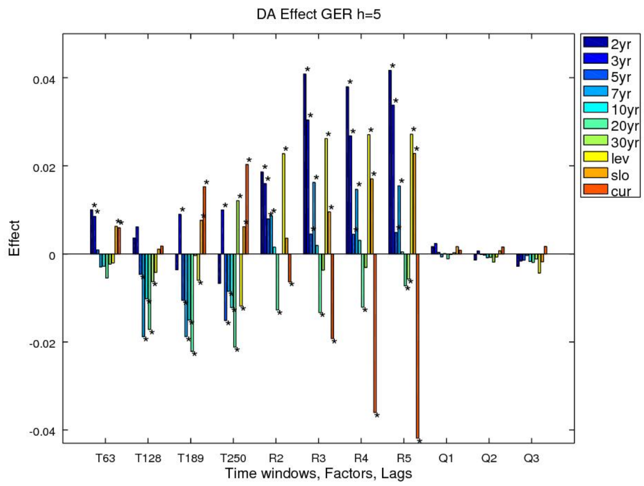
For the purpose of interpretation one should remember that MSFE values are squared percentage points. Moreover, it is noteworthy to highlight what effect time window size, factor number, and lag number selection can have on forecast performance across forecast horizons. Average MSFE values of strategies increase with larger forecast horizons. For GER for example MSFE values may range around 0.002 at $h = 1$, are just above 0.01 at $h = 5$, between 0.02 and 0.025 at $h = 10$, and around 0.035 at $h = 15$. The parameters of the ANOVA constants are around 0.005 at $h = 1$, 0.015 at $h = 5$, above 0.025 at $h = 10$, and above 0.04 at $h = 15$. Sign and size of the effects of time windows change over forecast horizons from 0.003 at $h = 1$ to 0.01 at $h = 15$. The effects of additional factors remains constant over the forecast horizons at 0.008. Furthermore, the effects of using autoregressive factor lags moves from < 0.0001 to 0.003. Regarding the choice of factors, we find that in additional ANOVA regressions, where the base strategy uses two factors, the effect of additional factors are mostly not significant under MSFE.

DA For GER we can see in Fig. 2.39 at the $h = 1$ forecast horizon larger time windows have significantly negative effects, i.e they produce forecasts which are outperformed under the DA criterion. Adding more factors significantly improves forecasts across all maturities and linear combinations. Here, curvature is the single notable exception. Using autoregressive lags improves forecasts. Strategies with larger time windows are still outperformed at $h = 5$ although results are more mixed. Additional factors improve forecasts, with the exception of the curvature and longer maturities. There are no relevant results for using lags. Using a $\tau = 63$ time window improves forecasts significantly at the $h = 10$ forecast horizon. In contrast, strategies that employ larger time windows are significantly outperformed. Employing more factors and lags produces results like those for $h = 5$. These features are also observed for $h = 15$.

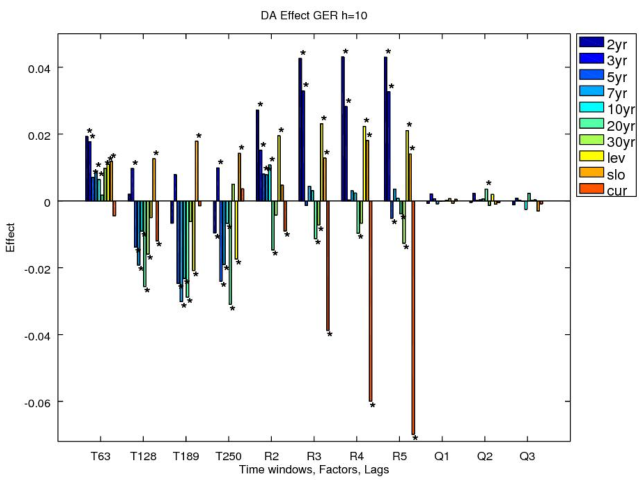
For the US seen in Fig. 2.40 at the $h = 1$ horizon larger time windows produce mixed results but give improvements for some maturities. More factors mostly improve forecasts. Adding lags produces mixed results. At $h = 5$ time windows of $\tau = 189$ and $\tau = 250$ improve forecasts. The effects for additional factors diverge. They improve $2yr$, $20yr$, and slope but are outperformed for $5yr$, $7yr$, level, and curvature. Using autoregressive lags produces no significant results. For $h = 10$ and $h = 15$ strategies with larger time windows are mostly outperformed. Here, $\tau = 250$ is the notable



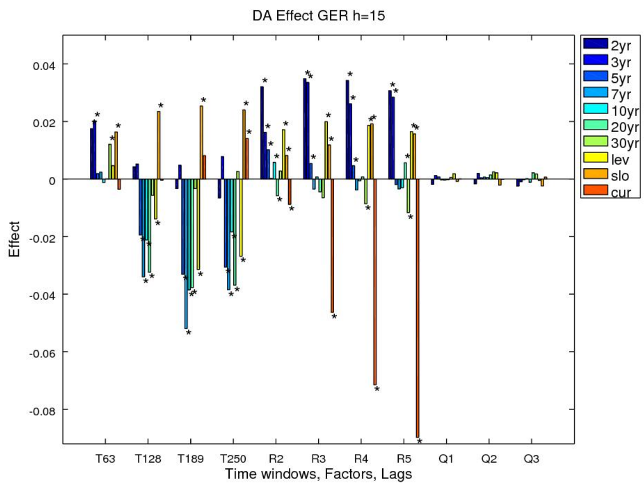
(a) $h = 1$



(b) $h = 5$

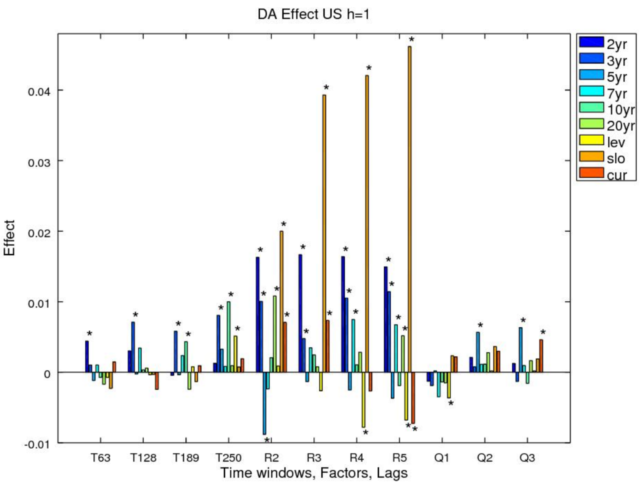


(c) $h = 10$

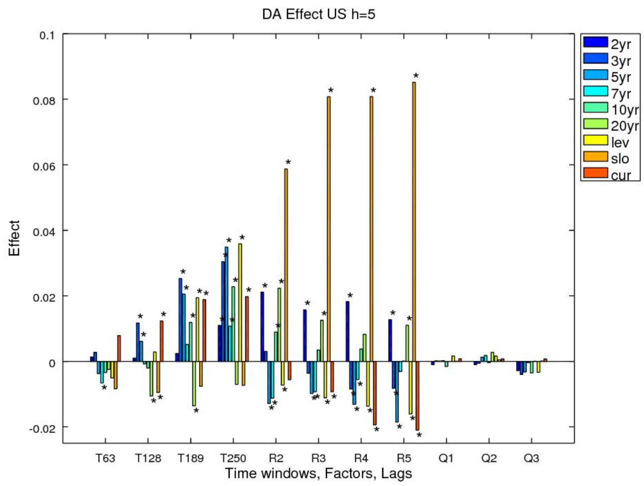


(d) $h = 15$

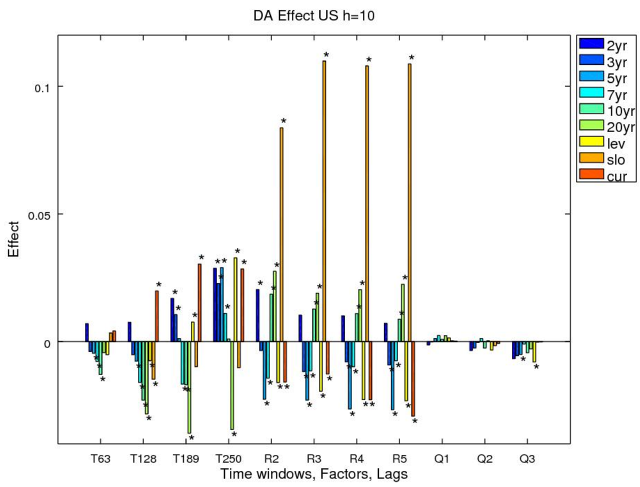
Figure 2.39: DA effects for Germany. For an explanation of the panels, see the notes under Fig. 2.35.



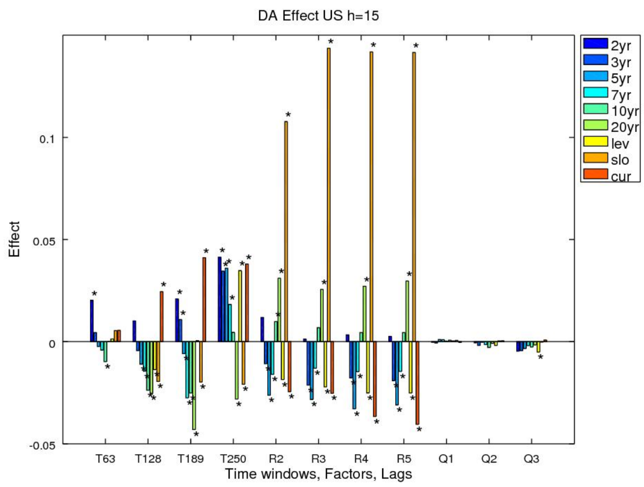
(a) $h = 1$



(b) $h = 5$

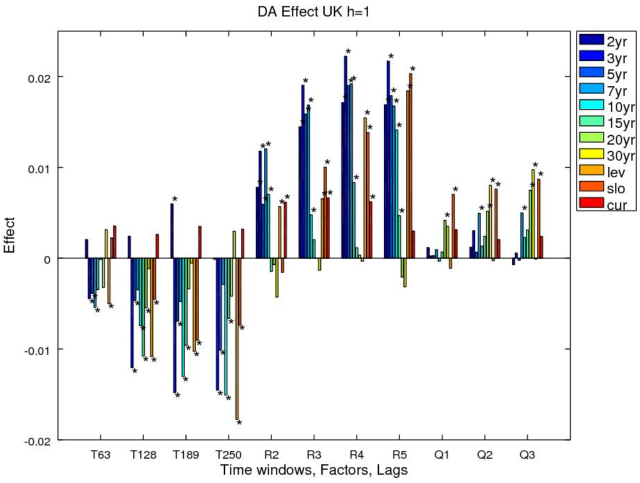


(c) $h = 10$

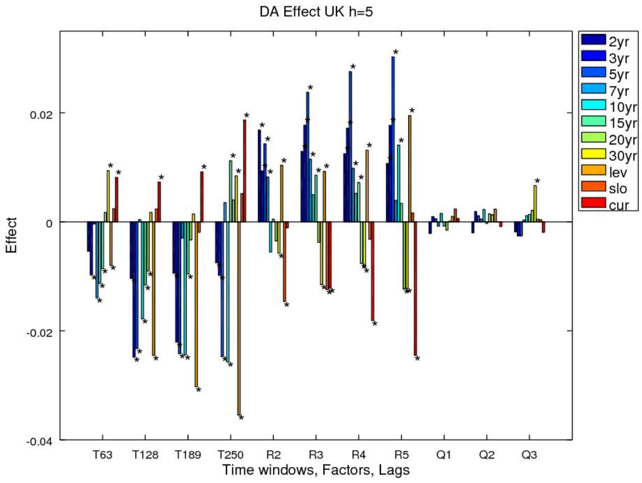


(d) $h = 15$

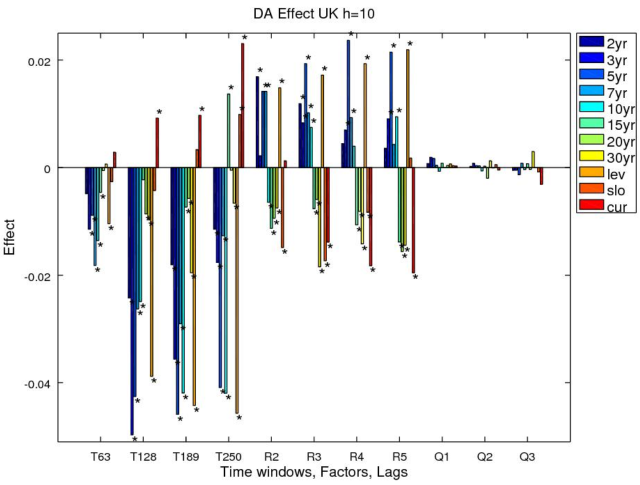
Figure 2.40: DA effects for the US. For an explanation of the panels, see the notes under Fig. 2.35.



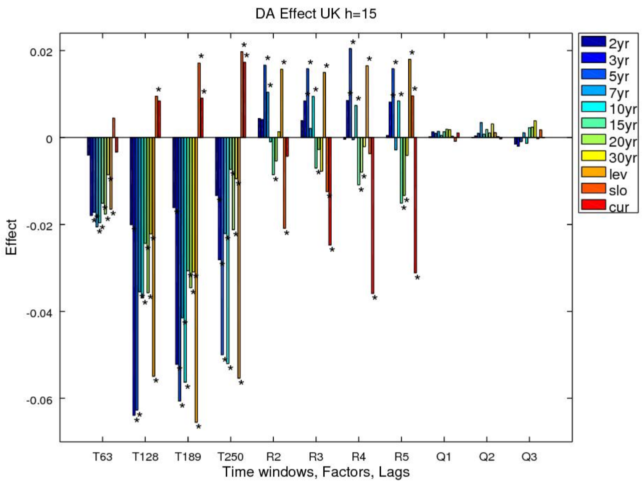
(a) $h = 1$



(b) $h = 5$



(c) $h = 10$



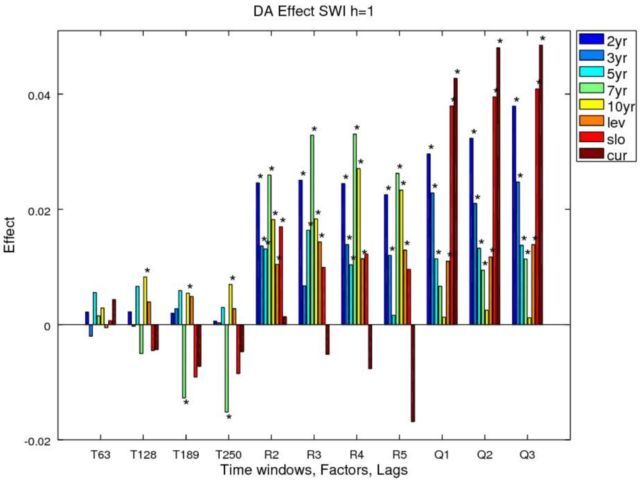
(d) $h = 15$

Figure 2.41: DA effects for the UK. For an explanation of the panels, see the notes under Fig 2.35.

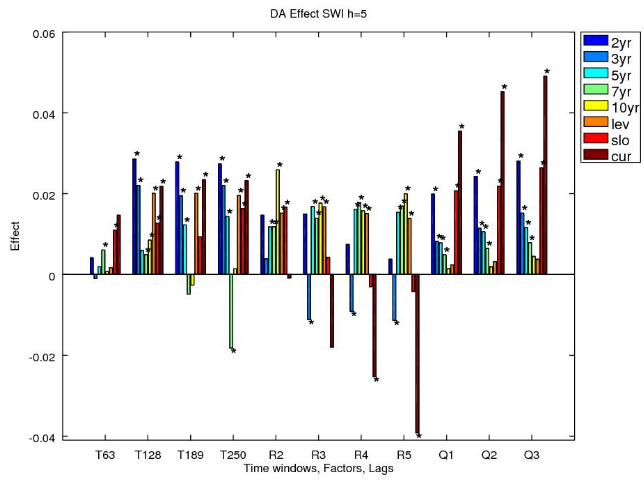
exception as it significantly improves forecasts. Effects of additional factors and lags are like those for $h = 5$.

Forecasts from larger time windows are outperformed while additional factors and lags improve forecasts for the UK at the $h = 1$ forecast horizon as seen in Fig. 2.41. At $h = 5$ strategies that employ larger time windows are mostly outperformed with some exceptions. The effects for more factors diverge. They improve for shorter maturities and level and are outperformed for $20yr$, $30yr$, and slope. There are no significant effects for adding lags. These observations hold for $h = 10$ and $h = 15$.

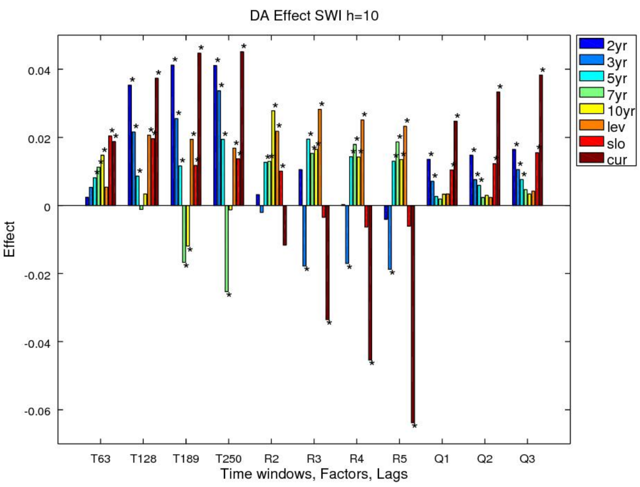
Larger time windows produce mixed results while additional factors and employing autoregressive lags are all beneficial to forecasts for SWI at the $h = 1$ horizon seen



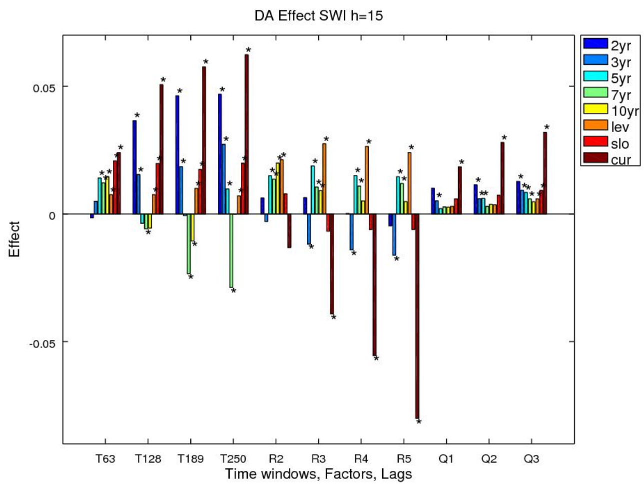
(a) $h = 1$



(b) $h = 5$



(c) $h = 10$



(d) $h = 15$

Figure 2.42: DA effects for Switzerland. For an explanation of the panels, see the notes under Fig 2.35.

in Fig. 2.42. For $h = 5$ larger time windows are beneficial as are more factors and lags with few exceptions. The $3yr$ maturity and curvature being the exception for additional factors. This holds for $h = 10$ and $h = 15$.

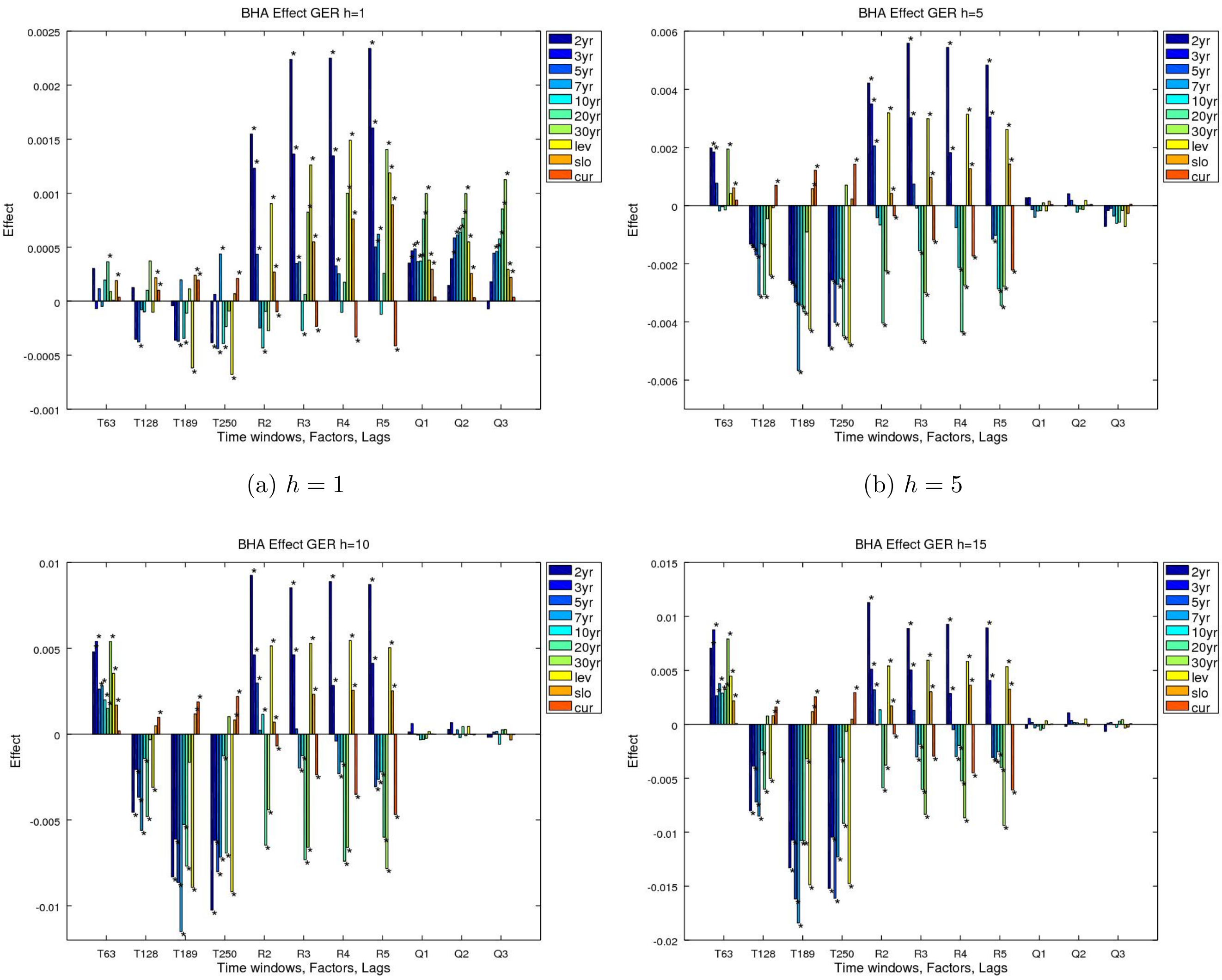
DA values are in fractions. The values for GER are around 0.5 or 50% at $h = 1$, at 0.52 and higher for $h = 5$, at 0.54 for $h = 10$, and around 0.55 for $h = 15$. DA values should not logically increase at larger forecast horizons like MSFE and BHA do. Parameter values of the ANOVA constants for $h = 1$ are below 0.5, around 0.5 for $h = 5$, and above 0.5 for $h = 10$ and $h = 15$. ANOVA effects for time windows, additional factors, and further lags range across forecast horizons from 0.01 to 0.04, 0.03 to 0.4, and from 0.01 to below 0.001 respectively in the case of GER. If two factors are chosen as the base strategy, the effects of additional factors are often not significant.

