Essays in Applied Time Series Analysis

Essays in Applied Time Series Analysis

Essays in toegepaste tijdreeksanalyse

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Friday, February 22, 2019 at 13:30 hours

by

Bartholemeus Jacob Leonard Keijsers
born in Buren, The Netherlands.

Erasmus University Rotterdam

Doctorate Committee

Promotor:	Prof.dr. D.J.C. van Dijk
Other members:	Dr. L.C.G. Pozzi
	Prof.dr. P.C. Schotman
	Dr. M. van der Wel
Copromotor:	Dr. H.J.W.G. Kole

ISBN: 978 90 3610 544 6

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 731 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Acknowledgments

Het is ruim 11 jaar geleden dat ik naar Rotterdam ben gekomen om econometrie te studeren. Dit is duidelijk uit de hand gelopen, aangezien je nominaal in vier jaar klaar kan zijn. Maar na de master kon ik nog geen afscheid nemen van het studeren, en hier heb ik geen spijt van gehad. Het promotietraject heeft me veel geleerd: over econometrie, de academische wereld, en ik heb er de halve wereld voor af mogen reizen. Uiteindelijk heeft dit gersulteerd in dit boekje. Dit was er natuurlijk niet gekomen zonder de hulp van anderen, die ik hier dan ook voor wil bedanken.

Allereerst mijn (co-)promoteren, bedankt dat jullie met mij aan de slag wilden en voor alle tijd en energie die jullie de afgelopen jaren in me gestoken hebben. Bart, ik vond het heel prettig om met je te werken. Ik kon altijd bij je langs voor (Bayesiaanse) econometrie, finance, en sportgerelateerde discussies. Hoewel we het niet samen hebben kunnen afmaken, ben ik blij dat je onderdeel kan zijn van de promotiecommissie. Dick, bedankt voor alle tips en adviezen. Ik waardeer ook vooral je hulp en vrijheid bij het uitzoeken van de plannen na de PhD, en de steun tijdens de job market. Erik, naar aanleiding van de discussies met jou tijdens het schrijven van mijn master thesis ben ik serieuzer over een PhD gaan denken. Tijdens mijn PhD was je voor mijn gevoel altijd al een copromotor. My master thesis has been the starting point of my academic career. The thesis has evolved – thanks to you, Bart and Erik – into a nice publication. I thank NIBC Bank and Global Credit Data for allowing us to work with their data set. In particular Michel van Beest, thanks for your support and knowledge.

Dank aan de commissieleden, Lorenzo Pozzi, Peter Schotman en Michel van der Wel, voor het doornemen van mijn thesis en het te voorzien van waardevolle feedback.

To my old colleagues at the Econometric Institute and Tinbergen Institute, for the pleasant work environment. In particular, I appreciated the openness and collegiality.

To my new colleagues at the University of Amsterdam, for the warm welcome and the trust in hiring me. I look forward to continue learning and developing new research projects. Thanks to Allan Timmermann, for your hospitality during my visit at UCSD. Also, I am grateful that you were willing to be one of my references. Herman, ook jij erg bedankt voor je referentie en hulp tijdens de job market.

Vooral wil ik mijn mede PhD studenten bedanken voor het delen van de ups en downs van een promotietraject. Dennis, voor het introduceren van de legendarische mini-tafeltennistafel in H08-11. Bruno, voor de tafeltenniswedstrijden (laten we het gelijkspel noemen), de Econometric Game, Game of Thrones en meer. Ik blijf je op de hoogte houden van de Terps! Didier, voor je enthousiasme (watermeloentjes!) en het uitspelen van de job market. Tom, voor de voorbereiding van mijn bezoek aan San Diego, met/zonder fiets. Francine en Sander, voor het meeverhuizen naar de 10e verdieping en de nodige finance input. Voor alle lunches, koffie en biertjes: Aiste, Albert-Jan, Anne, Anoek, Arash, Bert, Eran, Esmee, Gert-Jan, Gianluca, Koen, Matthijs, Max, Max, Megan, Michael, Myrthe, Sait, Victor en Xiao.

Verder bedankt aan de DEUR, nu inclusief dasjes. Pascal, door jou ben ik een PhD als optie gaan zien. Daarnaast (schoon-)familie en vrienden, voor de nodige afleiding in de vorm van lacrosse, perenijsjes, biertjes, reizen, festivals, spelletjes en andere zin en onzin.

Papa en mama, bedankt voor de onvoorwaardelijke steun en zorg. Jullie staan altijd voor me klaar, of het nou is voor een verhuizing of om op te passen. Mats, het is bijzonder om te zien hoe snel je je ontwikkelt en hoe nieuwsgierig je bent. Ik kan niet wachten om te zien hoe jij je weg in het leven vindt. Ten slotte, Loesje. Jij plaatst het onderzoek in perspectief. Bedankt voor je begrip als ik aan het nerden ben, en voor de avonturen die we al beleefd hebben. Ik houd superveel van je en kijk uit naar een mooie toekomst samen!

Contents

Chapter 1

Introduction

Many questions in economics and finance cannot be answered without analyzing time series variables. For example, time series analysis can be used to estimate the relationship between variables over time, and use that information to predict future values. Further, one needs to be careful that relationships might vary over time. It provides invaluable input for decision making by policy makers and investors. This thesis presents three essays in applied time series analysis, using a variety of methods to address several important issues in economics and finance.

Even though the chapters are seemingly unrelated, they are linked in the sense that they are examples of applications of time series analysis in finance and economics, that display relevance in practice. Chapter 2 shows the relationship between losses on bank loans and macroeconomic variables, which is of interest to banks and financial regulators. Chapter 3 shows that long-term investors can incur large losses if they do not take parameter instability in the relationship between stock returns and underlying state variables into account. This is relevant to pension funds for example. Chapter 4 examines the usefulness of economic uncertainty in predicting economic activity. This is important for policy makers who use forecasts on economic indicators as input for policy decisions.

Further, in each chapter, the issue at hand is answered from an applied econometrician's point of view, where the model is specified to accommodate typical features of the data. In chapter 2, we select a mixture model based on key characteristics of the distribution of losses. Chapter 3 uses a flexible model to estimate the instability in the relationship between predictors and the stock returns, rather than assuming the number of breaks in advance. Chapter 4 considers quantile prediction as well as mean forecasting, because economic uncertainty is likely to affect the lower tail of economic activity more heavily.

Though taking time to specify an appropriate econometric model may complicate the estimation procedure, it may yield a new perspective and interesting insights. These essays provide motivation that this is worth the effort in answering questions in economics and finance, both to practitioners and academics.

Below is an outline of the individual chapters.

Cyclicality in losses on bank loans

Chapter 2 is on the relationship between losses on bank loans and economic output. Banking regulations require banks that risk measures should "reflect economic downturn conditions" (Basel Committee on Banking Supervision, 2005). This is based on the assumption that losses on bank loans covary with the business cycle. However, this assumption builds mostly on evidence from research on losses on bonds, as data sets of bank loan losses are scarce and small. This raises the question whether this assumption is valid, because bank loans and bonds differ in several important ways. Bank loans are more senior and backed by collateral. Further, the bank is able to more closely monitor losses, and influence the recovery process for example in the form of an additional loan. Therefore, losses on bank loans are likely to be less cyclical.

Access to a unique data set from Global Credit Data, a consortium of banks that pools their credit data, allows us to formally investigate the cyclicality of bank loan losses in more detail. We combine the information from the default rate, the loss given default – the fraction of the exposure at default that is lost – and macroeconomic variables that represent the business cycle in a dynamic factor model that we estimate using Bayesian methodology.

The distribution of the loss given default has several key features that should be captured by an appropriate statistical model. First, it is bimodal with peaks near the extreme cases of zero and one. Second, the relative height of the modes varies over time – not their location. Third, while the loss given default for bonds are defined on the $[0, 1]$ interval, for a substantial portion of defaults – about 12 percent – the loss given default is outside this interval. This may be due to principal advances, which is an extra amount lent to the borrower to increase the likelihood of recovery. Combining these observations motivates us to propose a mixture of two normal distributions to model the loss given default. The mixture components have an intuitive interpretation of mild and severe losses. Further, the probability of a severe loss is allowed to vary over time and depends on a latent factor.

We find that a model specification with two factors describes the data best. One factor can be interpreted as a macro factor, and a second as a default-specific factor which captures credit movements unrelated to the business cycle. The results confirm that there is a negative relationship with the business cycle. That is, when macroeconomic conditions are bad, not only the number of defaults, but also the average loss increases. This leads to an increase in capital reserves required for loan portfolios: simulations from a hypothetical loan portfolio show that economic capital – the amount a bank needs to hold to remain solvent – increases by a factor two during recessions.

Long-term investing under uncertain parameter instability

Chapter 3 looks into the effects of parameter instability in the context of stock return predictability. Predicting the return on the stock market (in excess of a risk-free rate) is an endeavor that is probably as old as the stock market itself. As summarized by Rapach and Zhou (2013), a wealth of factors has been proposed that seem to hold predictive power. However, Welch and Goyal (2008) show that the in-sample predictability of a set of popular predictors does not translate to out-of-sample gains. A possible explanation is that the relationship between the predictors and stock returns is unstable, i.e. it is subject to structural breaks. There are a number of ways to model this instability, from assuming few breaks (Pettenuzzo and Timmermann, 2011; Henkel et al., 2011) to many breaks (Dangl and Halling, 2012; Johannes et al., 2014), but there is no consensus on which is most appropriate.

Therefore, we suggest a mixture innovation model, where we can estimate the instability, rather than assume it ex ante. Additionally, because we use Bayesian estimation methods, it provides insight into the uncertainty regarding this instability. The estimation results show that indeed there is substantial time-variation. It is more than suggested in the few breaks case, but also less than a break each time period. At the same time, there is quite some uncertainty regarding the break probability.

In the spirit of how Barberis (2000) investigates the impact of parameter uncertainty, we study the impact of (uncertainty regarding) parameter instability on the allocation of a long-term investor. We compare the performance under varying degrees of instability misspecification. The results show that while it is very costly to incorrectly assume that there is no parameter instability if there is so in reality, it does not seem to matter much what instability is assumed. Given the substantial time-variation that we find, this highlights the importance for the long-term investor to take parameter instability into account.

Does economic uncertainty help predict economic activity?

Chapter 4 examines the predictive power of economic uncertainty for real macroeconomic variables such as output and employment. Real options theory (Bernanke, 1983; Dixit and Pindyck, 1994) describes how economic uncertainty induces business cycle behavior. When uncertainty rises, firms delay investments and consumers postpone large investments, leading to lower economic activity. Bloom (2009) sparked the interest in testing this effect empirically by proposing measures of economic uncertainty. Since then, the literature has boomed and a large number of alternative measures have been proposed. All proposed measures aim to capture economic uncertainty, a latent – unobserved – entity that does not have an exact definition.

We gather a set of 15 measures and specify five categories in which each of them can be classified, based on how they estimate uncertainty. It can be based on (i) volatility, (ii) cross-sectional dispersion, (iii) news, (iv) surveys, and/or (v) forecast errors. Principal component analysis shows that the 15 measures have a fairly strong facture structure, with two common components that explain about 60% of the total variation in the uncertainty measures. The first is a general/average economic uncertainty measure, that loads fairly equally on most measures, with a slight emphasis on financial uncertainty. The second factor is a media/consumer confidence measure. It remains elevated after recessions.

Most of the empirical literature on economic uncertainty measures has focused on in-sample results. A perhaps more relevant question to policy makers is whether they can use these measures to obtain better forecasts, which may serve as input for policy decisions. Therefore, we conduct a real-time out-of-sample analysis, where we only use the data that was available to the forecaster at each point in time, taking into account data revisions of macroeconomic variables. The aim is to forecast the mean and various quantiles of four coincident variables: industrial production, employment, sales, and personal income. The coincident variables are used by the NBER dating committee to date peaks and troughs of the business cycle, thus linking our study to the real options theory.

We find that the uncertainty measures hold limited forecasting power for the mean compared to a benchmark model. The quantile forecasting results are somewhat better, in particular for employment, and at lower quantiles. These findings suggest a nonlinear relationship between economic uncertainty and economic activity.

Chapter 2

Cyclicality in losses on bank loans

Based on Keijsers, B., Diris, B., and Kole, E. (2018). Cyclicality in losses on bank loans. Journal of Applied Econometrics*, 33(4):533–552*

2.1 Introduction

Recent advances in the risk management of bank loans, such as stress tests for the banking sector, highlight the importance of investigating their risk in relation to the macroeconomic environment. As stated in the Basel II Accord, risk measures should "reflect economic downturn conditions where necessary to capture the relevant risks" (BCBS, 2005). Though loan defaults occur more frequently during economic downturns, it is neither clear whether the resulting losses also show cyclical behavior, nor whether they are related to the business cycle.

In this paper, we analyze the cyclical variation in bank loan losses, their relation to the business cycle and differences across loan categories, and show that this information can improve the risk management of banks. We base our analysis on a large sample of 22,080 defaults from Global Credit Data[1], spanning the period 2003–2010. Their databases contain loans and defaults, and information on the recovery, the seniority of the loans, and the industry, country and size of the borrowers. To exploit this detailed information, we build a model that can accommodate both time and cross-sectional

[1]In March 2015, the consortium changed its name to Global Credit Data from Pan European Credit Data Consortium.

variation in the default rate and the loss given default (LGD), and link them to macroeconomic variables. We show how the model can be used for risk management.

Our research brings new insights for two reasons. First, the cyclicality of bank loan losses might be different from the more commonly studied bond losses, whose LGD and default rates are cyclical, related to the business cycle and positively correlated with each other.[2] Bank loans differ in several fundamental respects from bonds. Banks monitor their loans more closely than bond owners. Their loans are often more senior and backed by collateral. Further, they can postpone the sale of collateral until a favorable economic state, hoping to receive a higher price.[3] As a consequence, the default rate and LGD for bank loans can be lower, less cyclical and less interrelated. Besides, our research is based on the actual workout LGD, whereas research on bond losses mostly uses the expected or market-implied LGD shortly after default.

Second, research on bank loans default is scarce because data are not easily available and typically constitute samples that are either short or focus on a single country or loan type, see, for example, Grunert and Weber (2009), Calabrese and Zenga (2010) and Hartmann-Wendels et al. (2014). We instead study a unique and rich data set that contains defaults for various countries and loan types over a period of eight years. Our model can reveal the influence of characteristics on time variation, in the form of different sensitivities to the same cycle or of adaption to different cycles, for example based on industry (see Shleifer and Vishny, 1992).

An important empirical difference between loan and bond LGD is their domain. Our data shows that bank loan LGD can exceed 100% or fall below 0%, whereas bond LGD always lies within this interval. If the LGD exceeds 100%, the bank loses more than the initial loan, for example because of principal advances (the bank lends an additional amount to the borrower for recovery). If the LGD falls below 0%, the bank recovers more than the initial loan, for example because of penalty fees, additional interest and recovered principal advances. Moreover, the LGD distribution is bimodal, with most loans being close to either a full recovery or a full loss. Schuermann (2004)

[2]See the surveys by Allen and Saunders (2003) and Schuermann (2004). Pesaran et al. (2006b), Duffie et al. (2007) and Azizpour et al. (2015) model the relation between the default rate and macro variables, whereas Frye (2000) and Creal et al. (2014) also include the LGD in their models.

[3]Acharya et al. (2007) shows the importance of the fire-sales effect for the LGD of bonds.

shows this bimodality for bond LGD. Because of the differences in domain, we cannot use existing models for bond LGD in our analysis.

Our model links macroeconomic variables, the default rate, and the LGD via latent factors that follow autoregressive processes. While this set-up has been used before (see e.g. Pesaran et al., 2006b; Koopman et al., 2012), the LGD component in our model is new. We model the LGD as a mixture of two normal distributions that differ in their means to capture the bimodality and LGD outside the $[0, 1]$ interval. So, losses can be mild or severe with a certain probability that depends on the factors.[4] The parameters that relate the LGD and the default rate to the latent factors can depend on loan characteristics.

We estimate our hierarchical model using a Bayesian Gibbs sampler. The main advantage of the Gibbs sampler is that it allows us to divide the complicated overall estimation problem into smaller subproblems (the different Gibbs steps) which are easier to deal with. The main complication in the estimation is that the probability of default and of severe losses depend in a nonlinear way on parameters and factors. We solve this issue using a new data augmentation technique proposed in Polson et al. (2013) which leads to easier to analyze Gaussian likelihoods conditional on the new Pólya-gamma latent variables.

Our results show the presence of a macro factor that captures the business cycle, and a default-specific factor that captures variation in the credit cycle unrelated to the business cycle.[5] The LGD distribution varies over time via the probability of a loss being severe or mild. We do not find evidence that the average LGD for either severe or mild losses varies over time. In line with earlier research (see e.g., Frye, 2000; Schuermann, 2004), default rates and LGD of bank loans go up during economic downturns. However, the default-specific factor has an opposite effect, indicating that increases in the default rate unrelated to the business cycle are typically caused by borrowers that miss a payment but catch up later, and hence the LGD is low or zero.

Loan and borrower characteristics influence the cyclical variation in the default rate and LGD. The LGD of a collateralized loan is on average lower, but fluctuates more strongly over the business cycle as in Bruche and González-Aguado (2010). Loans

[4] Knaup and Wagner (2012) also distinguish severe and mild losses to derive a bank's credit risk indicator.

[5] Koopman et al. (2012) refer to this default-specific factor as frailty factor.

to small and medium enterprises exhibit stronger fluctuations in both their default rates and LGDs compared to large corporates. While we find differences in the factor sensitivities of default rates and LGDs for different industries, we do not find evidence for industry-specific cycles.

We use our model to investigate the loss distribution of a fictional loan portfolio as in Miu and Ozdemir (2006) in a point-in-time setting. We calculate the expected loss and the economic capital (the difference the 99.9% quantile and the mean of the loss distribution). Both statistics show considerable variation over the business cycle. From peak to bottom of the cycle, the economic capital increases by a factor two. It is quite sensitive to the cyclical variation in the LGD. 22% of its increase over the cycle can be attributed to time-variation in the LGD. This result illustrates the importance of accounting for time-variation in the LGD that is related to the business cycle in risk management.

Our findings contribute to the literature on credit risk in two ways. First, we show that just as for bonds, the LGD of bank loans varies over the business cycle. The LGD for bank loans is generally much lower than for bonds, but can still double in times of distress. Whereas the average bond LGD varies from 25% to 80% as reported by Schuermann (2004), Altman et al. (2005), and Bruche and González-Aguado (2010), we find that loan LGD varies from 14% to 29% over time, though the periods that they study do not fully match with ours.[6] We also show how the cyclical behavior of the LGD is affected by characteristics. These results complements papers that only report how industry characteristics influence the (average) LGD (Schuermann, 2004; Acharya et al., 2007) or how the impact of seniority varies over the business cycle (Bruche and González-Aguado, 2010).

Second, our model exploits the panel structure of the LGD observations in a novel way. Though the time-varying mixture distribution for the LGD adds a layer of complexity to our model, it shows in detail how the LGD distribution changes. These insights would be lost if we would model the time-variation of the cross-sectionally averaged LGD or analyse the LGD distribution at each point in time. Our methods can be seen as an extension of Koopman and Lucas (2008), who only model the default

[6]Schuermann (2004) also reports variation in LGD between 20% and 55% for traded bank loans, whereas our loans are not traded.

rate, and Koopman et al. (2012), who add macro variables, by modeling the LGD as well. We also extend Calabrese (2014a), who only models the LGD, by linking the LGD to the default rate and macro variables. We deviate from Creal et al. (2014) and Bruche and González-Aguado (2010), who use the Beta distribution for the LGD, because the mixture of normals in our model can more easily accommodate observations outside $[0, 1]$. The default-specific factor also extends single Markov-switching business cycle of Bruche and González-Aguado (2010). Moreover, our model can flexibly include covariates and can easily be adopted in analyses of the risk of loan portfolios.

2.2 Data

We combine observations of macroeconomic variables, defaults of bank loans and their losses. Because we want to focus on the part for bank loans in our model, we make standard choices for the variables that represent the business cycle. In particular, we consider three macro variables that are also analyzed by Creal et al. (2014): the gross domestic product (GDP), industrial production (IP) and the unemployment rate (UR). The variables included are the growth rates with respect to the same quarter in the previous year. To match the mostly European loan data sets, we use macro variables of European countries. We provide an overview of the macro series in Appendix 2.A.

Besides the macroeconomic component, our model contains components for the default rate and the LGD. We calculate the default rate as the number of defaulted loans divided by the number of loans at the start of a year. The LGD is the amount lost as a fraction of the exposure at default (EAD). We have observations of the workout LGD (also known as economic LGD), which is based on the actual cash flows after default. They are timed to coincide with the default date. By analyzing workout LGD, we further complement studies of the LGD of bonds, which mostly use the market prices of bonds soon after default (see e.g. Schuermann, 2004). Our sources for default and LGD observations are unique databases from Global Credit Data, to which we have access via NIBC, a Dutch bank. We first discuss the databases, and then turn to the data sets that we analyze.

2.2.1 Global Credit Data

Global Credit Data (GCD) is an international cooperation of banks to support statistical research for the advanced internal ratings-based approach (IRB) under the Basel accords.[7] The members pool information on loans and defaults to create two anonymous databases, the LGD database with information about the losses on resolved defaults, and the loan database with information to analyze the default rate. GCD has been founded in 2004 by 11 banks, and has grown to 53 members (April 2017). It focused originally on LGD, but later expanded its focus to the default rate.

Data quality is a crucial issue for GCD. It sets specific and detailed rules with regard to the default information that its members should submit. Default definitions are based on the Basel accords, and GCD uses its own standards to characterize further aspects, such as the size, industry and region of the borrower. Before default data is included in the databases, GCD conducts regular audits to check whether the data that a member submitted complies with its standards. A methodology committee regularly reviews these rules. To stimulate participation, data is available to members on the give-to-get principle. To obtain default data from a given year and loan category, members have to submit their own default data for that given year and category.

The LGD database contains the cash flows of all defaulted loans of the member banks. Because defaults are only included after the recovery process has ended, theses cash flows are final and realized. It also contains the default and resolution dates, the seniority of the loan, the presence of collateral, the size or type of the borrower as well as the industry and country to which the borrower belongs. GCD aims at a representative database with defaults dating from 1998, but some defaults go back as far as 1983. Members are obliged to submit defaults dating from 2001 onwards. Table 2.B.1 in Section 2.B shows that a stable number of 40–45 banks contribute to the LGD database after 2001.

The loan database contains information about borrowers and defaults, in particular their size or type, industry and country. Information about the seniority and the presence of collateral is not available in this database, as these characteristics are not seen as default drivers. GCD started with the construction of this database in 2009.

[7] See `https://www.globalcreditdata.org/` for general information and `https://www.globalcreditdata.org/index.php?page=members` for an up-to-date overview of the members.

It aims at a representative database from 2000 onwards. Table 2.B.1 shows that the number of banks contributing to this database is considerably less than to the LGD database. Because the number of defaults per bank is generally small, the need for pooling is larger for LGD than for loan data.

GCD provides a new version of the LGD database semi-annually, and of the loan database annually. Via NIBC we have access to the June-2014 version of the LGD database and the June-2013 version of the default database. While the parts of the databases that are available to NIBC vary per characteristic, they represent a large proportion of the total databases.

2.2.2 Sample selection

NIBC's LGD data set contains 92,797 loans with 46,628 counterparties. We exclude non-representative observations based on Höcht and Zagst (2007) and NIBC's internal policy (see Appendix 2.B.2 for details). Following NIBC's practice, we discount all cash flows by the two-year swap rate plus the spread from the loan. When the contractual spread is unavailable, we use the average spread of all defaulted loans. We transform the resulting workout LGD to a percentage of the EAD. We order the LGDs by quarter in line with the frequency of the macro variables.

The loan data set consists of in total 2.80 million loans of which 37,385 go into default, leading to an average default rate of 1.34%. The number of loans and defaults is specified per year. Because the number of contributing banks to this database is lower, the number of defaults is lower than in the LGD data set. Though we filter outliers from the LGD data set based on the size of the loan (measured by EAD), we cannot do so for the loan data set because the base value of the loan is not recorded.

Because the loan database starts in 2003, and the most recent LGD observations may be biased, we restrict our analysis to the period 2003–2010. The LGD is positively correlated with the workout period, i.e. the period needed to resolve the default. The most recent observations are few, have a short workout period by construction, and their LGDs are therefore typically small. After filtering the raw LGD data set, our sample contains 22,080 LGD observations of mostly European defaults, one of the most comprehensive datasets for bank loan LGD studied thus far. The largest data set that is reported by Grunert and Weber (2009) in their summary of empirical studies on bank loan recovery rates contains 5,782 observations over the period 1992–1995. More recently, Calabrese and Zenga (2010) and Calabrese (2014a,b) study a portfolio of 149,378 Italian bank loan defaults resolved in 1999. Hartmann-Wendels et al. (2014) consider 14,322 defaulted German lease contracts from 2001–2009. Though large, these studies focus on defaults from a single country or a single loan type whereas our dataset is more extensive.

2.2.3 Sample characteristics

The average LGD and the default rate both exhibit cyclical behavior (see Figure 2.B.1 in Section 2.B). Both increase during the financial crisis, though the peak of the LGD (of 28.4%) falls in 2008, whereas the peak of the default rate (of 2.2%) falls in 2009. By 2010, the average LGD is back at its pre-crisis level, but the default rate remains higher.