BHA The results under the BHA criterion for GER are shown in Fig. 2.43. At the $h = 1$ forecast horizon time windows produce mixed results. Additional factors improve forecasts for short maturities and level and slope but are outperformed for longer maturities and curvature. Employing autoregressive lags significantly improves forecasts. For $h = 5$ the $\tau = 63$ time window improves forecasts. Strategies with time windows larger than $\tau = 63$ are significantly outperformed by those from a $\tau = 42$ time window. Additional factors are beneficial for shorter maturities and level and slope but are outperformed for longer maturities and curvature. The results diverge as they did for $h = 1$. Additional lags produce no statistically significant results. These features hold for $h = 10$ and $h = 15$.

For the US at $h = 1$ time windows of $\tau = 189$ and $\tau = 250$ improve results as seen in Fig. 2.44. Additional factors improve for $2yr$, $3yr$, $10yr$, $20yr$, and slope but are outperformed for $7yr$ and level. Using two ($q = 2$) or three ($q = 3$) autoregressive lags improves forecasts. For $h = 5$ the $\tau = 250$ time window improves forecasts. Adding more factors improves forecasts for $2yr$, $10yr$, $20yr$, and slope but are outperformed for $5yr$, $7yr$, level, and curvature. Additional lags produce no significant results. For $h = 10$ and $h = 15$ strategies from the $\tau = 125$ and $\tau = 189$ time windows are outperformed. The other effects are like those of $h = 5$.

Results for the UK as seen in Fig. 2.45 show that time windows are mostly outperformed at $h = 1$. In contrast, additional factors and lags improve forecasts. At $h = 5$ forecasts from larger time windows are outperformed. Additional factors improve for $2yr$, $3yr$, $5yr$, and level and are outperformed for $10yr$, $15yr$, $20yr$, $30yr$, slope, and curvature. Using autoregressive lags produces no results. These features hold for $h = 10$ and $h = 15$.

For SWI at the $h = 1$ forecast horizon we observe no statistically significant results,



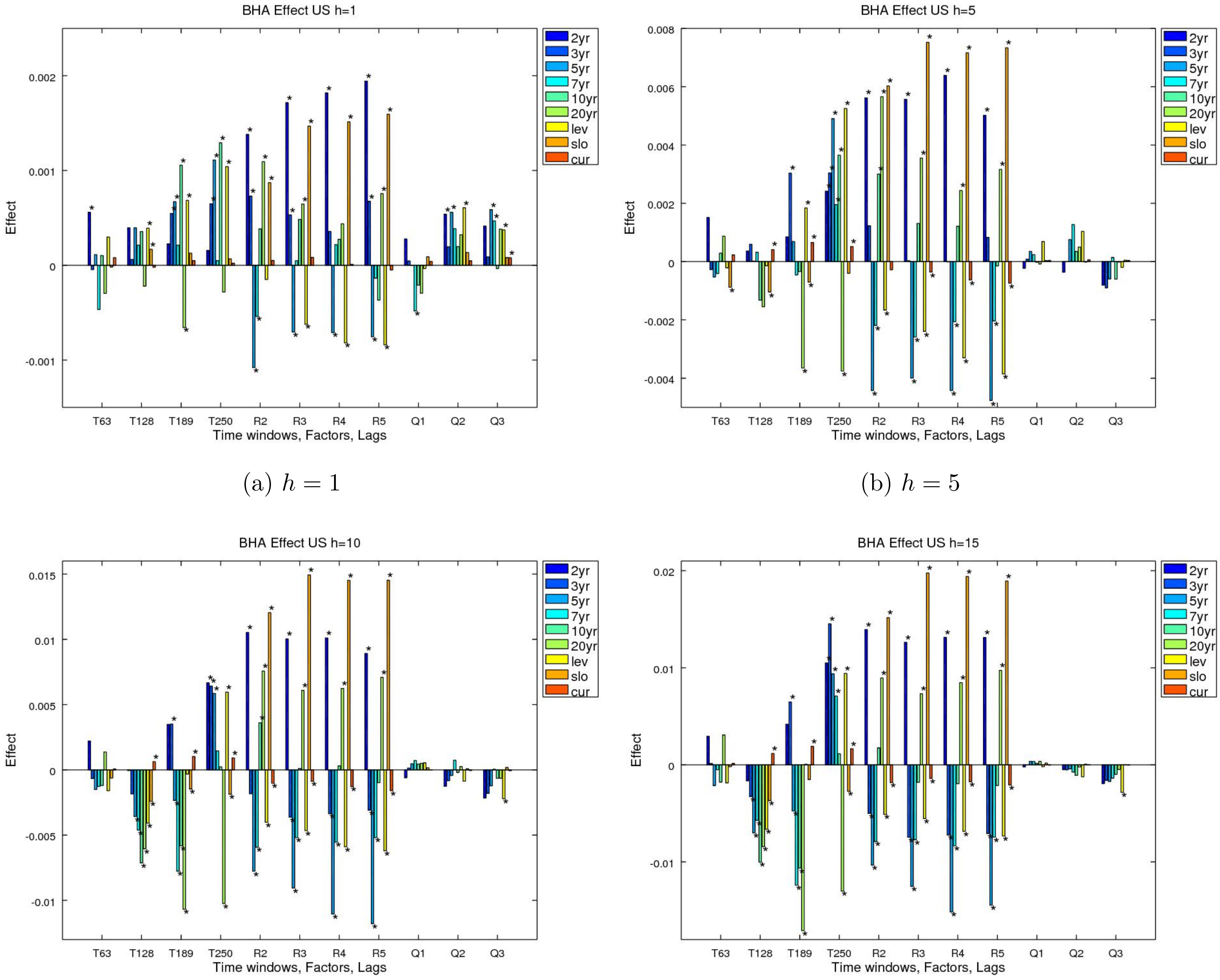
(a) $h = 1$

(b) $h = 5$

(c) $h = 10$

(d) $h = 15$

Figure 2.43: BHA effects for Germany. For an explanation of the panels, see the notes under Fig. 2.35.



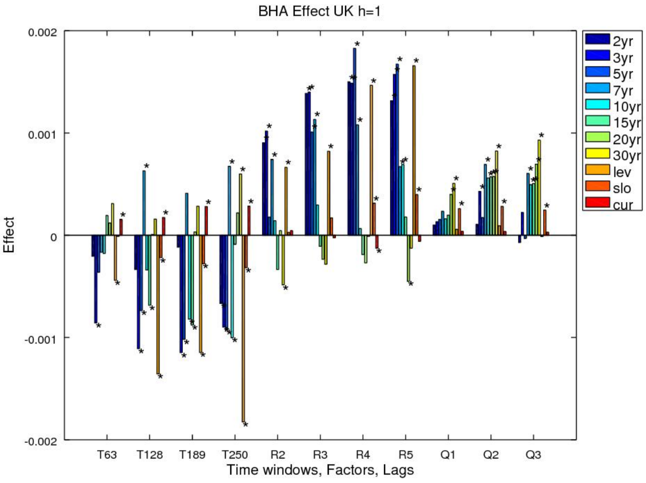
(a) $h = 1$

(b) $h = 5$

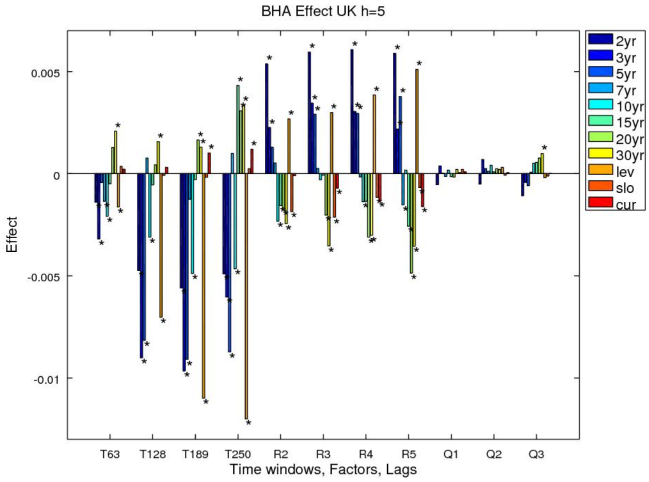
(c) $h = 10$

(d) $h = 15$

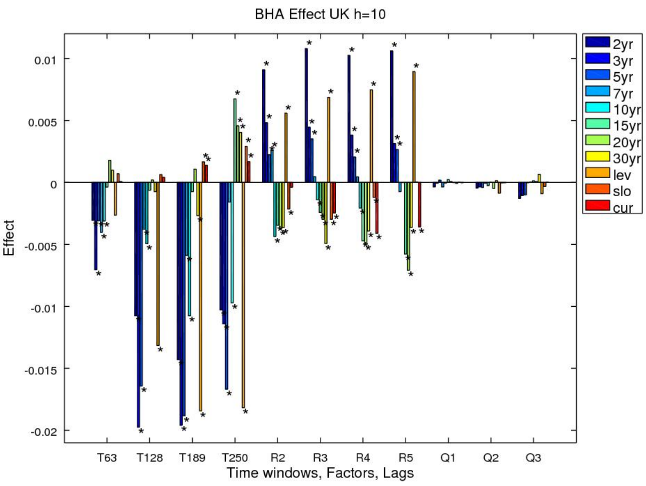
Figure 2.44: BHA effects for the US. For an explanation of the panels, see the notes under Fig. 2.35.



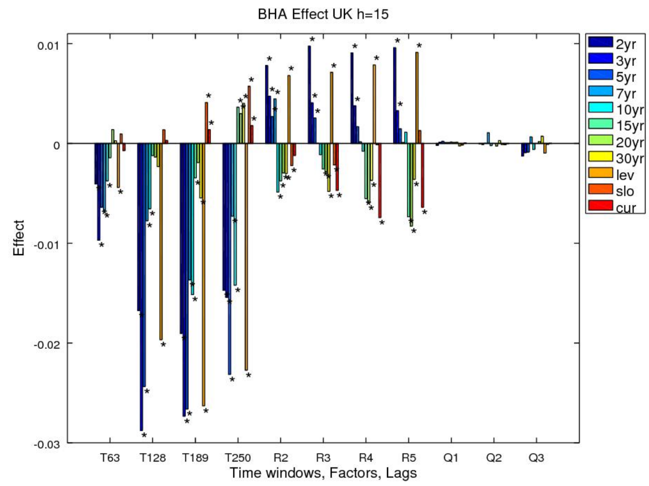
(a) $h = 1$



(b) $h = 5$

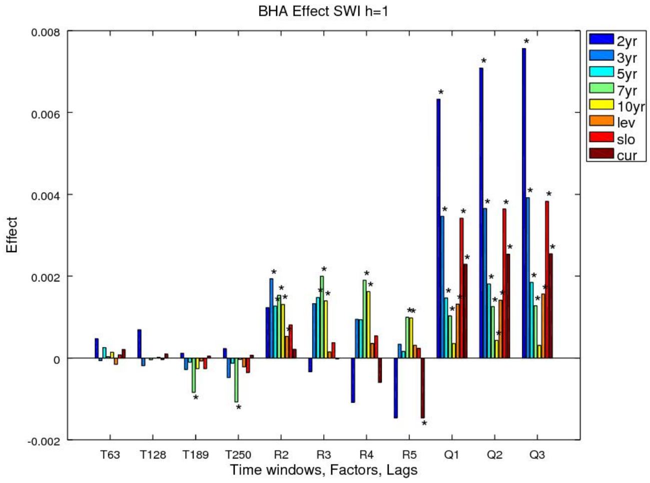


(c) $h = 10$

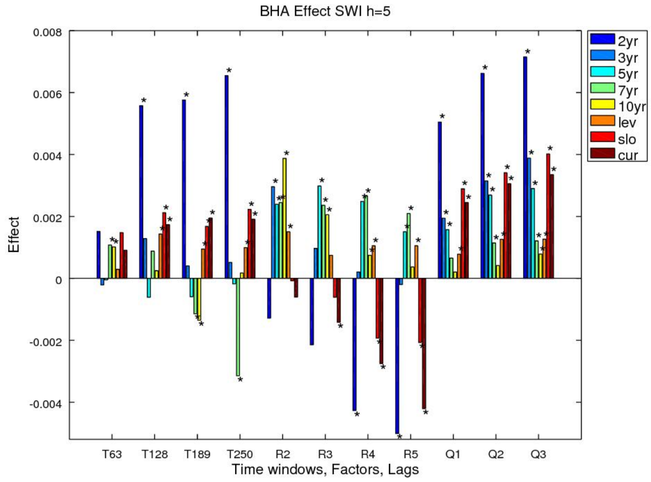


(d) $h = 15$

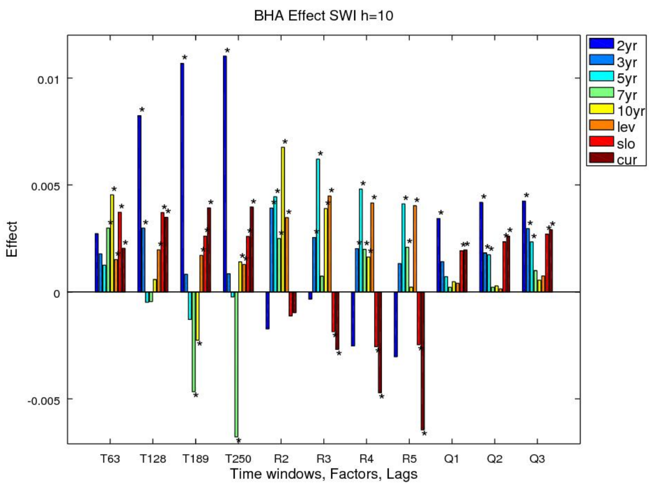
Figure 2.45: BHA effects for the UK. For an explanation of the panels, see the notes under Fig. 2.35.



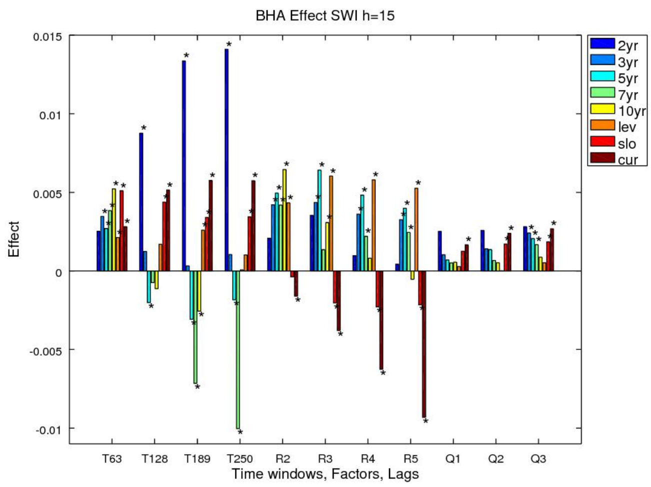
(a) $h = 1$



(b) $h = 5$



(c) $h = 10$



(d) $h = 15$

Figure 2.46: BHA effects for Switzerland. For an explanation of the panels, see the notes under Fig. 2.35.

with the exception of the $7yr$ maturity with the $\tau = 189$ and $\tau = 250$ time windows as can be seen in Fig. 2.46. Strategies with these time windows are outperformed. Additional factors and employing autoregressive lags improve forecasts. For $h = 5$ larger time windows, additional factors, and lags improve forecasts. The notable exceptions being the $7yr$ and $10yr$ maturities for the $\tau = 189$ and $\tau = 250$ time windows and slope and curvature for additional factors. This holds for $h = 10$ and $h = 15$.

BHA values are in percentage points. BHA values, similar to MSFE values, logically increase at larger forecast horizons, at least in absolute terms. For GER BHA values are around 0 or just above for $h = 1$, at $h = 5$ values are around 0.005, around 0.01 at $h = 10$, and at 0.015 at $h = 15$. The parameter values of the ANOVA constants are mostly above zero. For GER effects for time windows, additional factors, and using lags vary across forecast horizons between 0.0003 and 0.015, 0.0002 and 0.001, and from 0.01 towards zero respectively. If two factors are chosen as the base strategy, the effects of additional factors are often not significant. The results from this additional ANOVA over all three criteria indicate that a parsimonious model would mostly include no more than two factors.

2.3.3 Further Term Structures

There are multiple findings that imply that further variables outside the individual term structure may help to improve term structure forecasts (Mönch, 2008, Matthies, 2014). Matthies (2014) found that additional term structures and further financial data may improve forecasts under DA and BHA. The findings can be summed up as follows: Under DA and BHA further term structures and financial data improve forecasts for GER and SWI and produce significantly worse forecasts for the US and the UK.

The approach here differs in that additional financial data like stock market indices and exchange rates are not included. The two data sets that are compared are therefore first the individual term structure being forecasted and second the term structure being forecasted and the other three term structures. The results therefore examine the use of employing additional government term structures in other term structure forecasts. Logically, this extends the comparison to the basic AR-models. The comparison of DFM to AR-models examines the additional forecast value of further maturities of the same term structure. Furthermore, the approach of Matthies (2014) tests the additional forecasting value of further financial data. Ergo, the approach here tests the forecasting value of other government term structure and thereby fills a gap between the comparison of the AR-models and the results of Matthies (2014).

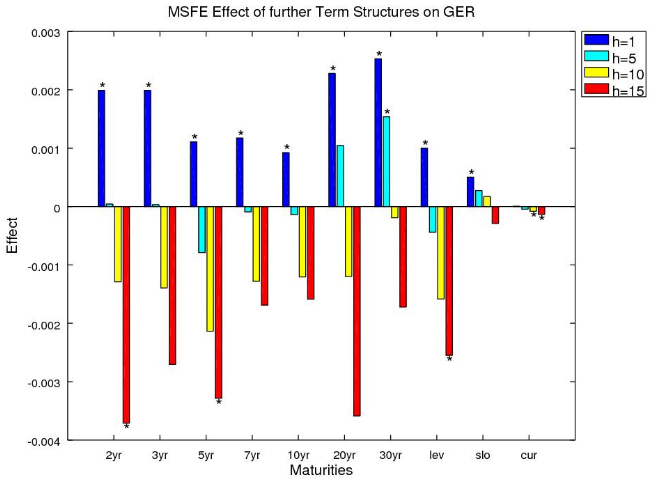
Under MSFE for GER strategies with additional term structures improve forecasts at longer forecast horizons. In contrast, for all maturities and linear combinations except for curvature, strategies using additional term structures are significantly outperformed at the $h = 1$ forecast horizon as seen in Figure 2.47 Panel (a). For the US in Panel (b), strategies that utilise additional term structures are significantly outperformed across all maturities and linear combinations and over all forecast horizons. There is only one case where more term structures improve, namely for $2yr$ at the $h = 15$ forecast horizon (but this effect is insignificant). For the UK in Panel (b) we find a similar picture as for GER. At the $h = 1$ and $h = 5$ forecast horizons strategies employing additional term structures are outperformed for all maturities and linear combinations while they improve for $h = 10$ and $h = 15$. For SWI in Panel (d) using more term structures mostly improves forecasts. The exceptions are the $2yr$ at $h = 1$, $h = 5$, and $h = 10$ and slope. The other exceptions are not statistically significant.

As seen in Figure 2.48 Panel (a) under the DA criterion the effects for GER are mostly significantly negative. This means that DFM strategies with additional term structures are outperformed. Exceptions are $20yr$, slope, and curvature, where forecasts are enhanced. For the US as seen in Panel (b), the effects are positive at the $h = 1$ forecast horizon. There are also improvements at longer forecast horizons for $2yr$, $3yr$, and curvature. In the other cases the strategies are significantly outperformed. For the UK in Panel (c), further term structures have the opposite effect under DA than under MSFE. At shorter forecast horizons forecasts are improved. The exceptions are $2yr$ and curvature. At $h = 15$ the effects are negative except for $2yr$. The effects are positive for SWI in Panel (d). The exceptions are $2yr$, $3yr$, $5yr$, and slope at $h = 1$ with significantly negative effects.

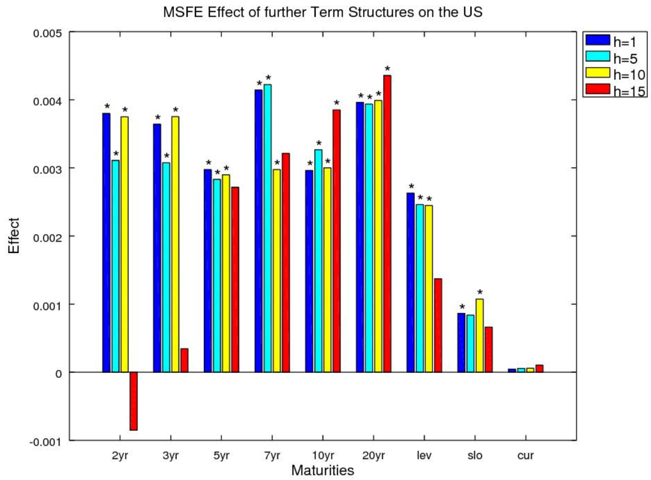
Under BHA there are significant positive effects for GER for $20yr$, $30yr$, and curvature and at the shorter forecast horizon for $5yr$, $7yr$, $10yr$, level, and slope as seen in Figure 2.49 Panel (a). In these cases employing additional term structures improves forecasts. In the other cases they significantly worsen forecasts. For the US in Panel (b) forecasts are improved at shorter forecast horizons and are outperformed at longer horizons for $3yr$, $5yr$, $7yr$, $10yr$, $20yr$, level, and slope. For curvature the effects are positive over all forecast horizons. In the case of the UK term structure seen in Panel (c), the effects are positive at short forecast horizons and negative at longer ones. The exceptions are $2yr$, $3yr$, and curvature. For SWI the effects are positive across all maturities and linear combinations as seen in Panel (d). The exceptions are the shorter forecast horizons of $3yr$ and for $5yr$ and $7yr$ at $h = 1$.