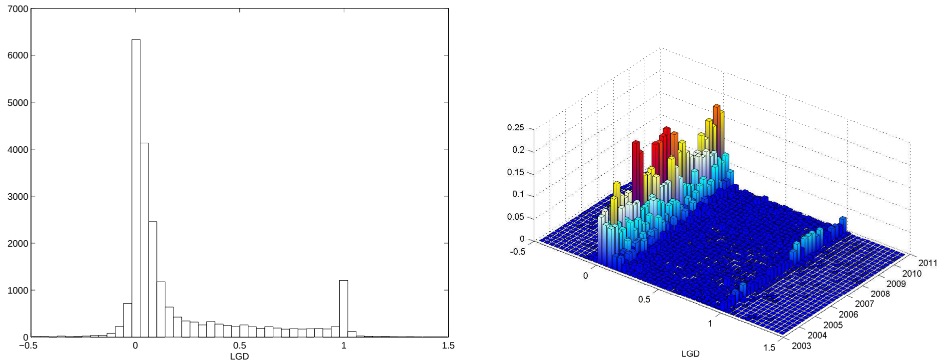
For our LGD sample we also investigate the cross-sectional distribution. When we pool all observations, the distribution shows bimodality, as is typical for LGD data (see e.g. Schuermann, 2004). Figure 2.1a shows that the LGD is either close to zero, or close to one when the complete value of the loan is lost. A substantial part (12.5%) of the LGD observations falls outside the $[0, 1]$ interval. These exceedances are related to principal advances, legal costs or penalty fees. A principal advance is an additional amount lent to aid the recovery of the defaulted borrower. If none of it is paid back, the losses exceed the EAD and the LGD exceeds one. If on the other hand the full debt is recovered, including penalty fees, legal costs and principal advances, the amount received during recovery exceeds the EAD and the LGD is negative. In line with Höcht and Zagst (2007) and Hartmann-Wendels et al. (2014), LGD observations below -0.5 or above 1.5 have been removed.

The bimodality in the distribution of the LGD is present in every quarter as shown by Figure 2.1b. In 2008, the financial crisis leads to higher peaks at zero and at one, indicating more defaults with either no loss or a full loss. In 2009, the peak around one is still present, but the peak at zero is substantially lower. The large proportion

Figure 2.1: Empirical distribution of LGD

(a) Histogram of pooled LGD observations

(b) Distribution of LGD observations per quarter



Panel a shows the histogram of the pooled set of LGD observations. Panel b shows the empirical distribution of the LGD observations per quarter. We use all LGD observations over the period 2003–2010 after applying the filters in Section 2.B.2.

of full losses explains the large average LGD in those years. Our modeling framework exploits both the bimodality and the time variation of the LGD.

We report the effect of loan and borrower characteristics on the LGD statistics in Table 2.1. The dip statistic of Hartigan and Hartigan (1985) indicates that all large subsamples are bimodal. Because some subsamples contain a small number of defaults, we limit our analysis to those subsets with at least 3,200 observations, which corresponds with 100 observations per quarter.

Panel A shows that, as can be expected, loans of lower seniority have on average a higher LGD, and a increased probability of an LGD above 0.5. Most loans in our sample are senior, and 44% have some form of collateral. The non-parametric Kruskal-Wallis (KW) test rejects the hypothesis that the distributions of the subsets have the same location.

In Panel B we split the sample based on the size or type of the borrower. GCD distinguishes SMEs, large corporates and some more specific types of financing, for example for real estate, aircraft, or shipping. While these specific types are interesting, the number of observations is too small, and we concentrate on loans to SMEs and large corporates. The differences between those two loan categories are small, but the KW-test still indicates that they are significant.

Panel C categorizes the loans according to the industry of the borrower as indicated by GCD. A large part of the loans (67%) is concentrated in three industries, being

Table 2.1: LGD statistics per loan and borrower characteristics

Characteristic	Defaults	Average	Fraction LGD > 0.5	HDS p-value
Total	22,080	0.204	0.170	0.000
Panel A: Seniority				
Senior secured*	9,723	0.175	0.138	0.000
Senior unsecured*	12,011	0.222	0.191	0.000
Subordinated secured	110	0.289	0.255	0.002
Subordinated	236	0.427	0.419	0.000
Panel B: Borrower size or type				
SME*	12,028	0.193	0.164	0.000
Large Corporate*	6,496	0.199	0.159	0.000
Real Estate Finance	2,068	0.326	0.284	0.000
Aircraft Finance	556	0.088	0.045	0.000
Shipping Finance	331	0.077	0.054	0.100
Project Finance	302	0.177	0.132	0.002
Banks	276	0.286	0.286	0.000
Public Services	23	0.246	0.174	0.234
Panel C: Industries				
Industrials*	6,944	0.178	0.150	0.000
Financials*	4,629	0.217	0.178	0.000
Consumer Staples*	3,232	0.186	0.162	0.000
Unknown	2,817	0.309	0.279	0.000
Information Technology	1,384	0.188	0.155	0.000
Consumer Discretionary	1,089	0.196	0.128	0.034
Other	606	0.147	0.102	0.000
Telecommunication Services	410	0.203	0.183	0.304
Utilities	391	0.145	0.079	0.280
Health Care	366	0.123	0.082	0.086
Materials	212	0.147	0.127	0.534

This table presents the number of defaults, the average LGD, the fraction of defaults with an LGD larger than 0.5 and the p-value of Hartigan and Hartigan's (1985) dip statistic (HDS) using 500 bootstraps, to test the null hypothesis of a unimodal distribution versus the alternative of a multimodal distribution, for different subsets of the LGD data set. Subsets with more than 3,200 observations (indicated by a *) are selected for analysis with our model in Section 2.4.2.

industrials, financials or consumer staples. Industrials have the lowest average LGD and proportion of defaults with LGDs below 0.5, followed by consumer staples, and then financials. The KW-test indicates again significant differences in the locations of the distributions.

In Section 2.B we compare our LGD data set to the LGD information of bonds in Moody's Ultimate Recovery Database (URD). The recovery of bank loans being different from bonds is an important motivation for our paper. Loans are typically more senior, more often have collateral, and lead to more closely monitoring (see Emery et al., 2004; Schuermann, 2004). We find that the LGDs on loans and bonds are bimodal, though bonds encounter more large LGDs. Our analysis will shed more light on the behavior of the LGD of bank loans, and the role that for example seniority plays.

2.3 Methods

2.3.1 Model specification

We propose a mixed measurement model in the style of Koopman et al. (2012) (whose notation we follow) and Creal et al. (2014), where the observations can follow different distributions and depend on latent factors. Our model contains a total of N variables at each point in time, though not all variables are always observed. We separate them in three sets being macro, loan and LGD variables, labeled m, l and d. We use y_{it}^{c} to denote the time t observation of variable i in category c, $c =$m, l, d. We use N^c to denote the size of a category. We collect all variables in the vector $\boldsymbol{y} = (\boldsymbol{y}^{\mathrm{m}\prime}, \boldsymbol{y}^{\mathrm{l}\prime}, \boldsymbol{y}^{\mathrm{d}\prime})'$.

The set of K latent factors $\boldsymbol{f}_t$ form the central part of our model, through which all observed processes are linked. We distinguish K^{m} macro factors $\boldsymbol{f}_t^{\mathrm{m}}$ that capture the business cycle, and affect all observed variables. Next to these macro factors, we introduce K^{l} loan factors $\boldsymbol{f}_t^{\mathrm{l}}$ that influence both the default and the LGD variables. The K^{d} LGD factors $\boldsymbol{f}_t^{\mathrm{d}}$ influence only the LGD variables. The factors $\boldsymbol{f}_t^{\mathrm{l}}$ and $\boldsymbol{f}_t^{\mathrm{d}}$ capture the dynamics of the credit cycle that are unrelated to the business cycle. Because of this general setup, we can investigate whether the LGD variables are related to the business cycle, the credit cycle, their own separate LGD cycle, or no cycle at all.

Following Koopman et al. (2012), we assume that $\boldsymbol{f}_t$ follows a VAR(1) process,

$$\boldsymbol{f}_{t+1} = \boldsymbol{\Phi} \boldsymbol{f}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{2.1}$$

where the coefficient matrix $\boldsymbol{\Phi}$ is a diagonal matrix. The innovations are serially uncorrelated. These restrictions ensure that the loan and LGD factors are independent of the macro factors, in line with the literature on credit risk (see e.g. Duffie et al., 2009; Koopman et al., 2012). We impose that $\boldsymbol{f}_t$ is stationary, so $|\phi_{kk}| < 1$. The initial state vector $\boldsymbol{f}_1$ follows the unconditional distribution of the latent process, that is $\boldsymbol{f}_1 \sim \mathrm{N}(0, \boldsymbol{\Sigma}_f)$ with $\boldsymbol{\Sigma}_f$ solving $\boldsymbol{\Sigma}_f = \boldsymbol{\Phi}\boldsymbol{\Sigma}_f\boldsymbol{\Phi}' + \boldsymbol{\Omega}$. For identification, we impose that the unconditional variance equals the identity matrix $\boldsymbol{\Sigma}_f = \boldsymbol{I}$.

The first variable set contains the N^{m} macro variables, which depend linearly on the latent macro factors,

$$\boldsymbol{y}_t^{\mathrm{m}} = \boldsymbol{\alpha}^{\mathrm{m}} + \boldsymbol{B}^{\mathrm{m}} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}^{\mathrm{m}}), \tag{2.2}$$

where $\boldsymbol{\alpha}^{\mathrm{m}}$ is a vector of size N^{m} containing the intercepts, and $\boldsymbol{B}^{\mathrm{m}}$ is a $N^{\mathrm{m}} \times K^{m}$ matrix with coefficients. The innovations in the macro variables follow a normal distribution with mean zero and variance matrix $\boldsymbol{\Sigma}^{\mathrm{m}}$ and are independent of the latent process. We standardize the macro variables to have zero mean and unit variance to ease the comparison of their relation with the latent factor (cf. Stock and Watson, 2002b). For identification, we impose that $\boldsymbol{B}^{\mathrm{m}}$ is lower triangular with a sign restriction on the diagonal elements, and $K^{\mathrm{m}} < N^{\mathrm{m}}$.

The second variable set contains the status of the loans, y_{it}^{l}. Loan i at time t can either be performing ($y_{it}^{\mathrm{l}} = 0$) or in default ($y_{it}^{\mathrm{l}} = 1$). Conditional on $\boldsymbol{f}_t$, y_{it}^{l} follows a Bernoulli distribution with default probability p_{it}^{l},

$$y_{it}^{\mathrm{l}} | \boldsymbol{f}_t \sim \mathrm{Bernoulli}(p_{it}^{\mathrm{l}}) \tag{2.3}$$

$$p_{it}^{\mathrm{l}} = \Lambda(\alpha_i^{\mathrm{l}} + \boldsymbol{\beta}_i^{\mathrm{l}\prime} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{l}\prime} \boldsymbol{f}_t^{\mathrm{l}}), \tag{2.4}$$

where, $\Lambda(z) = 1/(1 + \exp(-z))$ is the logistic function. The coefficients α_i^{l}, $\boldsymbol{\beta}_i^{\mathrm{l}}$ and $\boldsymbol{\gamma}_i^{\mathrm{l}}$ can depend on J^{l} loan-specific characteristics. Collecting them together with an intercept in the vector $\boldsymbol{x}_i^{\mathrm{l}}$, we obtain $\alpha_i^{\mathrm{l}} = \boldsymbol{\alpha}^{\mathrm{l}\prime} \boldsymbol{x}_i^{\mathrm{l}}$, $\boldsymbol{\beta}_i^{\mathrm{l}} = \boldsymbol{B}^{\mathrm{l}\prime} \boldsymbol{x}_i^{\mathrm{l}}$ and $\boldsymbol{\gamma}_i^{\mathrm{l}} = \boldsymbol{\Gamma}^{\mathrm{l}\prime} \boldsymbol{x}_i^{\mathrm{l}}$, where $\boldsymbol{\alpha}^{\mathrm{l}}$ is a vector and $\boldsymbol{B}^{\mathrm{l}}$ and $\boldsymbol{\Gamma}^{\mathrm{l}}$ are matrices. For identification, we impose that $\boldsymbol{\Gamma}^{\mathrm{l}}$ is lower triangular with sign-restricted diagonal elements. The number of loan factors should not exceed the number of characteristics, $K^{\mathrm{l}} \leq J^{\mathrm{l}}$.

We assume that conditional on $\boldsymbol{f}_t$, the default status of the loans are mutually independent. When no loan-specific characteristics are used, the default rate at time t is the same for all loans, $p_{it}^{\mathrm{l}} = p_t^{\mathrm{l}} = \Lambda(\alpha^{\mathrm{l}} + \boldsymbol{\beta}^{\mathrm{l}\prime} \boldsymbol{f}_t^{\mathrm{m}} + \gamma^{\mathrm{l}} \boldsymbol{f}_t^{\mathrm{l}})$, and the number of defaulted loans follows a binomial distribution. When the characteristics are categorical and separate the loans into groups, for example based on industry, country or borrower type, the number of defaulted loans within a particular group also follows a binomial distribution.

Whereas the structure of our model thus far is similar to Koopman et al. (2012), we propose a novel part for the final set of variables, which are the LGDs of a defaulted loan, y_{it}^{d}. Based on the empirical distribution in Figure 2.1, we distinguish defaulted loans with a severe loss (close to a full loss) from those with a mild loss (close to a full recovery). We model the default type by a latent binary variable s_{it} that takes a value zero (one) to indicate a mild (severe) loss.[8] Conditional on $\boldsymbol{f}_t$, s_{it} follows a Bernoulli distribution with parameter p_{it}^{d}. Conditional on s_{it}, y_{it}^{d} follows a normal distribution. Mathematically, this part of the model can be written as

$$y_{it}^{\mathrm{d}} \sim \begin{cases} \mathrm{N}(\mu_{i0}, \sigma_i^2) & \text{if } s_{it} = 0 \\ \mathrm{N}(\mu_{i1}, \sigma_i^2) & \text{if } s_{it} = 1 \end{cases} \tag{2.5}$$

$$s_{it} | \boldsymbol{f}_t \sim \mathrm{Bernoulli}(p_{it}^{\mathrm{d}}) \tag{2.6}$$

$$p_{it}^{\mathrm{d}} = \Lambda(\alpha_i^{\mathrm{d}} + \boldsymbol{\beta}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{l}} + \boldsymbol{\delta}_i^{\mathrm{d}\prime} \boldsymbol{f}_t^{\mathrm{d}}). \tag{2.7}$$

We assume that conditional on $\boldsymbol{f}_t$, the LGDs are independent.

Conditional on $\boldsymbol{f}_t$, y_{it}^{d} follows a mixture of two normal distributions[9] that differ in their means,

$$\mu_{is} = \boldsymbol{\mu}_s' \boldsymbol{x}_i^{\mathrm{d}}, \quad s = 0, 1 \tag{2.8}$$

with $\boldsymbol{\mu}_s$ a vector of size $J^{\mathrm{d}} + 1$. These means can again be a function of the J^{d} default characteristics that we collect together with an intercept in the vector $\boldsymbol{x}_i^{\mathrm{d}}$ of size $J^{\mathrm{d}} + 1$. To prevent label switching, we impose $\mu_{i0} < \mu_{i1}$ over the support of $\boldsymbol{x}_i^{\mathrm{d}}$. We also allow the variance to be a function of the loan characteristics,

$$\ln \sigma_i^2 = \boldsymbol{\lambda}' \boldsymbol{x}_i^{\mathrm{d}}, \tag{2.9}$$

with $\boldsymbol{\lambda}$ a vector of size $J + 1$. We do not allow the variances to depend on the default type s_{it}. Because of this restriction and $\mu_{i0} < \mu_{i1}$, the probability $\Pr[s_{it} = 0 | y_{it}^{\mathrm{d}}, \boldsymbol{f}_t]$ is a decreasing function of y_{it}^{d}. So, the probability of a default being labeled mild decreases for increasing LGD. Without this restriction, losses that become more and

[8] This model component of latent default types is similar to the latent distinction between good and bad loans or investment projects that is used to include asymmetric information in models of capital structure, see e.g. Flannery (1986).

[9] In section 2.4.4, we consider a mixture of Student's t distributions.

more extreme would be inferred with an increasing probability to be of the type with the largest variance.

The (conditional) probability of a severe loss, p^{d}_{it} depends via a logit transformation on the latent factors (cf. Equation (2.4)). The coefficients can depend on the default-specific characteristics, that is $\alpha^{\mathrm{d}\prime}_i = \boldsymbol{\alpha}^{\mathrm{d}}\boldsymbol{x}^{\mathrm{d}}_i$, $\boldsymbol{\beta}^{\mathrm{d}}_i = \boldsymbol{B}^{\mathrm{d}\prime}\boldsymbol{x}^{\mathrm{d}}_i$, $\boldsymbol{\gamma}^{\mathrm{d}}_i = \boldsymbol{\Gamma}^{\mathrm{d}\prime}\boldsymbol{x}^{\mathrm{d}}_i$ and $\boldsymbol{\delta}^{\mathrm{d}}_i = \boldsymbol{\Delta}^{\mathrm{d}\prime}\boldsymbol{x}^{\mathrm{d}}_i$, where $\boldsymbol{\alpha}^{\mathrm{d}}$ is a vector and $\boldsymbol{B}^{\mathrm{d}}$, $\boldsymbol{\Gamma}^{\mathrm{d}}$ and $\boldsymbol{\Delta}^{\mathrm{d}}$ are matrices. For identification, we impose that $\boldsymbol{\Delta}^{\mathrm{d}}$ is lower triangular with positive diagonal elements. The number of default factors should not exceed the number of characteristics, $K^{\mathrm{d}} \leq J^{\mathrm{d}}$.

The latent factors influence the LGD via the probability p^{d}_{it} with which the default is severe. We do not include an effect of the factors on the average LGD for a given type μ_{is}, because the location of the modes of the distributions in Figure 2.1b stay approximately constant. The variation in the relative heights of the peaks is driven via the mixture probability p^{d}_{it}.

2.3.2 Estimation

We use uninformative priors that are specified in Section 2.C. We impose the identification restrictions explained in the previous section using the priors.

We use a Gibbs sampler to estimate the model in a Bayesian way. This means that we simulate from the conditional posterior distributions for each parameter to obtain draws from the full posterior distribution of all parameters. The main advantage of the Gibbs sampler is that it allows us to divide our complicated estimation problem into smaller subproblems (the different Gibbs steps), which makes the estimation feasible. The following explains how we estimate the model without loan and default characteristics. The estimation of the model with these characteristics is a straightforward extension as explained in Section 2.C.

The main complication in the simulation is that the probabilities of default and of a severe loss depend in a nonlinear way on the parameters and the latent factors via the logistic function. We tackle this complication by using new results in Polson et al. (2013) and Windle et al. (2013), who show that one can easily sample the parameters and latent factors from their conditional posterior distributions once one adds auxiliary latent variables - denoted ω^{l}_t and ω^{d}_t here - to the model. The main idea is that we obtain an easier to analyze linear Gaussian (state space) model conditional on the auxiliary latent variables. This new result makes our analysis feasible.

For ease of notation, we define $\boldsymbol{A}^{\mathrm{m}}$ as the matrix that collects the intercepts and slopes in Equation (2.2), $\boldsymbol{\alpha}^{\mathrm{l}}$ as the vector which collects all parameters in Equation (2.4), and $\boldsymbol{\alpha}^{\mathrm{d}}$ as the vector which collects all parameters in Equation (2.7).

The Gibbs sampler consists of the following steps:

- Macro module
  Sample $\boldsymbol{A}^{\mathrm{m}}$ from the matricvariate normal distribution (re-draw until $\boldsymbol{A}^{\mathrm{m}}_{2,1} = \boldsymbol{B}^{\mathrm{m}}_{1,1} < 0$ for identification of the first factor) and $\boldsymbol{\Sigma}^{\mathrm{m}}$ from an inverse Wishart distribution.

- Loan status module
  Sample latent variable ω^{l}_t from a Pólya-gamma distribution and $\boldsymbol{\alpha}^{\mathrm{l}}$ from a multivariate normal distribution (re-draw until $\boldsymbol{\alpha}^{\mathrm{l}}_3 = \gamma^{\mathrm{l}} > 0$ for identification of the second factor).

- Loss given default module
  Sample latent variable ω^{d}_t from a Pólya-gamma distribution for all t and $\boldsymbol{\alpha}^{\mathrm{d}}$ from a multivariate normal distribution (re-draw until $\boldsymbol{\alpha}^{\mathrm{d}}_4 = \delta^{\mathrm{d}} > 0$ for identification of the third factor). Sample s_{it} from a Bernoulli distribution for all i, t, μ_0 and μ_1 from normal distributions (re-draw until $\mu_0 < \mu_1$) and σ^2 from an inverse gamma-2 distribution.

- Factor module
  Simulate the latent factors using the simulation smoother of Durbin and Koopman (2002b). Sample ϕ_{jj} for factors $j = 1, .., K$ using a Metropolis-Hastings step that imposes that $|\phi_{jj}| < 1$.

We refer to Section 2.C for the exact distributions and derivations. We retain 100,000 draws after a burn-in of 50,000 draws to obtain results. Increasing the number of simulations does not impact results.

2.3.3 Motivation of modeling choices

Our model is designed to get a detailed view of the variation in LGDs and default rates, both over time and in relation to loan characteristics, and also of the interplay between these two sources of variation. Though simpler analyses are available, they cannot satisfactorily answer our research questions, because they do not fully exploit the richness of our data set. We first highlight the appealing properties of our model, and then indicate how it deviates from the other advanced alternatives that have recently been proposed.

Arguably, our model has two layers of complexity. The first is the latent factor structure that drives the time-series dynamics and dependence of our variables. Alternatively, the probability of default and of a severe loss can be linked directly to the macro variables. An additional default-specific (frailty) factor would then be difficult to include. The literature on credit and default risk shows that a latent factor structure can accurately capture this issue.[10] We do not use a Markov-switching process as in Bruche and González-Aguado (2010) or Calabrese (2014a), because autoregressive processes more naturally link to the gradual changes in macro variables. Our model is more advanced than the models proposed by Frye (2000); Gordy (2003); Pykhtin (2003) as we can include more factors and explicitly model their behavior and effect on the probability of default and LGD.

The second layer stems from the panel structure of our LGD observations, which we model by a mixture distribution with time-varying weights. Easier solutions could consist of data reduction by modeling the time-variation of the cross-sectional average of the LGD, or a separate or two-step analysis of the LGD distribution at each point in time. Because we fully model the LGD distribution, we can incorporate the determinants of both the cross-sectional variation and the time-variation. We think that in particular the effect of the loan and default characteristics add interesting insights to our analysis. Besides, we circumvent a generated regressor problem as in Pagan (1984).

Our model component for the LGD differs from existing models. We do not use a standard Beta distribution for the LGD as in Creal et al. (2014) and Bruche

[10] See among others Frye (2000); Gordy (2003); Pykhtin (2003); Pesaran et al. (2006b); Koopman and Lucas (2008); Koopman et al. (2012); Creal et al. (2014); Azizpour et al. (2015).

and González-Aguado (2010) or a mixture of point masses at 0 and 1 and a Beta distribution as in Calabrese (2014b). These distributions only have support on the unit interval. Because over 10% of the LGD observations are outside the $[0, 1]$ interval (see Figure 2.1a), using these distributions would require a transformation of the data. The results for the discrete-continuous distribution are difficult to interpret, since the LGDs drawn from the Beta component can be arbitrarily close to 0 and 1.

2.4 Results

2.4.1 Models without loan and default characteristics

We start our analysis by investigating the general relation between the macro variables, the defaults and the LGDs. In the basic specification we investigate whether the defaults and LGDs exhibit cyclical behavior, and how many factors are needed to capture it. We do not include loan nor default characteristics in this analysis. Based on the evidence in Duffie et al. (2007); Koopman et al. (2012); Azizpour et al. (2015) that favor at least one macro and one default-specific factor, we take the two-factor model with a macro and a loan factor as our starting point. The macro factor can affect all processes, and the loan factor can only influence the probabilities of default and of a severe LGD. We then investigate the added value of an additional default factor. We present our estimation results in Table 2.2.

Our results for the two-factor specification show a persistent macro factor, and a loan factor with much quicker mean reversion. The macro variables show clear exposures to the macro factor. A high value signals a recessionary state of the world, with low growth rates for GDP and industrial production, and a high unemployment rate. Because the macro variables have been normalized with unconditional variances equal to one, an increase of the factor by one leads to changes equal to one standard deviation times their loadings, so decreases of GDP and IP by $0.969 \cdot 2.57\% = 2.49\%$ and $0.802 \cdot 6.41\% = 5.14\%$, and a increase of the unemployment rate by $0.879 \cdot 0.90\% = 0.79\%$. The factor has a slightly stronger effect on GDP growth and the unemployment rate than on the growth of industrial production.

Unsurprisingly, the business cycle factor positively affects the probability of default. On average its marginal effect is 0.11%, which is economically large, compared to the average default probability of 0.31% per quarter. The sensitivity to the loan factor is smaller than to the macro factor, and consequently the average marginal effect is smaller (0.080%) as well, though still sizable.