2.3.4 Estimation Methods

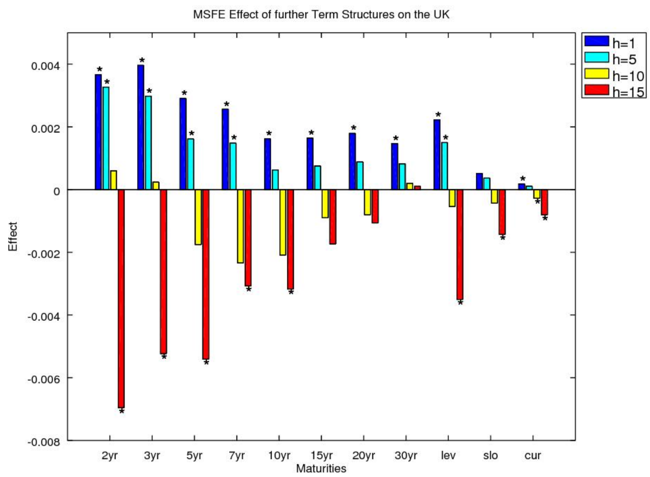
The alternative estimation methods presented in Section 2.1.2 compare the additional forecasting value of the one-step and two-step estimation methods. In Matthies (2014)



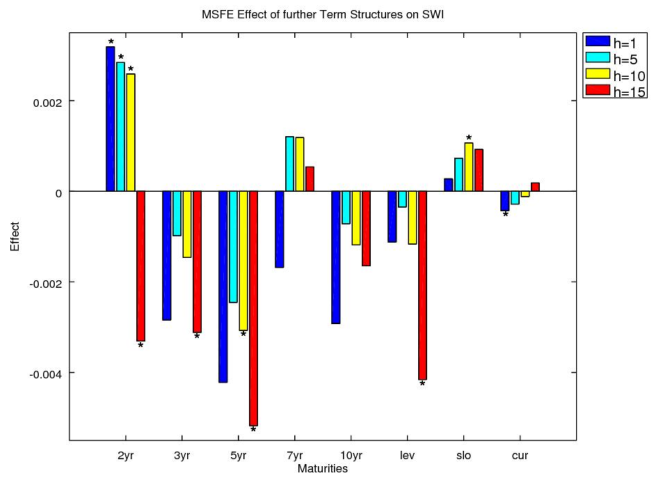
(a) GER



(b) US

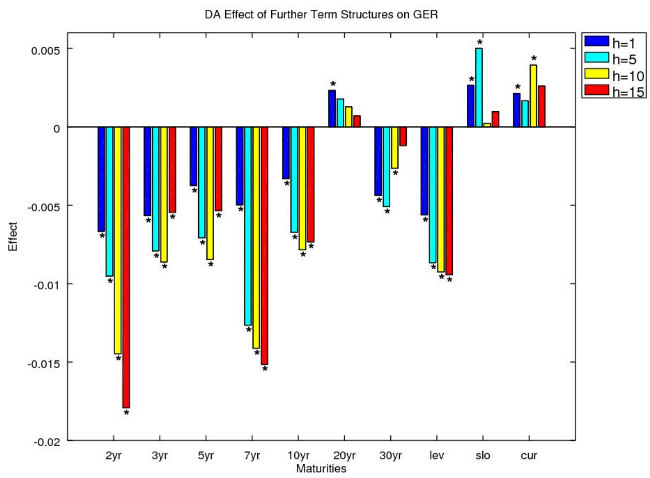


(c) UK

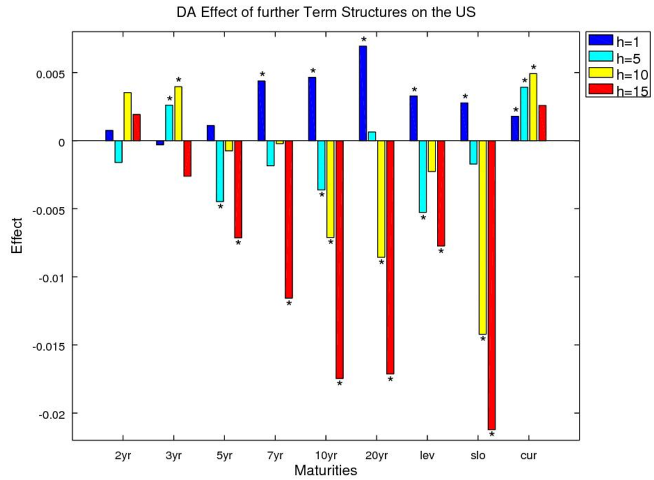


(d) SWI

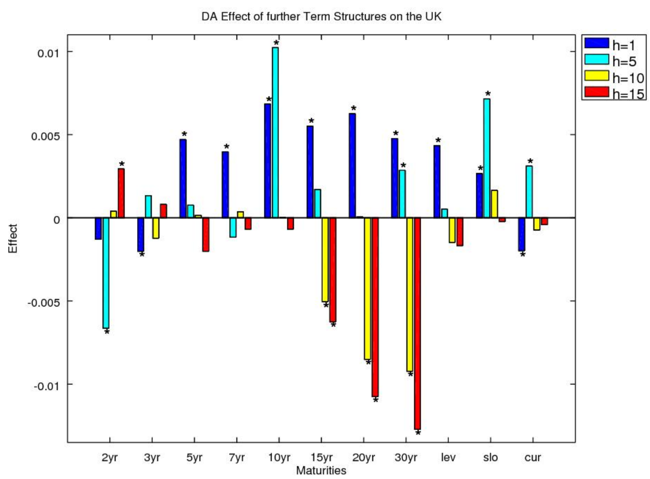
Figure 2.47: Effect of macroeconomic data w.r.t. MSFE for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



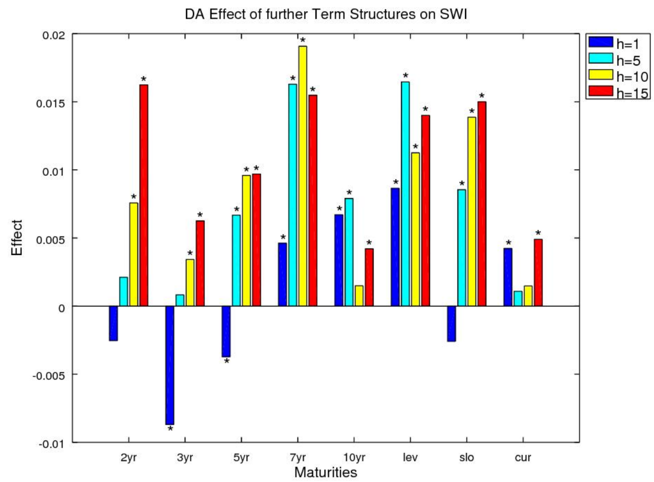
(a) GER



(b) US

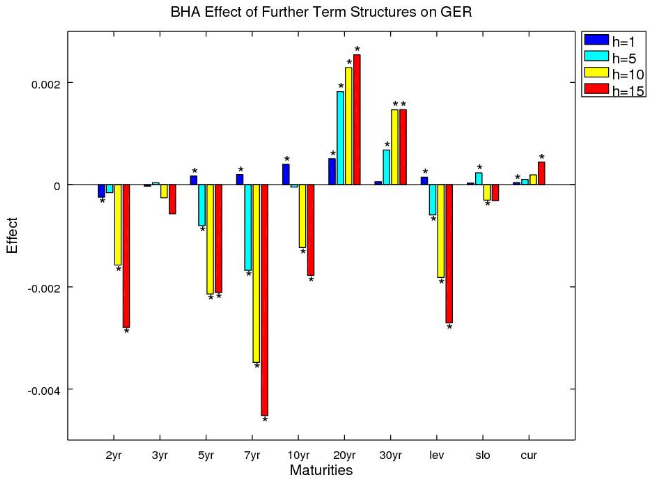


(c) UK

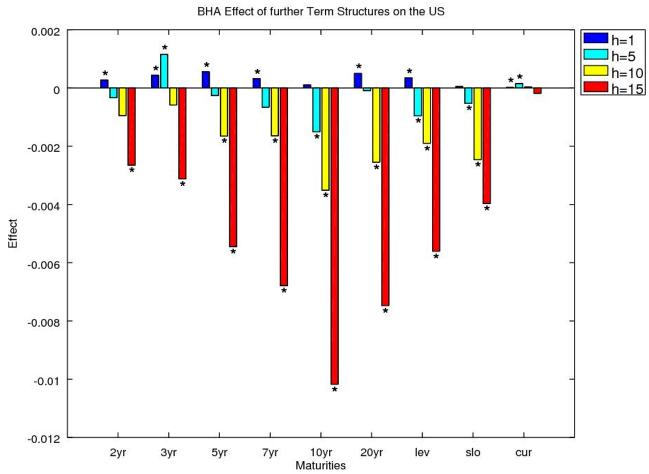


(d) SWI

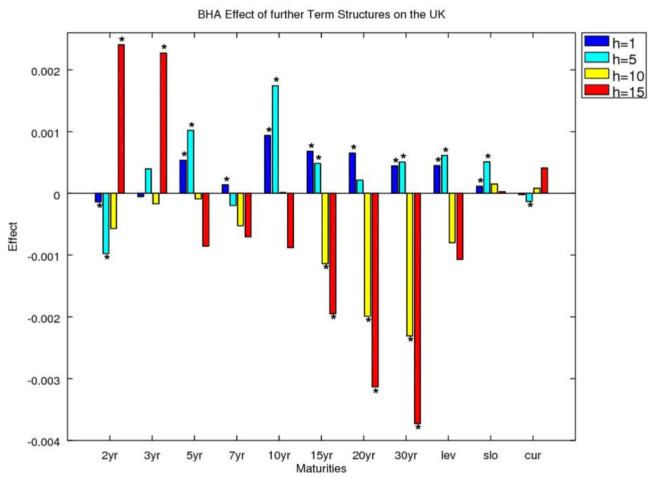
Figure 2.48: Effect of macroeconomic data w.r.t. DA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



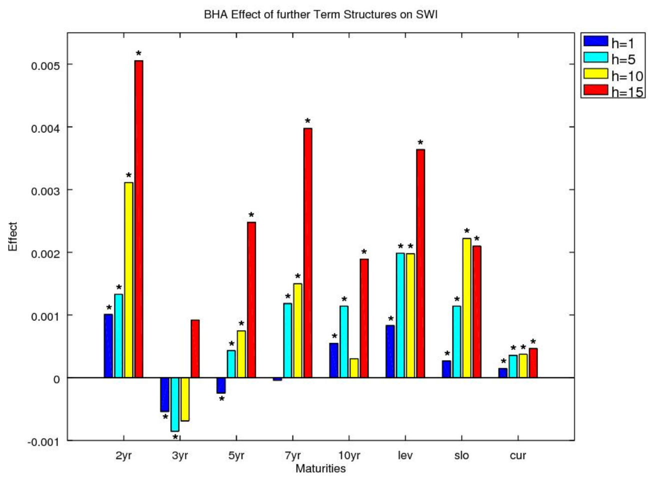
(a) GER



(b) US



(c) UK



(d) SWI

Figure 2.49: Effect of macroeconomic data w.r.t. BHA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.

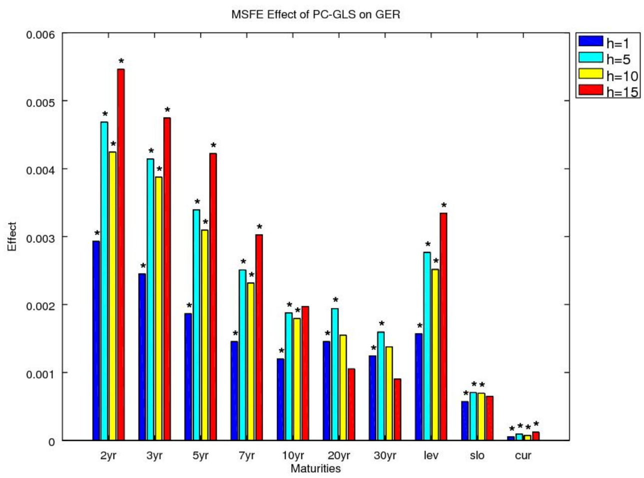
PC-GLS was mostly outperformed under DA and BHA. There, this comparison is conducted across a shorter time period (17.03.1999 to 12.03.2008). Furthermore, Matthies (2014) uses a different set of predictors for the the larger data set as it includes more financial time series. This is relevant as it relates to the statistical properties of different types of financial time series and their compatibility to the assumptions of the estimation methods. There is another aspect that is addressed in the comparison of PC-GLS and PC-OLS that differs to the results of Matthies (2014) apart from the robustness over a larger time period. The factors in the larger data set are only extracted from other government term structures in this application. In Matthies (2014) the factors are also extracted from stock indices, exchange rates, commodity prices, and volatility measures. Therefore, the two alternative estimation methods are not only compared in a longer time frame as in Matthies (2014) but also at extracting factors from different financial time series.

Under MSFE strategies that employ the PC-GLS, estimation methods are significantly outperformed for GER across all maturities and linear combinations at all forecast horizons as can be seen in Figure 2.50 Panel (a). This holds true also for the US in Panel (b) and SWI in Panel (d). In the UK seen in Panel (c) this is also mostly the case. The notable difference is that some effects at the $h = 15$ horizon are not significant and in one case (curvature) even negative (but not significant).

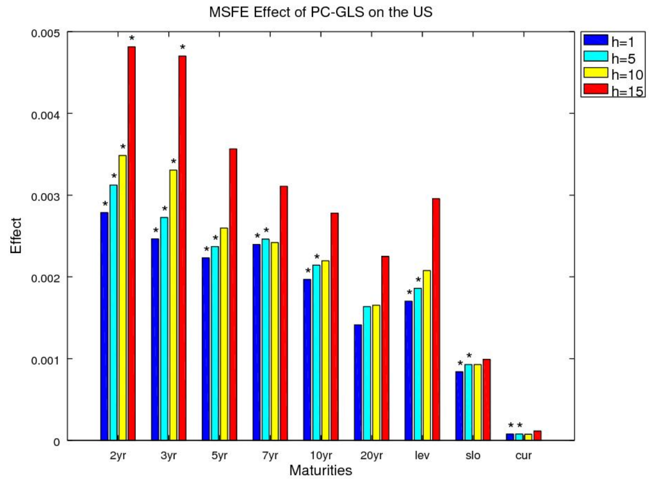
With respect to the DA criterion PC-GLS improves for $7yr$, $10yr$, $30yr$, and level for the GER term structure as seen in Figure 2.51 Panel (a). It is outperformed for the other maturities and slope and curvature. For the US, seen in Panel (b), there are positive but insignificant effects for $7yr$ and significantly positive effects for $10yr$ and $20yr$ at $h = 5$, $h = 10$, and $h = 15$. Significantly negative effects for all other maturities and linear combinations complete the picture. PC-GLS improves for the UK for the $15yr$ maturity and curvature at the longer forecast horizons as seen in Panel (c). The PC-GLS strategies are outperformed for the $2yr$, $5yr$, and $7yr$ maturities. The effects are mixed and mostly insignificant for the other maturities and linear combinations. Except for slope at the $h = 10$ and $h = 15$ forecast horizon all effects are significantly negative for SWI as seen in Panel (d). Under BHA the effects are mostly like those for DA and can be seen in Figure 2.52 Panels (a) to (d).

2.3.5 Forecasting Methods

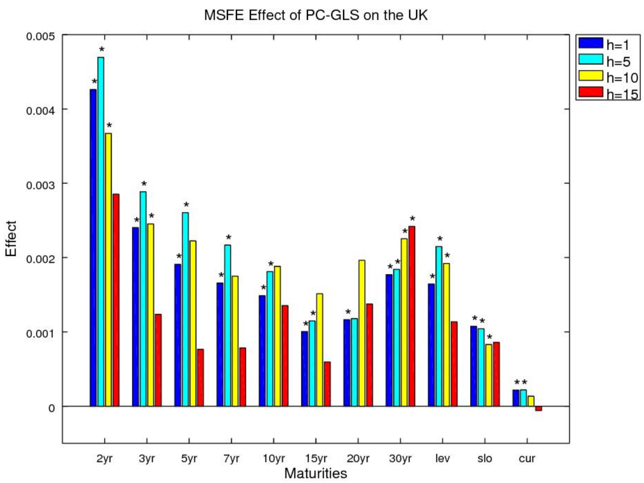
After factors are estimated they are employed to forecast future interest rates. The two alternative methods differ in their approach. BH exploits possible autoregressive features of factors, which are empirically described by Diebold & Li (2006), Blaskowitz & Herwartz (2009). SW uses the factors to directly explain future interest rates. In



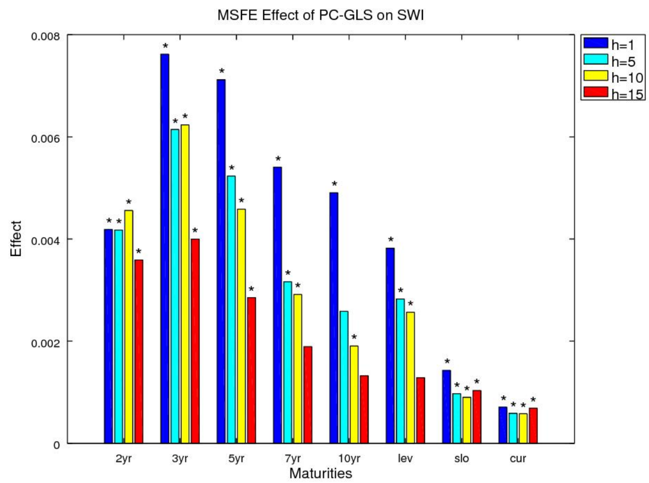
(a) GER



(b) US

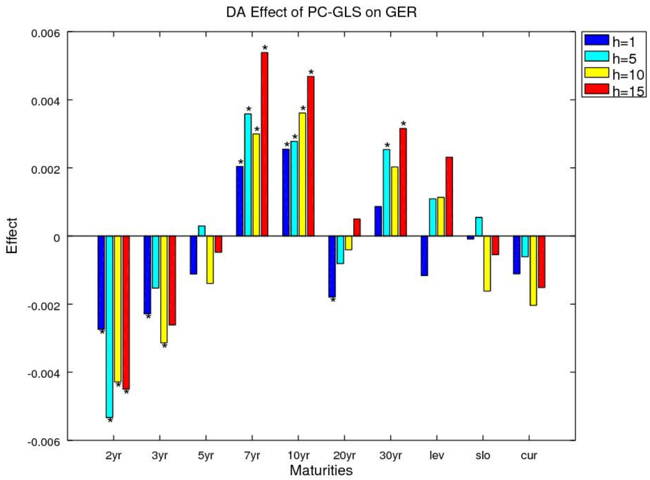


(c) UK

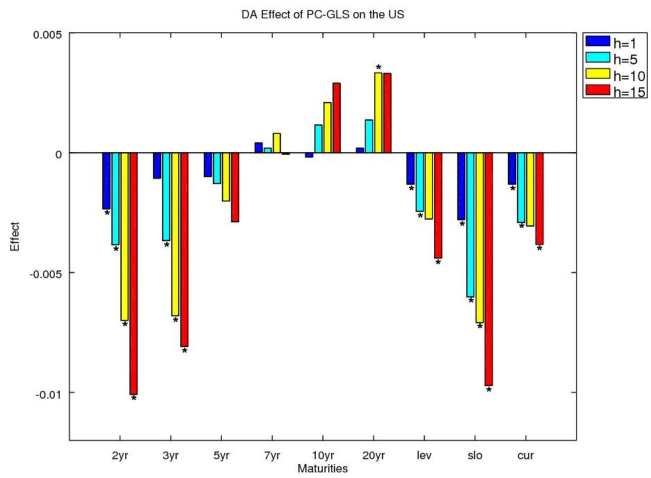


(d) SWI

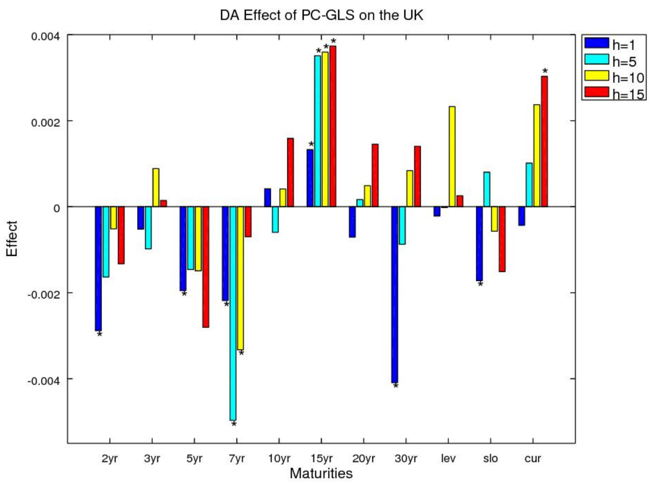
Figure 2.50: Effect of PC-GLS w.r.t. MSFE for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



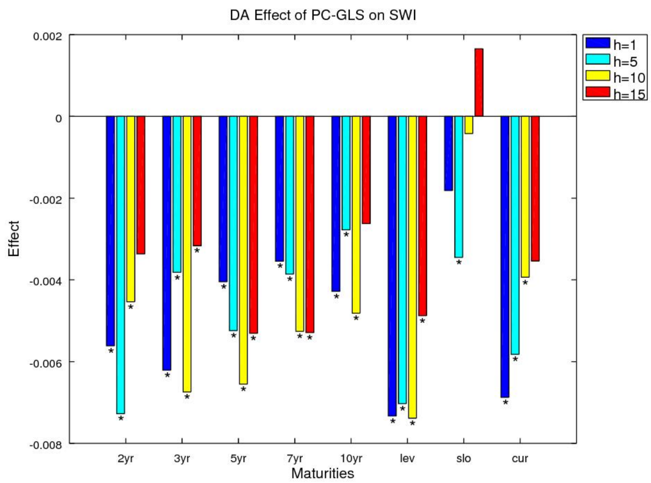
(a) GER



(b) US

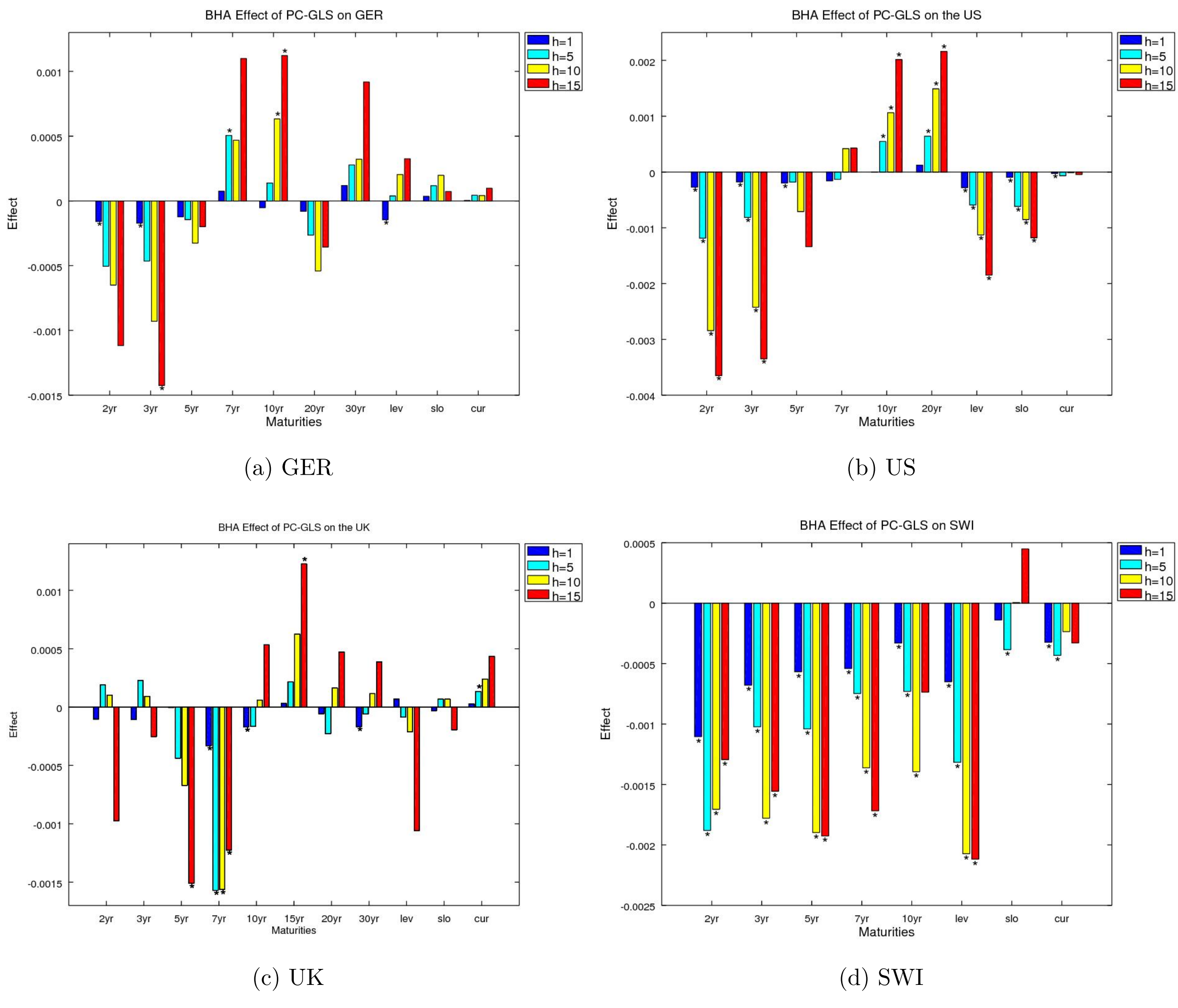


(c) UK



(d) SWI

Figure 2.51: Effect of PC-GLS w.r.t. DA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