Our main interest is the component for the LGD. Here we also see a clear effect of the business cycle. The posterior distribution of β^{d} has a mean of 0.328, but the spread is wide. So, during a recession when the latent factor is positive, the probability of defaulted loans with a severe loss increases. A mild LGD has a mean of 7.2%, whereas severe losses are on average much larger at 82.9%. The two loss types are clearly different, as indicated by the standard deviation of 13.1%. On average, the probability of a severe loss is 17.4%, and the average LGD equals 20.4%. The marginal effect of the macro factor on the severe loss probability is on average 4.5%, which translates to an increase of the average LGD by 3.4%. Though this effect is less strong than for the default rate, it is still quite substantial. The loan factor has a negative effect on the LGD, as indicated by the negative posterior mean and small standard deviation for γ^{d}. A positive shock to this factor leads to more defaults, but decreases the probability of a severe loss. These effects indicate that these defaults are related to firms that miss a loan payment (interest or repayment), but catch up afterwards.

To get a better understanding of the factors, we plot their evolution in Figure 2.2a. The macro factor starts negative, which indicates the benign economic environment of the first part of our sample period. After 2008, it shows a sharp increase, corresponding with the credit crisis. The loan factor shows more erratic behavior, and is less persistent than the macro factor. It is high around 2006 and low around 2008. So, given the sensitivities the factor has an upward effect on the default probabilities but downward on the LGDs around 2006. Around 2008, the effects are reversed, indicating fewer defaults related to temporary delays in payments.

The plot of the fit over time of the macro variables, the realized default rate, and the average LGD in Figure 2.3 shows which part of their variation is captured by the factors. The macro factor reasonably tracks the macro variables, in particular during the great recession. The model-implied and realized default rate series almost coincide. While the macro factor captures the long-term swings in the default rate,

Table 2.2: Parameter estimates, model without loan and default characteristics

	Macro and loan factor		Macro, loan and default factor		Macro and loan factor, constant LGD	
	Panel A: Factor					
ϕ_{11}	0.856	(0.078)	0.917	(0.050)	0.939	(0.045)
ϕ_{22}	0.322	(0.207)	0.271	(0.220)	0.290	(0.215)
ϕ_{33}			0.778	(0.130)		
	Panel B: Macro variables					
α_{GDP}	−0.008	(0.317)	−0.198	(0.391)	0.580	(0.668)
β_{GDP}	−0.969	(0.531)	−0.904	(0.562)	−1.102	(0.824)
α_{IP}	−0.007	(0.284)	−0.156	(0.338)	0.475	(0.572)
β_{IP}	−0.802	(0.465)	−0.725	(0.483)	−0.889	(0.714)
α_{UR}	0.008	(0.298)	0.253	(0.467)	−0.707	(0.801)
β_{UR}	0.879	(0.491)	1.125	(0.673)	1.376	(0.971)
WAIC2	110.3		−2.4		−9.3	
	Panel C: Loan status					
α^{l}	−5.850	(0.097)	−5.759	(0.122)	−6.057	(0.197)
β^{l}	0.371	(0.201)	0.318	(0.200)	0.360	(0.271)
γ^{l}	0.258	(0.058)	0.298	(0.065)	0.292	(0.070)
av. p^{l} ($\times 10^{-2}$)	0.309	(0.002)	0.310	(0.002)	0.310	(0.002)
m.e. of f^{m}	0.114	(0.062)	0.098	(0.092)	0.111	(0.084)
m.e. of f^{l}	0.080	(0.018)	0.062	(0.020)	0.090	(0.022)
WAIC2	323.2		319.8		320.5	
	Panel D: Loss given default					
μ_0	0.072	(0.001)	0.072	(0.001)	0.072	(0.001)
μ_1	0.829	(0.002)	0.829	(0.002)	0.829	(0.003)
σ	0.131	(0.001)	0.131	(0.001)	0.131	(0.001)
α^{d}	−1.643	(0.159)	−1.697	(0.302)	−1.560	(0.018)
β^{d}	0.328	(0.202)	0.031	(0.211)		
γ^{d}	−0.293	(0.075)	−0.153	(0.072)		
δ^{d}			0.512	(0.339)		
av. p^{d}	0.174	(0.003)	0.174	(0.003)	0.174	(0.003)
m.e. of f^{m}	0.045	(0.028)	0.004	(0.029)		
m.e. of f^{l}	−0.040	(0.010)	−0.021	(0.010)		
m.e. of f^{d}			0.070	(0.046)		
av. LGD	0.204	(0.002)	0.204	(0.002)	0.204	(0.002)
m.e. of f^{m}	0.034	(0.021)	0.003	(0.022)		
m.e. of f^{l}	−0.030	(0.008)	−0.016	(0.007)		
m.e. of f^{d}			0.053	(0.035)		
WAIC2	−25,978.3		−25,969.3		−25,958.4	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1. We report results for specifications with a macro and a loan factor, an additional default factor. and macro and loan factor that do not influence the LGD component. The specifications do not include loan or default characteristics. Panel A presents the elements of $\boldsymbol{\Phi}$ of the factor component. Panel B presents the macroeconomic component with the intercepts α and factor sensitivities β for the gross domestic product (GDP), industrial production (IP) and unemployment rate (UR). Panel C presents the loan status component. The probability of default has fixed effect α^{l} and factor sensitivities β^{l} for the macro factor, and γ^{l} for the loan factor. Panel D presents the LGD component. The LGD type can be either mild or severe. Conditional on the type, the LGD follows a normal distribution with mean μ_0 or μ_1, and volatility σ. The probability of severe loss has fixed effect α^{d} and factor sensitivities β^{d} for the macro factor, γ^{d} for the loan factor, and δ^{d} for the default factor. We also report time-series averages of the probability of default p^{l}, the probability of a severe loss p^{d} and the average LGD, the marginal effects that the factors have on these variables, and the Widely Applicable Information Criterion, version 2 (WAIC2) for each component. The number of observations, N, given by the sum of the macroeconomic, default rate, LGD observations is 22,208.

Figure 2.2: Latent factors

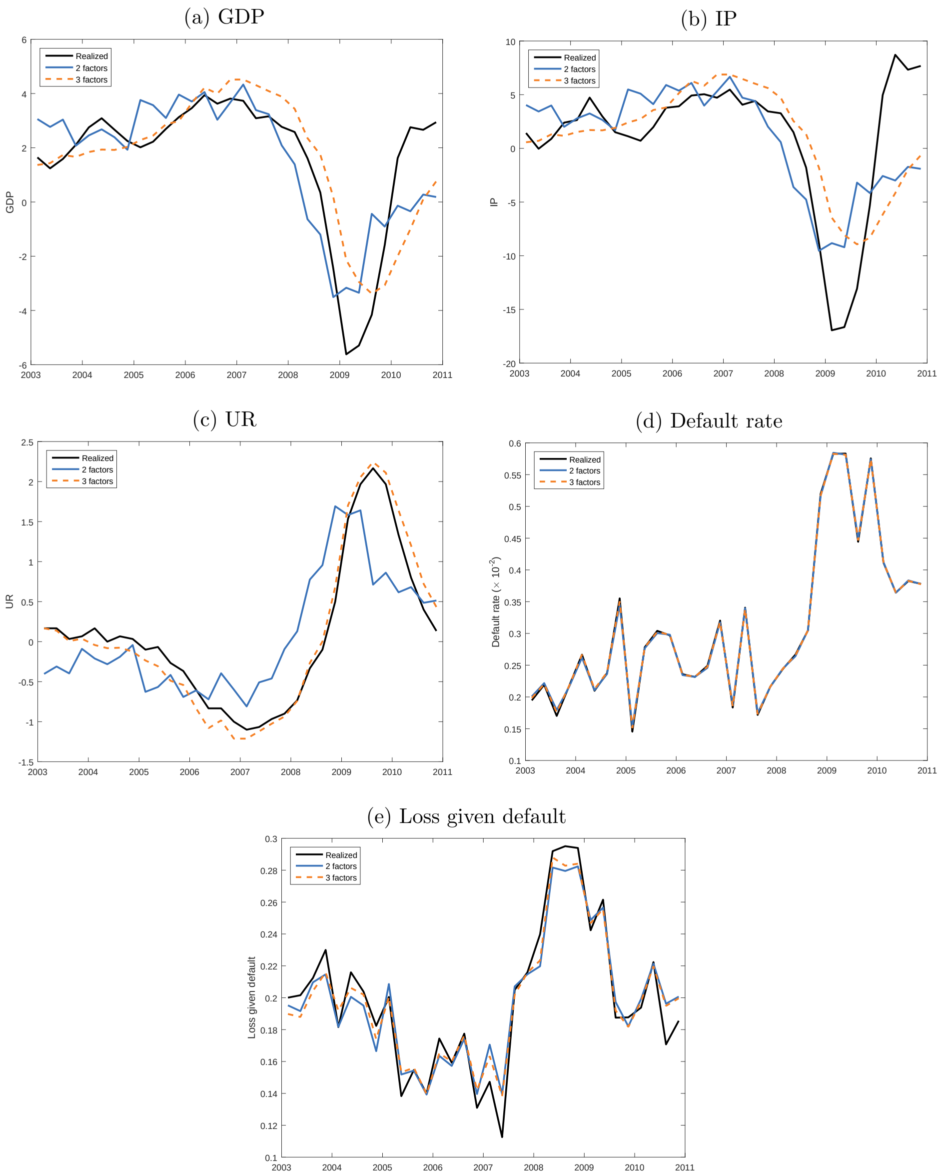
The figures show the smoothed latent factors for the model defined in Section 2.3.1. We report results for specifications with a macro and a loan factor (a), and an additional default factor (b). The specifications do not include loan or default characteristics.

the loan factor captures the more short-lived fluctuations. The deviations between the average model-implied and realized LGD are also relatively small, though larger than for the default rate. The combination of the factors captures the low average LGD in 2006-2007, and the subsequent pronounced upswing.

The addition of a third factor that can only influence the LGD decouples it from the other variables. The default factor is persistent, and strongly affects the LGD. The posterior mean of δ^{d} is 0.512, much larger than the mean of β^{d} of 0.031, which captures the effect of the macro factor. Compared to the two-factor specification, the influence of the macro factor decreases by a factor 10, and the effect of the loan factor is halved. Figure 2.2b shows that the default factor seems to lead the macro factor. A longer sample period may shed more light on this issue. Both the plot and the Widely Applicable Information Criterion, version 2 (WAIC2, Watanabe, 2010), which corrects for the number of parameters, indicate that the fit of the three-factor model is better for the macro variables, in particular for the unemployment rate. For the default rate changes are negligible. Though the fit for the average LGD in Figure 2.3e looks slightly better for the three-factor structure, the WAIC2 actually deteriorates.[11] Because we are mostly interested in capturing the LGD part in relation to the other variables, we do not favor the three-factor specification.

[11] Because of differences in the scales of the variables, the sum of the WAIC2 values cannot be used to evaluate the overall fit.

Figure 2.3: Time series fit



The panels show the time series fit of the model without cross-sectional differences for the growth rate of GDP (a), the growth rate of industrial production (b), the year-on-year change in the unemployment rate (c), the default rate (d) and the cross-sectional average of the LGD (e).

We investigate the importance of time-variation in the probability of a severe loss by estimating a two-factor specification with the restriction $\beta^{\mathrm{d}} = \gamma^{\mathrm{d}} = 0$. The WAIC2 value for the unrestricted specification is lower, indicating that the improvement in the fit outweighs the additional two parameters.

We conclude that the combination of a macro and a loan factor accurately captures the dynamics in the macro variables, probability of default and LGD. The macro factor captures the business cycle with relatively long swings, whereas the loan factor captures more short-lived fluctuations. During recessions, both the probability of default and the LGD increase, so the macro factor leads to positive dependence between default rates and LGD. However, the loan factor negatively affects their dependence, because it captures the defaults related to mere delays in loan payments.

2.4.2 The effects of loan and default characteristics

Loan and default characteristics affect the probability of default and the LGD of a loan. Their influence can take the form of a fixed effect, may influence the sensitivity to the latent factors, or give rise to a completely new latent factor. For example, Shleifer and Vishny (1992) argue that credit cycles are industry specific.

We use the richness of our data set to investigate the effect of seniority, and the size and industry of the borrower. Because this information in the GCD databases is categorical, we include it in our model by dummy variables. We require a minimum of 3,200 observations (100 per quarter) for a group to include the corresponding dummy variable.

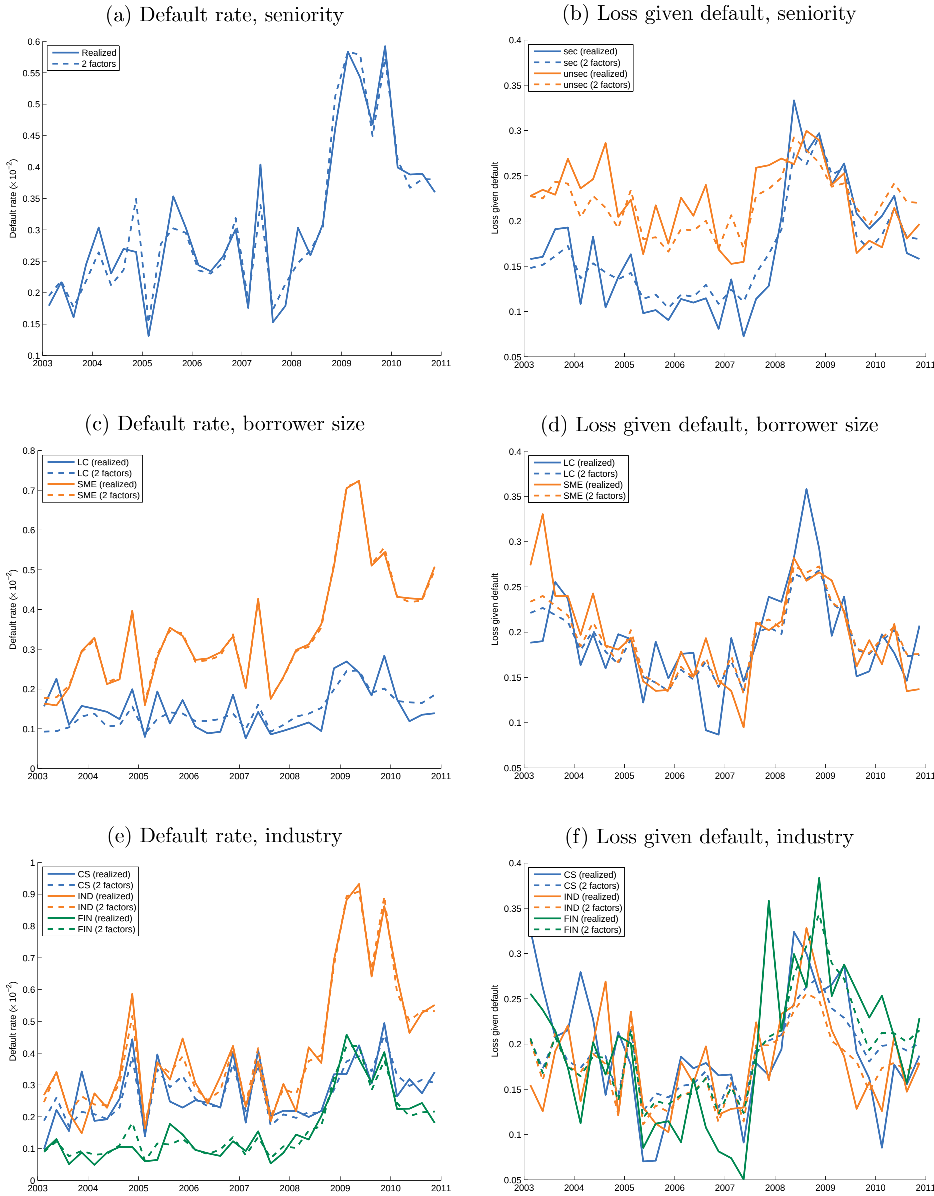
For each characteristic we estimate a two-factor model and extensions with additional loan and default factors. To save space we present the results for the two-factor models here. We show the full results in Section 2.D.

Effects of seniority

In our analysis of seniority, we distinguish defaults of senior secured and senior unsecured loans (see the number of observations in Table 2.1). Because seniority is only available for the LGD observations, it only influences the LGD component of our model.

The left panels of Table 2.3 show considerable differences between the LGD of senior secured and senior unsecured loans. First, unsecured loans suffer on average a larger

Figure 2.4: Time series fit, loan characteristics



The panels show the time series fit of the two-factor model with different loan and default characteristics for the default rate and the cross-sectional average of the LGD.

values for a mild LGD do not differ much. Together, these differences translate to an average LGD of 22.2% for an unsecured loan and 17.5% for a secured loan.

Table 2.3: Parameter estimates, model with loan and default characteristics

Panel A: Factor

	Seniority		Borrower size		Industry	
ϕ_{11}	0.813	(0.120)	0.845	(0.098)	0.910	(0.060)
ϕ_{22}	0.775	(0.138)	0.510	(0.217)	0.221	(0.269)

Panel B: Macro variables

	Seniority		Borrower size		Industry	
α_{GDP}	0.241	(0.631)	-0.433	(0.628)	0.526	(0.625)
β_{GDP}	-0.935	(0.694)	-1.225	(0.885)	−1.299	(0.880)
α_{IP}	0.187	(0.518)	-0.362	(0.543)	0.430	(0.532)
β_{IP}	-0.725	(0.583)	-1.025	(0.756)	−1.062	(0.753)
α_{UR}	-0.207	(0.556)	0.385	(0.572)	−0.490	(0.589)
β_{UR}	0.792	(0.612)	1.087	(0.814)	1.206	(0.832)

Panel C: Loan status

	Seniority		Large corp.		SME		Cons. staples		Industrials		Financials	
α^{l}	-5.665	(0.233)	−6.441	(0.174)	−5.533	(0.237)	−6.118	(0.173)	−5.929	(0.302)	−6.990	(0.377)
β^{l}	-0.179	(0.199)	0.404	(0.273)	0.559	(0.383)	0.203	(0.293)	0.602	(0.499)	0.787	(0.600)
γ^{l}	0.628	(0.354)	0.231	(0.101)	0.366	(0.143)	0.362	(0.103)	0.364	(0.106)	0.342	(0.108)
av. p^{l} ($\times 0.01$)	0.309	(0.002)	0.143	(0.003)	0.347	(0.002)	0.273	(0.005)	0.414	(0.005)	0.162	(0.004)
m.e. of f^{m}	-0.055	(0.061)	0.058	(0.039)	0.193	(0.132)	0.055	(0.080)	0.248	(0.205)	0.128	(0.097)
m.e. of f^{l}	0.193	(0.109)	0.033	(0.014)	0.126	(0.049)	0.099	(0.028)	0.150	(0.044)	0.055	(0.018)

Panel D: Loss given default

	Sen. Secured		Sen. Unsecured		Large corp.		SME		Cons. staples		Industrials		Financials	
μ_0	0.070	(0.002)	0.072	(0.001)	0.075	(0.002)	0.062	(0.001)	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)
μ_1	0.766	(0.004)	0.858	(0.003)	0.848	(0.005)	0.849	(0.003)	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)
σ	0.129	(0.001)	0.130	(0.001)	0.126	(0.001)	0.124	(0.001)	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)
α^{d}	−2.174	(0.580)	−1.665	(0.250)	−1.707	(0.291)	−1.564	(0.269)	−1.735	(0.265)	−1.786	(0.280)	−2.025	(0.465)
β^{d}	0.900	(0.619)	0.500	(0.320)	0.293	(0.302)	0.243	(0.272)	0.401	(0.314)	0.327	(0.335)	0.793	(0.549)
γ^{d}	0.041	(0.055)	−0.216	(0.133)	−0.404	(0.173)	−0.406	(0.167)	−0.327	(0.123)	−0.472	(0.147)	−0.492	(0.171)
av. p^{d}	0.151	(0.004)	0.191	(0.004)	0.160	(0.005)	0.166	(0.003)	0.163	(0.007)	0.156	(0.004)	0.185	(0.006)
m.e. of f^{m}	0.103	(0.070)	0.075	(0.048)	0.036	(0.037)	0.033	(0.037)	0.055	(0.043)	0.041	(0.042)	0.099	(0.068)
m.e. of f^{l}	0.005	(0.006)	−0.033	(0.020)	−0.050	(0.022)	−0.056	(0.023)	−0.045	(0.017)	−0.060	(0.019)	−0.061	(0.022)
av. LGD	0.175	(0.003)	0.222	(0.003)	0.199	(0.004)	0.193	(0.003)	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)
m.e. of f^{m}	0.072	(0.049)	0.059	(0.038)	0.028	(0.029)	0.026	(0.030)	0.043	(0.034)	0.032	(0.033)	0.070	(0.049)
m.e. of f^{l}	0.003	(0.004)	−0.026	(0.016)	−0.039	(0.017)	−0.044	(0.018)	−0.035	(0.014)	−0.047	(0.015)	−0.044	(0.015)

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1 for different loan and default characteristics. All model specifications have a macro and a loan factor. The horizontal panels correspond with the panels in Table 2.2. We distinguish defaults of senior secured and senior unsecured loans (left panels, labeled “Seniority”); loans and defaults of large corporates and SMEs (middle panels, labeled ”borrower size”); loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials (right panels, labeled “Industry”). The characteristics are included as dummy variables. Number of observations are 21,862, 18,684 and 14,997.

Second, secured loans are more sensitive to the business cycle. Its sensitivity is about 2 times as large, and consequently the marginal effect on the average LGD is 7.2% (5.9%) for the secured (unsecured) loans. The LGD of secured loans does not respond strongly to the loan factor. The posterior distribution of γ^{d} is wide and close to zero. The corresponding sensitivity of unsecured defaults is much larger and clearly negative. Apparently, delays in loan payments are concentrated in unsecured loans.

The behavior of the loan factor and the probability of default have also changed in comparison with the results without characteristics. The loan factor has become more persistent, and the default rate has lost is sensitivity to the macro factor. The posterior mean of β^{l} is negative, but its distribution is very wide. The plot in Figure 2.D.1a shows that the loan factor resembles the macro factor much more than in Figure 2.2a, except for the crisis period where it lags the macro factor, and looks again more like the loan factor of the model without characteristics. Because we do not know the seniority of each loan, we cannot further investigate this issue.

The fit over time of the default rate (Figure 2.4a) is similar to the model without characteristics in Figure 2.3d. Figure 2.4a clearly shows the differences in the LGD behavior. The average LGD is generally higher for unsecured loans than for secured loans, except during the credit crisis. These results are in line with Bruche and González-Aguado (2010) and extend Hamerle et al. (2011). The sensitivity to the macro factor explains the pronounced increase around 2008 in the LGD of secured loans. Its effect on the LGD of unsecured loans is partly offset by the loan factor. During the credit crisis, the LGD of both loan types increases, but less for unsecured loans. The dependence of the value of collateral on the business cycle may explain part of this effect.

Our results for a three factor specification, which has an additional default factor that only affects the LGD of unsecured loans do no support a different cycle for unsecured loans. While the fit of the average unsecured LGD improves, the WAIC2 actually increases. We conclude that a two-factor structure, with a strong effect of the macro factor on the secured LGD, and of the loan factor on the unsecured LGD is the best model.

Effects of borrower size

In our analysis of the effect of borrower size, we distinguish SMEs and large corporates, both for the loans and the defaults in our sample. Comparing the middle panels of Table 2.3 to the two-factor results in Table 2.2 shows that borrower size mostly affects the probability of default. It is much larger when lending to SMEs (0.35%) than to large corporates (0.14%). The sensitivities to both the macro and the loan factor are also higher for SMEs. This result is in line with the higher riskiness, both systematic and idiosyncratic that is documented for the equities of small firms. The distribution of the LGD does not vary in relation to borrower type. The differences in the coefficients and the implied statistics in the middle of panel D are small compared to their posterior standard deviations.

The macro and loan factors in Figure 2.D.2a largely resemble those in Figure 2.2a. The default rates implied by the two-factor structure track the realized rates closely for SMEs and a bit less for large corporates (Figure 2.4c). However, in particular in the first part of our sample period, there is room for improvement. Because of the large number of loans to SMEs (12,000 vs. 6,500 large corporates), the loan factor reflects the SME default rate more. The factors fit the dynamics of the average LGD for SME also better than for large corporates (Figure 2.4d).

In Section 2.D, we report the three-factor specification that has an additional SME loan factor. The comparison by WAIC2 values favors this specification, though the improvement over the two-factor model is small. The sensitivities of the SME probability of default and LGD to the general loan factor become less at the expense of the sensitivities to the new SME loan factor. The additional factor has most influence at the beginning and end of our sample period, but also points to different behavior during the credit crisis. We interpret these results as some evidence for separate default-specific factors that depend on size.

Effects of industry

The industry in which the borrower is active substantially influences the probability of default and the LGD. Due to the number of observations, we can distinguish the industries Consumer Staples (CS), Industrials (IND) and Financials (FIN). The results for this analysis in the right panels of Table 2.3 indicate that these differences pertain to both the components for the default probabilities and for the LGD.

For the probability of default, both the average and the factor sensitivities vary over the industries. Loans to borrowers in FIN (IND) have the lowest (highest) average default probability of 0.16% (0.41%). Sensitivities to the business cycle is highest for FIN, followed by IND. The average default probability for CS is in the middle (0.27%), but least sensitive to the business cycle. The sensitivities to the loan factor are about equal.