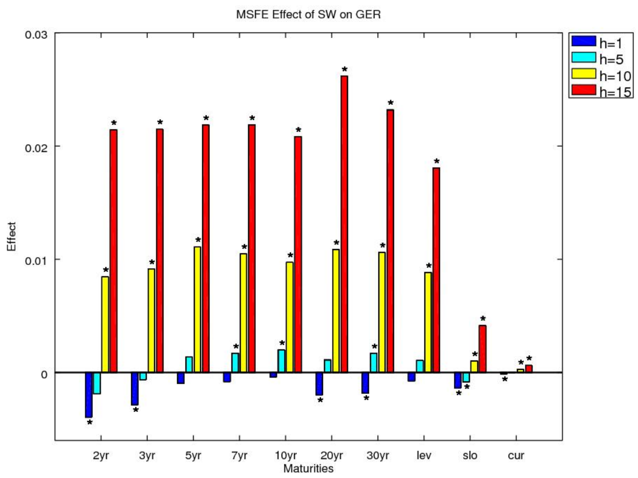
(a) GER

(b) US

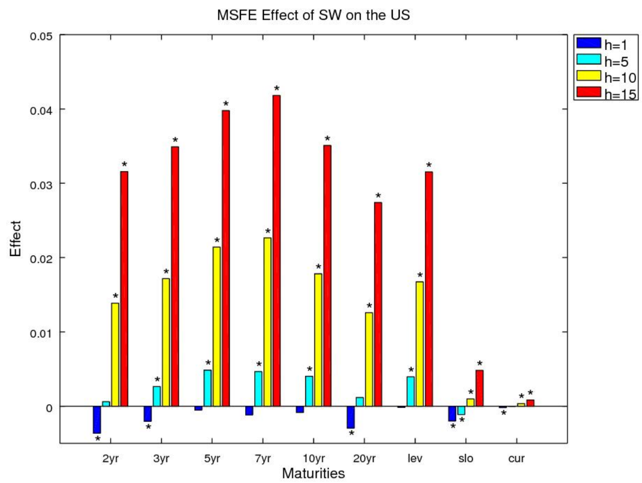
(c) UK

(d) SWI

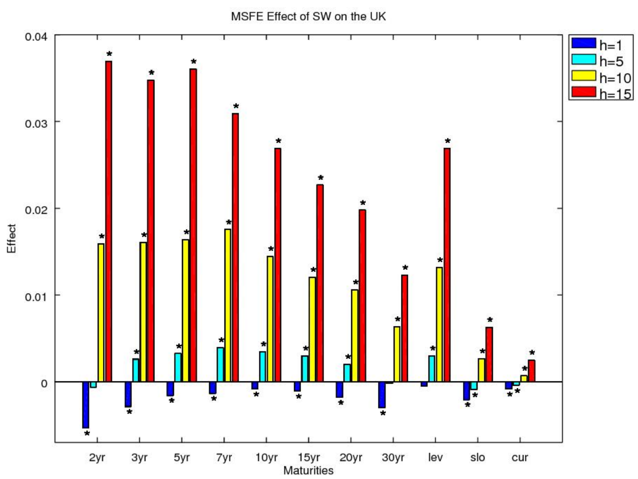
Figure 2.52: Effect of PC-GLS w.r.t. BHA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



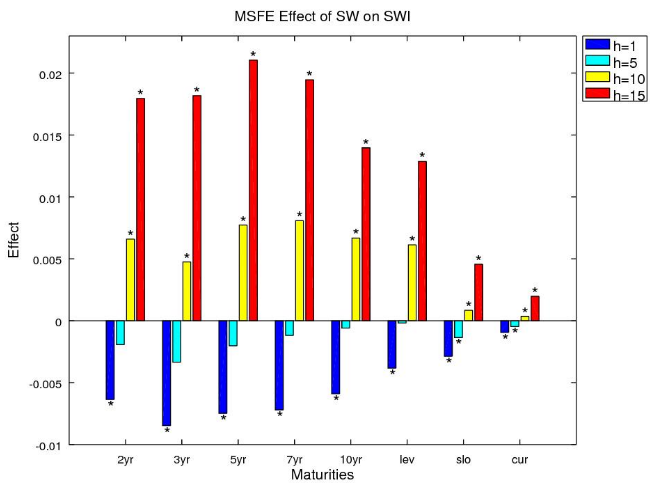
(a) GER



(b) US

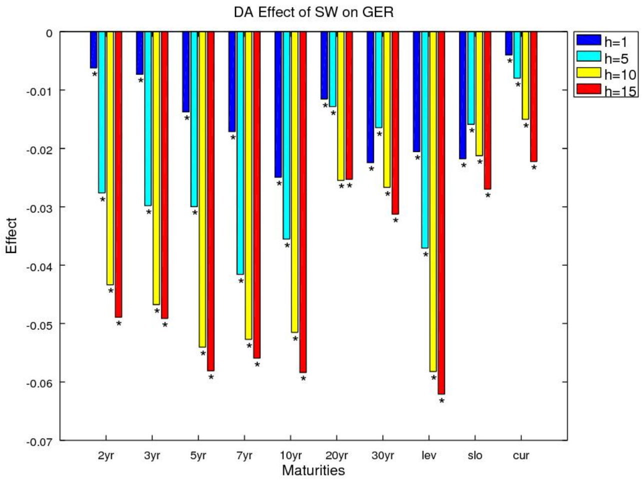


(c) UK

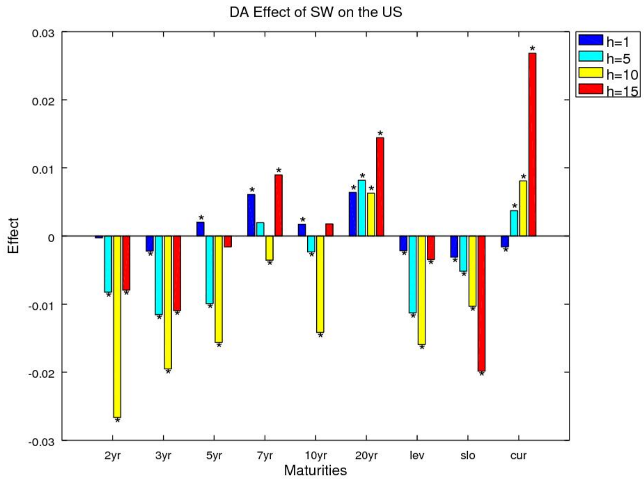


(d) SWI

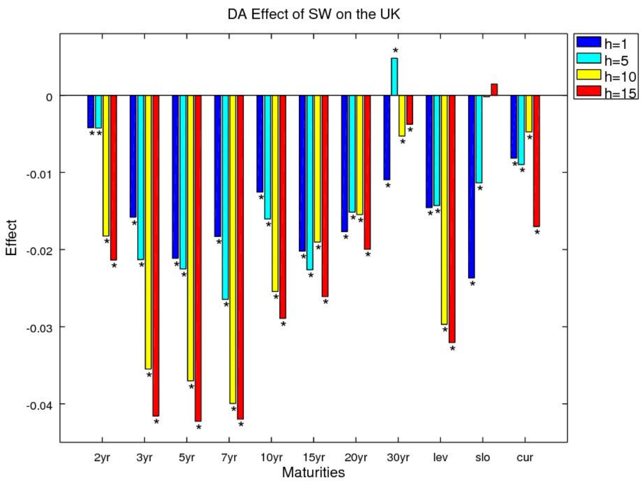
Figure 2.53: Effect of SW w.r.t. MSFE for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



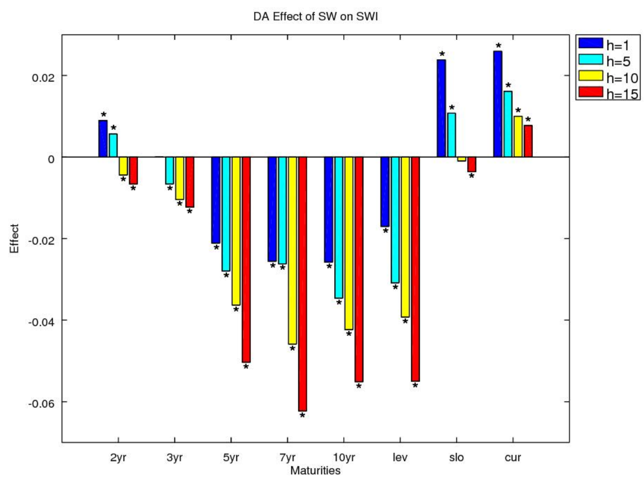
(a) GER



(b) US

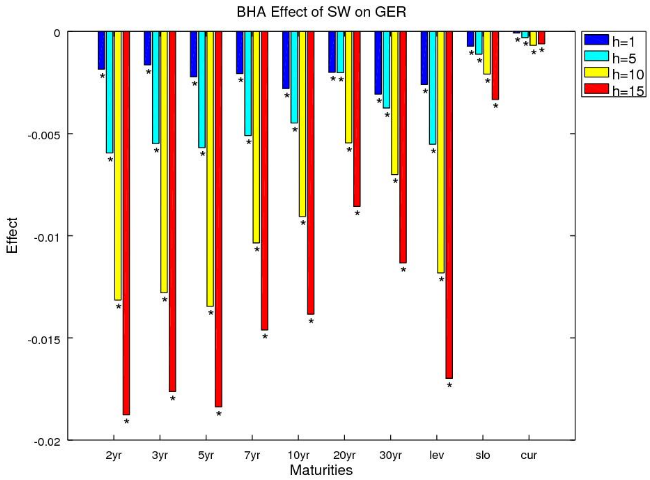


(c) UK

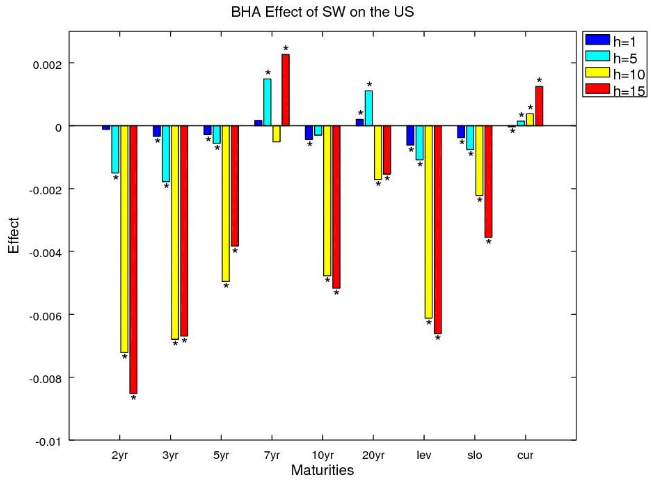


(d) SWI

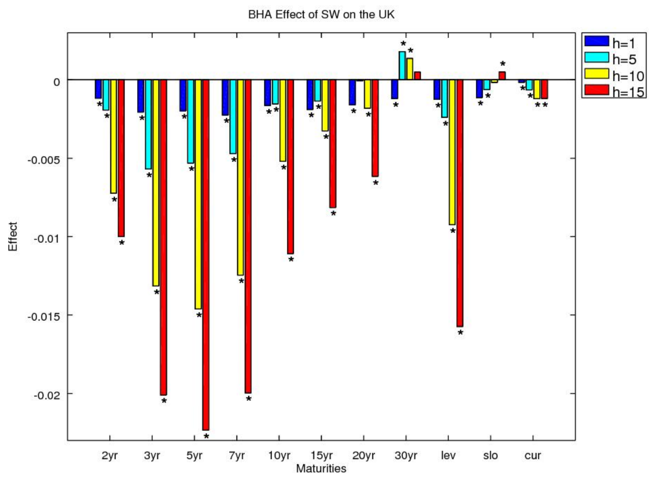
Figure 2.54: Effect of SW w.r.t. DA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.



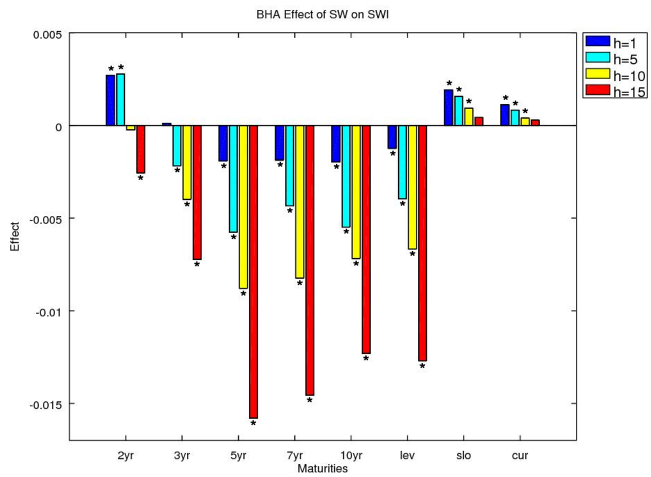
(a) GER



(b) US



(c) UK



(d) SWI

Figure 2.55: Effect of SW w.r.t. BHA for GER, SWI, the UK, and the US. Groups of four bars represent forecast horizons $h = 1, 5, 10, 15$. On the horizontal axis groups are from left to right the listed year maturities, and the level, slope, and curvature of the term structure. A star ($\star$) above a bar indicates that the estimate is significant at the 1% level.

Matthies (2014) SW was mostly outperformed under DA and BHA.

Under MSFE, SW has significantly negative effects for GER over all maturities and linear combinations at the $h = 1$ forecast horizon as seen in Figure 2.53 Panel (a). That means that SW improves forecasts in these cases. Furthermore, there are significant effects at $h = 5$ for $2yr$ and slope. Most effects at $h = 5$ are positive, and most of these are significant, these are those for $7yr$, $10yr$, $20yr$, and $30yr$. At $h = 10$ and $h = 15$, all effects are significantly positive. Therefore, SW can only improve at shorter forecast horizons for GER and is outperformed at longer forecast horizons. For the US and the UK the results are almost identical as seen in Panels (b) and (c). For SWI, there are no positive effects at $h = 5$ as seen in Panel (d). Otherwise, the effects are the same at $h = 1$, $h = 10$, and $h = 15$, as they are for GER, the US, and the UK.

Under DA and BHA, for GER strategies that employ SW are significantly outperformed across all maturities and linear combinations over all forecast horizons as seen in Figures 2.54 and 2.55 in Panel (a) respectively. For the US forecasts are improved under DA and BHA for $7yr$, $20yr$, and curvature as seen in Panel (b). The results are mixed or the SW strategies are outperformed in the other cases. For the UK, seen in Panel (c), strategies that employ SW are outperformed under DA and BHA with few exceptions. For SWI, the SW strategies improve for short forecast horizons for $2yr$ and for slope and curvature under the DA and BHA criteria as seen in Panel (d).

Comparing the sizes of effects of additional term structures, PC-GLS, and SW we find that SW most often has the larger effects across all three evaluation criteria. The smallest effects are usually those of the PC-GLS parameters. It follows that the choice of forecasting method has the strongest impact on forecast performance.

2.4 Concluding Remarks on Term Structure Forecasts

Modelling and forecasting government term structures is an important element of determining possible future risks of investing in a given country and in understanding the risk of alternative investments. Interest rates are a measure of default probability. To provide a basis for the modelling of the risk of many investments this thesis employs a large set of data driven models of the term structure of interest rates for ex-ante forecasting.

Factors analysis estimated via PCA supports the idea that the German, Swiss, US, and UK term structures are driven by level, slope, and curvature factors. Furthermore, we find that factors from a data set that covers all term structures indicate a global level factor.

A comparison with simple random walk and auto regressive forecasts show that random walk forecasts can only be improved on in a few instances with respect to statistical measures of forecast performance. On the other hand, dynamic factor models consistently improve on AR forecasts under both statistical and economic criteria. The choice of the size of time windows, factor numbers, and lag selection for DFM models depends on forecast horizons and evaluation criteria.

Additional term structures can improve forecasts for the Swiss term structure. For the German, US, and UK yield curve forecast results depend more on the forecast horizons. With regard to estimation and forecasting methods we find that the standard PC-OLS and employing autoregressive factors outperform the respective alternative PC-GLS and using the lagged correlation of factors and interest rates.

Overall, the results confirm the findigs of Diebold & Li (2006), Blaskowitz & Herwartz (2011), and Matthies (2014) that using the entire term structure as predictors improves interest rate forecasts. As an extension we find that additional term structures may improve forecasts, just as we found for additional financial data in Matthies (2014).

Government interest rates determine corporate interest rates as government credit ratings determine corporate credit ratings. Credit rating agencies provide estimates of corporate default probability and thereby a basis for investors to asses investment risks. The next chapter deals with the empirical literature on corporate credit ratings, the regulation of credit ratings, and the estimation, determinants, and empirical features of corporate credit ratings.

Chapter 3

Estimating Corporate Credit Ratings

This Chapter combines the main work of Matthies (2013a,c), and Matthies (2013b). We follow Matthies (2013c) for this introduction, for the discussion of the history and regulation of credit rating agencies, as well as different rating markets. In Section 3.1 we closely follow Matthies (2013b), and finally for the review of statistical methods in Section 3.2.1 we follow Matthies (2013a).

The ability of an entity (e.g. a corporation) to fully and punctually meet its debt obligations is the focus of credit ratings. Ratings represent the assessment of a credit rating agency (CRA) that measures the corporations fundamental creditworthiness, and thereby the risk of its default (Gonzales et al., 2004). CRAs are subject to regulation in the US and the EU (Hill, 2004). In contrast to the external ratings from CRAs banks develop internal or so-called 'shadow ratings' that they use to approve loans and measure the riskiness of investments. External ratings are generally solicited and paid by the issuers themselves[1] (i.e. issuer paid). The rating market can be regarded as an oligopoly. The widely known three 'big' agencies, Moody's, Standard & Poor's (S&P), and Fitch control around 95% of the market (see Asmussen (2005) and Wappenschmidt (2009, p. 13)).

Credit ratings are a mechanism of corporate governance. Issuers use them to signal transparency and low investment risk (Nordberg, 2011, p. 60). Possible principal agent problems can thereby be reduced and lead to a reduction of a corporations capital costs (Gonzales et al., 2004). From an investment perspective, a credit rating is a highly aggregated classification of an issuers debt. A reader is then able to inform himself on the possibility that an issuer will not be able to meet his debt obligations (Dilly & Mählmann, 2010). Potential investors and banks can use ratings as benchmarks and reference points as they conduct their own assessments (Erlenmaier, 2006, p. 39).

[1]This holds for most corporate ratings. Country ratings are unsolicited. Unsolicited ratings for corporations are rare and most ratings are solicited and withdrawn at the firms request.

Therefore, CRAs function as financial intermediaries on credit and financial markets. If the hypothesis of strictly efficient markets were to hold true, the existence of CRAs could not be justified (Ramakrishnan & Thakor, 1984).

It is quite natural that corporations that solicit credit ratings and pay for them would enjoy it if CRAs would provide them with favourable ratings. This poses a possible conflict of interest for CRAs, as they depend on corporations hiring them to publish credit ratings. A CRA's existence also depends on investors trusting their judgments. CRAs should therefore seek to gain reputational capital (Matthies, 2013c). On the other hand, CRAs are already incorporated in financial regulation. For example, in the US CRAs need to achieve the status of a Nationally Recognized Statistical Rating Organization (NRSRO) so they can be part of the regulation of financial markets. Partnoy (1999) therefore argues that CRAs need not gain reputaional capital, as the position granted to them through regulation allows them to sell regulatory licences without providing any additional information to markets. In this chapter, we present the individual market analysis from Matthies (2013b). It shows that the informational value of credit ratings depends on the supply and demand on rating markets, and that in certain cases the position of the CRAs might be undermined.

The highest rating category are the so called 'triple A' (AAA) ratings which are assigned to issuers with the highest credit quality. Ratings of lower quality (AA+ to D) follow and are placed on a discrete ordinal scale. Credit ratings and rating changes are of empirical importance for several reasons. They are the foundation of a large part of institutional investor regulation (Hill, 2004). Due to this regulation, these investors are restricted so they can only purchase or hold bonds that have a certain quality. Rating changes have a measured effect on the price of bonds (Katz, 1974) and on stock prices (Jorion & Zhang, 2007) of a corporation. They are furthermore employed in risk modelling (Nickel et al., 2000). Rating transitions are therefore relevant to determine the future development of the risk assessment of any given portfolio.

CRAs provide regular updates and re-evaluations on their ratings. Furthermore, they monitor developments of individual corporations. These developments can trigger CRAs to come to the conclusion that a change to the long term creditworthiness of an issuer has occurred. The CRA will change the rating and the rating will transition from one category into another one. Empirically, rating transitions exhibit non-Markov effects[2] (Lando & Skødeberg, 2002). Specifically, rating transitions are shown to be dependent on momentum and duration effects[3]. The studies of Nickel et al. (2000) and Lando & Skødeberg (2002) do not use firm specific risk factors that might account for

[2]This means that the probability of a rating transition is not only dependent on the state it is in.

[3]Momentum: An up- or downgrade occurred prior. Duration: Refers to how long a rating has been in a certain class.

such observations. Furthermore, Nickel et al. (2000) find that rating transitions vary across the business cycle. However, CRAs insist that their long term ratings follow a through-the-cycle approach, i.e. short term fluctuations of the business cycle should not cause ratings to change (Amato & Furfine, 2004).

As financial intermediaries CRAs should provide consistent long term information and have predictive power. With regard to credit ratings, predictive power requires that at a minimum higher rating categories have on average a lower default frequency than lower rating categories. The studies by Zhou (2001a) and Jorion & Zhang (2007) find that ratings have better default accuracy at longer time horizons (e.g. ten years) than shorter time horizons (one year). At ten year horizons, higher rating categories have lower default frequencies, and differences between default frequencies for rating categories increase down the rating scale. Therefore these findings suggest that ratings are distributed ordinally along the creditworthiness scale. Furthermore, the empirical observations provide an estimate of the default prediction which any given rating implies. In Jorion & Zhang (2007) the default frequencies are used to study market reactions to credit rating changes. Jorion & Zhang (2007) is one of many studies that test the additional information content of credit ratings beyond the basic information value for financial markets. This research focuses on information in credit ratings which capital markets have not already incorporated. The hypothesis of efficient bond, stock, and credit default swaps (CDS) markets is tested with regard to rating events[4]. An important finding is that markets exhibit asymmetric reactions to rating changes. Notably, downgrades cause stronger market reactions than upgrades (Jorion & Zhang, 2007). This might be due to the fact that downgrades present a larger change in creditworthiness.

A further important research question concerns to what degree publicly available information informs CRA's ratings. CRAs use public information such as financial data and ratios (Ederington, 1985, Blume et al., 1998, Altman & Rijken, 2004) and corporate governance characteristics (Bhojraj & Sengupta, 2003, Ashbaugh-Skaife et al., 2006, Jorion et al., 2009). Furthermore, macroeconomic developments are also empirically reflected in credit ratings (Nickel et al., 2000, Amato & Furfine, 2004).

Any assessment of corporate creditworthiness will incorporate measures of the firms earnings and their leverage, as well as some measure of the firms size. Basic intuition would inform any analyst that increasing earnings should have a positive and a larger leverage ratio should have a negative effect on ratings ceteris paribus[5]. Corporate governance mechanisms may have some further important effects on credit ratings. For instance, a principal agent problem between stockholders and management may arise

[4]For a review see Gonzales et al. (2004) or Matthies (2013a).

[5]See Matthies (2013a) for a detailed discussion.

(Bhojraj & Sengupta, 2003). Furthermore, wealth transfer effects from bondholders to shareholders may have a negative effect on credit ratings (Ashbaugh-Skaife et al., 2006). Corporate governance also impacts spending on accounting. Counterintuitively higher spending on accounting and hiring the most qualified accountants has a negative impact on accounting quality, as it increases the possibility of earnings management (Jorion et al., 2009). As the possibilities of earnings management increase, CRAs' standards will change as well on a macro level. The changing agency standards were first observed by Blume et al. (1998). The impact of macroeconomic factors on ratings can be furthermore understood through their relation to the business cycle (Nickel et al., 2000, Amato & Furfine, 2004).

For the purpose of estimating credit ratings, the ordered probit model is the method most often used (Matthies, 2013a)[6]. The assumptions that ratings are ordered and are on an ordinal scale are the two main advantages of the ordered probit model. In contrast, a weak point is the implicit assumption that agencies apply a point-in-time perspective to credit ratings (Altman & Rijken, 2004). As a consequence, an ordered probit approach might then predict more rating changes than are observed. The through-the-cycle approach is designed to achieve rating stability. Duration and hazard methods as employed in Du & Suo (2005) might address such issues. Furthermore, it is possible to use ordinary least squares (OLS), multivariate discriminate analysis (MDA), and unordered logit (Ederington, 1985), as well as machine learning methods, e.g. neuronal networks (e.g. Dutta & Shekhar (1988)).

A drawback of the OLS approach is that it disregards the ordinal structure of credit ratings. The OLS method is, of course, simple and and easy to estimate. Therefore we use the method developed in Matthies (2013c) to address the problem of the ordinal structure. In an OLS regression the ratings are replaced by their corresponding default probabilities, i.e. the default frequencies calculated by Jorion & Zhang (2007). The ordered probit approach assumes an unobserved continuous variable that underlies the discrete observable variable. This OLS approach assumes that the unobservable variable can be estimated using the default frequencies of the corresponding rating categories. From a methodological perspective this approach uses an independent estimate of the unobservable continuous variable that underlies the probit model to proxy the discrete observable variable. Furthermore in this chapter we perform panel OLS and probit regressions as in Blume et al. (1998), Amato & Furfine (2004), Jorion et al. (2009), and Alp (2013). Estimation for the S&P ratings of the 100 largest US, UK, German, Japanese, French, Australian, and Canadian non-financial firms in 2005 in the time span from 1990 to 2009 is performed. This is an extension of the analysis in Matthies (2013c), where only the US-firms were employed. The results

[6]Alternatively, it is also possible to use the ordered logit model.

are discussed with respect to the possible increasing stringency of rating agencies' standards. We discuss the implications for CRAs via the explanation of these findings by Jorion et al. (2009). Furthermore, we have an international approach and discuss the possible implications of CRAs operating in different market settings.

This chapter continues as follows: Section 3.1 provides a short overview of the history of CRAs and their regulation, as well as an analysis of the different markets in which they operate. In Section 3.2 we provide a review of different statistical methods that can be employed to estimate and forecast credit ratings and rating changes. Furthermore, we present the empirical methods used in this chapter. Section 3.3 presents the ratings and the data used to estimate them. In Section 3.4 the results are presented and discussed. Section 3.5 provides a short conclusion to credit ratings.

3.1 Rating Agency Regulation and Rating Markets

Regulation is an important aspect of how CRAs function on financial markets. The legal framework is essential in understanding the alternative views of reputational capital and regulatory licenses of CRAs. The regulation together with differences of the supply and demand on corporate, sovereign, and structured finance rating markets, as well as the CRAs business model, determine how CRAs function on markets.

3.1.1 Rating Agency Regulation

In the aftermath of the 1929 crash and subsequent economic crisis concerns grew about defaulting bonds (Partnoy, 1999). Starting in 1936 regulation for ratings began when 'recognised' rating manuals became the basis for certain bank regulations (White, 2010). Then in the 1970s demand for ratings grew after numerous defaults led to a growing focus on the safety of debt (Hill, 2004). Since then, ratings play an important role for insurance, pensions, banking, securities, and real estate regulation. In 1973 the concept of NRSRO ratings was incorporated for the first time in SEC regulation. More recently CRA regulation was motivated as a reaction to problems in the CRA industry. The Sarbanes-Oxley Act of 2002 came in response to the Enron crisis of 1999/2000. The Act ultimately led to the Credit Rating Agency Reform Act of 2006 with the aim of increasing competition between CRAs (Coskun, 2008). The sub-prime mortgage crisis of 2008 pushed US legislation to write the Dodd-Frank Act in 2010. In contrast to the Credit Rating Agency Reform Act, the main aim was to reduce conflicts of interest (Altman et al., 2010). Compensation of CRAs is an important factor of how they are governed. Since the 1970s the largest CRAs have changed their business model to mostly issuer-paid. Prior to this change CRAs were solely investor-paid. The possible conflicts of interest that could result from the issuer-pays model have been blamed among other issues for the credit rating crisis of 2008 (White, 2010).

An important regulating factor on the rating market is the demand side for ratings driven by investors request. The so called two-rating norm reduces competition between the large CRA as investors demand at least two ratings. Thereby, a CRA like S&P can be almost certain to be solicited. This situation limits the possibility of issuers to perform rating shopping. The additional ratings can then confirm the first rating (Bongaerts et al., 2012).

Reputational Capital and Regulatory Licenses Based on the regulating history, Partnoy (1999) perceives two conflicting interpretations for the raison d'être of CRAs. These are the reputational capital view of CRAs and the regulatory license view. Following the reputational capital view of CRAs, agencies add additional information to capital markets and reduce information asymmetries by producing accurate ratings. Thus they gain reputation. Investors and issuers are both willing to pay for ratings. Investors gain additional information and issuers can signal superior rating quality.

In contrast, Partnoy (1999) questions the additional informational value of ratings. The regulatory license view focuses on the legally privileged position of CRAs. This position allows them to sell regulatory licenses to issuers. According to Partnoy (1999) the regulatory changes that started in the 1970s gave CRAs the opportunity to establish themselves as gatekeepers and sell regulatory licenses.

3.1.2 Rating Markets

The most important markets for ratings are those for corporate, sovereign, and structured finance products. We here provide an analysis of these three markets that closely follows Matthies (2013b). It is important to asses the aspects of each market that determines how CRAs function on that market.

Corporate Ratings Corporate bonds are the traditional market for CRAs. This market has some important features. Here a large number of issuers provide bonds to a large number international investors (Rosner, 2009). Consequently the market has low liquidity risk (Rosner, 2009). Information regarding the creditworthiness of corporations is publicly available and corporate ratings are issuer-paid. Investors in these markets expect corporations to solicit more than one rating. Furthermore, there are alternative approaches to determine default probability, so investors need not solely rely on credit ratings for their investment decisions[7].

[7] See for example the the Z-Score of Altman (1968)

The market structure of corporate ratings does not make conflicts of interest a pressing issue. The large number of issuing corporations, the demands of investors, combined with the oligopoly of CRAs gives CRAs a strong market position. There is hardly any possibility for CRAs to be substituted by another agency. CRAs do not depend on any individual issuer. This holds, in particular, if investors' investment decisions are not subject to rating restrictions (as is the case for private investors) and ratings are used as confirmation of their own analysis. Therefore, without regard to the issue whether ratings provide additional information, agencies can provide the service of confirming and assisting investor analysis.

The issue of reputational capital is more difficult with respect to the informational content and market reactions. The observation that credit rating changes can often be predicted using market indicators does not necessarily imply that ratings do not provide additional information to markets. The most important aspect being that credit ratings follow a through-the-cycle perspective while market indicators are point-in-time measures. All measures of default, wether market based, credit ratings, or Altman's Z-Score will correlate and overlap to a certain degree with each other but are distinct alternative measures. The argument put forward by Partnoy (1999) that ratings are inaccurate because not all bond spreads in a single rating category are equal is nonsensical. It is based on the assumption that all useful information is contained in credit spreads. Market measures might reflect all the information investors have, yet ratings should not reflect speculation which can be an important element of market prices. If CRAs provide accurate ratings without skewing their judgment through a conflict of interest, they provide the market with information that is free of market features like speculation. The two rating regime and investors expectations about ratings are crucial. In a country where these things will not hold, CRAs might be compromised.

Sovereign Ratings CRAs have rated more and more sovereign governments because of investors' demand and governments seeking access to capital markets. There is a strong correlation between yield spreads and sovereign ratings (Cantor & Packer, 1996). Sovereign rating changes affect the stock markets (Kaminsky & Schmukler, 2002). Typical important variables that are correlated with sovereign ratings are per capita income, GDP growth, inflation rates, and government external debt, as well as measures of economic development and government effectiveness (Cantor & Packer, 1996, Afonso et al., 2011). For sovereign ratings some studies explore differences between CRAs assessments. There is a set of common variables, but also some variables for individual CRAs that cause slight differences for sovereign ratings across CRAs (Hill et al., 2010).

Given that sovereign ratings are unsolicited, CRAs can not be pressured by bond issuers (at least financially). In this market ratings are investor paid. The conflict of interest between issuers and CRAs caused by the issuer-pays model is thus avoided. On this market CRAs would need to rely on their reputation. There are some aspects that could become an issue that might compromise ratings or put their informational value into question. Firstly, it is debatable that ratings add information to markets if they solely rely on publicly available information. Furthermore, the investor-pays model might lead to conflicts of interest between investors and CRAs. CRAs might feel compelled to give out lower ratings and thereby increase yield spreads on issues of low risk and thereby deflate ratings (Richardson & White, 2009).

Structured finance products Structured finance products are supplied by a small number of investment banks while the demand for structured finance products comes mostly from institutional investors who are restricted by NRSRO ratings. As for corporate bonds ratings in this market ratings for structured finance products are issuer-paid. In contrast to the corporate market, bonds of structured finance products often only have one rating (Benmelech & Dlugozs, 2010). Institutional investors only require one rating for regulatory purposes. In contrast to sovereign and corporate ratings, experience with structured finance products given their short history is limited. For structured finance products information is not publicly available and hard to obtain in order to produce alternative and independent assessments from investors. Furthermore, the market for structured finance products has a high liquidity risk (Rosner, 2009).

The default probabilities of the individual mortgages that underlie the structured finance products determine their ratings (Standard & Poor's, 2009). Rosner (2009) points out some crucial points that distinguish structured finance products from sovereign and corporate bonds. An important aspect is that structured finance products have more complex legal structures. The assets of structured finance products are static in contrast to the actively managed and dynamic assets of corporations. The corporate make up of structured finance products are designed according to CRA criteria unlike corporations who determine their own corporate structure.

In 2008 downgrades exceeded upgrades for structured finance products to a far greater degree than they did for corporate ratings at any time. The issues with the highest downgrade ratios were those with only one rating (Benmelech & Dlugozs, 2010). This suggests that rating shopping by issuers of structured finance products caused ratings to be heavily inflated. CRAs put more effort into evaluating default risk if mortgages had more than one rating (Morkoetter et al., 2017). Further evidence comes from Griffin & Tang (2011) who found that CRAs had positively adjusted ratings beyond their own models.

The empirical evidence is highly suggestive and the market structure in the build up to the 2008 crisis indicates that CRAs were involved in selling regulatory licences, and this has become the consensus in the literature (Altman et al., 2010, Calomiris, 2009, White, 2010, Levine, 2010). Bonds with high yields in combination with high ratings are very attractive for institutional investors. Structured finance products are put out by a small number of issuers, so that an individual issuer had power to pressure CRAs. This suggests that rating shopping can occurre, where an issuer would pressure one CRA by threatening to not solicit that CRA as another CRA would provide a better rating. S&P note that after the 2008 crisis many issuers who did not receive a AAA rating were unwilling to publish their ratings (Standard & Poor's, 2012).

3.2 Statistical Methods

We provide a review of the empirical methods used to estimate credit ratings and rating transitions and discuss their properties. Then we present the methods used in this application.

3.2.1 Review of Methods

In order to estimate and forecast credit ratings a number of statistical methods can be employed. OLS estimation and ordered probit methods are the standard approaches. Furthermore, multivariate discriminate analysis, unordered logit, and neural networks can be used to estimate ratings. Additionally, duration and hazard models can be employed to model rating transitions. The underlying assumptions of these differ in some important respects and should be considered while estimating ratings.

Classical Methods

Classical methods can be distinguished with regard to the assumption of how ratings relate to each other. The *ordinary least squares (OLS)* and the *ordered probit* are ordered methods. This assumption acknowledges that ratings are ranked. In contrast, the *unordered logit* and *linear discriminant analysis* methods make no such assumption. This means that neither of these applications 'know' that for instance a AA rating is between a AAA and an A rating. In Ederington (1985) we find a comprehensive theoretical comparison of these four methods. Furthermore, the study includes an empirical application.

In the *OLS* regression application in Ederington (1985) integers are assigned to each rating class. This allows the ratings to be ordered but also makes the implicit assumption that ratings are ordered on an interval scale, meaning that the differences in terms of creditworthiness between rating classes are the same.

The scale problem in the OLS method can be more adequately addressed by the *ordered probit* model. Here, the existence of an unobservable continuous variable (e.g. non default probability or creditworthiness) is assumed. Additionally the method estimates partition points between which the continuous variable falls. Accordingly, the continuous variable corresponds to the observable discrete variable (i.e. rating class). The intervals between the partition points can vary in size. In both the OLS and the ordered probit method variables are assumed to influence credit ratings equally over rating classes. The following two methods disregard the ordered structure of ratings but allow the independent variable effects to vary across rating classes.

The multinomial or *unordered logit* and the *multivariate discriminate analysis (MDA)* models both allow the variable's impact to vary across rating categories. Unordered logit and MDA have different basic assumptions but use the same classification equation. The unordered logit assumes that the error terms have a Weibull distribution. In contrast, in the unordered probit model the, disturbances are assumed to be normally distributed. MDA is a conjoint method in contrast to the unordered logit model. In the logit and probit approach, ratings are modeled as independent variables, while MDA assumes the distribution parameters of the firm's characteristics to be dependent on the bond ratings.

Ederington (1985) provides a comparison of the ordered probit, the OLS, the unordered logit, and the MDA in an empirical application. The unordered logit and ordered probit perform superior to the OLS and MDA methods. Specifically, for in-sample estimation, the unordered logit has the best fit while the out-of-sample prediction is best performed by the ordered logit. The ordered logit method has become the established method in most studies (Blume et al., 1998, Amato & Furfine, 2004, Ashbaugh-Skaife et al., 2006, Matthies, 2013c, Alp, 2013). The ordered structure of the ratings and the specific features of the rating scale are adequately addressed with this approach. However the method disregards possible changes in the influence of individual factors across rating classes. Expanding from the cross sectional approach in Ederington (1985) to panel data, the ordered probit regression assumes a point-in-time perspective (Altman & Rijken, 2006). This issue could cause problems when using the probit method to forecast rating changes as in Amato & Furfine (2004). Other rating transition studies use duration or hazard models (Du & Suo, 2005, Koopman et al., 2008).

Learning Methods

Beyond the approach to estimate ratings in business and economic studies, there are further interests in estimating credit ratings. The classification of a credit rating to a set of firm specific variables is at its core a categorisation problem. Artificial intelligence methods can be 'trained' on samples of ratings and financial data for the purpose of forecasting exercises. Studies that follow such an approach focus on the development and application of alternative methods in computer science. These are methods such as neural networks (Dutta & Shekhar, 1988, Kwon et al., 1997), support vector machines (SVM) (Huang et al., 2004, Härdle et al., 2005, Ahn et al., 2005, Cao et al., 2006, Shih & Chen, 2006, Lee, 2007, Ye et al., 2008), fuzzy logic (Shin et al., 2004, Liu & Liu, 2005), and π-grammatical evolution (Brabazon & O'Neill, 2008).

Rating Transition Methods

Estimating ratings can be done either cross-sectionally or with panel data. Any approach to rating transitions on the other hand needs to have a time series element. An important element of rating transition estimation is the prior rating. Rating transitions forecasts are usually conditioned on the prior rating.

Some of the studies mentioned above (e.g. Amato & Furfine (2004)) use firm specific factors in rating transition predictions. In contrast other studies solely rely on a set of more general predictors. Transition probabilities can be estimated via rating transition matrices. The empirical properties of transition probabilities and what determines them are approached in other studies.

Rating transition matrices are a central feature of modern risk management. Many credit risk models rely on transition probabilities (Lando & Skødeberg, 2002). Here the risk in a given portfolio is measured using the distribution of rating transitions of the underlying bonds.

Nickel et al. (2000) make the assumption that a constant transition probability p_{ij} exists that reflects the likelihood in a given time period that a rating from class i will change to class j. An unconditional estimator of rating transition probabilities can be simply obtained by dividing the number of firms (or bonds) that change from i in time period t to j in $t+1$ with the total number of firms in class i at t. All p_{ij} put into their row and collum position then make up the transition matrix. It is common for transition matrices to be diagonally dominated. This means that the diagonal and those fields closest to it are heavily loaded while fields far from the diagonal converge to zero. It practical terms, this reflects the fact that most ratings do not change and if they change, they tend to only do so over one or a few notches (Nickel et al., 2000, Lando & Skødeberg, 2002, Kim & Sohn, 2008).

Nickel et al. (2000) provides different conditional and unconditional transition matrices. In conditioned transition matrices, the precision of probability estimation is reduced for lower rated bonds. This is due to the fact that there are fewer speculative bonds and higher volatility in the unconditional transition matrices. Furthermore, transition matrices depend on the state of the business cycle, the regional origin of the issuer, and the industry of the issuer.

Transition matrices are applied in numerous rating studies (Lando & Skødeberg, 2002, Kim & Sohn, 2008, Koopman et al., 2008). In Nickel et al. (2000) it is assumed that rating transitions probabilities are produced by a Markov process. This means that all ratings in a category have the same up and downgrade probabilities.

Duration and momentum effects address this basic assumption of transition matrices. More efficient estimates of transition matrices employ continuous-time estimations of transition probabilities (Lando & Skødeberg, 2002, Du & Suo, 2005, Koopman et al., 2008). In contrast to other studies that use yearly transition rates, they consider monthly (Lando & Skødeberg, 2002, Du & Suo, 2005) and even daily (Koopman et al., 2008) observations. Lando & Skødeberg (2002) motivate this procedure by the fact that observations for large transitions are rare or do not occur at all. For example, AAA to D transitions are rarely observed in one year, but a bond or firm can still, in the time of one year, be downgraded from AAA to A and from A to D. In contrast to discrete-time methods, continuous-time methods capture effects like this one. Moreover, they test credit rating transitions for so called non-Markov effects, in particular, duration and momentum. In a Markov chain a transition probability depends solely on the current state an object is in and not on how it reached it or how long it has been in that state. In terms of rating transitions that would mean that the probability of, for instance, an AA rated bond to be downgraded does not depend on how long it has been rated AA (duration) or if it reached its state through a down- or upgrade (momentum).

Lando & Skødeberg (2002) and Du & Suo (2005) find that both momentum and duration effects are affecting rating transitions. In Lando & Skødeberg (2002) a strong downgrade momentum and upgrade momentum for lower rated bonds is shown. This means that a previous downgrade increases the probability to be downgraded again, but the equivalent for upgrades only holds for lower rated bonds. Furthermore, with respect to duration, the longer a firm occupies a rating class the less likely it will be up- or downgraded.

The asymmetric effect measured for momentum fits together with the results of Jorion & Zhang (2007) with regard to the information content of credit ratings. Part of the price adjustment process around rating changes might include the expectations of market participants of further reclassifications. Furthermore, the reluctancy of agencies to issue an upgrade shortly after a downgrade (Altman & Rijken, 2006) and the higher possibility of a further downgrade might contribute to the asymmetric price adjustments that are empirically observed.

3.2.2 Applied Methods

The standard method to estimate and predict credit ratings is the ordered probit model (Blume et al., 1998, Amato & Furfine, 2004). The specific features of the rating scale mentioned above can be addressed by an ordered probit model. In this case, an ordered probit panel approach for cross-sectional time series analysis is employed. Furthermore, ratings are also estimated employing the corresponding default frequencies estimated by Jorion & Zhang (2007) in an OLS regression, as well as an unordered logit panel model and a logit herachical logit model that allows for interaction effects. Specifically, the default frequency approach tests if the assumigly unobservable variable in the probit model can be estimated. Furthermore, the two logit models test the assumption that effects are constant over all rating classes.

Ordered probit estimation

Most studies measuring the determinants of credit ratings (Blume et al., 1998, Amato & Furfine, 2004, Jorion et al., 2009, Alp, 2013) employ the ordered probit approach. Altman & Rijken (2004) use an ordered logit model, which does not differ much in application or result[8].

In the ordered probit approach, the observed discrete ratings are regressed on explanatory variables via an unobserved continuous variable. This unobserved variable is assumed to underly the ratings. The unobserved variable is defined here as the likelihood of not defaulting. Partitioning the range of the unobserved variable then sorts it into the discrete categories. The unobserved variable is a linear function of the observed explanatory variables.

The ratings of firm n at the end of year t is denoted as $R_{n,t}$ and encoded as 5 for AAA down to 1 for all non-investment grade ratings, so there are $K = 5$ categories[9]. Then, $y_{n,t}$ is the unobservable variable 'likelihood of not defaulting' or 'creditworthiness' that underlies the $R_{n,t}$, for which μ_k for $k = 1, \ldots, K-1$ are the partition points independent of n and t. The model for $y_{n,t}$ is then:

$$y_{n,t} = \alpha_t + \boldsymbol{X}_{n,t}\boldsymbol{\beta}' + \epsilon_{n,t}, \tag{3.1}$$

where α_t is the intercept for year t[10], $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients, $\boldsymbol{X}_{n,t}$ is a $1 \times p$ vector of the n-th firm's specific risk factors in year t, and $\epsilon_{n,t}$ is a Gaussian error term with $\mathbf{E}[\epsilon_{n,t}|\boldsymbol{X}_{n,t}] = 0$.

[8] See Ederington (1985) for a more detailed discussion. The logit and probit method differ in the error distribution. The probit approach assumes a Gaussian distribution, while the logit assumes a logistic distribution.

[9] See Table 3.1 for more details.

[10] As in Blume et al. (1998) we set α_1 to zero.

The most probable rating category for any observation given $\boldsymbol{X}_{n,t}$ is then the estimated $\hat{y}_{n,t}$:

$$Pr(R_{n,t} = k|\theta) = \begin{cases} Pr(\hat{y}_{n,t} \geq \mu_4|\theta) & \text{for} \quad k = 5 \\ Pr(\mu_k > \hat{y}_{n,t} \geq \mu_{k-1}|\theta) & \text{for} \quad k = 2,3 \\ Pr(\mu_1 > \hat{y}_{n,t}|\theta) & \text{for} \quad k = 1. \end{cases}$$

Estimation is done via maximum likelihood (ML) estimation. Blume et al. (1998) allow for heteroscedasticity in the error terms. The $\epsilon_{n,t}$ can be assumed to be autocorrelated in panel data. Therefore, the ML-estimates are consistent but the standard errors are not. Blume et al. (1998) reestimate the $\boldsymbol{\beta}$ covariance matrix with the method of West & Newey (1987) to obtain consistent standard errors. Jorion et al. (2009) calculate clustered standard errors adjusted for the clustering of firms. Here, similar to Alp (2013) we correct heteroscedasticity and autocorrelation via the Newey-West method (West & Newey, 1987)[11].