The LGDs differ mainly in their sensitivity to the business cycle. It is highest for FIN, followed by CS and lowest for IND. The differences in marginal effects on the average LGD, which equal 7.0%, 4.3% and 3.2%, are substantial. The differences in the other parameters are less consequential, and the average LGD is about the same for the different industries.

The factor estimates are again not much different than in our baseline specification. The ability of the model to fit the default rates remains remarkably good, as deviations of the model-implied series from the realisations are small (Figure 2.4e). For the LGD series, the deviations are a bit larger.

Our analysis of a four-factor specification with two additional loan factors does not provide evidence in favor of separate credit cycles for each industry, as proposed by Shleifer and Vishny (1992). The three loan factors are difficult to distinguish from white noise, and the factor loadings do not indicate that each industry has its own factor. The four-factor specification improves the fit, though perhaps less than expected.

We conclude that industry characteristics have an important effect on default rates and LGD. Loans to industrials are generally most risky, as their probability of default is highest, but loans to financials vary most related to business cycle fluctuations. Loans to CS firms are more in the middle.

2.4.3 Implications for risk management

We illustrate the implications of our model for risk management by the calculation of the expected loss and the economic capital, which measures the risk of unexpected losses on a loan portfolio. We calculate it as the difference between 99.9% quantile and the mean of the loss distribution. We show how they change during our sample period for a portfolio that corresponds with our data set and based on the latent factors we have inferred. Though other papers[12] have already shown the importance of the positive dependence between default rates and LGD for economic capital in an unconditional setting, our analysis gives additional insights in the dynamics of the loss distribution.

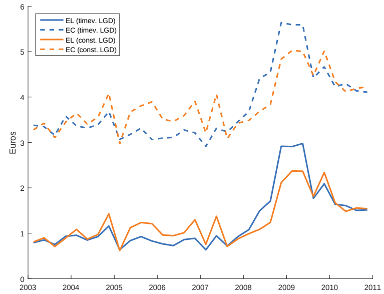
At each point in time we consider a portfolio of 2,000 loans, each with an exposure at default of 1 euro, as in Miu and Ozdemir (2006). After every full iteration of the Gibbs sampler, we simulate the loss on the portfolios, conditional on the values drawn for the parameters and factors in that iteration. This yields the posterior loss distribution per time period, from which we get the quarterly expected loss and economic capital.

Figure 2.5 shows the cyclical variation in the expected loss distribution and the economic capital, both for our base two-factor model and the restricted version with constant LGD as presented in of Section 2.4.1. We clearly see that both the expected loss and the economic capital fluctuate stronger when the LGD is also time-varying. When the LGD cannot vary over time, the expected loss varies from 0.61 (0.03%) to 2.37 (0.12%), but when the LGD can vary, the maximum is at 2.98 (0.15%), a substantial increase. In good times, time variation in the LGD leads to a slightly lower expected loss, but in bad times to a pronounced increase. These effects carry over to the economic capital, which is clearly lower from 2005 to 2006, and rapidly increases during 2008. During 2008, economic capital based on time-varying LGD is 5.64 (0.28%) compared to 5.03 (0.25%) with constant LGD. While these number may seem small, it means that 0.60 of the cyclical increase of 5.64 - 2.91 = 2.73 comes from time-variation in the LGD, so 22%. The increase of 0.60 is similar to the results in Bruche and González-Aguado (2010, Table 3).

These results illustrate how our model can be used in a risk management setting. By using values for the latent factors at a specific point in time, or prespecified values, stress tests can be conducted. Based on the point-in-time estimates of portfolio losses, or by constructing an unconditional loss distribution, our model can be used to construct through-the-cyle economic capital (see Miu and Ozdemir, 2006).

[12] See Frye (2000); Pykhtin (2003); Gordy (2003); Düllmann and Gehde-Trapp (2004); Miu and Ozdemir (2006) among others.

Figure 2.5: Portfolio simulation results



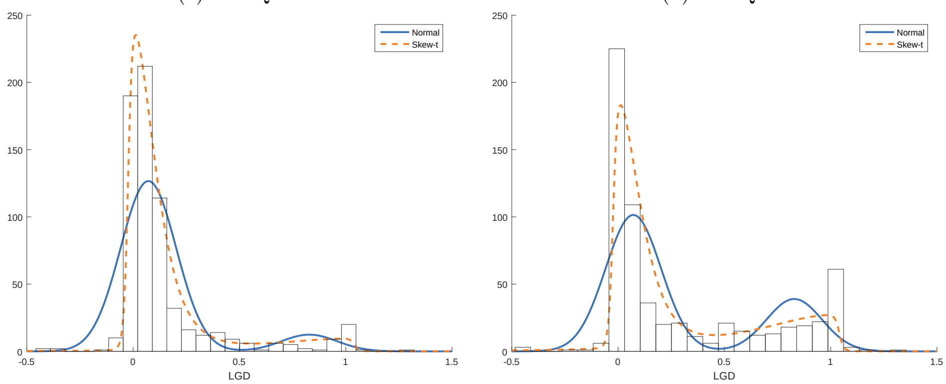
The figure presents the portfolio loss (EL) and economic capital (EC) for a portfolio of loans, each quarter consisting of 2,000 loans and each loan with an exposure of 1 euro. The EL and EC are based on the two-factor model without cross-sectional variation, with a time-varying LGD (solid line) and a constant LGD (dashed lines).

2.4.4 Alternative specifications

Figure 2.6 shows the fit of our model for the LGD distributions at two points in time, Q4 of 2005 and Q2 of 2008. The probability of a severe loss is much higher in 2008Q2, and accordingly, we see an increase in the right mode and a decrease in the left mode. While this effect captures some of the changes in the empirical distributions, other more flexible distributions may lead to a better fit. We therefore investigate a replacement of the normal distribution in Equation (2.8) by the skewed Student's t of Azzalini and Capitanio (2003). We show the full results in Section 2.E and discuss the main consequences here.

Figure 2.6: Mixture fit

(a) 2005Q4 (b) 2008Q2



Panel a shows the cross-sectional fit on the LGDs of the mixture of normals and mixture of skew-t for the fourth quarter of 2005, and panel b shows the fit for the second quarter of 2008. The fitted distributions are based on the posterior means of the parameters of the two-factor model without cross-sectional variation.

The dotted lines in Figure 2.6 show an improvement in the fit. The low degrees of freedom leads to more peakedness for the mild losses. Both distributions show considerable skewness, which helps to fit the observations in the middle. Though we observe an increase in the average probability of a severe loss, the overall factor structure, and the factor sensitivities change only marginally (see Figure 2.E.1).

Though the mixture of skewed Student's t distributions offers a better cross-sectional fit, it has an important theoretical disadvantage. The mixture of normal distributions with the same variance has the property that $\Pr[s_{it} = 0 | y^{\mathrm{d}}_{it}, \boldsymbol{f}_t]$ is a decreasing function of y^{d}_{it}. So, if the LGD grows larger, the probability with which it is inferred as mild decreases. The (skewed) Student's t distribution does not have this property because of its fat tails, even when the degrees of freedom of the mixture components are equal.[13] For this reason, we do not replace the normal distributions in our specifications. Because the consequences for the other model components are small, we conclude that our results are robust to this choice.

Our specification does not allow for lead-lag relations between the macro variables and the loan and default variables. While we leave a full investigation of models with lead-lag effects, as well as richer VAR dynamics for future research, we briefly investigate the potential of such extensions by simply leading and lagging the macro series. We find that using past values of the macro variables does not lead to a better fit of the default rates and LGD series. Using future values slightly increases the fit of the LGD series but not of the default rate. This is likely related to the workout period. Though interesting, a model that needs future information is of course less useful in practice.

2.5 Conclusion

The loss given default and the default rate on bank loans are both cyclical. We show that this variation stems from a macro factor capturing the business cycle and a default-specific factor that captures variations in the credit cycle on top of the business cycle. The time variation in the LGD is explained by changes in the probability of a

[13] We illustrate this effect in Figure 2.E.2.

severe versus a mild loss. While different from bonds, bank loans are also sensitive to the business cycle with higher default rates and LGD during downturns.

Our model describes the stylized facts of the LGD on bank loans well. It captures the bimodal shape of the empirical distribution and provides an interpretation of the components, by explicitly modeling the extremes of no and full loss. It is flexible enough to include the differences across loan characteristics. Further, the model has applications in risk management, such as the calculation of economic capital.

2.A Macroeconomic variables

Table 2.A.1: Macroeconomic variables description

Abbreviation	Subject	Measure	Country	Transformation
GDP	B1_GE: Gross domestic product - expenditure approach	GYSA: Growth rate compared to the same quarter of previous year, seasonally adjusted	OECD - Europe	-
IP	Industrial production, s.a.	Growth on the same period of the previous year	OECD - Europe	-
UR	Harmonised unemployment rate (monthly), Total, All persons	Level, rate or quantity series, s.a.	European Union (28 countries)	Difference with same quarter of previous year

The table presents the macroeconomic variables as defined in the OECD database (see `http://stats.oecd.org/`) and possible transformations.

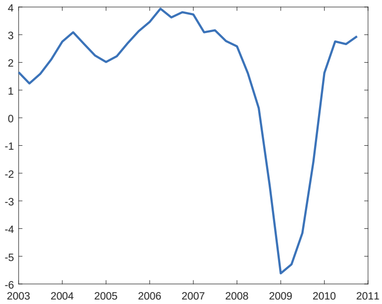
Table 2.A.2: Macroeconomic variables statistics

	GDP	IP	UR
Mean	1.575	0.968	0.070
Median	2.621	2.806	0.017
Maximum	3.940	8.715	2.167
Minimum	-5.617	-16.945	-1.100
Standard deviation	2.567	6.412	0.903
Skewness	-1.755	-1.692	0.877
Kurtosis	4.981	5.086	3.084
AR(1)	0.894	0.861	0.938
AR(2)	0.646	0.535	0.786
AR(3)	0.343	0.135	0.581
AR(4)	0.070	-0.212	0.354

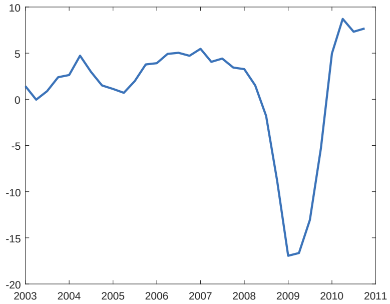
The table presents descriptive statistics for the macroeconomic variables gross domestic product, industrial production, and unemployment rate, in differences with the same quarter of the previous year, as defined in Table 2.A.1. AR(x) is the x-th order autocorrelation.

Figure 2.A.1: Macroeconomic variables time series

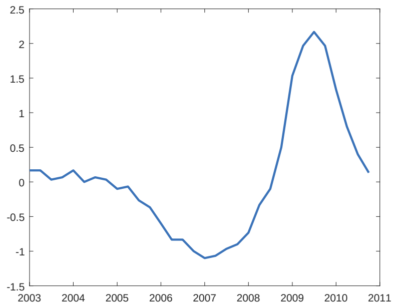
(a) GDP



(b) IP



(c) UR



The figures present the time series of the macroeconomic variables gross domestic product (a), industrial production (b) and unemployment rate (c), in differences with the same quarter of the previous year, as defined in Table 2.A.1.

2.B Loan and default data

2.B.1 GCD databases

2.B.2 Data Filter

Following Höcht and Zagst (2007), who also use data from the Global Credit Data Consortium, and NIBC's internal policy, we apply the following filters to the LGD database.

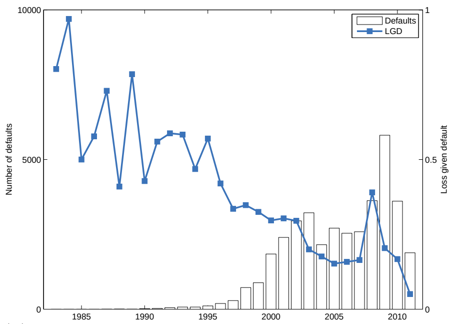
- *EAD* $\geq$ *€100,000.* The paper focuses on loans where there has been an actual (possible) loss, so EAD should be at least larger than 0. Furthermore, there are some extreme LGD values in the database for small EAD. To account for this noise, loans with EAD smaller than € 100,000 are excluded.

Table 2.B.1: Number of banks contributing to the databases

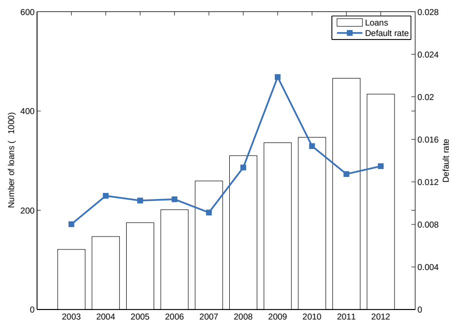
Year	LGD	Loan
2000	33	NA
2001	38	NA
2002	41	NA
2003	43	7
2004	39	9
2005	41	10
2006	45	10
2007	47	11
2008	46	14
2009	46	16
2010	43	17
2011	40	17
2012	42	16
2013	39	NA
2014	37	NA

This table shows how many banks contribute loans and defaults to the LGD and the loan databases for a given year. The versions of the databases correspond with June 2014 (LGD) and June 2013 (Loan).

Figure 2.B.1: Evolution of LGD and default rate observations



(a) Number of defaults and average LGD over time



(b) Default rate

Panel a presents the average LGD and the number of observations per year for the period 1983–2011 from the Global Credit Data default database. Panel b presents the number of loans and the observed default rate per year for the period 2003–2012 from the Global Credit Data loan database.

- $-10\% < \big((CF+CO)-(EAD-EAR)\big)/(EAD+PA) < 10\%$, *where CF cash flows, CO charge-offs and PA principal advances.* The cash flows that make up the LGD should be plausible, because they are the major building blocks of the LGD. A way of checking this is by looking at under-/overpayments. The difference between the EAD and the exposure at resolution (EAR), where resolution is the moment where the default is resolved, should be close to the sum of the cash flows and charge-offs. The cash flow is the money coming in and the charge-off is the acknowledgement of a loss in the balance sheet, because the exposure is expected not to be repaid. Both reduce the exposure and should explain the difference between EAD and EAR. There might be an under- or overpayment, resulting in a difference. To exclude implausible cash flows, these loans are excluded when they are more than or equal to 10% of the EAD and principal advances (PA). The 10% is a choice of the Global Credit Data Consortium.

- $-0.5 \leq LGD \leq 1.5$. Although theoretically, LGD is expected between 0 and 1, it is possible to have an LGD outside this range, e.g. due to principal advances or a profit on the sale of assets. Abnormally high or low values are excluded. They are implausible and influence LGD statistics too much.

- *No government guarantees.* The database contains loans with special guarantees from the government. Most of the loans are subordinated, but due to the guarantee, the average of the subordinated LGD is lower than expected. Because the loans are very different from others with the same seniority and to prevent underestimation of the subordinated LGD, these loans are excluded from the dataset.

Some consortium members also filter for high principle advances ratios, which is the sum of the principal advances divided by the EAD. Even though high ratios are plausible, they are considered to influence the data too much and therefore exclude loans with ratios larger than 100%. NIBC does include these loans, because they are supposed to contain valuable information and the influence of outliers is mitigated because they cap their LGD to 1.5. The data shows that the principal advances ratio does not exceed 100%, so applying the filter does not affect the data and is therefore not considered.

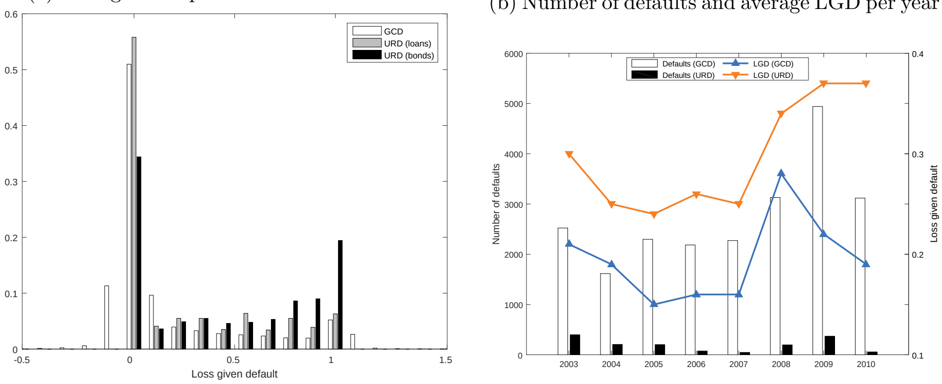
2.B.3 Comparison with Moody's Ultimate Resolved Database

Because we do not have direct access to Moody's URD, we use its discussion in Altman and Kalotay (2014) and Bastos (2014) to construct a comparison. Moody's URD contains information about some 5,200 resolved defaults of bonds (around 60%) and bank loans (around 40%), and about 1,000 borrowers. Figure 2.B.2a shows that the LGD distribution of bonds is still bimodal, but has more probability mass at large losses. The LGD distributions of bank loans of both data sets are quite similar, even though URD focuses on "US non-financial corporations holding over $50 million in debt at the time of default" (Bastos, 2014). Consequently, the average LGD of bonds is much higher (55.1%) than of loans (19.5% based on URD and 20.1% in our data set). Figure 2.B.2b shows that the average LGD of bonds exceeds that of loans. The behavior of both series is similar before the credit crisis, but differs after it. It also shows that our LGD data set is much larger than the URD. These results confirm that the LGD of bank loans differ substantially from bonds.

Figure 2.B.2: Comparison of LGD in our data set with Moody's URD

(a) Histogram of pooled LGD observations

(b) Number of defaults and average LGD per year



Panel a shows the pooled distribution of the LGD, and panel b shows the number of defaults and the average LGD per year, both for our GCD data set and Moody's Ultimate Recovery Database. We use all LGD observations in the GCD data set over the period 2003–2010 after applying the filters in Section 2.B.2. The URD distributions in panel a are based on Altman and Kalotay (2014), and the evolution in panel b on Bastos (2014).

2.C Bayesian estimation procedure

This section provides a description of the Bayesian estimation of the model in Section 2.3. The likelihood, priors and posteriors are derived for the model without loan and default characteristics. We explain in Section 2.C.4 how to extend this to include loan and default characteristics, and how to replace the LGD component by a mixture of skew-t distributions.

2.C.1 Likelihood and latent variables

The likelihood consists of a macro, loan status and LGD component. First, we consider the macro component of the likelihood. Define $\boldsymbol{Y}^{\mathrm{m}}$ as the $T \times N^{\mathrm{m}}$ matrix with observations on $\boldsymbol{y}_t^{\mathrm{m}}$ and define $\boldsymbol{X}^{\mathrm{m}}$ as the $T \times (K^{\mathrm{m}}+1)$ matrix with a constant and observations on the macro factors $\boldsymbol{f}_t^{\mathrm{m}}$. We can write the likelihood as

$$p(\boldsymbol{Y}^{\mathrm{m}}|\boldsymbol{A}^{\mathrm{m}}, \boldsymbol{\Sigma}^{\mathrm{m}}, \boldsymbol{f}_t) \propto |\boldsymbol{\Sigma}^{\mathrm{m}}|^{-T/2} \exp\left(-\frac{1}{2}\,\mathrm{tr}((\boldsymbol{\Sigma}^{\mathrm{m}})^{-1}(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})'(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})\right).$$

Second, we derive the loan status component of the likelihood. Define $\boldsymbol{y}^{\mathrm{l}}$ as the vector with all loan indicators y_{it}^{l}, the vector $\boldsymbol{\psi}^{\mathrm{l}}$ with elements $\psi_t^{\mathrm{l}} = \alpha^{\mathrm{l}} + \beta^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{m}} + \gamma^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{l}}$, D_t as the number of defaulted loans in period t and L_t as the number of total loans in period t. We can write the likelihood as

$$\begin{aligned} p(\boldsymbol{y}^{\mathrm{l}}|\boldsymbol{\psi}^{\mathrm{l}}) &= \prod_{i,t}^{N,T} (p_t^{\mathrm{l}})^{y_{it}^{\mathrm{l}}}(1-p_t^{\mathrm{l}})^{1-y_{it}^{\mathrm{l}}} \\ &= \prod_{t}^{T} \Lambda(\psi_t^{\mathrm{l}})^{D_t}\big(1-\Lambda(\psi_t^{\mathrm{l}})\big)^{L_t-D_t} \\ &= \prod_{t}^{T} \frac{\exp(\psi_t^{\mathrm{l}})^{D_t}}{\big(1+\exp(\psi_t^{\mathrm{l}})\big)^{L_t}}. \end{aligned}$$

Third, we analyze the LGD component given the severe loss indicator s_{it}. Define $\boldsymbol{y}^{\mathrm{d}}$ as the vector with all realized LGDs and define $\boldsymbol{s}$ as the vector which contains all s_{it} for all i, t. We can write the likelihood as

$$p(\boldsymbol{y}^{\mathrm{d}}|\boldsymbol{s}, \mu_0, \mu_1, \sigma^2) \propto \prod_{i,t}^{N,T} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right).$$

We use three types of latent variables in our model next to the factors. First, we use the severe loss indicators s_{it} for all i, t. Define the vector $\boldsymbol{\psi}^{\mathrm{d}}$ with elements $\psi_t^{\mathrm{d}} = \alpha^{\mathrm{d}} + \beta^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{m}} + \gamma^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{l}} + \delta^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{d}}$, T_t as the total number of LGD observations in period

t and N_t as the total number of severe losses in period t.

$$\begin{aligned} p(\boldsymbol{s}|\boldsymbol{\psi}^{\mathrm{d}}) &= \prod_{it}^{NT} (p_t^{\mathrm{d}})^{s_{it}} (1-p_t^{\mathrm{d}})^{(1-s_{it})} \\ &= \prod_{t}^{T} \Lambda(\psi_t^{\mathrm{d}})^{N_t} \big(1-\Lambda(\psi_t^{\mathrm{d}})\big)^{(T_t-N_t)} \\ &= \prod_{t}^{T} \frac{\exp(\psi_t^{\mathrm{d}})^{N_t}}{\big(1+\exp(\psi_t^{\mathrm{d}})\big)^{T_t}}. \end{aligned}$$

Second, we follow Polson et al. (2013) and use the auxiliary latent variables ω_t^{l} and ω_t^{d} to make the sampling of respectively the loan status and LGD components easier

$$p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) = \mathrm{PG}(L_t,\psi_t^{\mathrm{l}}),$$
$$p(\omega_t^{\mathrm{d}}|T_t,\psi_t^{\mathrm{d}}) = \mathrm{PG}(T_t,\psi_t^{\mathrm{d}}),$$

where the definition of the Pólya-gamma distribution is given in equation (1) of Polson et al. (2013).

Finally, we derive some useful results that help us with deriving the posterior distribution. Windle et al. (2013) show that the Pólya-gamma distribution has the special form

$$p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) = \cosh^{L_t}(\psi_t^{\mathrm{l}}/2)\exp(-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2)p(\omega_t),$$

and that the following holds

$$\cosh^{L_t}(\psi_t^{\mathrm{l}}/2)/\big(1+\exp(\psi_t^{\mathrm{l}})\big)^{L_t} \propto \exp(-\psi_t^{\mathrm{l}}L_t/2).$$

This implies the following result that we use in the next sections

$$
\begin{aligned}
p(\boldsymbol{y}^{\mathrm{l}}|\boldsymbol{\psi}^{\mathrm{l}})\prod_{t}^{T}p(\omega_t^{\mathrm{l}}|L_t,\psi_t^{\mathrm{l}}) &= \prod_{t}^{T}\frac{\exp(\psi_t^{\mathrm{l}})^{D_t}}{\big(1+\exp(\psi_t^{\mathrm{l}})\big)^{L_t}}\cosh^{L_t}(\psi_t^{\mathrm{l}}/2)\exp(-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\big(D_t\psi_t^{\mathrm{l}}-\psi_t^{\mathrm{l}}L_t/2-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2\big)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\big(\kappa_t^{\mathrm{l}}\psi_t^{\mathrm{l}}-\omega_t^{\mathrm{l}}(\psi_t^{\mathrm{l}})^2/2\big)p(\omega_t^{\mathrm{l}}) \\
&\propto \prod_{t}^{T}\exp\left(-\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}-\psi_t^{l}\right)^2\right)\exp\left(\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}\right)^2\right)p(\omega_t^{\mathrm{l}}),
\end{aligned}
$$

where $\kappa_t^{\mathrm{l}} = D_t - L_t/2$.

Similarly, the following holds

$$
p(\boldsymbol{y}^{\mathrm{d}}|\boldsymbol{\psi}^{\mathrm{d}})\prod_{t}^{T}p(\omega_t^{\mathrm{d}}|T_t,\psi_t^{\mathrm{d}}) \propto \prod_{t}^{T}\exp\left(-\frac{\omega_t^{\mathrm{d}}}{2}\left(\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}}-\psi_t^{d}\right)^2\right)\exp\left(\frac{\omega_t^{\mathrm{d}}}{2}\left(\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}}\right)^2\right)p(\omega_t^{\mathrm{d}}),
$$

where $\kappa_t^{\mathrm{d}} = N_t - T_t/2$.

Refer to page 8-9 of Windle et al. (2013) for more details.

2.C.2 Prior

First, we consider the macro parameters

$$
p(\boldsymbol{A}^{\mathrm{m}},\boldsymbol{\Sigma}^{\mathrm{m}}) \propto \mathrm{iW}(0.01\boldsymbol{I}_{N^{\mathrm{m}}},N^{\mathrm{m}})I(\boldsymbol{A}^{\mathrm{m}}),
$$

where $\boldsymbol{I}_{N^{\mathrm{m}}}$ is an identity matrix of dimension N^{m}. Second, we consider the loadings on the factors for the loan status and LGD components

$$
p(\boldsymbol{\alpha}^{\mathrm{l}},\boldsymbol{\alpha}^{\mathrm{d}}) \propto I(\boldsymbol{\alpha}^{\mathrm{l}},\boldsymbol{\alpha}^{\mathrm{d}}).
$$

Third, we impose priors for the parameters of the LGD component

$$
p(\mu_0,\mu_1,\sigma^2) \propto \mathrm{iG2}(0.01,0.01)I(\mu_0,\mu_1).
$$

Finally, we use a prior for the persistence of the factor

$$p(\phi_{jj}) \propto I(\phi_{jj}) \text{ for all } j.$$

The indicator functions $I(\cdot)$ impose the identification restrictions mentioned in the main text. To be precise, they impose that the first macro variable loads negatively on the macro factor, that the probability of default loads positively on the second default factor and that the probability of a severe loss loads positively on the third loan factor. The functions also impose that $\mu_0 < \mu_1$ and $-1 < \phi_{jj} < 1$ for all j.

2.C.3 Posterior

Macro component

We collect the terms involving $\boldsymbol{\Sigma}^{\mathrm{m}}$ and $\boldsymbol{A}^{\mathrm{m}}$ from the likelihood and prior and get

$$\begin{aligned} p(\boldsymbol{A}^{\mathrm{m}}, \boldsymbol{\Sigma}^{\mathrm{m}}|\ldots) \propto{}& |\boldsymbol{\Sigma}^{\mathrm{m}}|^{-(T+N^{\mathrm{m}}+N^{\mathrm{m}}+1)/2} \\ &\times \exp\left(-\frac{1}{2}\operatorname{tr}\left((\boldsymbol{\Sigma}^{\mathrm{m}})^{-1}\left[0.01\boldsymbol{I}_{N^{\mathrm{m}}} + (\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})'(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})\right]\right)\right) I(\boldsymbol{A}^{\mathrm{m}}) \end{aligned}$$

Using standard results for multivariate regression models we see that $\boldsymbol{A}^{\mathrm{m}}$ can be drawn from a matricvariate normal distribution and $\boldsymbol{\Sigma}^{\mathrm{m}}$ from an inverse Wishart distribution

$$\begin{aligned} p(\boldsymbol{A}^{\mathrm{m}}|\ldots) &= \mathrm{MN}\left((\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{X}^{\mathrm{m}})^{-1}(\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{Y}^{\mathrm{m}}), \boldsymbol{\Sigma}^{\mathrm{m}} \otimes (\boldsymbol{X}^{\mathrm{m}\prime}\boldsymbol{X}^{\mathrm{m}})^{-1}\right), \\ p(\boldsymbol{\Sigma}^{\mathrm{m}}|\ldots) &= \mathrm{iW}\left(0.01\boldsymbol{I}_{N^{\mathrm{m}}} + (\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}})'(\boldsymbol{Y}^{\mathrm{m}} - \boldsymbol{X}^{\mathrm{m}}\boldsymbol{A}^{\mathrm{m}}), T + N^{\mathrm{m}}\right), \end{aligned}$$

where we redraw until $\boldsymbol{A}^{\mathrm{m}}$ satisfies the identification restrictions.

Loan status component

Since ω_t^{l} only occurs in its own distribution (and not in the priors or likelihoods), we sample ω_t^{l} from its Pólya-gamma distribution

$$p(\omega_t^{\mathrm{l}}|\ldots) = \mathrm{PG}(L_t, \psi_t^{\mathrm{l}}).$$

We collect the terms involving $\boldsymbol{\alpha}^l$ from the likelihood, prior and latent variable distributions

$$p(\boldsymbol{\alpha}^{\mathrm{l}}|\ldots) \quad \propto \quad \prod_{t}^{T} \exp\left(-\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}} - \psi_t^l\right)^2\right) \exp\left(\frac{\omega_t^{\mathrm{l}}}{2}\left(\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}\right)^2\right) p(\omega_t^{\mathrm{l}}) I(\boldsymbol{\alpha}^{\mathrm{l}}),$$

where ψ_t^{l} is a function of $\boldsymbol{\alpha}^{\mathrm{l}}$.

We see that we can interpret the single term in the product as the likelihood of a pseudo data point $\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}}$ drawn from a normal distribution with mean ψ_t^{l} and variance $1/\omega_t^{\mathrm{l}}$ as in Windle et al. (2013). Following Polson et al. (2013) and using the standard results for a linear regression with heteroscedasticity, we simulate $\boldsymbol{\alpha}^{\mathrm{l}}$ from a normal distribution

$$\boldsymbol{\alpha}^{\mathrm{l}} \sim \mathrm{N}(\boldsymbol{m}^{\mathrm{l}}, \boldsymbol{V}^{\mathrm{l}}),$$

where

$$\begin{aligned}
\boldsymbol{V}^{\mathrm{l}} &= (\boldsymbol{X}^{\mathrm{l}\prime} \boldsymbol{\Omega}^{\mathrm{l}} \boldsymbol{X}^{\mathrm{l}})^{-1}, \\
\boldsymbol{m}^{\mathrm{l}} &= (\boldsymbol{X}^{\mathrm{l}\prime} \boldsymbol{\Omega}^{\mathrm{l}} \boldsymbol{X}^{\mathrm{l}})^{-1} (\boldsymbol{X}^{\mathrm{l}\prime} \boldsymbol{\kappa}^{\mathrm{l}}),
\end{aligned}$$

where $\boldsymbol{X}^{\mathrm{l}}$ is a matrix that contains the constant and the relevant factors, where $\boldsymbol{\kappa}^{\mathrm{l}}$ is a vector that collects the elements κ_t^{l} and where $\boldsymbol{\Omega}^{\mathrm{l}}$ is a diagonal matrix with ω_t^{l} as diagonal elements. We redraw $\boldsymbol{\alpha}^{\mathrm{l}}$ until it fulfills the identification restrictions.

LGD component

Using a similar reasoning as above, we sample ω_t^{d} from the Pólya-gamma distribution

$$p(\omega_t^{\mathrm{d}}|\ldots) = \mathrm{PG}(T_t, \psi_t^{\mathrm{d}})$$

and sample $\boldsymbol{\alpha}^{\mathrm{d}}$ from a normal distribution

$$\boldsymbol{\alpha}^{\mathrm{d}} \sim \mathrm{N}(\boldsymbol{m}^{\mathrm{d}}, \boldsymbol{V}^{\mathrm{d}}),$$

where

$$\begin{aligned} \boldsymbol{V}^{\mathrm{d}} &= (\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\Omega}^{\mathrm{d}}\boldsymbol{X}^{\mathrm{d}})^{-1}, \\ \boldsymbol{m}^{\mathrm{d}} &= (\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\Omega}^{\mathrm{d}}\boldsymbol{X}^{\mathrm{d}})^{-1}(\boldsymbol{X}^{\mathrm{d}\prime}\boldsymbol{\kappa}^{\mathrm{d}}), \end{aligned}$$

where $\boldsymbol{X}^{\mathrm{d}}$ is a matrix that contains the constant and the relevant factors, where $\boldsymbol{\kappa}^{\mathrm{d}}$ is a vector that collects the elements κ_t^{d} for all t and where $\boldsymbol{\Omega}^{\mathrm{d}}$ is a diagonal matrix with ω_t^{d} as diagonal elements. We redraw $\boldsymbol{\alpha}^{\mathrm{d}}$ until it fulfills the identification restrictions.

We collect the terms involving s_{it} in the prior and likelihood and obtain

$$p(s_{it}|\ldots) \propto (p_t^{\mathrm{d}})^{s_{it}}(1-p_t^{\mathrm{d}})^{(1-s_{it})} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right).$$

Hence we can sample s_{it} from the following Bernoulli distribution

$$p(s_{it}=1|\ldots) = \frac{p_t^{\mathrm{d}}\mathrm{N}(\mu_1,\sigma^2)}{(1-p_t^{\mathrm{d}})\mathrm{N}(\mu_0,\sigma^2)+p_t^{\mathrm{d}}\mathrm{N}(\mu_1,\sigma^2)}.$$

We collect the terms involving μ_1 from the posterior and get

$$p(\mu_1|\ldots) \propto \prod_{i,t}^{N,T} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}(y_{it}^{\mathrm{d}} - \mu_1 s_{it})^2\right).$$

Hence, we can sample μ_1 (and μ_0 as well) from normal distributions

$$\begin{aligned} p(\mu_1|\ldots) &= \mathrm{N}\left(\bar{y}_1, \frac{\sigma^2}{N_1}\right), \\ p(\mu_0|\ldots) &= \mathrm{N}\left(\bar{y}_0, \frac{\sigma^2}{N_0}\right), \end{aligned}$$

where $\bar{y}_1$ is the sample mean for the N_1 observations with a latent indicator of 1 and where $\bar{y}_0$ is the sample mean for the N_0 observations with a latent indicator of 0. We redraw μ_0 and μ_1 from their posterior distributions until they fulfill the identification restrictions.

We collect the terms involving σ^2 from the prior and likelihood and get

$$p(\sigma^2|\ldots) \propto \sigma^{-\frac{NT+0.01+2}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(0.01 + \sum_{it}^{NT}(y_{it}^{\mathrm{d}} - \mu_0(1-s_{it}) - \mu_1 s_{it})^2\right)\right),$$

which means that we can sample σ^2 from an inverse gamma-2 distribution

$$p(\sigma^2|\ldots) = \mathrm{iG2}\left(0.01 + \sum_{it}^{NT}(y_{it}^{\mathrm{d}} - \mu_0(1 - s_{it}) - \mu_1 s_{it})^2, 0.01 + NT\right).$$

Factor component

If we collect the terms involving the factors from the prior, likelihood and latent variable distributions, we see that we have a linear Gaussian state space model as in Windle et al. (2013).

Our model has the following transition equation

$$\boldsymbol{f}_{t+1} = \boldsymbol{\Phi}\boldsymbol{f}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \boldsymbol{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}'$ because of the restriction on the unconditional covariance matrix.

We obtain the following observation equations

$$\begin{aligned}
\boldsymbol{y}_t^{\mathrm{m}} &= \boldsymbol{\alpha}^{\mathrm{m}} + \boldsymbol{B}^{\mathrm{m}}\boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{m}}),\\
\frac{\kappa_t^{\mathrm{l}}}{\omega_t^{\mathrm{l}}} &= \alpha_i^{\mathrm{l}} + \boldsymbol{\beta}_i^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{l}\prime}\boldsymbol{f}_t^{\mathrm{l}} + \zeta_t^{\mathrm{l}}, & \zeta_t^{\mathrm{l}} &\sim \mathrm{N}(0, 1/\omega_t^{\mathrm{l}}),\\
\frac{\kappa_t^{\mathrm{d}}}{\omega_t^{\mathrm{d}}} &= \alpha_i^{\mathrm{d}} + \boldsymbol{\beta}_i^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{m}} + \boldsymbol{\gamma}_i^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{l}} + \boldsymbol{\delta}_i^{\mathrm{d}\prime}\boldsymbol{f}_t^{\mathrm{d}} + \zeta_t^{\mathrm{d}}, & \zeta_t^{\mathrm{d}} &\sim \mathrm{N}(0, 1/\omega_t^{\mathrm{d}}),
\end{aligned}$$

where the second and third observations are pseudo data points as explained in the derivations of $\boldsymbol{\alpha}^{\mathrm{l}}$ and $\boldsymbol{\alpha}^{\mathrm{d}}$. We sample the latent factor $\boldsymbol{f}_t$ using the simulation smoother of Durbin and Koopman (2002b).

We collect the terms involving ϕ_{jj} from the equation for the latent factor and obtain

$$p(\phi_{jj}|\ldots) \propto (1 - \phi_{jj}^2)^{-1/2}\exp\left(-\frac{1}{2(1 - \phi_{jj}^2)}(f_{j,t+1} - \phi_{jj}f_{jt})^2\right), \text{ for all } j.$$

Since ϕ_{jj} occurs in both the mean and variance, we cannot derive the posterior analytically. We use a Metropolis-Hastings step instead where we use a normal distribution as proposal density and where we calculate the acceptance probability in the usual way for the independence Metropolis-Hastings sampler. To calculate the mean and variance of the proposal density we maximize the log PDF of the above expression and use the mode as mean and the inverse of the negative Hessian as the covariance matrix. We redraw from the proposal density until the proposed draw satisfies the identification restriction.

2.C.4 Extensions

We consider a couple of alternative models in the main paper and appendix.

First, we consider models with loan and default characteristics that affect the probability of default and the LGD of a loan. It is straightforward to extend our method to allow for these differences. The only difference is that we need to draw all parameters and latent variables except for $\boldsymbol{f}_t$, ϕ_{jj}, $\boldsymbol{A}^{\mathrm{m}}$ and $\boldsymbol{\Sigma}^{\mathrm{m}}$ per group of characteristics.

Second, we consider a model with a mixture of two skewed Student's t distributions instead of a mixture of normals. The only difference is that we draw the parameters in the LGD component (except for ω_t^{d} and $\boldsymbol{\alpha}^{\mathrm{d}}$) based on the derivations in Frühwirth-Schnatter and Pyne (2010). Please refer to the online appendix of Frühwirth-Schnatter and Pyne (2010) for more details on the conditional posterior distributions.

2.D Results for models with loan and default characteristics

Figure 2.D.1: Latent factors, seniority

(a) Two factors (b) Three factors

The figures show the smoothed latent factors for the model defined in Section 2.3.1. We report results for specifications with a macro and a loan factor (a), and an additional loan factor (b). We distinguish defaults of senior secured and senior unsecured loans. The characteristics are included as dummy variables.

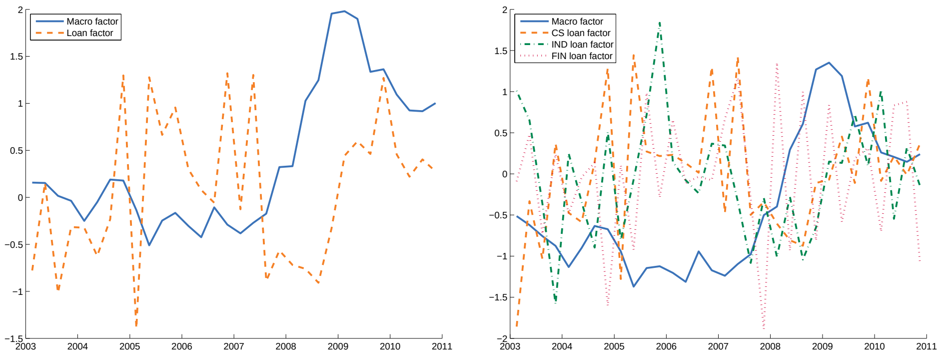
Figure 2.D.2: Latent factors, borrower size

(a) Two factors (b) Three factors

The figures show the smoothed latent factors for the model defined in Section 2.3.1. We report results for specifications with a macro and a loan factor (a), and an additional loan factor (b). We distinguish loans and defaults of large corporates and SMEs. The characteristics are included as dummy variables.

Figure 2.D.3: Latent factors, industry

(a) Two factors (b) Four factors



The figures show the smoothed latent factors for the model defined in Section 2.3.1. We report results for specifications with a macro and a loan factor (a), and an additional two loan factors (b). We distinguish loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials. The characteristics are included as dummy variables.

Table 2.D.1: Parameter estimates, model with seniority

	Macro and loan factor				Macro and two loan factors			
			Panel A: Factor					
ϕ_{11}	0.813	(0.120)			0.803	(0.115)		
ϕ_{22}	0.775	(0.138)			0.720	(0.231)		
ϕ_{33}					0.477	(0.309)		
			Panel B: Macro variables					
α_{GDP}	0.241	(0.631)			-0.035	(0.418)		
β_{GDP}	-0.935	(0.694)			-0.843	(0.487)		
α_{IP}	0.187	(0.518)			-0.027	(0.346)		
β_{IP}	-0.725	(0.583)			-0.633	(0.411)		
α_{UR}	-0.207	(0.556)			0.031	(0.388)		
β_{UR}	0.792	(0.612)			0.756	(0.456)		
WAIC2	129.4				123.3			
			Panel C: Loan status					
α^{l}	-5.665	(0.233)			-5.780	(0.175)		
β^{l}	-0.179	(0.199)			-0.174	(0.320)		
γ_1^{l}	0.628	(0.354)			0.574	(0.292)		
av. p^{l} ($\times 0.01$)	0.309	(0.002)			0.309	(0.002)		
m.e. of f^{m}	-0.055	(0.061)			-0.054	(0.099)		
m.e. of f_1^{l}	0.193	(0.109)			0.177	(0.090)		
WAIC2	321.3				320.0			
			Panel D: Loss given default					
	Sen. secured		Sen. unsecured		Sen. secured		Sen. unsecured	
μ_0	0.070	(0.002)	0.072	(0.001)	0.070	(0.002)	0.072	(0.001)
μ_1	0.766	(0.004)	0.858	(0.003)	0.766	(0.004)	0.858	(0.003)
σ	0.129	(0.001)	0.130	(0.001)	0.129	(0.001)	0.130	(0.001)
α^{d}	−2.174	(0.580)	−1.665	(0.250)	−1.930	(0.393)	−1.542	(0.315)
β^{d}	0.900	(0.619)	0.500	(0.320)	0.849	(0.450)	0.385	(0.247)
γ_1^{d}	0.041	(0.055)	−0.216	(0.133)	−0.037	(0.050)	0.146	(0.126)
γ_2^{d}							−0.257	(0.240)
av. p^{d}	0.151	(0.004)	0.191	(0.004)	0.151	(0.004)	0.191	(0.004)
m.e. of f^{m}	0.103	(0.070)	0.075	(0.048)	0.097	(0.051)	0.058	(0.037)
m.e. of f_1^{l}	0.005	(0.006)	−0.033	(0.020)	−0.004	(0.006)	0.022	(0.019)
m.e. of f_2^{l}							−0.039	(0.036)
av. LGD	0.175	(0.003)	0.222	(0.003)	0.175	(0.003)	0.222	(0.003)
m.e. of f^{m}	0.072	(0.049)	0.059	(0.038)	0.067	(0.036)	0.046	(0.029)
m.e. of f_1^{l}	0.003	(0.004)	−0.026	(0.016)	−0.003	(0.004)	0.017	(0.015)
m.e. of f_2^{l}							−0.030	(0.028)
WAIC2	−11,645.0		−14,438.7		−11,648.0		−14,434.2	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional loan factors. The horizontal panels correspond with the panels in Table 2.2. We distinguish defaults of senior secured and senior unsecured loans. The characteristics are included as dummy variables.

Table 2.D.2: Parameter estimates, model with borrower size

	Macro and loan factor				Macro and two loan factors			
Panel A: Factor								
ϕ_{11}	0.845	(0.098)			0.861	(0.090)		
ϕ_{22}	0.510	(0.217)			0.245	(0.262)		
ϕ_{33}					0.618	(0.190)		
Panel B: Macro variables								
α_{GDP}	-0.433	(0.628)			-0.627	(0.492)		
β_{GDP}	-1.225	(0.885)			-1.265	(0.886)		
α_{IP}	-0.362	(0.543)			-0.535	(0.440)		
β_{IP}	-1.025	(0.756)			-1.083	(0.790)		
α_{UR}	0.385	(0.572)			0.568	(0.455)		
β_{UR}	1.087	(0.814)			1.149	(0.782)		
WAIC2	114.9				109.7			
Panel C: Loan status								
	Large corp.		SME		Large corp.		SME	
α^{l}	-6.441	(0.174)	-5.533	(0.237)	-6.394	(0.127)	-5.474	(0.132)
β^{l}	0.404	(0.273)	0.559	(0.383)	0.335	(0.256)	0.426	(0.323)
γ_1^{d}	0.231	(0.101)	0.366	(0.143)	0.303	(0.088)	0.168	(0.080)
γ_2^{d}							0.392	(0.160)
av. p^{l} ($\times 0.01$)	0.143	(0.003)	0.347	(0.002)	0.147	(0.004)	0.346	(0.002)
m.e. of f^{m}	0.058	(0.039)	0.193	(0.132)	0.049	(0.038)	0.147	(0.111)
m.e. of f_1^{l}	0.033	(0.014)	0.126	(0.049)	0.045	(0.013)	0.058	(0.028)
m.e. of f_2^{l}							0.135	(0.055)
WAIC2	349.9		333.4		252.1		319.2	
Panel D: Loss given default								
	Large corp.		SME		Large corp.		SME	
μ_0	0.075	(0.002)	0.062	(0.001)	0.075	(0.002)	0.062	(0.001)
μ_1	0.848	(0.005)	0.849	(0.003)	0.849	(0.005)	0.849	(0.003)
σ	0.126	(0.001)	0.124	(0.001)	0.126	(0.001)	0.124	(0.001)
α^{d}	-1.707	(0.291)	-1.564	(0.269)	-1.630	(0.268)	-1.517	(0.270)
β^{d}	0.293	(0.302)	0.243	(0.272)	0.333	(0.389)	0.348	(0.295)
γ_1^{d}	-0.404	(0.173)	-0.406	(0.167)	-0.490	(0.143)	-0.186	(0.110)
γ_2^{d}							-0.370	(0.159)
av. p^{d}	0.160	(0.005)	0.166	(0.003)	0.160	(0.005)	0.166	(0.003)
m.e. of f^{m}	0.036	(0.037)	0.033	(0.037)	0.042	(0.049)	0.048	(0.041)
m.e. of f_1^{l}	-0.050	(0.022)	-0.056	(0.023)	-0.062	(0.018)	-0.026	(0.015)
m.e. of f_2^{l}							-0.051	(0.022)
av. LGD	0.199	(0.004)	0.193	(0.003)	0.199	(0.004)	0.193	(0.003)
m.e. of f^{m}	0.028	(0.029)	0.026	(0.030)	0.033	(0.038)	0.038	(0.032)
m.e. of f_1^{l}	-0.039	(0.017)	-0.044	(0.018)	-0.048	(0.014)	-0.020	(0.012)
m.e. of f_2^{l}							-0.040	(0.017)
WAIC2	$-8{,}207.3$		$-15{,}655.8$		$-8{,}194.0$		$-15{,}661.0$	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional loan factors. The horizontal panels correspond with the panels in Table 2.2. We distinguish loans and defaults of large corporates and SMEs. The characteristics are included as dummy variables.

Table 2.D.3: Parameter estimates, model with industries

	Macro and loan factor						Macro and three loan factors					
Panel A: Factor												
ϕ_{11}			0.910	(0.060)					0.882	(0.071)		
ϕ_{22}			0.221	(0.269)					−0.052	(0.300)		
ϕ_{33}									0.260	(0.272)		
ϕ_{44}									−0.274	(0.328)		
Panel B: Macro variables												
α_{GDP}			0.526	(0.625)					−0.412	(0.331)		
β_{GDP}			−1.299	(0.880)					−1.136	(0.675)		
α_{IP}			0.430	(0.532)					−0.336	(0.299)		
β_{IP}			−1.062	(0.753)					−0.928	(0.579)		
α_{UR}			−0.490	(0.589)					0.376	(0.316)		
β_{UR}			1.206	(0.832)					1.037	(0.632)		
WAIC2			100.0						97.7			
Panel C: Loan status												
	Cons. staples		Industrials		Financials		Cons. staples		Industrials		Financials	
α^{l}	−6.118	(0.173)	−5.929	(0.302)	−6.990	(0.377)	−5.884	(0.075)	−5.388	(0.140)	−6.328	(0.184)
β^{l}	0.203	(0.293)	0.602	(0.499)	0.787	(0.600)	0.202	(0.180)	0.545	(0.361)	0.717	(0.457)
γ_1^{d}	0.362	(0.103)	0.364	(0.106)	0.342	(0.108)	0.405	(0.087)	0.297	(0.089)	0.272	(0.109)
γ_2^{d}									0.233	(0.075)	0.216	(0.099)
γ_3^{d}											0.173	(0.048)
av. p^{l} ($\times 0.01$)	0.273	(0.005)	0.414	(0.005)	0.162	(0.004)	0.274	(0.005)	0.415	(0.005)	0.162	(0.004)
m.e. of f^{m}	0.055	(0.080)	0.248	(0.205)	0.128	(0.097)	0.055	(0.049)	0.225	(0.149)	0.116	(0.074)
m.e. of f_1^{l}	0.099	(0.028)	0.150	(0.044)	0.055	(0.018)	0.110	(0.024)	0.123	(0.037)	0.044	(0.018)
m.e. of f_2^{l}									0.096	(0.031)	0.035	(0.016)
m.e. of f_3^{l}											0.028	(0.008)
WAIC2	350.7		321.9		281.9		268.3		293.2		241.9	

This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1 for different loan and default characteristics. We report results for specifications with a macro and a loan factor, and an additional two loan factors. The horizontal panels correspond with the panels in Table 2.2. We distinguish loans and defaults of borrowers in the sectors Consumer Staples, Industrials, and Financials. The characteristics are included as dummy variables. The number of observations is 14,997.