Unordered logit

A critical assumption in the ordered probit estimation is the stable effect of variables across all rating classes. To test this assumption we follow Ederington (1985) who suggest an unordered logit to estimate credit ratings as an alternative. Furthermore, we treat ratings as hierarchical data and allow for interaction effects. In the case of hierarchical data, we follow Liao (1994). Both the ordered probit and the unordered logit are able to incorporate the ordinal structure of credit ratings. The significant differences are that the unordered logit can allow the coefficients of the variables to differ over rating classes but it disregards the ordered structure of credit ratings. Hence, the choice between ordered probit and unordered logit is one between structure and flexibility (Ederington, 1985). The unordered logit model can therfore be employed to test the stability of the determinants of the coefficients. The results of Ederington (1985) suggest that coefficients vary over rating classes. Yet, the flexibility reduces the predictive power compared to an ordered approach.

The unordered logit is mostly constructed like the ordered probit with the exception that a logit estimation assumes a logistic distribution for the error terms $\epsilon_{n,t}$.

Employing default rates

Here we develop a method based on the idea that rating categories relate to specific default categories. Based on empirical observations ratings are translated into corresponding default rates D_k estimated by (Jorion & Zhang, 2007). Then, $1-D_k$ represents

[11]See Campbell et al. (1997, p.530,ff) for more details.

the likelihood of not defaulting. Using historical default rates we can obtain an estimate of the unobserved variable. Taking as unobserved variable the probit of this likelihood, one may perform an ordinary least squares (OLS) regression on the individual default rates according to the observed rating. Then one may use the arithmetic half-way distances between the default frequencies as thresholds in order to assign the estimated likelihood of not defaulting values to credit ratings.

It is possible to perform the regression on either the default probabilities or on the corresponding values of the inverse of the standard normal cumulative distribution function, i.e. the probit values. Using the probit values has the advantage that the dependent variable has a range between minus infinity and infinity. Using the actual default rates has the advantage that the coefficients of the regression can be easily interpreted as a change in probability of not defaulting.

The model then has the same structure as in (3.2.2) except that the $y_{n,t}$ are now observable. The partition points μ_k can then be calculated as $(1-D_{k-1})+\frac{1}{2}(D_k-D_{k-1})$, where the D_k are taken from Tab. 3.1[12]. Estimation of the model is then performed as in Greene (2003, p.283 ff).

Using default frequencies, Jorion & Zhang (2007) show that the price-adjustment processes is better explained if they are conditioned on the prior rating[13]. Similarly, if credit ratings behaved like other default measures that employ a more point-in-time perspective such as Merton-type methods, this approach should improve rating prediction. Yet agencies insist that they use a through-the-cycle approach. This is motivated so as to ensure a degree of rating stability (Altman & Rijken, 2006). Therefore, the default measure approach should produce a rating transition matrix with larger transition probabilities (Altman & Rijken, 2004, Kim & Sohn, 2008). The transition frequencies of this method should hence produce an upper bound and the actual rating transition matrix a lower bound for the purpose of model evaluation. The estimated transition probabilities of an ordered probit or an unordered logit should lie between these values.

3.3 Data

This sections details the set of ratings available from the 100 selected firms of the US, the UK, Japan, Germany, France, Canada, and Australia. Following this the explanatory variables and their possible effects on credits ratings are discussed. Thereafter some sample properties are examined.

[12]Here, the fitted default rates are employed.
[13]Prior research had failed to find significant price reactions after upgrades.

3.3.1 Ratings

S&P ratings range from AAA (the highest debt quality) to D (firms in default). In order to indicate further distinctions in creditworthiness ratings from AA to CCC can furthermore have a '+' or a '-'. Investment grade ratings are those from AAA to BBB-, while BB+ and lower ratings are called speculative ratings. The cutoff from BBB- and BB+ is important. Institutional investors, for instance, may not purchase bonds rated BB+ or lower (Hill, 2004). We group all AA, A, BBB ratings together for the purpose of estimation. Furthermore, all non-investment grade ratings (BB, B, and CCC, and D) are grouped together. Hence we have $K = 5$ rating categories.

Table 3.1 depicts each rating category and its 10 year default frequency as calculated by (Jorion & Zhang, 2007) as well as fitted default frequencies. The frequencies are fitted so that for instance the rate for AAA is lower than that of AA+ and they can be interpreted as default probabilities. We find differences in default frequencies increase down the rating scale. The difference between BBB- and BB+ is tenfold larger than the difference from AAA to AA- of four notches.

Rating	D_k	fittet D_k	$R_{n,t}$
AAA	0.005	0.003	5
AA+	0.004	0.005	4
AA	0.007	0.007	4
AA-	0.012	0.010	4
A+	0.016	0.015	3
A	0.017	0.022	3
A-	0.023	0.031	3
BBB+	0.047	0.045	2
BBB	0.055	0.065	2
BBB-	0.109	0.092	2
BB+	0.140	0.128	1
BB	0.187	0.177	1
BB-	0.265	0.238	1
B+	0.315	0.313	1
B	0.396	0.400	1
B-	0.492	0.493	1
CCC	0.572	0.586	1
D	1	1	1
Time period	1981 – 2002		

Table 3.1: Default frequencies and fitted default frequencies of rating categories of Standard & Poor's (Source Jorion & Zhang (2007)) and the classification for model estimation according to (3.2.2).

In this chapter S&P foreign long term issuer ratings from 1990 to 2009 are collected[14]. This sample excludes banks and other financial firms in contrast to Blume et al. (1998) and Amato & Furfine (2004). This distinction may lead to different empirical results, as financial companies have different capital structures. The sample then comprises the ratings of the 100 largest US, UK, Japanese, German, French, Canadian, and Australian non-financial publicly tradable stock companies of 2005.

The distribution of rating classes in this sample across time are depicted in Table 3.2. Matthies (2013c) only employed the US firms. The data set therefore contained far less lower rated firms. Here the rating distribution is more like in Amato & Furfine (2004), Jorion et al. (2009), and Alp (2013). Sample selection is very important in this regard. Of the 100 largest firms in the US most are rated at some point. In the other economies, the fraction of rated firms is lower. For these six countries Germany has the lowest fraction with 25 of 100 firms rated. The firms for which there was financial data available and that had no rating are collected in the not rated (NR) column. A firm's size ceteris paribus improves ratings. Jorion et al. (2009) note that in their sample on average investment firms are nine times as large as speculative firms. Therefore the choice of the 100 largest firms in 2005 might result in fewer lower rated firms.

Table 3.3 displays how many firms in any of the six countries had a S&P rating over countries over time. The highest number of firms are in the US at all times. In Japan beginning in the new millennia the number of ratings increases massively. From 2002 to 2003 the number of rated firms increases from 29 to 78. After 2008 the number decreases down to 46. In contrast the number of ratings increases continuously in the Canada, the UK, France, Australia, and Germany.

3.3.2 Explanatory Variables

Three different groups of variables can be used to determine credit ratings. Firstly, financial data and ratios that measure factors such as leverage, liquidity, and profitability (see e.g. Blume et al. (1998)) are essential. Secondly, corporate governance characteristics will feature in an assessment of credit worthiness. Studies that employ corporate governance characteristics aim to measure the effects of corporate governance mechanisms on credit ratings that influence principal agent problems between management and stockholders (Bhojraj & Sengupta, 2003) and the redistribution of wealth from bondholders to shareholders (Ashbaugh-Skaife et al., 2006). Thirdly, macroeconomic variables that measure business-cycle effects and fundamental changes influence ratings. CRA insist that they employ a through-the-cycle approach that estimates the long term credit worthiness of a firm independent of short term business cycle-effects.

[14]Issuer and issue ratings can have different determinants. A issue rating determines the creditworthiness of one specific obligation, while the issuer rating reflects the firm's creditworthiness in general.

Rating Distribution						
	AAA	AA	A	BBB	BB—D	NR
1990	15	21	27	3	9	50
1991	16	22	30	8	9	102
1992	18	25	38	10	11	99
1993	18	28	44	11	14	106
1994	16	30	49	14	13	113
1995	17	30	54	27	12	113
1996	17	32	57	35	12	218
1997	16	36	66	45	10	328
1998	14	39	71	50	9	342
1999	12	42	82	59	14	336
2000	11	42	87	79	18	322
2001	10	41	89	106	27	302
2002	9	35	95	109	35	310
2003	8	40	97	129	67	269
2004	6	41	99	141	70	267
2005	6	40	104	143	72	269
2006	6	39	103	151	68	267
2007	6	36	100	148	60	278
2008	6	35	107	141	59	275
2009	3	30	102	130	59	289

Table 3.2: Distribution of US, UK, Japan, Germany, France, Canada, and Australia sample ratings 1990 – 2009.

				Distribution				
Year	Australia	Germany	France	Japan	Canada	USA	UK	Sum
1990	2	0	2	9	2	50	10	75
1991	2	1	3	10	3	56	10	85
1992	2	1	5	12	5	65	12	102
1993	5	2	5	15	7	67	14	115
1994	8	2	7	16	7	68	14	122
1995	13	4	9	16	12	72	14	140
1996	15	4	11	17	15	74	17	153
1997	16	5	14	21	21	77	19	173
1998	17	5	16	20	24	80	21	183
1999	18	7	18	23	27	86	31	210
2000	19	15	22	24	31	90	37	238
2001	22	17	26	26	45	95	43	274
2002	22	18	27	29	45	96	47	284
2003	24	19	27	78	49	98	47	342
2004	25	21	30	87	51	97	47	358
2005	27	24	32	86	52	97	48	366
2006	29	23	32	88	51	97	48	368
2007	28	23	32	76	49	96	46	350
2008	28	23	31	75	49	96	46	348
2009	29	22	32	46	51	95	49	324
Sum	351	236	381	774	596	1652	620	4610

Table 3.3: Yearly distribution of ratings over countries.

Nevertheless, credit ratings are found to correlate with the business-cycle (Amato & Furfine, 2004, Nickel et al., 2000).

Here, 8 variables extracted from financial statements that fall into the first category are employed. We use pretax interest coverage, operating income to sales, a long term debt ratio, a total debt ratio, and total assets as in Blume et al. (1998). Furthermore, 2 ratios suggested by Altman (1968) to predict corporate bankruptcy and later used in credit rating studies by Altman & Rijken (2004) and Kim & Sohn (2008) are employed. These are the retained earnings to total assets and the earnings before interest and taxes (EBIT) to total assets ratios. Furthermore, we use the return on assets ratio (ROA) as additional variable.

3.3.3 Definitions and Possible Effects

We now discuss the individual ratios and their possible impact on ratings. Pretax interest coverage (IC) is operating income after depreciation plus interest expenses divided by interest expenses. A decline in interest expenses should improve ratings if operating income after depreciation is positive. Operating income should be positively related to ratings. Blume et al. (1998) highlight the non-linear influence of interest coverage, i.e. for small values of interest coverage changes are relevant while for large values changes become negligible. Furthermore, negative values are not meaningful as a negative operating income and a decline in interest expenses would cause a positive development at the margin although the variable would become negative (Amato & Furfine, 2004). Accordingly, we follow Blume et al. (1998), Amato & Furfine (2004), Jorion et al. (2009), and Alp (2013), and set all negative values to zero and set all values larger than 100 to 100. Furthermore, as in Blume et al. (1998) in order to account for non-linear effects IC, is partitioned into four variables as in Table 3.4.

	IC_1	IC_2	IC_3	IC_4
$IC \in [0, 5)$	IC	0	0	0
$IC \in [5, 10)$	5	$IC - 5$	0	0
$IC \in [10, 20)$	5	5	$IC - 10$	0
$IC \in [20, 100]$	5	5	10	$IC - 20$

Table 3.4: Non linear partitioning of Interest Coverage.

The operating income to sales ratio (OI NS) is the operating income before depreciation divided by net sales. This ratio can be used as a proxy of both earnings and cash flow. A firm's earnings margin indicates if it can generate the necessary cash to service its debt obligations. The value of a firm's assets are measured by its earnings. Here we exclude extreme values as values that are larger than 1 in absolute terms are set to -1 and 1 accordingly.

The long term debt ratio (LTD/TA) and the total debt ratio (TD/TA) are measures of leverage. LTD/TA is defined as long term debt divided by total assets while TD/TA is defined as the sum of long term debt, debt in current liabilities, and short term debt divided by total assets. Issuer ratings aim to measure a firm's ability to serve all its financial responsibilities but are also linked to the issue rating of unsecured long-term debt. Both ratios are typically negatively related to credit ratings.

Firm size (TA) is considered to be a measure of business risk (Amato & Furfine, 2004). Firm size will correlate with firm age, more diversified product lines and a higher variety of revenues. Altman & Rijken (2004) use firm age as a determinant for similar reasons. Therefore, firm size should be positively related to credit ratings. Furthermore, by modelling the residual variance as a function of firm size, Blume et al. (1998) find that financial ratios are more informative for larger firms. This result implies that variables for larger firms are more stable over time or alternatively that relevant variables are missing in their analysis. Firm size measured by total assets is deflated using a CPI taken from the US Department of Labour dated 2011.04.15, downloaded 2011.04.28: `ftp://ftp.bls.gov/pub/special.request/cpi/cpiai.txt` . Then the natural logarithm is taken.

Following Altman's Altman (1968) default prediction model (Altman & Rijken, 2004, Kim & Sohn, 2008), we furthermore include a retained earnings to total assets ratio (RE/TA). Retained earnings serve as proxies for the historic profitability of a firm. They furthermore implicitly measure the age of a firm, as older firms usually have a higher retained earnings in relation to their size. Retained earnings can be used in less profitable times to ensure the payment of obligations. Hence, they are positively related to credit ratings. Indeed Kim & Sohn (2008) find that the retained earnings total assets ratio is strongly related to future upgrades.

A proxy of the firm's current profitability is the earnings before interests and taxes to total assets (EBIT/TA) ratio (Altman & Rijken, 2004). It can measure the true productivity of a firm's assets (Altman, 1968). A fundamental factor of a firm's existence including its creditworthiness is dependent on the earnings power of its assets. This ratio should therefore be positively related to credit ratings.

For individual firms the proxy of profitability used by the rating agency may differ. A further measure of profitability is Return on assets (ROA). Furthermore, we include the five year arithmetic average of ROA (ROA 5yr) to test if we can thereby capture changes in the fundamental creditworthiness of firms.

Certain studies (e.g. Blume et al. (1998), Bhojraj & Sengupta (2003), Amato & Furfine (2004)) use market based measures of risk, i.e. the market β, as further determinant of credit ratings. The measured relationship between β and credit ratings varies if different characteristics of corporate governance are used (Bhojraj & Sengupta, 2003). We therefore conclude that the effect of market based measures and credit ratings is not yet fully understood. Hence they are omitted.

Three measures of the business cycle are used by Amato & Furfine (2004) to test the procyclicality of credit ratings. These are the NBER recession indicator, the real and potential GDP output gap, and a discrete measure of that same output gap. In Matthies (2013c) only the actual GDP output gap (GDP GAP) was employed, as it is the only business cycle measure with a significant impact in the Amato & Furfine (2004) study. As the GDP growth gap is positively correlated with real GDP, GDP GAP should be positively related to credit ratings. Here, unlike in Matthies (2013c) no GDP output gap is used as this is an international study with credit ratings from multiple countries. To measure a possible individual country risk effect dummy variables are employed for each non-US country.

3.3.4 Sample Properties

The mean values of the explanatory variables for each combined rating class (AAA, AA, A, BBB, and BB-D) and the not rated firms (NR) are given in Table 3.5. Examining the mean values for non-rated firms, we find that on average these firms are smaller than rated ones and have an average OI/NS ratio similar to that of non-investment grade firms. Furthermore, all earnings ratios (RE/TA, EBIT/TA, and ROA) are higher than those of non-investment grade firms. They have a LTD/TA ratio lower than that of AAA rated firms. This might be due to interest payments being higher for non-rated firms than for rated firms. IC and TD/TA values are similar to the mean values of AAA and A rated firms respectively.

There is a clear distinction between investment grade firms and non-investment grade firms for all variables. For the four combined investment grade categories mean values do not always increase or decrease as expected. There is a clear upward trend for TA so that average firm size increases for each combined class of rating categories. This is also the case for IC, RE/TA, and EBIT/TA except that the mean AAA value is smaller than the mean AA value.

Other firm specific factors may also help to determine differences in creditworthiness. Macroeconomic factors must be ruled out, as they would affect all firms equally. Yet differences in corporate governance characteristics like those employed by Bhojraj & Sengupta (2003), Ashbaugh-Skaife et al. (2006) could be more adequate.

3.4 Results

The estimation of corporate credit ratings using the ordered probit method assumes that ratings have ordered and ordinal structure. Using an international sample of firms

	Variables							
	IC	OI/NS	LTD/TA	TD/TA	TA	RE/TA	EBIT/TA	ROA
AAA	13.539	0.105	0.217	0.292	12.848	0.246	0.094	0.066
AA	16.346	0.119	0.215	0.28	12.364	0.263	0.098	0.062
A	14.6	0.116	0.197	0.266	11.644	0.238	0.094	0.066
BBB	11.708	0.118	0.241	0.31	11.149	0.186	0.09	0.066
BB-D	6.418	0.08	0.301	0.377	10.77	0.074	0.053	0.035
NR	13.588	0.083	0.182	0.265	9.324	0.121	0.084	0.057

Table 3.5: Sample variables' means conditioned on rating class.

we test if the US CRA S&P might have a home bias or if S&P is perhaps more sceptical of US accounting data. Keeping the effects of the explanatory variables constant over time, we can test for structural shifts in the global assessment of creditworthiness.

3.4.1 Country Risk Factors, Default Rates, and Structural Shifts

Firstly, we compare the estimation results of the ordered probit with the OLS estimation using default rates. Then, with an unordered logit and yearly regressions, we test the stability of the effects over rating classes and across time.

In Table 3.6 the results of four different regressions are shown. We consider two alternative ordered probit models and two alternative OLS models. For both statistical methods we examine the model with and without country dummies[15]. We test for country specific risk factors and international distinctions to US corporate risk assessments.

For interest coverage we find the non linear effect in all models. The effect for small values (IC_1) is stronger. For larger values we find some negative effects. These are not significant in the probit models. The operating income to net sales ratio has an unexpected negative effect. The effects are not significant in the probit models.

The two leverage ratios for long term debt and total debt have opposite effects. While LTD/TA has a counterintuitive positive effect, TD/TA has the expected negative effect. The effects are stronger in the models that use country dummies. These opposite effects for the total and long term ratios can also be observed in Blume et al. (1998), Amato & Furfine (2004), Jorion et al. (2009), and Alp (2013). In the other studies, these effects are the other way around.

Firm size has the expected positive effect in all models. The earnings ratios should have a positive effect on credit ratings. In the probit regressions the retained earnings

[15] Only the coefficients of the variables are shown. The yearly constants are presented in Table 3.7 and in Figure 3.1 below.

	Probit		Probit C		OLS			OLS C		
	Coefficient	CorrSE	Coefficient	CorrSE	Coefficient	STD b	t-stat	Coefficient	STD b	t-stat
IC1	0.045	0.012	0.036	0.013	0.002	0.001	2.984	0.003	0.001	4.468
IC2	-0.032	0.017	-0.029	0.0177	0.001	0.001	1.11	0.004	0.001	4.018
IC3	0.037	0.008	0.036	0.008	-0.001	0.001	-1.939	0.000	0.001	0.133
IC4	-0.003	0.001	-0.003	0.001	0.000	0.000	3.939	0.001	0.000	6.014
OI	-0.113	0.204	-0.257	0.218	-0.106	0.014	-7.581	-0.199	0.013	-15.415
LTD	0.397	0.285	1.534	0.303	0.116	0.019	6.013	0.102	0.018	5.703
TD	-1.176	0.26	-2.133	0.281	0.027	0.018	1.56	0.063	0.017	3.776
TA	0.333	0.015	0.39	0.018	0.038	0.001	50.493	0.046	0.001	63.84
RE	0.297	0.0934	0.521	0.096	-0.013	0.006	-2.116	0.007	0.006	1.215
EBIT	-0.404	0.582	0.261	0.59	0.132	0.04	3.324	0.159	0.036	4.433
ROA	0.356	0.65	0.096	0.661	0.076	0.045	1.711	-0.034	0.04	-0.852
AUS			0.565	0.084				0.095	0.005	19.522
CAN			-0.124	0.074				0.088	0.004	21.129
FR			0.925	0.077				0.098	0.004	22.163
GER			0.204	0.088				0.02	0.005	3.67
JAP			0.484	0.059				-0.01	0.004	-2.924
UK			0.367	0.058				0.029	0.004	8.173

Table 3.6: Coefficients of ordered probit panel and OLS regressions with fixed time effects. The first two columns represent the coefficients and the corrected Newey-West errors of the ordered probit model without country dummies. Columns three and four represent ordered probit model with country dummies. Columns five, six, and seven contain the coefficients, their respective standard errors, and their t-statistic of the OLS model without country dummies respectively. Columns eight to ten represent the OLS model with country dummies.

ratio has the expected positive effect. In the OLS regressions the effect is not significant with country dummies and negative and significant without country dummies. The EBIT ratio is not significant in the probit models, while it has the expected significant positive effect in the OLS models. ROA has no significant effect in any regression.

The country dummies are positive and significant for Australia, France, Germany, and the UK in both the probit and the OLS model. For Canada, the country dummy has a negative but not significant effect in the probit model and a positive significant effect in the OLS model. For Japan there is also a switch in signs from the probit to the OLS model, in this case from significantly positive to significantly negative.

Although rating agencies apply a through-the-cycle approach, studies empirically observe the procyclicality of credit ratings. Blume et al. (1998) suggest that the decreasing constants of the panel model indicate that agencies standards have become more stringent over time. When discussing their results, the use of a measure of the business cycle as a control variable therefore seems appropriate. Jorion et al. (2009) argue that the effect measured by Blume et al. (1998) disappears if a firm specific measure for accounting quality is used. More specifically, accounting quality of investment grade firms seems to have declined over time.

We analyse the evolution of the constants from the ordered probit and the OLS panel models in comparison to Blume et al. (1998), Amato & Furfine (2004), and (Jorion et al., 2009). The development of the four groups of constants from the unordered logit estimation are shown in Section 3.4.2, and compared with the findings of Jorion et al. (2009), which stated that the constants only had a negative trend for speculative ratings.

The yearly constants of the probit and OLS regressions are displayed in Table 3.7. The dummies show a negative trend and decrease in time from zero and become significantly negative. The first year (1990) dummy is set to zero. Figure 3.1 displays the evolution of the dummy variables of our sample along with the results of Blume et al. (1998), Amato & Furfine (2004), and Jorion et al. (2009). The yearly dummies show a similar trend to those of the other studies and continue the trend.

The dummies of the probit model with country dummies are more negative than the year dummies of the probit model without country dummies. The Blume et al. (1998) sample contains only investment grade ratings. In contrast, Amato & Furfine (2004) also include speculative ratings, while the study of Jorion et al. (2009) performs separate regressions for investment and speculative ratings. The yearly constants of both probit models and the Matthies (2014) dummies behave similar to the investment grade estimation of Jorion et al. (2009).

The negative trend in the yearly dummies is seen as evidence by Blume et al. (1998) that agencies' standards have become more stringent. Using a yearly individual

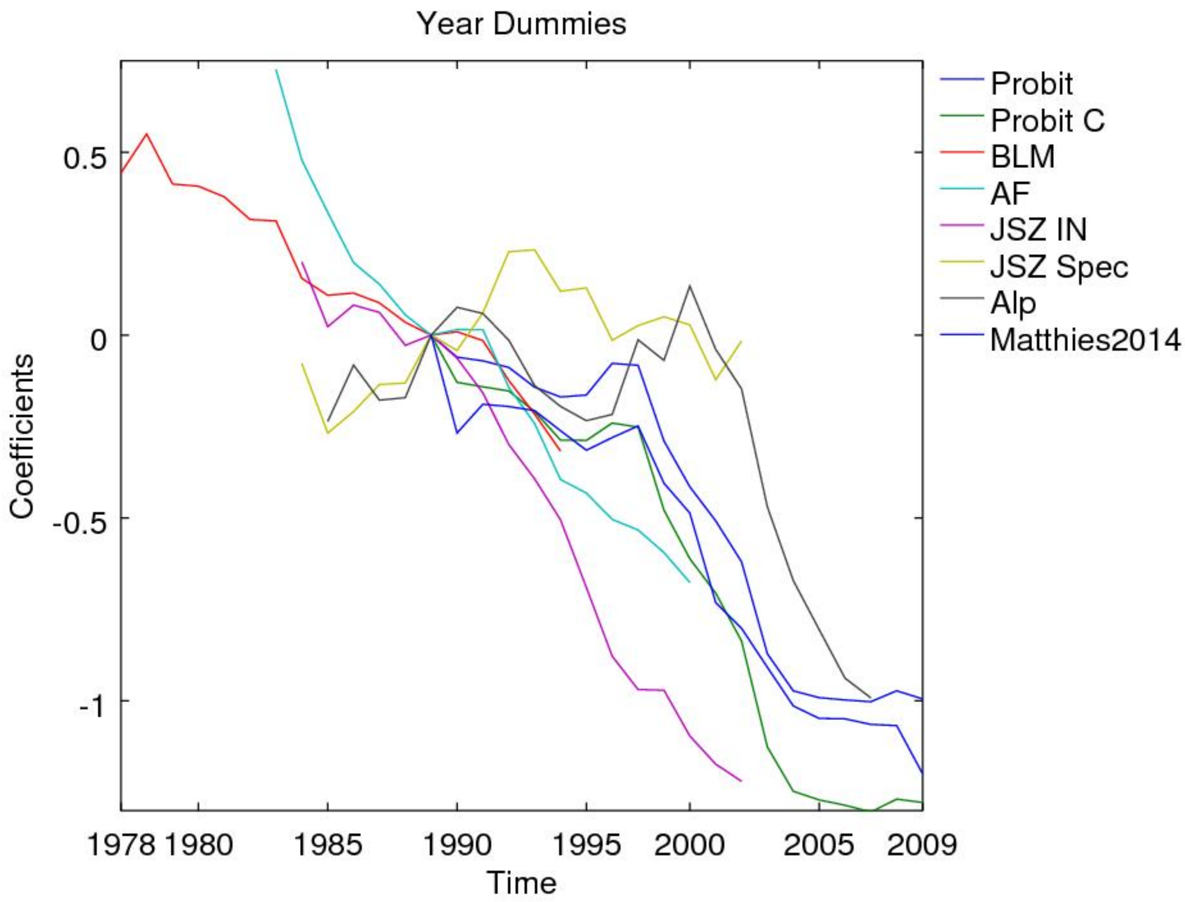


Figure 3.1: Evolution of probit constants over time from different studies. Probit: The Probit model, Probit C: The probit model with country dummies, BLM: Blume et al. (1998), AF: Amato & Furfine (2004), JSZ IN: Probit regression for investment grade ratings of Jorion et al. (2009), JSZ Spec: Probit regression for speculative grade ratings of Jorion et al. (2009), A: Alp (2013), and Matthies2014: Matthies (2014): The sample in this study.

	Year dummies									
		corr SE		corr SE		t-stat	STD B		t-stat	STD B
1990	0	-	0	-	0	-	-	0	-	-
1991	-0.06	0.213	-0.129	0.214	-0.457	0.013	-36.45	-0.567	0.012	-48.343
1992	-0.07	0.206	-0.141	0.207	-0.455	0.012	-38.151	-0.569	0.011	-50.538
1993	-0.088	0.204	-0.153	0.205	-0.456	0.012	-38.847	-0.569	0.011	-51.318
1994	-0.142	0.202	-0.209	0.203	-0.459	0.012	-39.455	-0.573	0.011	-52.043
1995	-0.169	0.196	-0.288	0.197	-0.459	0.011	-41.202	-0.576	0.011	-54.197
1996	-0.163	0.191	-0.288	0.193	-0.455	0.01	-42.632	-0.58	0.01	-56.101
1997	-0.077	0.183	-0.241	0.185	-0.462	0.01	-46.202	-0.588	0.01	-59.876
1998	-0.083	0.181	-0.252	0.183	-0.463	0.01	-46.864	-0.591	0.01	-60.617
1999	-0.289	0.179	-0.478	0.181	-0.465	0.01	-48.034	-0.596	0.01	-61.808
2000	-0.415	0.176	-0.611	0.179	-0.462	0.009	-49.121	-0.593	0.01	-62.913
2001	-0.508	0.174	-0.705	0.177	-0.455	0.009	-49.969	-0.592	0.01	-63.836
2002	-0.62	0.174	-0.836	0.177	-0.458	0.009	-50.201	-0.596	0.01	-64.03
2003	-0.871	0.172	-1.126	0.175	-0.463	0.009	-51.562	-0.597	0.01	-65.166
2004	-0.973	0.172	-1.247	0.175	-0.467	0.009	-51.889	-0.602	0.01	-65.403
2005	-0.991	0.172	-1.271	0.175	-0.466	0.009	-51.98	-0.601	0.01	-65.451
2006	-0.998	0.172	-1.285	0.175	-0.465	0.009	-51.921	-0.6	0.01	-65.384
2007	-1.002	0.172	-1.304	0.176	-0.471	0.009	-51.823	-0.608	0.01	-65.326
2008	-0.973	0.172	-1.269	0.176	-0.47	0.009	-51.771	-0.609	0.01	-65.332
2009	-0.995	0.174	-1.278	0.177	-0.468	0.009	-50.944	-0.609	0.01	-64.6

Table 3.7: Year dummies of the probit and OLS panel estimation with and without country dummies from 1990 – 2009, where 1990 is set to zero.

estimations, they show that the coefficients of the explanatory variables are stable over time[16]. In contrast, the main results stem from a panel model. Based on these results Blume et al. (1998) come to the conclusion that CRAs have become more critical when they judge a firms default probability. Furthermore, Amato & Furfine (2004) show that these results are robust to the inclusion of measures of the business cycle.

In their discussion, Blume et al. (1998) point out that the literature does not suggest further important determinants of credit ratings at the time of their study. They therefore dismiss the critique that the trend might indicate the development of factors which were omitted from their study. Yet, notably corporate governance characteristics have since then been shown to determine credit ratings (Bhojraj & Sengupta, 2003, Ashbaugh-Skaife et al., 2006). It is important to point out that both Bhojraj & Sengupta (2003) and Ashbaugh-Skaife et al. (2006) only use cross sectional data and therefore do not asses a possible impact of corporate governance mechanisms on rating stability.

Alternatively, Jorion et al. (2009) interpret the negative trend as CRA reaction to declining accounting standards. CRAs assessments are critically reliant on the quality of firms accounting data, as the agencies do not themselves collect data (Hill, 2004). Jorion et al. (2009) higlight two important aspects of their results. Firstly, the negative

[16] We perform this assumption in Section 3.4.3

trend is restricted to investment grade ratings (see Figure 3.1). Secondly, the negative trend disappears after including a measure of firm specific accounting quality. They conclude that the observed decline is determined by changes in accounting quality.

3.4.2 Stability Over Rating Classes

	Unordered Logit Regression			
	BB-BBB	BB-A	BB-AA	BB-AAA
IC_1	0.134	-0.036	-0.051	-0.058
(P-Value)	(0.076)	(0.602)	(0.45)	(0.413)
IC_2	-0.077	-0.043	-0.048	0.036
(P-Value)	(0.495)	(0.672)	(0.632)	(0.738)
IC_3	0.113	0.135	0.03	0.013
(P-Value)	(0.035)	(0.003)	(0.498)	(0.786)
IC_4	0.015	-0.005	0.008	0.015
(P-Value)	(0.152)	(0.523)	(0.249)	(0.055)
OI/NS	-5.446	-5.464	-4.982	-6.573
(P-Value)	(0.000)	(0.000)	(0.000)	(0.000)
LTD/TA	7.569	7.976	6.82	-1.63
(P-Value)	(0.000)	(0.000)	(0.000)	(0.392)
TD/TA	-8.017	-7.143	-4.695	3.028
(P-Value)	(0.000)	(0.000)	(0.001)	(0.079)
TA	1.859	1.361	0.856	0.588
(P-Value)	(0.000)	(0.000)	(0.000)	(0.000)
RE/TA	2.391	1.431	1.093	0.3673
(P-Value)	(0.000)	(0.02)	(0.07)	(0.568)
$EBIT/TA$	-7.337	-9.544	-8.434	-15.824
(P-Value)	(0.022)	(0.001)	(0.003)	(0.000)
ROA	10.675	10.639	11.646	19.144
(P-Value)	(0.001)	(0.000)	(0.000)	(0.000)
AUS	-11.942	-14.323	-14.943	-14.283
(P-Value)	(0.975)	(0.97)	(0.969)	(0.97)
CAN	-13.396	-14.032	-13.703	-10.937
(P-Value)	(0.966)	(0.965)	(0.965)	(0.972)
FR	3.855	2.865	1.764	0.435
(P-Value)	(0.000)	(0.000)	(0.000)	(0.251)
GER	-14.577	-15.465	-15.856	-16.259
(P-Value)	(0.971)	(0.969)	(0.968)	(0.967)
JAP	-0.091	-0.144	-0.562	-2.418
(P-Value)	(0.808)	(0.679)	(0.097)	(0.000)
UK	1.088	-0.146	-0.798	-0.97
(P-Value)	(0.004)	(0.652)	(0.011)	(0.004)

Table 3.8: Unordered Logit estimation.

We now test the assumption that effects over rating classes are constant. An unordered logit panel estimation and a estimation in which ratings are treated as hierarchical is conducted for this purpose. Table 3.8 presents the coefficient results of the unordered logit estimation and Table 3.9 presents the results of the hierarchical logit estimation[17]. Here the $K-1$ groups of coefficients determining ratings between the base class BB and all other classes are displayed in Table 3.8 and the $K-1$ groups of coefficients determining ratings between rating classes are presented in Table 3.9.

As can be seen in Table 3.8 and Table 3.9 interest coverage has mostly no significant effect. The OI/NS has no effect in the hierarchical model but a counterintuitive negative effect in the nominal model. For the unordered logit and the hierarchical model we find stable effects for most variables. Variables have the same stable significant effects as in the ordered probit estimations. We find significant effects for the proxies of profitability over almost all rating classes. The coefficients of the retained earnings ratio and ROA are both positive in the unordered setting. In contrast, the EBIT ratio has an unexpected negative effect like OI NS.

In Ederington (1985) results of a similar analysis indicate that coefficients might not necessarily be constant over rating classes. Yet, in other studies variables have stable effects Blume et al. (1998), Jorion et al. (2009). A functional coefficient model might perhaps address such issues. It might solve the problem posed by certain variations of effects across rating classes found in the hierarchical model. Moreover financial ratios are more informative for larger firms than for smaller ones (Blume et al., 1998) and corporate governance characteristics have a larger effect for lower rated firms (Bhojraj & Sengupta, 2003). As mentioned above, larger firms have on average better ratings. In a functional coefficient model firm size might therefore be used as a factor to determine the coefficients of credit rating with firm size.

In Table 3.10 the constants of the unordered logit estimation are displayed and Table 3.11. We find a negative trend for all sets of coefficients. This suggests that effects are mostly stable over rating classes.

3.4.3 Stability Over Time

The ordered probit and the OLS estimation made the basic assumption that coefficients are constant over time. In order to test this we perform yearly ordered probit regressions independently of each other. The results of these estimations are displayed in Table 3.12 without country dummies and in Table 3.13 with country dummies.

In Table 3.12 we can see that the constants become more negative over time. Prior to 1996 there are few significant effects due to the small sample size. TA has a con-

[17] As above for the ordered probit panel estimation, we discuss the results of the constants in detail in Table 3.10 and Table 3.11.

Hierarchical Logit Regression				
	BB-BBB	BBB-A	A-AA	AA-AAA
IC_1	0.175	0.012	-0.0147	-0.036
(P-Value)	(0.000)	(0.672)	(0.721)	(0.645)
IC_2	-0.048	-0.022	-0.042	0.043
(P-Value)	(0.4)	(0.586)	(0.431)	(0.715)
IC_3	0.045	0.099	0.0	0.036
(P-Value)	(0.155)	(0.000)	(0.999)	(0.495)
IC_4	0.012	-0.014	0.002	0.027
(P-Value)	(0.123)	(0.000)	(0.702)	(0.004)
OI/NS	-0.221	0.057	0.871	-13.457
(P-Value)	(0.749)	(0.908)	(0.184)	(0.000)
LTD/TA	0.689	2.233	5.84	-0.041
(P-Value)	(0.389)	(0.003)	(0.000)	(0.981)
TD/TA	-2.34	-3.54	-5.184	2.499
(P-Value)	(0.003)	(0.000)	(0.000)	(0.153)
TA	0.753	0.641	0.397	0.786
(P-Value)	(0.000)	(0.000)	(0.000)	(0.000)
RE/TA	1.135	0.68	0.599	-0.426
(P-Value)	(0.000)	(0.004)	(0.085)	(0.553)
$EBIT/TA$	1.923	-0.045	3.96	-17.44
(P-Value)	(0.277)	(0.976)	(0.055)	(0.000)
ROA	-0.303	-1.863	-3.339	24.413
(P-Value)	(0.875)	(0.26)	(0.14)	(0.000)
AUS	2.735	0.464	-1.11	-100.714
(P-Value)	(0.000)	(0.013)	(0.002)	(0.999)
CAN	0.618	-0.659	-3.18	-101.175
(P-Value)	(0.001)	(0.000)	(0.002)	(0.999)
FR	1.638	1.396	1.184	0.564
(P-Value)	(0.000)	(0.000)	(0.000)	(0.135)
GER	1.073	0.231	-0.358	-103.114
(P-Value)	(0.000)	(0.226)	(0.223)	(0.999)
JAP	0.434	0.754	1.131	-2.717
(P-Value)	(0.015)	(0.000)	(0.000)	(0.000)
UK	1.521	0.598	-0.235	-0.897
(P-Value)	(0.000)	(0.000)	(0.144)	(0.013)

Table 3.9: Logit estimation with hieracical data.

Unordered Logit Regression				
	BB-BBB	BB-A	BB-AA	BB-AAA
1991	0.075	-0.518	0.651	0.755
(P-Value)	(0.932)	(0.575)	(0.345)	(0.294)
1992	0.209	-0.316	0.667	1.111
(P-Value)	(0.806)	(0.728)	(0.312)	(0.115)
1993	-0.092	-0.276	0.416	0.863
(P-Value)	(0.912)	(0.764)	(0.525)	(0.218)
1994	-0.042	-0.667	0.365	0.745
(P-Value)	(0.961)	(0.46)	(0.582)	(0.29)
1995	0.114	-1.07	0.232	0.989
(P-Value)	(0.893)	(0.22)	(0.721)	(0.156)
1996	0.313	-1.16	0.162	0.81
(P-Value)	(0.714)	(0.179)	(0.803)	(0.242)
1997	0.753	-1.231	0.175	0.727
(P-Value)	(0.367)	(0.145)	(0.78)	(0.27)
1998	0.914	-1.504	-0.032	0.348
(P-Value)	(0.285)	(0.075)	(0.96)	(0.598)
1999	-0.384	-2.377	-0.646	-0.043
(P-Value)	(0.646)	(0.005)	(0.319)	(0.95)
2000	-1.013	-2.807	-0.857	-0.054
(P-Value)	(0.215)	(0.001)	(0.187)	(0.937)
2001	-1.657	-3.481	-1.222	-0.485
(P-Value)	(0.043)	(0.000)	(0.068)	(0.489)
2002	-2.153	-3.692	-1.399	-0.316
(P-Value)	(0.009)	(0.000)	(0.039)	(0.658)
2003	-3.268	-4.227	-1.62	-0.414
(P-Value)	(0.000)	(0.000)	(0.019)	(0.567)
2004	-3.755	-4.556	-1.779	-0.452
(P-Value)	(0.000)	(0.000)	(0.012)	(0.543)
2005	-3.79	-4.515	-1.743	-0.342
(P-Value)	(0.000)	(0.000)	(0.014)	(0.646)
2006	-3.859	-4.688	-1.813	-0.371
(P-Value)	(0.000)	(0.000)	(0.011)	(0.62)
2007	-3.954	-4.887	-1.931	-0.485
(P-Value)	(0.000)	(0.000)	(0.007)	(0.521)
2008	-3.694	-4.674	-1.826	-0.237
(P-Value)	(0.000)	(0.000)	(0.01)	(0.751)
2009	-4.519	-5.583	-2.735	-1.243
(P-Value)	(0.000)	(0.000)	(0.001)	(0.149)

Table 3.10: Yearly dummies of the unordered Logit estimation.

	BB-BBB	BBB-A	A-AA	AA-AAA
1991	-0.208	-1.299	0.193	1.323
(P-Value)	(0.742)	(0.096)	(0.693	(0.099)
1992	-0.178	-1.178	-0.087	1.662
(P-Value)	(0.774)	(0.129)	(0.851)	(0.033)
1993	-0.312	-0.919	-0.136	1.524
(P-Value)	(0.606)	(0.246)	(0.767)	(0.049)
1994	-0.183	-1.344	-0.204	1.562
(P-Value)	(0.765)	(0.076)	(0.657)	(0.044)
1995	0.112	-1.826	-0.476	1.712
(P-Value)	(0.854)	(0.012)	(0.289)	(0.027)
1996	0.287	-1.843	-0.439	1.68
(P-Value)	(0.642)	(0.01)	(0.321)	(0.031)
1997	0.756	-1.929	-0.364	1.463
(P-Value)	(0.212)	(0.006)	(0.391)	(0.045)
1998	1.106	-2.037	-0.321	0.954
(P-Value)	(0.079)	(0.003)	(0.445)	(0.19)
1999	0.54	-2.363	-0.611	0.549
(P-Value)	(0.355)	(0.001)	(0.146)	(0.462)
2000	0.242	-2.629	-0.807	0.623
(P-Value)	(0.664)	(0.000)	(0.053)	(0.408)
2001	0.124	-2.925	-0.777	0.467
(P-Value)	(0.817)	(0.000)	(0.063)	(0.541)
2002	-0.207	-3.031	-1.129	0.252
(P-Value)	(0.694)	(0.000)	(0.008)	(0.749)
2003	-0.947	-3.396	-1.26	0.011
(P-Value)	(0.062)	(0.000)	(0.003)	(0.989)
2004	-1.183	-3.634	-1.453	-0.125
(P-Value)	(0.02)	(0.000)	(0.000)	(0.879)
2005	-1.266	-3.647	-1.559	0.022
(P-Value)	(0.013)	(0.000)	(0.000)	(0.978)
2006	-1.2	-3.76	-1.571	0.063
(P-Value)	(0.019)	(0.000)	(0.000)	(0.939)
2007	-1.128	-3.839	-1.545	0.097
(P-Value)	(0.028)	(0.000)	(0.000)	(0.907)
2008	-1.052	-3.748	-1.693	-0.143
(P-Value)	(0.04)	(0.000)	(0.000)	(0.862)
2009	-0.983	-3.754	-1.781	-1.466
(P-Value)	(0.057)	(0.000)	(0.000)	(0.116)

Table 3.11: Yearly dummies of the Logit estimation with hierarchical data.

	Const	IC_1	IC_2	IC_3	IC_4	OI/NS	LTD/TA	TD/TA	TA	RE/TA	$EBIT/TA$	ROA
1990	-0.929	-0.017	-0.131	0.119	-0.011^{**}	-6.349	-0.225	1.952	0.211	-0.127	-25.167^{*}	37.189^{*}
1991	-0.189	-0.078	-0.044^{***}	0.128	-0.021^{**}	-5.121	-1.308	1.992	0.178	-1.473	-16.644^{**}	26.584^{***}
1992	-0.234	-0.125	0.0546	-0.014	-0.004	-5.233^{**}	-1.112	1.294	0.179	-0.897	-4.408	12.815^{***}
1993	0.344	-0.073	0.22	0.016	-0.013	-1.661	0.633	0.99	0.083	-1.411	-14.877^{**}	19.834^{***}
1994	1.09	-0.128	0.021	0.044	-0.013	-1.627	-0.585	0.035	0.05	0.195	-1.198	1.213
1995	1.407	-0.056	-0.059	0.056	-0.012	1.56	-1.03	-1.089	0.042	0.07	-6.103	4.866
1996	0.637	0.077	-0.152	0.069	-0.015^{*}	-1.935	-0.993	-1.429	0.111	-1.016	1.728	1.458
1997	-1.203	0.089	-0.104	0.083^{*}	-0.005	-1.754	0.351	-1.05	0.26	-0.276	0.263	-1.995
1998	-1.315	0.104	-0.2^{**}	0.131^{***}	-0.006	0.248	0.216	-2.557^{*}	0.346^{***}	-0.999^{**}	-0.087	-4.574
1999	-3.18^{***}	0.13^{**}	-0.117	0.078^{*}	-0.011	-2.491^{**}	0.607	-1.513	0.428^{***}	-0.631	-0.68	2.407
2000	-2.724^{***}	0.119^{**}	-0.038	0.056	-0.012^{*}	-2.139^{**}	-2.838^{**}	1.503	0.349^{***}	0.026	-1.964	2.876
2001	-2.373^{***}	0.063	-0.01	0.043	-0.016^{*}	-0.098	-1.91^{*}	0.323	0.307^{***}	-0.329	1.836	-0.46
2002	-3.017^{***}	0.069	-0.025	0.07	-0.017^{*}	0.184	-0.677^{**}	-0.198^{***}	0.339^{***}	-0.122	0.925	0.249
2003	-3.689^{***}	0.026	0.013	0.014	-0.009^{*}	0.994	2.595^{**}	-3.321^{***}	0.38^{***}	0.548	0.389	1.409
2004	-4.273^{***}	0.078^{*}	-0.079	0.038	-0.006	0.512	2.741^{**}	-3.841^{***}	0.433^{***}	0.835^{**}	3.169	-3.241
2005	-4.879^{***}	0.111^{**}	-0.051	0.012	0.001	1.053	2.71^{**}	-3.333^{***}	0.465^{***}	0.834^{***}	1.036	-2.211
2006	-5.028^{***}	0.078^{*}	-0.036	0.026	0.004	0.907	0.558	-1.232	0.469^{***}	0.846^{**}	1.485	-2.413
2007	-4.94^{***}	0.04	-0.005	0.038	-0.001	-0.197	0.329	-0.979	0.48^{***}	0.841^{**}	-0.971	-0.154
2008	-4.987^{***}	0.05	-0.035	0.035	0.003	-0.413	-0.485	0.103	0.477^{***}	0.516^{*}	-0.644	0.683
2009	-4.966^{***}	0.138^{***}	-0.053	-0.011	0.005	0.037	0.06	-0.861	0.484^{***}	0.412	1.892	-4.783^{*}

Table 3.12: Coefficients of year by year regressions using Probit. Coefficients significant at the 10% *, at the 5% with **, and at 1% level with ***.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Const	0.126	-1.789	−0.694***	−3.379***	−3.156***	−3.612***	−4.367***	−4.879***	−5.984***	−6.706***	−6.662***	−6.316***	−6.046***	−5.617***
IC_1	0.021	0.086	0.071	0.123*	0.118**	0.039	0.068	0.029	0.057	0.078*	0.057	0.013	0.042	0.146***
IC_2	-0.131	-0.067	−0.223**	-0.125	-0.056	-0.009	-0.057	-0.013	-0.1	-0.057	-0.024	0.031	-0.015	-0.088
IC_3	0.062	0.095**	0.147***	0.089**	0.05	0.036	0.069*	0.0152	0.024	0.005	0.029	0.027	0.032	-0.008
IC_4	−0.017*	-0.006	-0.01	−0.014**	−0.012*	-0.012	-0.013	−0.01*	−0.009*	-0.001	0.004	0	0.003	0.002
OI/NS	-2.527	-2.047	0.095	−2.627**	−2.009*	-0.124	0.299	0.954	0.559	0.582	0.33	-0.541	-0.673	0.039
LTD/TA	-0.141	2.216	2.322*	1.836	−2.214*	-0.52	1.191	4.743***	4.877***	4.415***	2.049*	1.155	-0.183	2.077
TD/TA	-2.286	-2.153	−4.556***	−2.622*	0.877	-0.604	-1.742	−5.248***	−6.174***	−5.285***	−2.497**	-1.799	-0.088	−2.937**
TA	0.158	0.283***	0.308***	0.433***	0.372***	0.38***	0.424***	0.458***	0.559***	0.609***	0.59***	0.583***	0.553***	0.531***
RE/TA	-0.673	0.18	-0.541	-0.296	0.03	0.107	0.138	0.84**	1.144***	1.136***	1.184***	1.126***	0.653**	0.438
$EBIT/TA$	5.525	0.667	-0.744	-1.076	-0.271	3.532	3.554	-0.672	5.166*	3.327	1.804	-0.796	-1.46	2.296
ROA	-2.088	-3.584	-3.884	3.217	2.449	-0.988	-1.711	4.325	-2.909	-3.35	-2.833	-0.637	1.768	-3.925
AUS	0.018	0.471	0.123	0.674*	0.257	0.623*	0.631*	0.625*	0.842**	0.871***	0.888***	0.831***	0.591**	0.609**
CND	-0.668	-0.534	−1.202***	-0.433	-0.248	0.105	0.115	-0.091	-0.031	0.078	0.078	0.006	-0.097	-0.214
FRA	1.329***	1.334***	0.927**	0.906**	0.779**	0.992***	1.003***	1.146***	1.315***	1.368***	1.156***	0.919***	0.808***	0.671**
GER	-0.484	0.288	-0.081	0.268	0.21	0.449	0.386	0.4198	0.234	0.526*	0.34	0.057	-0.138	-0.066
JAP	0.083	0.544*	0.644**	0.533*	0.545*	0.765***	0.946***	0.624***	0.844***	0.702***	0.438**	0.392**	0.28	0.812***
UK	0.413	0.812**	0.555*	0.484*	0.153	0.304	0.448**	0.489*	0.397*	0.416*	0.394*	0.379*	0.108	0.162

Table 3.13: Coefficients of year by year regressions with country dummies. Coefficients significant at the 10% *, at the 5% with **, and at 1% level with ***.

sistently significant positive effect in any given year. This further highlights the importance of firm size in the agencies' assessment of a firm's creditworthiness. The total debt ratio and the retained earnings ratios have the expected negative and positive effect respectively. For IC, we find that small values have a significant positive effect in certain years while large values may have a negative effect. This supports the view that interest coverage has non linear effect.