Table 2.D.3: Parameter estimates, model with industries (Continued)

	Macro and loan factor						Macro and three loan factors					
	Panel D: Loss given default											
	Cons. staples		Industrials		Financials		Cons. staples		Industrials		Financials	
μ_0	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)	0.056	(0.002)	0.056	(0.002)	0.085	(0.003)
μ_1	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)	0.851	(0.006)	0.837	(0.004)	0.796	(0.006)
σ	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)	0.121	(0.002)	0.119	(0.001)	0.145	(0.002)
α^{d}	−1.735	(0.265)	−1.786	(0.280)	−2.025	(0.465)	−1.516	(0.141)	−1.690	(0.200)	−1.578	(0.255)
β^{d}	0.401	(0.314)	0.327	(0.335)	0.793	(0.549)	0.351	(0.250)	0.256	(0.296)	0.664	(0.443)
γ_1^{d}	−0.327	(0.123)	−0.472	(0.147)	−0.492	(0.171)	−0.293	(0.103)	−0.238	(0.137)	−0.350	(0.167)
γ_2^{d}									−0.490	(0.149)	−0.215	(0.158)
γ_3^{d}											−0.313	(0.099)
av. p^{d}	0.163	(0.007)	0.156	(0.004)	0.185	(0.006)	0.163	(0.007)	0.157	(0.004)	0.185	(0.006)
m.e. of f^{m}	0.055	(0.043)	0.041	(0.042)	0.099	(0.068)	0.048	(0.034)	0.033	(0.038)	0.084	(0.056)
m.e. of f_1^{l}	−0.045	(0.017)	−0.060	(0.019)	−0.061	(0.022)	−0.040	(0.014)	−0.030	(0.017)	−0.044	(0.021)
m.e. of f_2^{l}									−0.063	(0.019)	−0.027	(0.020)
m.e. of f_3^{l}											−0.040	(0.013)
av. LGD	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)	0.186	(0.006)	0.178	(0.004)	0.217	(0.004)
m.e. of f^{m}	0.043	(0.034)	0.032	(0.033)	0.070	(0.049)	0.038	(0.027)	0.025	(0.029)	0.060	(0.040)
m.e. of f_1^{l}	−0.035	(0.014)	−0.047	(0.015)	−0.044	(0.015)	−0.032	(0.011)	−0.024	(0.014)	−0.032	(0.015)
m.e. of f_2^{l}									−0.049	(0.015)	−0.019	(0.014)
m.e. of f_3^{l}											−0.028	(0.009)
WAIC2	−4,361.6		−9,569.7		−4,385.9		−4,361.0		−9,574.1		−4,382.3	

See table note on previous page.

2.E Results for model with mixture of skew-t

In Section 2.4.4, we discuss replacing the LGD component's mixture of normals with a mixture of skewed Student's t, or skew-t, distributions.

Following Azzalini and Capitanio (2003), if a random variable X is skew-t distributed, $X \sim St(\xi, \omega, \alpha, \nu)$, then

$$f(x; \xi, \omega, \alpha, \nu) = \frac{2}{\omega} t_\nu(q_x) T_{\nu+1}\left(\alpha q_x \sqrt{\frac{\nu+1}{q_x^2+\nu}}\right), \tag{2.E.1}$$

with location parameter ξ, scale parameter ω, shape parameter α, and degrees of freedom ν, where $q_x = (x-\xi)/\omega$, and t_ν and T_ν the PDF and CDF of a standard Student's t distribution with ν degrees of freedom.

The expected value of X is not ξ, but is shifted to the left or right, depending on the skewness. It is given by

$$\mathrm{E}[X] = \xi + \omega\delta\sqrt{\frac{\nu}{\pi}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \tag{2.E.2}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. The normal distribution is nested, by setting $\alpha = 0$ and $\nu \to \infty$.

For the Bayesian estimation procedure, we follow the reparametrization by Frühwirth-Schnatter and Pyne (2010), $g(x; \boldsymbol{\theta}) = f(x; \xi, \omega, \alpha, \nu)$, with parameter vector $\boldsymbol{\theta} = (\xi, \sigma, \psi, \nu)$, where $\sigma^2 = \omega^2(1-\delta^2)$ and $\psi = \omega\delta$. The estimates for the model with the mixture of skew-t distributions are presented in Table 2.E.1. Plugging in the posterior mean of $\boldsymbol{\theta}$, we find that the mean of the mild (severe) loss is 0.090 (0.605).

The PDF of a mixture is given by the sum of the PDFs of the mixture components, weighted by the mixture probability, or mathematically, $h(x; p_t^{\mathrm{d}}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = (1 - p_t^{\mathrm{d}})g(x; \boldsymbol{\theta}_0) + p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)$. From this, we can calculate the posterior mixture probabilities. The likelihood that, conditional on the latent factors $\boldsymbol{f}_t$, observation y_{it}^{d} is a severe loss is

$$\Pr[s_{it} = 1 | y_{it}^{\mathrm{d}}, \boldsymbol{f}_t] = \frac{p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)}{(1-p_t^{\mathrm{d}})g(x; \boldsymbol{\theta}_0) + p_t^{\mathrm{d}} g(x; \boldsymbol{\theta}_1)}, \tag{2.E.3}$$

and $\Pr[s_{it} = 0|y^{\mathrm{d}}_{it}, \boldsymbol{f}_t] = 1 - \Pr[s_{it} = 1|y^{\mathrm{d}}_{it}, \boldsymbol{f}_t]$. The mixture fit in Figure 2.6 and the posterior mixture probabilities in Figure 2.E.2 are based on the posterior means of $\boldsymbol{\theta}$ and of the implied probability of a severe loss p^{d}_t.

Table 2.E.1: Parameter estimates, mixture of skew-*t*

	Mixture of normals		Mixture of skew-t	
	Panel A: Factor			
ϕ_{11}	0.856	(0.078)	0.878	(0.075)
ϕ_{22}	0.322	(0.207)	0.230	(0.235)
	Panel B: Macro variables			
α_{GDP}	−0.008	(0.317)	−0.342	(0.355)
β_{GDP}	−0.969	(0.531)	−1.033	(0.611)
α_{IP}	−0.007	(0.284)	−0.273	(0.313)
β_{IP}	−0.802	(0.465)	−0.824	(0.517)
α_{UR}	0.008	(0.298)	0.341	(0.355)
β_{UR}	0.879	(0.491)	1.032	(0.617)
	Panel C: Loan status			
α^{l}	−5.850	(0.097)	−5.700	(0.118)
β^{l}	0.371	(0.201)	0.428	(0.260)
γ^{l}	0.258	(0.058)	0.254	(0.059)
av. p^{l} ($\times 10^{-2}$)	0.309	(0.002)	0.310	(0.002)
m.e. of f^{m}	0.114	(0.062)	0.132	(0.080)
m.e. of f^{l}	0.080	(0.018)	0.078	(0.018)
	Panel D: Loss given default			
ξ_0	0.072	(0.001)	−0.024	(0.001)
ξ_1	0.829	(0.002)	1.041	(0.002)
σ	0.131	(0.001)	0.017	(0.001)
ψ_0			0.119	(0.003)
ψ_1			−0.399	(0.008)
ν_0			4.891	(0.172)
ν_1			3.065	(0.143)
α^{d}	−1.643	(0.159)	−1.242	(0.187)
β^{d}	0.328	(0.202)	0.399	(0.264)
γ^{d}	−0.293	(0.075)	−0.308	(0.104)
av. p^{d}	0.174	(0.003)	0.220	(0.004)
m.e. of f^{m}	0.045	(0.028)	0.064	(0.043)
m.e. of f^{l}	−0.040	(0.010)	−0.050	(0.017)
av. LGD	0.204	(0.002)		
m.e. of f^{m}	0.034	(0.021)		
m.e. of f^{l}	−0.030	(0.008)		

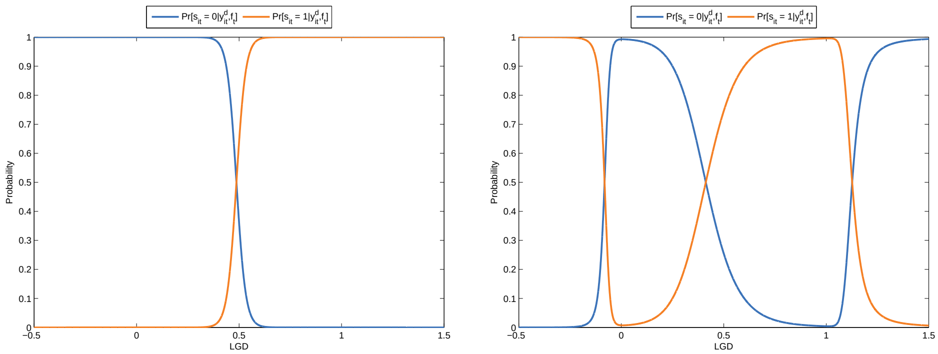
This table presents the posterior mean and standard deviation (in parentheses) of the parameters of the model in Section 2.3.1 with different distributions for the LGD component. All model specifications have a macro and a loan factor. The specifications do not include loan or default characteristics. The horizontal panels correspond with the panels in Table 2.2. The LGD type can be either mild or severe. In the left column, conditional on the type, the LGD follows a normal distribution with mean ξ_0 or ξ_1 (corresponding to μ_0 and μ_1 in Table 2.2), and volatility σ. In the right column, conditional on the type, the LGD follows a skew-t distribution with location parameter ξ_0 or ξ_1, scale parameter σ, shape parameter ψ_0 or ψ_1 and degrees of freedom ν_0 or ν_1.

Figure 2.E.1: Latent factors, skew-t

The figure shows the latent factors of the two-factor model without cross-sectional variation, with the LGD distributed as a mixture of normals or mixture of skew-t.

Figure 2.E.2: Smoothed posterior state probabilities

(a) Normal (b) Skew-t



The figures present the smoothed state probabilities $\Pr[s_{it} = 0|y_{it}^{\mathrm{d}}, \boldsymbol{f}_t]$ (blue line) and $\Pr[s_{it} = 1|y_{it}^{\mathrm{d}}, \boldsymbol{f}_t]$ (orange line) for the two-factor model , with a mixture of normals (a) and with a mixture of skew-t distributions (b) for the LGD. The mixture probabilities are for the first quarter of 2003 and based on the posterior mean estimates of the parameters of the two-factor model without cross-sectional variation.

Chapter 3

Long-term investing under uncertain parameter instability

3.1 Introduction

One of the challenges in economic research is that relationships change over time. These changes could be due to policy changes, macroeconomic shocks, or learning by economic agents, among others. Instabilities are widely recognized to be present in both macroeconomic (Stock and Watson, 1996, 2003) and financial time series (Ang and Timmermann, 2012). If instabilities are ignored, inference can be distorted and predictive performance negatively impacted (Rossi, 2013). It is important to recognize time-variation in modeling economic time series, as these models are input for decisions by, for example, central bankers and investors.

We consider the consequences of time-varying relationships for a long-term investor, such as a pension fund. Specifically, we compare different types of time-variation and assess the risk involved in misspecifying the dynamic process. The asset allocation of long-term investors crucially depends on the model for asset returns. In case of return predictability, returns are mean reverting, and stocks safer in the long run. The optimal portfolio composition for the long-term investor then differs from that of a myopic, i.e. short-term, investor due to a hedge term that protects against negative changes in future investment opportunities (Campbell and Viceira, 2002; Brandt, 2010).

Following the famous work by Welch and Goyal (2008), there has been an increased interest into time-varying parameters in the return predictability literature. They show that for a large set of popular predictor variables, the out-of-sample predictive ability is poor and unstable. Predictive models often do not outperform the historical average. One of the explanations for the poor performance is that the relationship between predictors and the equity premium varies over time. Timmermann (2008) argues that predictive content does exist occasionally. But when it appears, it is quickly exploited by market participants, thus creating short-lived "pockets of predictability."[1]

Evidence of instability in return predictability is not limited to Welch and Goyal (2008). In a recursive estimation, Pesaran and Timmermann (1995) find that the fit of their predictive model with the dividend yield varies, and improvements are associated with relatively volatile periods. The earliest known formal test of structural breaks in the relationship between the dividend yield and stock returns is Viceira (1997). Interestingly, he does not find significant evidence for a structural break. Subsequent papers do report substantial evidence of structural instabilities using standard structural break tests. Rapach and Wohar (2006) find medium to strong evidence of structural breaks for many predictors, and Paye and Timmermann (2006) show that instabilities are present across international equity markets. Lettau and Van Nieuwerburgh (2008) detect one or two breaks in the dividend yield using yearly data, and show that correcting for these increases the predictive ability. Ang and Bekaert (2007) also find that the predictive power of the dividend yield is unstable. The predictive value covaries positively with interest rates and deteriorates in the 1990s.

Rapach and Zhou (2013) argue that there are four ways to improve on out-of-sample performance, despite the instability. Three options (constraints, forecast combinations, and diffusion indices) rely on shrinkage, the fourth is to explicitly model the parameter instability. We focus on the last option. Several studies have suggested methods that allow for time-varying parameters, such as change point models, assuming few breaks (Henkel et al., 2011; Pettenuzzo and Timmermann, 2011), or time-varying parameter models, assuming breaks each period (Dangl and Halling, 2012; Johannes et al., 2014).

[1] Farmer et al. (2018) look into this phenomenon further, reconciling it with investors' incomplete learning about a persistent growth component in the cash flows that is subject to regime switches.

However, there is no consensus on what type of time-variation should be used to best capture these dynamics.

Given the lack of consensus, our first goal is to find out what the *type of time-variation* supported by the data is, similar to how Dangl and Halling (2012) aim to capture the "*degree of time-variation*." Instead of assuming the number of breaks, or type of instability, ex ante, we estimate it using a mixture innovation model (McCulloch and Tsay, 1993; Giordani and Kohn, 2008). It is a flexible and general type of state space model, where each time-varying parameter is modeled as a mixture of innovations. The type of time-variation depends explicitly on a break probability and size. Parameters can change or remain constant each period, subject to a stochastic break process.

The mixture innovation model provides an ideal framework to answer the questions at hand, because (i) the time-variation is governed by the break probability, a natural and intuitive measure for the type of instability, (ii) it nests multiple types of time-variation and allows for periods without breaks, (iii) we do not have to set the number of breaks, but can infer an estimate and uncertainty of the break probability from the data, (iv) it is easy to define independent break processes for parameter subsets, such that we can apply stochastic volatility without imposing any break process on the mean, and (v) it is computationally efficient, with $O(T2^K)$ operations, with T the number of observations, and K the number of break processes, whereas for example the change point model requires $O(T^2)$ operations (Koop and Potter, 2007).

Based on US data from 1946 to 2015, and using Bayesian methodology, we show that there is substantial time-variation in the relationship between the dividend price ratio and stock returns. However, it is hard to clearly identify large breaks, and we find quite some uncertainty regarding the break probability due to the low signal-to-noise ratio. The results are robust to changes in the break probability prior.

Our second goal is to investigate the economic consequences of misspecifying the nature of instability of the relationship between the dividend price ratio and excess returns. This is achieved by interpreting the prior on the break probability as the investor's views regarding the instability. We economically evaluate the different views for a long horizon investor with power utility preferences by calculating the loss in certainty equivalent return from assuming an incorrect type of time-variation. This is motivated by the work of Barberis (2000) and other research on long-term investing, see e.g. Campbell and Viceira (2002) and Brandt (2010).

We find substantial differences regarding the risk on stock returns depending on the type of time-variation, i.e. few large or many small breaks. We show that it can be extremely costly to assume a stable relationship between the dividend price ratio and excess stock returns. When the allocation is optimized under the assumption of a static model, and the data generating process does contain instabilities, the potential loss at the long horizon is very large, up to a difference of 17% in certainty equivalent return. The downside risk of assuming constant parameters is large, even if the per period break probability is small. This is because at the long horizon, the probability of encountering a break is still large. Furthermore, we find that when taking uncertainty regarding the break probability into account, the predictive density widens. Ignoring that the break probability is unknown can lead to about a 1% loss in certainty equivalent return. Seemingly in contrast to Johannes et al. (2014), we find that it is more important to take instability in the intercepts and loadings into account, rather than in the variances. This difference can be explained the fact that I focus on the long horizon.

Given the substantial time-variation in the relationship between the dividend price ratio and the excess returns, and the potentially large downside risk we find of wrongly assuming a stable relationship, we recommend that investors take this instability into account when optimizing their long-term portfolio. An investor with min-max utility (Hansen and Sargent, 2001) selects a model with many breaks. We show that an allocation based on this type of instability is most robust to instability misspecification in terms of certainty equivalent return.

Our main contribution is that we estimate the break probability, instead of assuming ex ante that the relationship between the dividend price ratio and excess returns is subject to few or many breaks. There is much interest in how to model the instability between return and predictor variables, but it is not obvious which of the methods works best, as the type of time-variation is unknown. On the one hand, it has been suggested to characterize the instability through a small number of breaks of relatively large magnitude, for example due to exogenous shocks such as policy changes or conditional on the state of the economy. This type of time-variation can be captured by a Markov switching model (Hamilton, 1989) or a change point model (Chib, 1998). The parameters then change when the economy is in a new, possibly recurring, regime. The regime may depend on a latent or observed variable. Following this strategy, Ang and Bekaert (2002), Guidolin and Timmermann (2007) and Henkel et al. (2011) estimate a Markov switching model with two to four regimes. Their findings suggest that the predictive power is countercyclical, it is stronger during recessions. In line with this evidence, Gonzalo and Pitarakis (2012) find that predictability depends on industrial production in a regime switching model. Guidolin and Timmermann (2007) show that the allocation to stocks is larger in bull markets. The costs of regime switches are about 1.3% annualized return at the long horizon. Pettenuzzo and Timmermann (2011) consider a change point model, such that it is not possible to return to a past regime. They find about eight breaks between the dividend yield and the equity premium in the 1926 to 2005 period. As the investment horizon increases, the costs from ignoring the breaks grow up to 7.8% annualized return at the 120 month horizon. This is because the probability of a break occurring increases with the horizon. Even with a small break probability, the probability of a break over the 10 year investment period is substantial.

On the other hand, it has been suggested that parameters change continuously, such that they vary smoothly over time, for example due to learning or a relationship with business cycle fluctuations. Modeling these dynamics can be achieved by assuming that there is a (small) break each period, using time-varying parameter models. Dangl and Halling (2012), Johannes et al. (2014) and Diris (2014) implement such models. Dangl and Halling (2012) exclude time-varying volatility, but do allow for time-varying loadings. As with the regime switching models, they also find a countercyclical relationship. Johannes et al. (2014) allow investors to learn the predictability coefficient, and find utility gains of about 2% annualized returns at the 2 year horizon. They stress that a combination of model components needs to be included to achieve out-of-sample performance, including parameter uncertainty and instability, stochastic volatility and predictable returns.

Our research adds to the long-term investment literature (Campbell and Viceira, 2002; Brandt, 2010), in particular the effects of modeling choices and associated uncertainties. The consequences of parameter uncertainty (Kandel and Stambaugh, 1996; Barberis, 2000) and model uncertainty (Avramov, 2002; Cremers, 2002) have been studied. Pettenuzzo and Timmermann (2011), Johannes et al. (2014) and others show the impact of parameter instability, but ignore the impact of uncertain parameter instability, where the type of time-variation is uncertain.

Furthermore, we add to the literature applying mixture innovation models. They have been applied mostly to macroeconomic time series, among others by Giordani et al. (2007), Giordani and Kohn (2008), Koop et al. (2009), Groen et al. (2013). Our methodology is most closely related to Koop et al. (2009), who, in a monetary policy application, estimate a mixture innovation model with a similar break process specification, but they ignore persistence. They share our motivation of trying to let the data speak on the parameter instability. We contribute to this literature by applying this model on data with a low signal-to-noise ratio. To our knowledge, the only other application of the mixture innovation model to financial data is by Ravazzolo et al. (2008). They use it to characterize both parameter and model uncertainty regarding stock return predictability in a univariate setting predicting one period ahead, whereas we have a multivariate setting and focus on longer horizons. Moreover, they ignore time-varying volatility, whereas we explicitly take this into account and allow for different break processes for the mean and variance.

Finally, our research is related to work comparing different types of time-varying models instead of just comparing to a static model. Despite that Elliott and Müller (2006) show that it is difficult to distinguish between different models such as a regime switching or time-varying parameter model from a testing perspective, there have been some recent attempts.[2] Clark and Ravazzolo (2015), Bauwens et al. (2015) and Pettenuzzo and Timmermann (2017) statistically compare (mostly univariate) models on macroeconomic time series. Our approach differs from these studies, because we economically evaluate the performance for different types of time-variation in a multivariate model for financial time series, with a lower signal-to-noise ratio. One example that compare types of time-variation in financial time series is Pesaran and Timmermann (2002). They use a time-varying parameter model and methods based on recursive testing procedures, such as Bai and Perron's (1998) break test, to predict the sign of the equity premium. Instead, we focus on the long-term investor and

[2] I specifically consider the many versus few breaks cases here. In particular in the financial volatility literature, comparing GARCH and stochastic volatility with specific components is more common, see e.g. Hansen and Lunde (2005) or Nakajima (2012). But this excludes a model with few breaks.

the predictive density, rather than short horizon market timing. Second, we use a mixture innovation model which incorporates the few breaks case, explicitly specifying a dynamic process for the parameters.

The paper is structured as follows. Section 3.2 defines the mixture innovation model, and details the priors and estimation procedure. Section 3.3 describes the data. Section 3.4 presents the estimation results, followed by the economic evaluation in Section 3.5. Section 3.6 contains a sensitivity analysis to the break probability prior, and Section 3.7 concludes.

3.2 Methodology

We follow the return predictability and portfolio choice literature (Brandt, 2010; Rapach and Zhou, 2013) and use a restricted vector autoregressive (VAR) model for the excess log return r_t,

$$r_t = \beta_{1t} + \beta_{2t} z_{t-1} + \varepsilon_{1t}, \tag{3.1}$$

$$z_t = \beta_{3t} + \beta_{4t} z_{t-1} + \varepsilon_{2t}, \tag{3.2}$$

for $t = 1, \ldots, T$, with T the sample size, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$, and z_t the predictor at time t. The intercepts and loadings $\boldsymbol{\beta}_t = (\beta_{1t}, \ldots, \beta_{4t})'$ and variance-covariance matrix $\boldsymbol{\Omega}_t$ are potentially time-varying.

We follow the time-varying parameter VAR (TVP-VAR) literature (Primiceri, 2005) and decompose the variance-covariance matrix $\boldsymbol{\Omega}_t$ as follows,

$$\boldsymbol{A}_t \boldsymbol{\Omega}_t \boldsymbol{A}_t' = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t', \tag{3.3}$$

where $\boldsymbol{A}_t$ is a lower triangular matrix

$$\boldsymbol{A}_t = \begin{bmatrix} 1 & 0 \\ \alpha_t & 1 \end{bmatrix}, \tag{3.4}$$

with covariance term α_t and $\boldsymbol{\Sigma}_t = \mathrm{diag}(\sigma_{1t}, \sigma_{2t})$, a diagonal matrix with the (structural) volatilities σ_{1t} and σ_{2t} on the diagonal. This decomposition simplifies inference, because

α_t is not restricted between zero and one,[3] and it allows us to rewrite (parts of) the model in (conditionally) Gaussian state space form to easily draw the variance and covariance terms.

3.2.1 Modeling parameter instability

The formulation in Equations (3.1)–(3.4) is general in the sense that it does not specify the parameters' dynamics. We employ a mixture innovation (MI) model, introduced by McCulloch and Tsay (1993), to describe the time-variation. It is a flexible and general (conditionally Gaussian) state space model, where each time-varying parameter θ_t is modeled as,

$$\theta_t = \theta_{t-1} + \kappa_t \eta_t, \tag{3.5}$$

where $\eta_t \sim \mathrm{N}(0, \sigma_\eta)$, such that the break size depends on the break process κ_t, and we ignore a persistence parameter for illustrative purposes. We focus on a binary break process, $\kappa_t \in \{0, 1\}$, which is i.i.d. with break probability $\Pr[\kappa_t = 1] = \pi$.

To see how the process in Equation (3.5) works, consider what happens for different values of the break process κ_t. If $\kappa_t = 0$, then $\theta_t = \theta_{t-1}$, so there is no break and the parameter stays in the current regime. If $\kappa_t = 1$, then $\theta_t = \theta_{t-1} + \eta_t$, there is a break and we are in a new regime. This break process provides flexibility and differentiates it from alternative models. The break probability π and break size σ_η define the type of instability. The name mixture of innovations becomes clear when rewriting Equation (3.5) as $\theta_t = \theta_0 + \sum_{s=1}^{t} \kappa_t \eta_t$, where the number of mixture components is equal to the number of values κ_t can take.

Roughly, there are two other ways to model the parameter instability. One option is to use a change point (CP) model (Chib, 1998; Pesaran et al., 2006a; Koop and Potter, 2007) or regime switching model to allow for a limited number of regimes, where transitions to a new regime follow a Markov process. The second option is to use a time-varying parameter (TVP) model (Cogley and Sargent, 2001, 2005; Primiceri,

[3]Note that α_t is not the covariance or correlation, but a transformation thereof, because $\mathrm{corr}(\varepsilon_{1t}, \varepsilon_{2t}) = -\alpha_t \sigma_{1t} / \sigma_{2t}$.

2005), where the parameters are modeled as a random walk or autoregressive process and change each period.

We use an MI model for our analysis because of its attractive properties.[4] First, the break probability π is intuitive to interpret, which facilitates prior elicitation and provides a measure for the risk of breaks or parameter instability. Second, the MI model nests many types of time-variation, such as few (large) breaks or many (small) breaks and as extreme cases the constant parameter case, if $\pi = 0$, and the continuously time-varying case, if $\pi = 1$. The prevailing methods to model time-variation are nested in this framework. The TVP model is the extreme case of an MI model with $\pi = 1$. The CP model is approximately captured by a small value for π. It is not exactly the classic CP model (Chib, 1998), because we cannot fix the number of breaks. Rather, we approximate the CP model with an unknown number of breaks (Koop and Potter, 2007). Hence, we can study multiple time-variation settings in one model framework, which allows for easy comparison.[5] Third, there is no need to select the type of time-variation, or the number of breaks, ex ante. Estimating the break probability, we infer the type of instability supported by the data and the posterior variance provides the uncertainty regarding this instability. Fourth, we can easily specify independent break processes for different parameters, which allows us to specify different dynamics for the mean and variance, for example. Fifth, inference is computationally efficient due to the algorithm of Gerlach et al. (2000),[6] applicable to any conditionally Gaussian state space model. In state space terminology, Equations (3.1) and (3.2) are the observation or measurement equations (Durbin and Koopman, 2012). The efficiency of Gerlach et al.'s (2000) algorithm comes from drawing κ_t without conditioning on the states θ_t. Drawing the break process κ_t is linear in the number of values the break process can take. It requires $O(T2^K)$ operations, with T the sample size and K the number of binary break processes (Giordani and Kohn, 2008).

[4]See Giordani and Kohn (2008) for a general argument in favor of the mixture innovation model.

[5]To illustrate the flexibility, the MI model can for example be combined with a CP model (Giordani et al., 2007), or written as a regime switching model (Giordani and Kohn, 2008).

[6]The algorithm can be improved in specific cases. Giordani and Kohn (2008) propose an adaptive Metropolis Hastings step to more efficiently draw for periods where the break process takes the same value (almost) every draw. Further, Fiorentini et al. (2014) suggest block-sampling the break process, which is more efficient if the number of possible states is more than two.

A TVP model does not provide us with the desired answers for a number of reasons. First, it does not provide a measure for the type of instability as the number of breaks is fixed. Second, a TVP model cannot capture large shifts, because stable periods skew the break size estimate downwards. Third, TVP models can be interpreted as assuming $T-1$ breaks, an extreme estimate of the instability. This increases the risk of fitting noise rather than signal.

Change point models also suffer from some drawbacks that make it less suitable for our analysis. First, it is difficult to allow for independent break processes, because the number of regimes grows exponentially with the number of independent break processes, as each combination of realizations is defined as a separate regime.[7] Second, the number of breaks is assumed to be known in Chib's (1998) CP model. Koop and Potter (2009) show that fixing the number of breaks implies a very restrictive prior and creates the problem of break probability piling up at the end of the sample. This happens because an exact number of break locations needs to be drawn. If not all breaks have been drawn before the final periods, the remaining breaks are forced to occur at the end of the sample. Pesaran et al. (2006a) propose Bayesian model averaging in a hierarchical CP model to avoid choosing the number of breaks.[8] However, this does not solve the backpiling problem (Koop and Potter, 2007), and requires a maximum number of breaks. Furthermore, the model has to be estimated for each number of breaks under consideration, increasing computation time. Koop and Potter (2007) suggest a CP model with an unknown number of breaks, but it is computationally demanding. Chib's (1998) CP model requires $O(T^2)$ operations, and with Poisson distributed durations requires an additional $O(T)$ evaluations (Koop and Potter, 2007).[9]

A semi-parametric alternative to the MI model is the infinite hidden Markov-switching (IHMS) model. It assumes that parameter instability is governed by an underlying hierarchical Dirichlet process (Teh et al., 2006). Fox et al. (2011) extends

[7]An example that does allow for different break processes is Maheu and Song (2014).

[8]Similarly, Maheu and Gordon (2008) use Bayesian model averaging over a set of submodels, where each submodel is based on a different sample size.

[9]The regime duration in a MI model is geometrically distributed (Giordani and Kohn, 2008), which implies a monotonically decreasing distribution. Koop and Potter (2007) suggest Poisson distributed durations as they offer more flexibility in the distributional shape. Giordani and Kohn (2008) show that it is also possible to incorporate this into the MI model, but we do not pursue this.

this principle to a 'sticky' version to allow for path dependence. The IHMS model shares some of the advantages of the MI model, such as an infinite number of possible breaks, and Bauwens et al. (2017) show how to define different break processes for different parameters. Crucial for our application however, the MI model is intuitive to interpret and the break probability π provides a measure for easy comparison between different types of parameter instability.

3.2.2 Mixture innovation model

The time-varying parameters can naturally be divided into three groups, that may be subject to different types of time-variation: (i) intercepts and loadings, (ii) residual variances, and (iii) residual covariance. To capture the grouping, we implement a MI model with three independent binary break processes, one for each parameter type. Koop et al. (2009) apply a similar specification to macroeconomic data.

The intercepts and loadings $\boldsymbol{\beta}_t$ follow a mixture innovation process that share one break process. To model persistence of $\boldsymbol{\beta}_t$, the state equation slightly deviates from the specification in Equation (3.5). The state equation for the intercepts and loadings $\boldsymbol{\beta}_t$ is

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_1 + f(\kappa_{1t}, \boldsymbol{\Phi}_1)(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_1) + \kappa_{1t}\boldsymbol{\eta}_t, \tag{3.6}$$

for $t = 2, \ldots, T$, where $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{4t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_1)$ with break size matrix $\boldsymbol{Q}_1$, κ_{1t} is a binary break process which is one if there is a break at time t and zero otherwise that is i.i.d. with break probability $\Pr[\kappa_{1t} = 1] = \pi_1$, $\boldsymbol{\mu}_1 = (\mu_{1,1}, \ldots, \mu_{1,4})'$ is a vector of long run means, and $f(\kappa_{1t}, \boldsymbol{\Phi}_1)$ is an autoregressive (AR) function,

$$f(\kappa_{1t}, \boldsymbol{\Phi}_1) = \begin{cases} \boldsymbol{I}_{n_k} & \text{if } \kappa_{1t} = 0 \\ \boldsymbol{\Phi}_1 & \text{if } \kappa_{1t} = 1, \end{cases} \tag{3.7}$$

where $\boldsymbol{I}_{n_k}$ is the $n_k \times n_k$ identity matrix, with $n_k = 4$ the parameter vector length, and $\boldsymbol{\Phi}_1 = \mathrm{diag}(\phi_{1,1}, \ldots, \phi_{1,4})$ a diagonal matrix with autoregressive parameters. We refrain from estimating the off-diagonal elements, to limit the number of parameters.

The outcome of the function in Equation (3.7) is either the identity matrix if there is no break, or a diagonal matrix with autoregressive parameters if there is a break at time t. Hence, if there is no break we have the same value as in the previous period, but if there is a break, we have an autoregressive process. The autoregressive MI (ARMI) process in Equation (3.6) nests the simple AR process as a special case, if $\pi_1 = 1$, and the MI process in Equation (3.5) if $\boldsymbol{\Phi}_1 = \boldsymbol{I}_{n_k}$.[10]

The autoregressive function is used to model the parameters' persistence. Many applications of the MI model (or TVP model) assume a random walk by setting $f(\kappa_{1t}, \boldsymbol{\Phi}_1) = \boldsymbol{I}$, to reduce the number of parameters, see e.g. Primiceri (2005), Koop et al. (2009) or Groen et al. (2013). The parameters then follow a non-stationary process, and the variance in the limit is unbounded. Primiceri (2005) argues that this assumption is "undesirable," but "innocuous" when thinking of the time period as finite. However, non-stationary parameters imply that the predictive density's variance grows linearly with the forecast horizon and is unbounded in the limit. It complicates our economic evaluation, as this requires draws from the predictive density up to 240 months into the future, as we will explain in Section 3.5. Although 240 periods is not infinity, the predictive density at this horizon is very wide and can lead to numerically infeasible draws. Therefore, we choose to model the persistence, and exclude the random walk case using our prior (see Section 3.2.3). Note that even though many ignore the persistence, it is included in the mixture innovation framework considered by Gerlach et al. (2000). Further, the distinguishing property of the mixture innovation process in Equation (3.5) that the parameters are unchanged if $\kappa_{1t} = 0$ and do change if $\kappa_{1t} = 1$ is preserved in Equation (3.6).

[10]The ARMI process in Equation (3.6) is stationary if $|\phi_i| < 1$, for all i. The unconditional distribution under stationarity is equal to that of an AR process with the same mean, variance and autoregressive parameter, as long as $\pi > 0$. In other words, the unconditional distribution of the process does not depend on the break probability π as long as it is larger than zero. The intuition is that if we ignore the breaks, an ARMI process has the same long run distribution as an AR process. The no break draws, if $\kappa_t = 0$, are (in the long run) equally distributed over the break draws, and hence do not influence the shape of the long run distribution. The difference is that convergence to the long run distribution will be slower if $\pi < 1$. See Appendix 3.A.2 for a proof.

The state equations for the variances $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2, \sigma_{2t}^2)'$ and the covariance term α_t are

$$\log \boldsymbol{\sigma}_t^2 = \boldsymbol{\mu}_2 + f(\kappa_{2t}, \boldsymbol{\Phi}_2)(\log \boldsymbol{\sigma}_{t-1}^2 - \boldsymbol{\mu}_2) + \kappa_{2t}\boldsymbol{\zeta}_t, \tag{3.8}$$

$$\alpha_t = \mu_3 + f(\kappa_{3t}, \phi_3)(\alpha_{t-1} - \mu_3) + \kappa_{3t}\xi_t, \tag{3.9}$$

for $t = 2, \ldots, T$, where $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_2)$ and $\xi_t \sim \mathrm{N}(0, q_3^2)$ independent from $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ and each other, κ_{2t} and κ_{3t} are (variance and covariance specific) binary break processes that are i.i.d. with break probabilities $\Pr[\kappa_{2t} = 1] = \pi_2$ and $\Pr[\kappa_{3t} = 1] = \pi_3$, $\boldsymbol{\mu}_2 = (\mu_{2,1}, \mu_{2,2})'$ and μ_3 are the long run means, and $f(\cdot,\cdot)$ the autoregressive function as defined in Equation (3.7), with $\boldsymbol{\Phi}_2 = \mathrm{diag}(\phi_{2,1}, \phi_{2,2})$ and ϕ_3 the autoregressive parameters. We model the log of the variance terms to ensure positivity, following the stochastic volatility literature (Kim et al., 1998).

The break processes κ_{1t}, κ_{2t} and κ_{3t} are independent and allowed to, but not restricted to, break at different points in time. It can be extended to allow for dependence between the breaks, or to allow each parameter to have its own break process. We refrain from this (i) due to the computational costs associated with including an extra break process, and (ii) because we are most interested in differentiating between breaks in the first and second moment, which is captured by the current setup.

3.2.3 Priors

We employ Bayesian methodology to estimate the MI model in Equations (3.1)–(3.4) and **??**–**??** This is a natural way to include parameter uncertainty, a relevant risk for the long-term investor (Barberis, 2000). Moreover, it allows us to assess the degree of uncertainty on the break probability. Given the Bayesian methodology, we need to specify prior distributions for the static parameters and initial conditions for the state equations. The priors are mostly standard conjugate empirically Bayesian priors, based on Primiceri (2005), Koop et al. (2009) and Diris (2014). The location and scaling are based on ordinary least squares (OLS) estimates of the restricted VAR in Equations (3.1)–(3.2) using the first ten years of data. This period is excluded from the estimation sample. We use fairly informative priors regarding the initial values of the state variables, but less so for the long run mean and break probability.

As prior for the long run means $\boldsymbol{\mu}$, we take a normal distribution,

$$\boldsymbol{\mu}_1 \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, 10^6 \operatorname{var}(\hat{\boldsymbol{\beta}})), \tag{3.10}$$

$$\boldsymbol{\mu}_2 \sim \mathrm{N}(\widehat{\log \boldsymbol{\sigma}^2}, 10^6 \operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})), \tag{3.11}$$

$$\boldsymbol{\mu}_3 \sim \mathrm{N}(\hat{\alpha}, 10^6 \operatorname{var}(\hat{\alpha})), \tag{3.12}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator on the first ten years of data and $\operatorname{var}(\hat{\boldsymbol{\beta}})$ its variance covariance matrix. To ensure that the prior is uninformative, the variance is multiplied by a large number, 10^6. We do not have OLS estimates for the structural log variances $\log \boldsymbol{\sigma}^2$ and covariance term α and their variances, so we follow Koop et al. (2009) and Diris (2014) and draw a 1,000 times from an inverse Wishart distribution,

$$\boldsymbol{\Omega}^* \sim \mathrm{IW}(\hat{\boldsymbol{\Omega}}\tau, \tau), \tag{3.13}$$

with $\hat{\boldsymbol{\Omega}}$ the OLS estimate of the residual covariance matrix and degrees of freedom $\tau = 120$ set to the number of time periods used to calibrate the prior. Note that the degrees of freedom in an inverse Wishart distribution can indeed be interpreted as the number of observations to calibrate the location. For each draw of $\boldsymbol{\Omega}^*$, we use the decomposition in Equation (3.3) to obtain a draw of the structural log variances $\log \boldsymbol{\sigma}^2$ and covariance term α. The sample mean and variance of the draws are the estimates $\widehat{\log \boldsymbol{\sigma}^2}$ and $\hat{\alpha}$ and their variances $\operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})$ and $\operatorname{var}(\hat{\alpha})$.

Following Diris (2014), the off-diagonal elements of $\boldsymbol{\Phi}_k$ are set to zero and we take a truncated normal distribution as prior for the diagonal of the autoregressive parameter matrix $\boldsymbol{\Phi}_k$,

$$\operatorname{diag}(\boldsymbol{\Phi}_k) \sim \mathrm{N}(\underline{\boldsymbol{m}}_k, \underline{\boldsymbol{V}}_k) I(|\phi_{ki}| < 1, \forall i = 1, \ldots, n_k), \tag{3.14}$$

for $k = 1, 2, 3$, where $I(A)$ is one if condition A holds and zero otherwise, with $\underline{m}_{ki} = 0.9$ and $\underline{V}_{k(i,i)} = 0.2^2$ and $\underline{V}_{k(i,j)} = 0$ for $i, j = 1, \ldots, n_k$, $i \neq j$ and $k = 1, 2, 3$. The prior is truncated to ensure stationary parameters, as discussed in 3.2.2. We use the notation of adding an underscore bar to prior hyperparameters. The mean and variance reflect our belief in a persistent process for the parameters.

As prior for the break probability π_k, we take a beta distribution,

$$\pi_k \sim \text{Beta}(\underline{a}_k, \underline{b}_k), \tag{3.15}$$

for $k = 1, 2, 3$. The hyperparameters $\underline{a}_k$ and $\underline{b}_k$ reflect the investor's views on the presence of breaks (or lack thereof).

For the intercepts and loadings, we set the hyperparameters $\underline{a}_k$ and $\underline{b}_k$ such that the prior's expectation is between a few breaks (about one break every ten years) and many breaks (about one break per year) case. In particular, we set $\underline{a}_1 = 1$ and $\underline{b}_1 = 59$, which implies an average duration of 60 months, or five years, between breaks, and relatively small prior strength, in line with the prior strength used by Ravazzolo et al. (2008) and Groen et al. (2013). The hyperparameters in the beta prior can be thought of as adding $\underline{a} + \underline{b}$ observations: $\underline{a}$ ones and $\underline{b}$ zeros, or $\underline{a}$ breaks. If we define the prior strength as $(\underline{a} + \underline{b})/T$, with T the sample size, we have a prior strength of 7.1%.[11] In Section 3.4.1, we assume that the break probability is known, and restrict π_1 to a single value. In Section 3.5, we investigate the influence of the prior's expected value on the results.

For the variances and covariance's break processes, we take $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. It is uninformative in π, but puts most probability mass at the many breaks case ($\pi > 1/12$) to favor stochastic volatility. Ignoring the stylized fact of volatility clustering might bias the amount of time-variation in the loadings upwards.[12]

As prior for the break size matrix $\boldsymbol{Q}_k$ we have an inverse Wishart distribution,

$$\boldsymbol{Q}_k \sim \text{IW}(\underline{\boldsymbol{W}}_k \underline{\nu}_k, \underline{\nu}_k), \tag{3.16}$$

[11] Ravazzolo et al. (2008) and Groen et al. (2013) employ priors with prior strengths of 7.4% and 5.2%.

[12] We have also experimented with setting the hyperparameters to an intermediate break process for all break processes, i.e. $\underline{a}_i = 1$ and $\underline{b}_i = 59$, for $i = 1, 2, 3$. This would be in line with for example Pettenuzzo and Timmermann (2011), who estimate a change point model with simultaneous breaks in the mean and the variance. The posterior estimates are implausible as they imply a (nearly) constant variance, but volatile loadings. It seems that this is due to an identification issue, where the time-variation is explained by the volatility and/or the mean. Time-variation is then mostly explained by the mean, because this has the largest impact on the fit.

for $k = 1, 2, 3$, with $\underline{\nu}_k$ degrees of freedom and

$$\underline{\boldsymbol{W}}_1 = \frac{c_1}{\mathrm{E}[\pi_1]} \operatorname{var}(\hat{\boldsymbol{\beta}}), \tag{3.17}$$

$$\underline{\boldsymbol{W}}_2 = \frac{c_2}{\mathrm{E}[\pi_2]} \operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2}), \tag{3.18}$$

$$\underline{w}_3 = \frac{c_3}{\mathrm{E}[\pi_3]} \operatorname{var}(\hat{\alpha}), \tag{3.19}$$

where c_k is a scaling factor, $\mathrm{E}[\pi_k]$ is the expected prior break probability, and $\operatorname{var}(\hat{\boldsymbol{\beta}})$, $\operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})$, and $\operatorname{var}(\hat{\alpha})$ are the variance covariance matrices of the OLS estimators of $\boldsymbol{\beta}$, $\log \boldsymbol{\sigma}^2$, and α on the first ten years of data as in Equations (3.10)–(3.12).

The model is known to be sensitive to the choice of c_k, see e.g. Primiceri (2005) for a discussion. We set $c_k = 0.001$ for $k = 1, 2, 3$, but setting $c_k = 0.01$ or $c_k = 0.0001$ does not affect the results qualitatively.

The presence of $\mathrm{E}[\pi_k]$ in Equations (3.17)–(3.19) reflects the assumption that the time-variation can be characterized either as a small number of large breaks or a large number of small breaks. Following Koop et al. (2009), if the prior for the break probability changes, the prior for the break size should change accordingly.

We should be careful when choosing the degrees of freedom $\underline{\nu}_k$. To see why, consider the case where π_1 is close to 0. The full conditional posterior for the break size $\boldsymbol{Q}_1$ is an inverse Wishart distribution,

$$\boldsymbol{Q}_1 | \ldots \sim \mathrm{IW}\left(\underline{\boldsymbol{W}}_1 \underline{\nu}_1 + \sum_{t=2}^{T} \kappa_{1t} \boldsymbol{\eta}_t \boldsymbol{\eta}_t', \underline{\nu}_1 + \sum_{t=2}^{T} \kappa_{1t} \right), \tag{3.20}$$

with κ_{1t} the break process and $\boldsymbol{\eta}_t$ the residuals from Equation (3.6) of the current draw, and we use the notation "$| \ldots$" to denote conditioning on the data and all other parameters. Because π_1 is close to zero, it is likely that the number of breaks in the current draw is small or even zero. Therefore, $\sum_{t=2}^{T} \kappa_{1t} \approx 0$ and the break size posterior is (approximately) equal to the prior. This has two consequences. First, the prior degrees of freedom needs to be sufficiently large to ensure that the (finite) break size variance exists, i.e. $\underline{\nu}_k > n_k + 3$, with n_k the matrix diagonal's size. Actually, we need more degrees of freedom, because even if the variance exists, for small degrees of freedom, the posterior break size variance still explodes if the break probability is small and inference is complicated due to unrealistic draws. Second, if the break probability π_k is small, the break size prior is very informative, even if the prior degrees of freedom $\underline{\nu}_k$ is small. This is because if π_1 is small, the sum $\sum_{t=2}^{T} \kappa_{1t}\boldsymbol{\eta}_t\boldsymbol{\eta}_t'$ will be close to zero, and the mean of $\boldsymbol{Q}_1$ is (almost) fully determined by the prior's location $\underline{\boldsymbol{W}}_1\underline{\nu}_1$. Hence, one should be careful when interpreting the break size results in that case. In a TVP model, the degrees of freedom is often set equal to the minimum for the (finite) prior break size variance to exist. Instead of one, we add ten degrees of freedom, such that $\underline{\nu}_k = n_k + 3 + 10$, to limit the prior's informativeness, while ensuring that the break size variance does not explode if the break probability is small.

As priors for the initial conditions of the state equations we take a normal distribution,

$$\boldsymbol{\beta}_1 \sim \mathrm{N}(\boldsymbol{0}, \underline{\nu}_1 \operatorname{var}(\hat{\boldsymbol{\beta}})), \tag{3.21}$$

$$\log \boldsymbol{\sigma}_1^2 \sim \mathrm{N}(\boldsymbol{0}, \underline{\nu}_2 \operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})), \tag{3.22}$$

$$\alpha_1 \sim \mathrm{N}(0, \underline{\nu}_3 \operatorname{var}(\hat{\alpha})), \tag{3.23}$$

where $\operatorname{var}(\hat{\boldsymbol{\beta}})$, $\operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})$, and $\operatorname{var}(\hat{\alpha})$ are the variance covariance matrices of the OLS estimators of $\boldsymbol{\beta}$, $\log \boldsymbol{\sigma}^2$, and α on the first ten years of data as in Equations (3.10)–(3.12), and $\underline{\nu}_1$, $\underline{\nu}_2$, and $\underline{\nu}_3$ equal to the degrees of freedom in the prior of the break size in Equations (3.17)–(3.19). This is quite informative, and consistent with the initial value probably being approximately equal to the OLS estimate of the ten preceding years.

3.2.4 Inference

The MI model is estimated using a Gibbs sampler that mostly follows the algorithm of Koop et al. (2009), which is based on Primiceri (2005) with the addition of drawing the break processes $\boldsymbol{\kappa}_t$ and probabilities $\boldsymbol{\pi}$. We add a step to draw the autoregressive parameters $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2$ and ϕ_3. This sections provides a brief outline of the algorithm. A detailed description is in Appendix 3.B.

For each of the state variables, we rewrite the VAR such that we can apply methods for (conditionally) linear Gaussian state space models. Then, we first draw the break process using the algorithm by Gerlach et al. (2000). Second, we draw the latent variable using the simulation smoother of Durbin and Koopman (2002a). After drawing the time-varying parameters, we draw the static parameters from their full conditional posterior distributions. We iterate until we have 10,000 retained draws, after removing a burn-in sample of 2,000 draws. Increasing the number of draws does not affect results. Appendix 3.C provides a convergence analysis.

3.3 Data

We use the monthly S&P 500 index return and the one-month T-bill rate as risk-free rate to construct the excess log return. Our sample consists of 960 monthly observations from January 1936 to December 2015, of which the first ten years, January 1936 to December 1945, are used for prior calibration, see Section 3.2.3. As predictor, we take the (log) dividend price ratio, defined as the ratio of the sum of the dividends over the last 12 months to the current stock price. It is one of the most popular predictors in the literature and shown to hold predictive power (Campbell and Shiller, 1988), and has been used in most studies focusing on parameter instability. Data is from the extended Welch and Goyal (2008) data set.[13] Table 3.1 presents summary statistics of both variables, and Figure 3.1 plots the time series.

To motivate modeling the possibility of unstable parameters, we highlight two properties of the data that can be captured by the mixture innovation model. First, a stylized fact of stock returns is volatility clustering. The null hypothesis of homoskedasticity is strongly rejected by Engle's (1982) ARCH test for our sample, see the last row of Table 3.1. The possibility of time-varying volatility in Equation (3.8) captures this heteroskedasticity. Second, Lettau and Van Nieuwerburgh (2008) document the presence of one break (in 1991) or two breaks (in 1954 and 1994) in the dividend price ratio, using yearly data from 1927 to 2004. Consistent with their finding, an inspection of the dividend price ratio in Figure 3.1b suggests a level shift in the 1990s. This break in the data could disrupt the predictive relationship with the equity premium, which the MI model captures by allowing for instability in the intercepts and loadings in Equations (3.1)–(3.2).

[13] Available on Amit Goyal's website: `http://www.hec.unil.ch/agoyal/`.

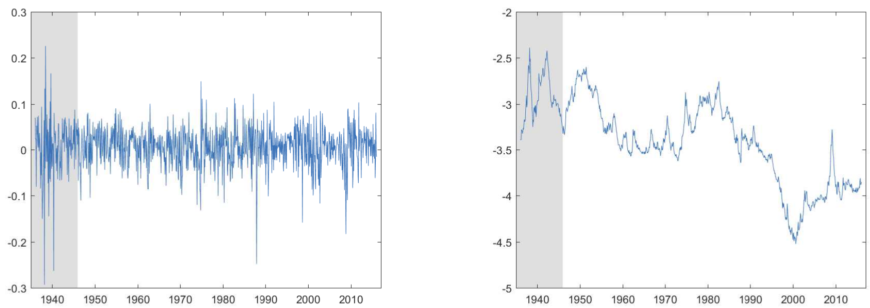
Table 3.1: Summary statistics

	Excess returns	Dividend price ratio
Mean	0.0052	−3.4719
Median	0.0087	−3.4476
Standard deviation	0.0420	0.4342
Skewness	−0.6460	−0.2592
Kurtosis	5.2180	2.4071
Minimum	−0.2482	−4.5240
Maximum	0.1492	−2.5979
AR(1)	0.0516	0.9952
AR(2)	−0.0234	0.9897
AR(3)	0.0370	0.9844
AR(12)	0.0503	0.9289
ARCH test (p-value)	0.0000	0.0000

The table presents summary statistics of the excess log return and log dividend price ratio for the period January 1946 – December 2015. AR(x) is the x-th order autocorrelation. Engle's (1982) ARCH test is based on residuals from a static VAR(1) model, estimated using maximum likelihood.