The effects of the explanatory variables are mostly stable over time. In certain cases there are switches or shifts possibly caused by the sample selection. The small sample size at the beginning of the data set can make consistent estimation difficult.

Prior to 1996 there was not a sufficient number of non-US firms in the data set to estimate the country dummies. Table 3.13 starts in 1996. In comparison to the yearly regressions without country dummies in Table 3.12 we find less significant effects for IC. The long term debt ratio often has a positive effect. TA, RE, and TD have the same effect. Furthermore, the constants have a negative trend. For the country dummies we find positive effects for Australia, France, Japan, and the UK.

3.5 Concluding Remarks on Credit Ratings

In this thesis we provide an extensive review of empirical research on corporate credit ratings as discussed in Matthies (2013a). It is important to highlight the ordinal structure of credit ratings. We give an analysis of different rating markets and how different market setups could lead to possible conflicts of interests. Based on the findings of the literature review we combine the assumption of the most frequent method used to estimate credit ratings, i.e. ordered probit (Blume et al., 1998), with the findings in Jorion & Zhang (2007), to propose an OLS estimation where credit ratings are replaced with their respective default rates. Thereby we provide an observable variable for the otherwise unobservable continuous variable that is assumed to underly the standard ordered probit model. This study analyses the statistical properties and stability of S&P ratings from a sample made up of the 100 largest non-financial firms in 2005 from the US, Australia, Canada, France, Germany, Japan, and the UK.

The main contributions of this thesis for credit ratings are firstly, that from a methodological perspective OLS regressions can be used as an alternative to the established ordered probit analysis if ratings are replaced with their respective default frequencies. A comparison of the standard ordered probit method with the OLS method shows that these two methods produce similar estimation results. Secondly, the probit and OLS analysis confirm previous results (Blume et al., 1998, Amato & Furfine, 2004, Jorion et al., 2009) that suggest a possible commitment of CRAs to obtain reputational capital. This interpretation can be drawn from the continuing decline of yearly constants of the panel probit and OLS models, as it implies that declining quality in accounting standards (Jorion et al., 2009) necessitate CRAs to give out more stringent ratings in order to provide accurate risk assessments. Thirdly, in contrast we find evidence that a CRA[18] may give out more lenient ratings in markets like Japan and France, where credit ratings are not an established feature of corporate governance, in comparison to the US. This may be deduced from positive country specific effects in the panel probit models.

A central requirement for the incorporation of credit ratings in an investment decision or investment guideline is that CRAs attempt to accurately represent default probability. Only then can they serve as a reliable tool to asses the risk of an investment. If CRAs pursue to accurately assess default probabilities, they will automatically gain reputational capital. Yet, the primary goal of CRAs as private firms is of course their bottom line, i.e. their earnings. Since their main source of earnings stems from issuer paid solicited credit ratings, this might entail a possible conflict of interest.

The data of this thesis shows that for a comparative group of non-financial firms from seven developed countries credit ratings are not equally common on an international level[19]. In the US, credit ratings are a central feature of corporate governance for large publicly traded firms. In the UK, France, Germany, and Japan ratings are not as present or not at all times. In Canada and Australia, comparatively many more firms have a credit rating considering the relatively small size of their economies.

A low demand for credit ratings in a particular market might weaken a CRAs position in that market. CRAs might thus be pressured to give out less stringent ratings in order to increase demand from firms to solicit their ratings. However, in the long run the existence and economic success of CRAs is predicated on gaining reputational capital. Less stringent ratings might therefore reflect short term strategies of credit rating agencies to establish themselves in non-US financial markets.

The macroeconomic properties of the sample are consistent with the findings of other studies (Blume et al., 1998, Amato & Furfine, 2004, Jorion et al., 2009, Alp, 2013). We show that the trend of decreasing year dummies α_t first observed by Blume et al. (1998) from 1978 to 1995 continues up to 2009 in our sample. This continuing trend may be interpreted as the CRAs' response to declining quality of accounting data. This would suggest that in the long run CRAs pursue to obtain reputational capital.

Furthermore, for Japan and France we find country specific positive effects in comparison to the US in a probit panel analysis. Intuitively, one might interpret this as there being a higher average risk for US firms, as the positive effect reflects that a US

[18] In this case S&P.

[19] See Matthies (2013b) for a more extensive overview.

firm has a lower rating than a French or Japanese firm with the same accounting data. Yet, these positive effects need not reflect a lower risk in countries such as Japan and France, but are perhaps due to the CRA approach caused by rating market pressure on the CRA in these economies, where corporate governance standards do not require a rating. The positive effects for the French and Japanese dummy might indicate that the CRA could possibly give out more favourable ratings to be more often solicited. Another reason for the positive effect is perhaps differing accounting standards in other countries. S&P might have more trust in the accuracy of accounting data than it has in US accounting data.

The OLS regression developed in this thesis underscores the idea that ratings reflect real risk assessments of default probability. It supports the notion that ratings of CRAs can provide useful information for investors, as default frequencies of different rating categories align as if they carry actual information. Moreover, we note that firm size is the most consistent determinant for credit quality in all models.

Chapter 4

Conclusion

This thesis addresses the modelling of risk in financial economics in two major aspects. Firstly, government term structures of interest rates are forecasted via dynamic factor models at a daily frequency. This deals with the short term fluctuations as to how markets regard the long term possibility of countries defaulting on their debt. Secondly, corporate credit ratings are estimated with yearly data from financial statements. Here, the credit rating agencies' estimate of the probability of default of a single business entity operating within the larger economies is put into focus.

Concerning government bond term structures, we provide several forecast evaluations of data-driven Nelson-Siegel type dynamic factor models. Thereby, the statistical and economic information value of different data sets, estimation and forecasting methods, and model characteristics are compared. Daily government term structure data from Germany, the US, the UK, and Switzerland from 2000 – 2017 is used. Principal component analysis shows that factors for individual term structures represent level, slope, and curvature. For all term structure data the first factor represents a global level factor. Further factors capture the correlation of country specific slope and curvature factors.

The results confirm that factor models using the yield curve improve forecast approaches that only rely on single maturities forecast, i.e. univariate auto-regressive models. Furthermore, multi factor models with two or more factors are beneficial for modelling and forecasting term structures. We find that additional term structures can be used as predictors and improve forecasts, thereby filling a gap in the investigation of different predictors for term structure forecasts. These predictors may be viewed as lying between univariate maturity models, as extensively tested here, and term structure models (Diebold & Li, 2006, Blaskowitz & Herwartz, 2009, Matthies, 2014) at one end, and at the other end a data set with a large number of financial variables as in Matthies (2014). We show that models with additional term structures perform similar to models with additional financial data in comparison to models with a single term structure. Lastly, we find that principal component generalised least squares estimation as developed by Breitung & Tenhofen (2011) and the Stock & Watson (2002) forecasting approach are mostly outperformed by principal component ordinary least squares estimation and the Blaskowitz & Herwartz (2011) forecasting method.

With regard to credit ratings this work analyses the statistical properties and stability of ratings from a sample made up of the 100 largest non-financial firms of the US, the UK, Germany, France, Japan, Canada, and Australia in 2005. The analysis includes data from 1990 to 2010. We perform an ordered probit panel estimation and an ordinary least squares regression using default frequencies of the respective rating categories. The ordinary least squares approach developed here tests the assumption that credit ratings reflect actual default probability and thereby the risk of investing in any given firm. This new modelling of credit ratings confirms the information content of credit ratings for non-financial firms.

This thesis confirms previous findings (Blume et al., 1998, Amato & Furfine, 2004, Jorion et al., 2009) that a firm with the same accounting data can receive lower and lower ratings over time. These findings can be interpreted as credit rating agencies trying to gain reputational capital.

A new finding is that a firm, with the same accounting data, may receive a lower rating in the US than a firm in France or Japan. This suggests that market pressure in non-US markets may compromise the risk assessments of a credit rating agency due to a conflict of interest.

Bibliography

Abad, P., & Novales, A. (2005). An error correction factor model of term structure slopes in international swap markets. *Journal of International Financial Markets, Institutions and Money*, *15*, 229–254.

Adelino, M., & Ferreira, M. A. (2016). Bank ratings and lending supply: Evidence from sovereign downgrades. *The Review of Financial Studies*, *29*, 1709–1746.

Afonso, A., Gomes, P., & Rother, P. (2011). Short-and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, *16*, 1–15.

Ahn, H., Kim, K.-J., & Han, I. (2005). Combining pairwise SVM classifiers for bond rating. In *2005 KMIS Internationazl Conference* (p. 104).

Almeida, H., Cunha, I., Ferreira, M. A., & Restrepo, F. (2017). The real effects of credit ratings: The sovereign ceiling channel. *The Journal of Finance*, *72*, 249–290.

Alp, A. (2013). Structural shifts in credit rating standards. *The Journal of Finance*, *68*, 2435–2470.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*, 589 – 609.

Altman, E., & Rijken, H. (2004). A point-in-time perspective on through-the-cycle ratings. *Financial Analysts Journal*, *62*, 54 – 70.

Altman, E., & Rijken, H. (2006). How rating agencies achieve rating stability. *Journal of Banking & Finance*, *28*, 2679 – 2714.

Altman, E. I., Oncu, T. S., Richardson, A., & White, L. (2010). Regulation of rating agencies. *Regulating Wall Street, John Wiley & Sons Inc, Hoboken, NJ*, (pp. 443–467).

Amato, J., & Furfine, C. (2004). Are credit ratings procyclical? *Journal of Banking & Finance*, *28*, 2641 – 2677.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, *50*, 745–787.

Ashbaugh-Skaife, H., Collins, D., & LaFond, R. (2006). The effects of corporate governance on firms credit ratings. *Journal of Accounting and Economics*, *42*, 203 – 243.

Asmussen, J. (2005). Rating-Agenturen und Finanzaufsicht. *Betriebswirtschaftliche Forschung und Praxis*, *57*, 246 – 255.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*, 135–171.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191–221.

Benmelech, E., & Dlugozs, J. (2010). The credit rating crisis. In *NBER 2010*.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, *120*, 387–422.

Bhojraj, S., & Sengupta, P. (2003). Effect of corporate governance on bond ratings and yields: The role of institutional investors and outside directors. *Journal of Business*, *76*, 455 – 475.

Blaskowitz, O., & Herwartz, H. (2009). Pca-based ex-ante forecasting of swap term structures. *International Journal of Theoretical and Applied Finance*, *12*, 465–489.

Blaskowitz, O., & Herwartz, H. (2011). On economic evaluation of directional forecasts. *International Journal of Forecasting*, *27*, 1058–1065.

Blume, M. E., Lim, F., & MacKinlay, A. C. (1998). The declining credit quality of us corporate debt: Myth or reality? *The journal of finance*, *53*, 1389–1413.

Bongaerts, D., Cremers, K., & Goetzmann, W. N. (2012). Tiebreaker: Certification and multiple credit ratings. *The Journal of Finance*, *67*, 113–152.

Brabazon, A., & O'Neill, M. (2008). Bond rating with π-grammatical evolution. *Studies in Computational Intelligence*, *102*, 17 – 30.

Breitung, J., & Eickmeier, S. (2006). Dynamic factor models. *Modern econometric analysis*, (pp. 25–40).

Breitung, J., & Tenhofen, J. (2011). Gls estimation of dynamic factor models. *Journal of the American Statistical Association*, *106*, 1150–1166.

Calomiris, C. W. (2009). A recipe for ratings reform. *The Economists' Voice*, *6*.

Campbell, J. Y., Lo, A. W.-C., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. princeton University press.

Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *The Journal of Fixed Income*, *6*, 76–91.

Cao, L., Guan, L., & Jingqing, Z. (2006). Bond rating using support vector machine. *Intelligent Data Analysis*, *10*, 285 – 296.

Carstensen, K. (2003). Nonstationary term premia and cointegration of the term structure. *Economics Letters*, *80*, 409–413.

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica (pre-1986)*, *51*, 1281.

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, *164*, 4–20.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *The American economic review*, *95*, 138–160.

Coskun, D. (2008). Credit rating agencies in a post-Enron world: Congress revisits the NRSRO concept. *Journal of Banking Regulation*, *9*, 264–283.

Dewachter, H., & Lyrio, M. (2006). Macro factors and the term structure of interest rates. *Journal of Money, Credit and Banking*, (pp. 119–140).

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, *130*, 337–364.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*, 134–144.

Diebold, F. X., Piazzesi, M., & Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review*, *95*, 415–420.

Dilly, M., & Mählmann, T. (2010). Ableitung markt-impliziter Ratings aus Credit Default Swap-Spreads. *Zeitschrift für Bankrecht und Bankwirtschaft*, *22*, 487 – 507.

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, *164*, 188–205.

Du, Y., & Suo, W. (2005). An empirical study on credit rating change behavior. Unpublished manuscript.

Duffee, G. R. (2011). Information in (and not in) the term structure. *Review of financial studies*, *24*, 2895–2934.

Dutta, S., & Shekhar, S. (1988). Bond rating: a non-conservative application of neural networks. In *IEEE Int Conf on Neural Networks*. Publ by IEEE.

Ederington, L. (1985). Classification models and bond ratings. *The financial review*, *20*, 237 – 261.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). Explaining the rate spread on corporate bonds. *the journal of finance*, *56*, 247–277.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, (pp. 251–276).

Erlenmaier, U. (2006). The shadow rating approach — experience from bank practice. In *The Basel II Risk Parameters* (pp. 39 – 77). Berlin: Springer.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The journal of Finance*, *46*, 555–576.

Giese, J. V. et al. (2008). Level, slope, curvature: Characterising the yield curve in a cointegrated var model. *Economics: The Open-Access, Open-Assessment E-Journal*, *2*, 1–20.

Gonzales, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Zins, C., & Wieland, M. (2004). Market dynamics associated with credit ratings: A literature review. *Banque de France Financial Stability Review*, *4*, 53 – 76.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Griffin, J. M., & Tang, D. Y. (2011). Did credit rating agencies make unbiased assumptions on CDOs? *The American Economic Review*, *101*, 125–130.

Hall, A. D., Anderson, H. M., & Granger, C. W. (1992). A cointegration analysis of treasury bill yields. *The review of Economics and Statistics*, (pp. 116–126).

Härdle, W., Moro, R., & Schäfer, D. (2005). Predicting bankruptcy with support vector machines. In *Statistical Tools for Finance and Insurance* (pp. 225–248). Springer.

Hartzmark, M. L. (1991). Luck versus forecast ability: Determinants of trader performance in futures markets. *Journal of Business*, (pp. 49–74).

He, J., Hu, W., & Lang, L. H. (2000). Credit spread curves and credit ratings, .

Hill, C. (2004). Regulating the rating agencies. *Washington University Law Review*, *82*, 43 – 95.

Hill, P., Brooks, R., & Faff, R. (2010). Variations in sovereign credit quality assessments across rating agencies. *Journal of Banking & Finance*, *34*, 1327–1343.

Huang, J.-Z., & Huang, M. (2012). How much of the corporate-treasury yield spread is due to credit risk? *The Review of Asset Pricing Studies*, *2*, 153–202.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit ratings analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, *37*, 543 – 558.

Jorion, P., Shi, C., & Zhang, S. (2009). Tightening credit standards: The role of accounting quality. *Review of Accounting Studies*, *14*, 123 – 160.

Jorion, P., & Zhang, G. (2007). Information effects of bond rating changes: The role of the rating prior to the announcement. *The Journal of Fixed Income*, *16*, 45 – 59.

Kaminsky, G., & Schmukler, S. L. (2002). Emerging market instability: do sovereign ratings affect country risk and stock returns? *The World Bank Economic Review*, *16*, 171–195.

Katz, S. (1974). The price adjustment process of bonds to rating reclassification: A test of bond market efficiency. *The Journal of Finance*, *29*, 551 – 559.

Kim, Y., & Sohn, S. (2008). Random effects model for credit rating transitions. *European Journal of Operational Research*, *184*, 561 – 573.

Koopman, S., Lucas, A., & Monteiro, A. (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, *142*, 399 – 424.

Kwon, Y., Han, I., & Lee, K. (1997). Ordinal pairwise partitioning (OPP) approach to neural networks training in bond raiting. *Intelligent Systems in Accounting, Finance & Management*, *6*, 23 – 40.

Lando, D., & Skødeberg, T. (2002). Analyzing rating transition and rating drift with continuous observations. *Journal of Banking & Finance*, *26*, 423 – 444.

Lee, Y.-C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, *33*, 67 – 74.

Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, (pp. 580–590).

Levine, R. (2010). An autopsy of the us financial system: accident, suicide, or negligent homicide. *Journal of Financial Economic Policy*, *2*, 196–213.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. 101. Sage.

Litterman, R. B., & Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, *1*, 54–61.

Liu, X., & Liu, W. (2005). Credit rating analysis with AFS fuzzy logic. In *Lecture Notes in Computer Science* (pp. 1198 – 1204). Berlin: Springer volume 3612.

Matthies, A. B. (2013a). *Empirical research on corporate credit-ratings: A literature review*. Technical Report SFB 649 Discussion Paper.

Matthies, A. B. (2013b). Rating agencies and reputational capital–a survey of current evidence with new empirical results. *Global Economy and Finance Journal*, *6*, 29 – 52.

Matthies, A. B. (2013c). *Statistical properties and stability of ratings in a subset of US firms*. Technical Report SFB 649 Discussion Paper.

Matthies, A. B. (2014). Validation of term structure forecasts with factor models. *The Journal of Risk Model Validation*, *8*, 65.

Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, *146*, 26–43.

Morkoetter, S., Stebler, R., & Westerfeld, S. (2017). Competition in the credit rating industry: Benefits for investors and issuers. *Journal of Banking & Finance*, *75*, 235–257.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, (pp. 473–489).

Nickel, P., Perraudin, W., & Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, *24*, 203 – 227.

Nordberg, D. (2011). *Corporate Governance: Principles And Issues*. London: Sage Publication.

Partnoy, F. (1999). The siskel and ebert of financial markets: Two thumbs down for the credit rating agencies. *Wash. ulq*, *77*, 619.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes 3rd edition: The art of scientific computing* volume 3. Cambridge university press Cambridge.

Ramakrishnan, R., & Thakor, A. (1984). Information reliability and a theory of financial intermediation. *Review of Economic Studies*, *51*, 415 – 432.

Recchioni, M. C., & Tedeschi, G. (2017). From bond yield to macroeconomic instability: A parsimonious affine model. *European Journal of Operational Research*, .

Richardson, M., & White, L. (2009). The rating agencies: is regulation the answer? *Financial Markets, Institutions & Instruments*, *18*, 146–148.

Rosner, J. (2009). Toward an understanding: NRSRO failings in structured ratings and discreet recommendations to address agency conflicts. *The Journal of Structured Finance*, *14*, 7–22.

Shih, J.-Y., & Chen, W.-H. (2006). A study of taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, *30*, 427 – 435.

Shin, K.-S., Kim, H.-J., & Kwon, S.-B. (2004). A GA-based fuzzy decision tree approach for corporate bond rating. In *Lecture Notes in Computer Science* (pp. 505 – 514). Berlin: Springer volume 3157.

Standard, & Poor's (2009). *U.S. CMBS Rating Methodology and Assumptions for Conduit/Fusion Pools*. New York, NY: McGraw-Hill.

Standard, & Poor's (2012). *Should 'AAA' Be The Only Way For Structured Finance?*. New York, NY: McGraw-Hill.

Stock, J. H., & Watson, M. (2011). Dynamic factor models. *Oxford Handbook on Economic Forecasting*, .

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*, 1167–1179.

Stock, J. H., & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis*. Technical Report National Bureau of Economic Research.

Wappenschmidt, C. (2009). *Ratinganalyse durch internationale Ratingagenturen*. Frankfurt am Main.

West, K. D., & Newey, W. K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

White, L. J. (2010). Markets: The credit rating agencies. *The Journal of Economic Perspectives*, *24*, 211–226.

Ye, Y., Liu, S., & Li, J. (2008). A multiclass machine learning approach to credit rating prediction. In *2008 International Symposiums on Information Processing (ISIP)* (pp. 57 – 61).

Yu, W.-C., Salyards, D. M. et al. (2009). Parsimonious modeling and forecasting of corporate yield curve. *Journal of Forecasting*, *28*, 73.

Yu, W.-C., & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields using dynamic nelson-siegel models. *International Journal of Forecasting*, *27*, 579–591.

Yue, V. Z. (2010). Sovereign default and debt renegotiation. *Journal of international Economics*, *80*, 176–187.

Zhang, H. (1993). Treasury yield curves and cointegration. *Applied Economics*, *25*, 361–367.

Zhou, C. (2001a). Credit rating and corporate default. *Journal of Fixed Income*, *3*, 30 – 40.

Zhou, C. (2001b). The term structure of credit spreads with jump risk. *Journal of Banking & Finance*, *25*, 2015–2040.

Eidesstattliche Versicherung

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 22.04.2019 Alexander B. Matthies

(Alexander B. Matthies)