Figure 3.1: Time series of data

(a) Excess returns (b) Dividend price ratio



The figures present the monthly time series of the excess log returns (a) and the log dividend price ratio (b) for the period January 1936 – December 2015, where the shaded area (January 1936 – December 1945) is used for prior calibration.

3.4 Estimation results

This section presents the estimation results. We start with models where the break probability is known, after which we assume the break probability is unknown and estimate it.

3.4.1 Few versus many breaks

First, we estimate the mixture innovation model under a known break probability. We assume various values for the break probability $\boldsymbol{\pi}$, from no to few to many breaks. In particular, we consider $\pi_1 \in \{0, 1/600, 1/120, 1/60, 1\}$. This includes the extreme cases of a static, i.e. no breaks ($\boldsymbol{\pi} = \mathbf{0}$), and a TVP-VAR ($\boldsymbol{\pi} = \boldsymbol{\imath}$) model. The few breaks cases are motivated by previous studies in this context. Specifically, $\pi_1 = 1/120$ (one break expected every 10 years) matches the 8 to 10 breaks reported by Pettenuzzo and Timmermann (2011), and $\pi_1 = 1/600$ (one break expected every 50 years) is in line with the one to two breaks found by Lettau and Van Nieuwerburgh (2008).[14] Fixing the break probability at $\pi_1 = 1/60$ (one break expected every 5 years) is an intermediate case to explore what happens if we move from the few breaks case to the many breaks case. The break probability for the variances and covariance term is set at $\pi_2 = \pi_3 = 1$, i.e. it is modeled as stochastic volatility, to capture the volatility clustering. We do this for all values of π_1, except for $\pi_1 = 0$. For $\pi_1 = 0$, we consider a fully static model, ignoring parameter instability, so $\pi_i = 0$, for $i = 1, 2, 3$.

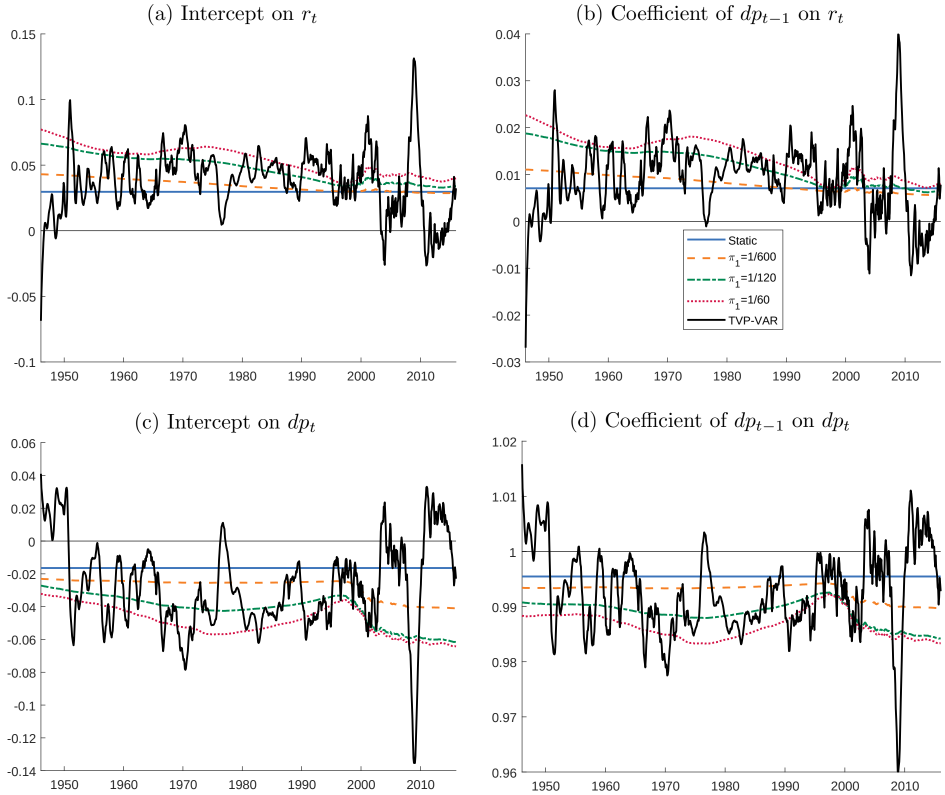
The posterior estimates of the time-varying parameters in Figures 3.1 and 3.2 are in line with previous research. First, Figure 3.1b confirms that the weak but positive predictive power of the dividend price ratio has decreased from the 1970s, and in particular small in the 1990s (see e.g. Ang and Bekaert, 2007). Second, the static model estimates the predictability coefficient in Figure 3.1b to be smaller compared to models that do allow for time-varying parameters. The conditional level of predictive power is probably underestimated by the static model, as it is a smoothed estimate over a sample with periods where predictive power fluctuates. Including the periods without predictive power shrink the estimate downwards. Third, the estimates illustrate why the dividend price ratio is often used in long-term investing literature. The high persistence (see Figure 3.1d), combined with a positive predictability coefficient in Figure 3.1b and strong negative residual correlation in Figure 3.2c implies strong mean reversion, making stocks safer in the long run and ensuring a sizable hedge term. Fourth, Figure 3.2a shows that volatility clustering is captured by all models, other than the static model. We recognize periods of high volatility, such as the Oil Crisis in the 70s, Black Monday on October 1987 and the recent financial crisis in 2008. We shall discuss the fit in terms of an information criterion in Section 3.4.2.

Comparing across the break probabilities, we notice that the results in Figure 3.1 are in line with the interpretation of π_1. As we increase the break probability, the

[14] This is an approximation of a change point model with an unknown number of breaks, but with a prior on the number of breaks, as we do not fix the number of breaks, but the break probability.

posterior estimate displays more instability. The TVP-VAR's estimate is much more volatile than for any of the other break probabilities, and suggests fitting noise rather than signal. The estimates for the other types of time-variation are largely similar to each other. If we expect one or two breaks, the intercepts and loadings are quite stable. Interestingly, Figures 3.1c and 3.1d provide some evidence of a break in the dividend price ratio at the end of the 1990s or start of 2000s, conform Lettau and Van Nieuwerburgh's (2008) finding. There is smooth time-variation when $\pi = 1/120$ and $\pi = 1/60$, indicating uncertainty regarding the location of additional breaks.

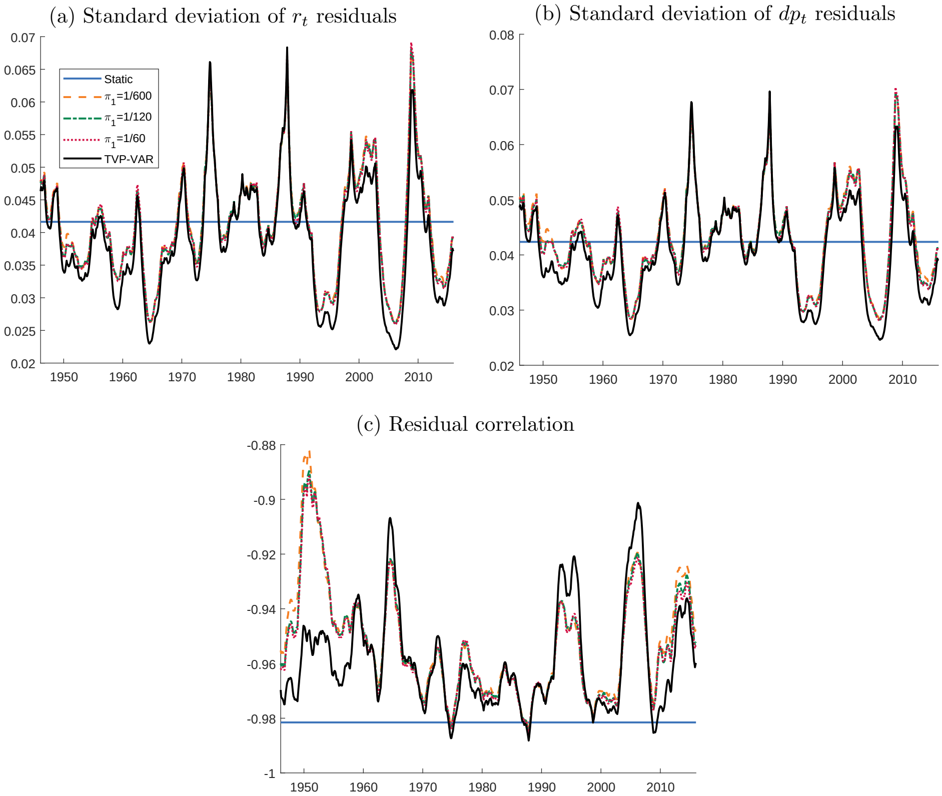
Figure 3.1: Posterior mean of intercepts and loadings under known π



(a) Intercept on r_t (b) Coefficient of dp_{t-1} on r_t (c) Intercept on dp_t (d) Coefficient of dp_{t-1} on dp_t

The figures present the posterior mean of the time-varying intercepts and loadings from the mixture innovation (MI) model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9), with break probability for intercepts and loadings fixed at $\pi_1 \in \{0, 1/600, 1/120, 1/60, 1\}$. The break probability for the variances and the covariance term is fixed at $\pi_2 = \pi_3 = 1$ for all models, except for the static case ($\pi_1 = 0$), where $\pi_2 = \pi_3 = 0$. TVP-VAR is the model with $\pi_1 = \pi_2 = \pi_3 = 1$. See Section 3.2.3 for the prior specifications.

Figure 3.2: Posterior mean of residual variance and correlation under known π

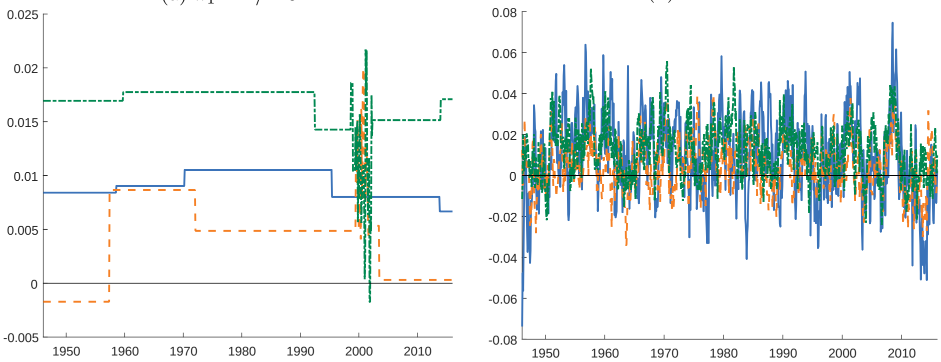


The figures present the posterior mean of the residual variance and correlation from the mixture innovation (MI) model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9), with break probability for intercepts and loadings fixed at $\pi_1 \in \{0, 1/600, 1/120, 1/60, 1\}$. The break probability for the variances and the covariance term is fixed at $\pi_2 = \pi_3 = 1$ for all models, except for the static case ($\pi_1 = 0$), where $\pi_2 = \pi_3 = 0$. TVP-VAR is the model with $\pi_1 = \pi_2 = \pi_3 = 1$. See Section 3.2.3 for the prior specifications.

To illustrate further that the different break probabilities imply different statistical properties, Figure 3.3 shows individual draws for the MI model when $\pi = 1/120$ and the TVP-VAR model. Draws from the TVP-VAR model are very noisy, with a break each period. In contrast, the draws under a small break probability exhibit large periods of no change, and relatively large breaks. Some draws display noisy periods near the end of the 1990s, perhaps because of uncertainty regarding a break around that time. These draws are representative of the individual draws that feed directly into the predictive density. Their large differences motivate us to assess the consequences for the asset allocation in Section 3.5.

Figure 3.3: Three posterior draws of predictability coefficient

(a) $\pi_1 = 1/120$ (b) TVP-VAR



The figures present three random posterior draws of β_{2t}, the coefficient of dp_{t-1} on r_t, for the mixture innovation model, with break probability fixed at $\pi_1 = 1/120$ (a), and the TVP-VAR model (b).

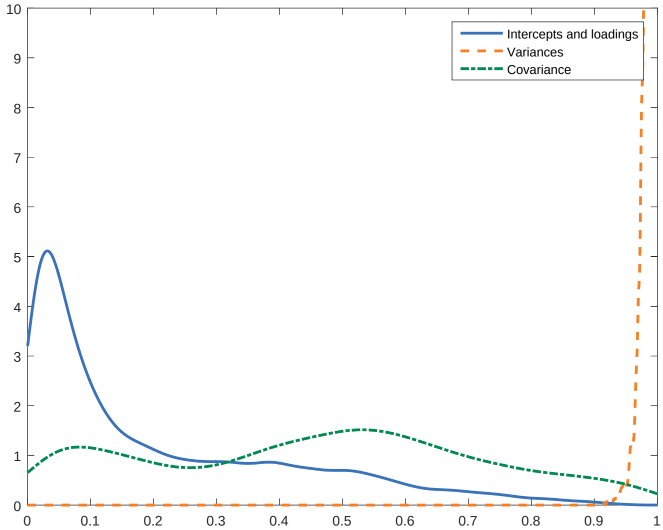
3.4.2 Estimating the parameter instability

In reality, the break probability is unknown, emphasized by the lack of consensus in the literature. Therefore, we specify the prior for $\boldsymbol{\pi}$ as in Section 3.2.3, and let the data speak on the type of instability, and the uncertainty of it.

Figure 3.4 presents the posterior break probability. Most probability mass is near $\pi_1 = 0$, and the mode is smaller than 0.05. The posterior mean of the break probability of the coefficients' break process is 0.206, such that the expected duration between breaks is about 5 months. This is much larger than the expected prior value of 0.017, and implies that the data suggests a substantial number of breaks. At the same time, the density is quite wide, with a standard deviation of 0.211. It is substantially larger than the prior standard deviation of 0.016. This indicates quite some uncertainty regarding the break probability, likely due to the data's low signal-to-noise ratio.

One important note is that the break probability might be skewed upward due to the parameterization of the mixture innovation model. Because we model the break probability, relatively much of the support of π is associated with a large number of breaks, and even extreme priors put some weight on this part. For example, our prior with hyperparameters $\underline{a}_1 = 0.1$ and $\underline{b}_1 = 59.9$ implies an expected break probability of only $1/600$. Still, there is a non-negligible 2.4% probability that π_1 is larger than $1/60$, equivalent to one break every five years on average. Furthermore, this is the break probability, not the number of breaks. It is not possible to set the prior to exactly one or two breaks. This might be less so if we estimate a change point model, where we would estimate the number of breaks, and it is easier to allocate zero probability mass to the many breaks cases. However, this would not provide us with the direct measure for instability, and an idea of the uncertainty regarding the instability.[15]

Figure 3.4: Posterior of break probability



The figure presents the posterior densities of the break probabilities from the mixture innovation model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9), where the break probability is assumed to be unknown. The hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_1 = 1$ and $\underline{b}_1 = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. See Section 3.2.3 for the other prior specifications.

The posterior mean of the time-varying parameters is presented in Figures 3.5 and 3.6. There is substantial time-variation in the intercepts and loadings according to the MI model, more than implied by the few breaks case, i.e. $\pi_1 = 1/120$ or $\pi_1 = 1/600$, see Figure 3.1. It is more stable than suggested by the TVP-VAR model. Although the posterior mean seems like a smoother version of the TVP-VAR model's, the individual draws in Figure 3.7 highlight the difference between the models. Some draws resemble those of the TVP-VAR model, whereas other draws display periods of stability, in line

[15] An alternative to reduce the upward skew is to use a threshold mixture innovation model, proposed by Huber et al. (2017). The model combines the MI model with the threshold model (Nakajima and West, 2013). A threshold needs to be exceeded for a break to occur, shrinking small breaks to zero. Applying the MI model without a threshold is more natural in our application, because we are also interested in the many (small) breaks case.

with draws from a model with a small break probability. The variety between the draws is a reflection of the wide posterior break probability distribution in Figure 3.4.

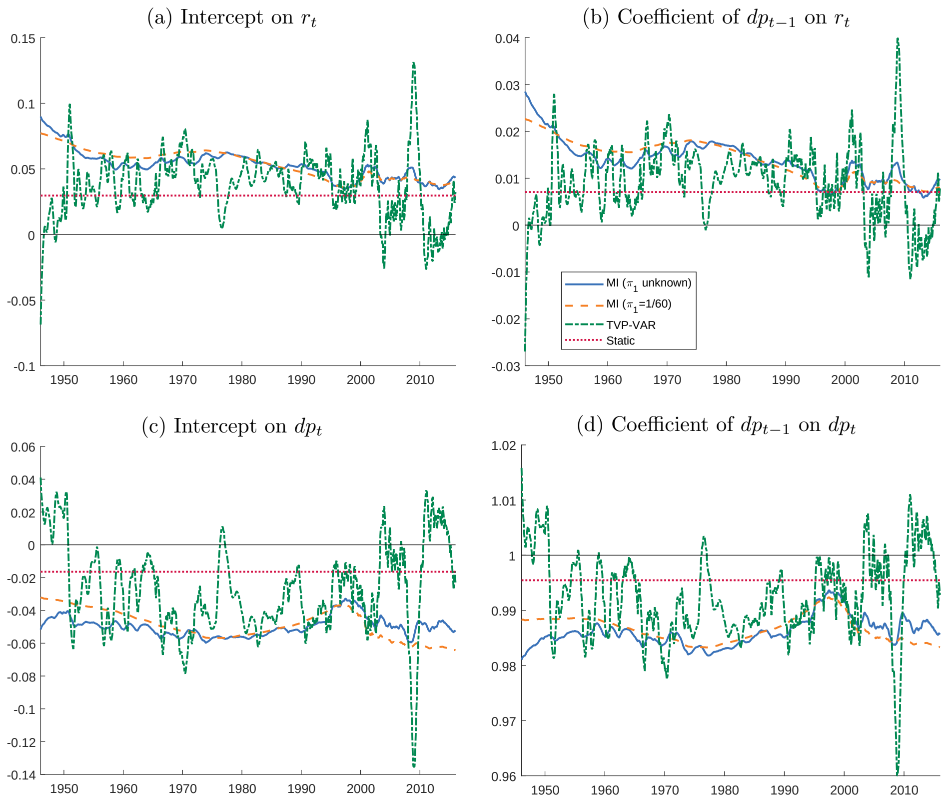
The combination of only few large breaks and a wide posterior distribution for the break probability suggests that it is hard to identify the break probability with high certainty, probably due to the low signal-to-noise ratio. This highlights the difficulty in predicting stock returns.

The variance for the MI model with unknown break probability in Figure 3.6 is practically the same as for the TVP-VAR model and for the MI model with fixed break probability. The MI model confirms the presence of heteroskedasticity, as the posterior of the break probability in Figure 3.4 strongly suggests stochastic volatility rather than a few regime switches. Johannes et al. (2014) also find that including stochastic volatility is important. To investigate this, we include a model, labeled MI-b, with the intercepts and loadings restricted to be constant.

We compare the fit of the models using the Watanabe Information Criterion (WAIC, Watanabe, 2010), a Bayesian measure for the fit that penalizes the number of parameters. A smaller WAIC value indicates a better model fit. The estimated effective number of parameters varies depending on the time-variation in the model. It is calculated as the variance of the observation level loglikelihood over the posterior draws, summed over all observations. This is known as WAIC type 2, which is quite stable and has the attractive property of being related to leave-one-out cross-validation (Gelman et al., 2014).

Table 3.1 shows the following results. First, allowing for time-varying volatility indeed improves the model fit, because the WAIC is lower for MI-b than the static model. Second, the model with unknown break probability provides a slightly better fit over the models with fixed small break probability. Given the improved fit if the break probability increases, it suggests that, according to the WAIC, the few breaks case underestimates the break probability. Third, the TVP-VAR model provides the best fit. It is somewhat surprising that the estimated number of parameters is larger for the MI model than for the TVP-VAR model, even though the parameters in the TVP-VAR model are more volatile. This might be due to uncertainty of the break locations in the MI model. Therefore, we interpret this result with some caution.

Figure 3.5: Posterior mean of intercepts and loadings under unknown π

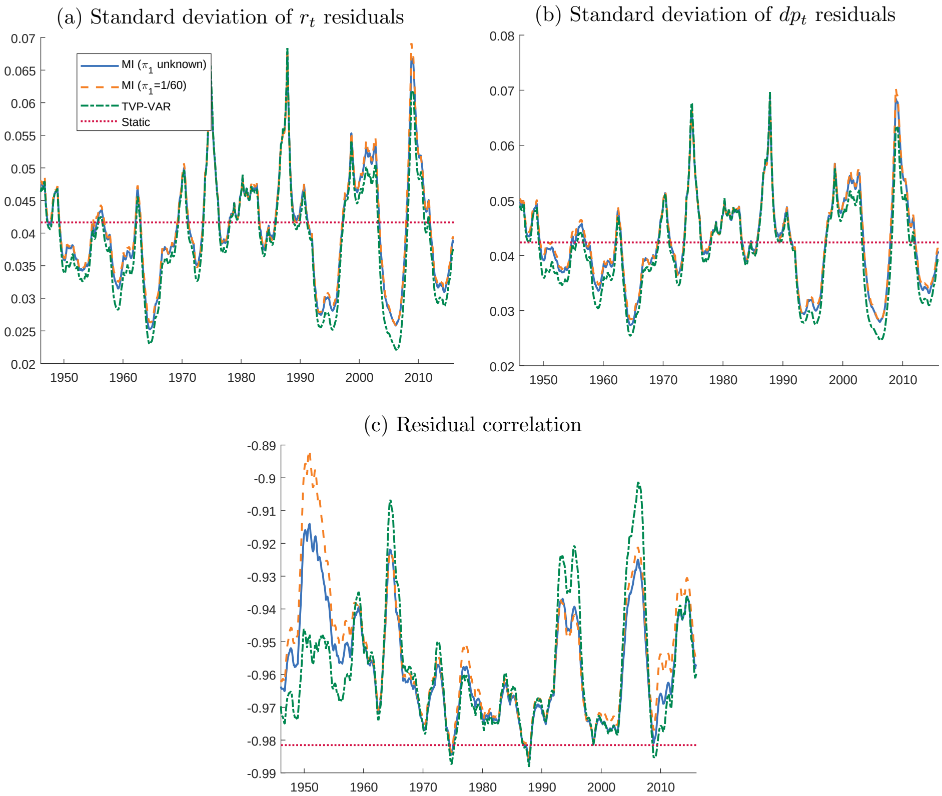


The figures present the posterior mean of the time-varying intercepts and loadings from the mixture innovation (MI) model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9). If π_1 is assumed unknown, the hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_1 = 1$ and $\underline{b}_1 = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. If the break probability is assumed known, it is fixed at the prior mean, such that $\pi_1 = 1/60$ and $\pi_2 = \pi_3 = 1$. TVP-VAR is the model with $\pi_1 = \pi_2 = \pi_3 = 1$, and static is the model where $\pi_1 = \pi_2 = \pi_3 = 0$. See Section 3.2.3 for the other prior specifications.

3.5 Economic evaluation

In the spirit of Barberis (2000) we analyze the influence of the break probability on the portfolio allocation of a long-term investor. This provides insight into whether the statistical differences in the previous sections translate to economic differences in terms of portfolio risk and returns. Instability in the relationship between predictors and the excess returns is arguably more important to the long-term investor than for the short-term investor. Even if the break probability is small, the probability of a break occurring in the period of holding the portfolio can still be high. For example,

Figure 3.6: Posterior mean of residual variance and correlation under unknown π

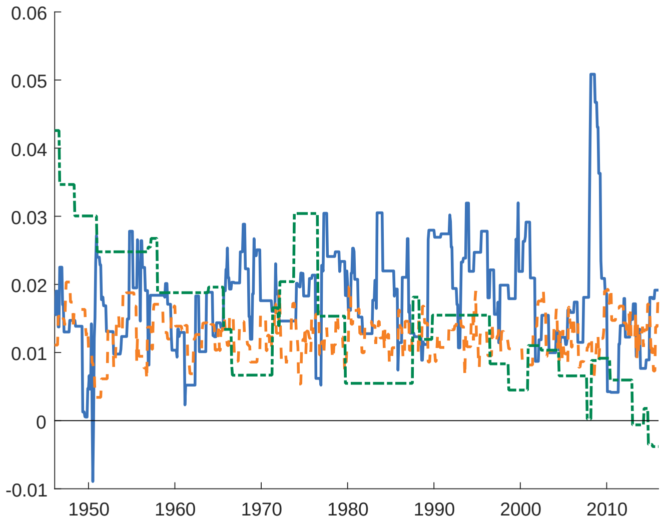


The figures present the posterior mean of the residual variance and correlation from the mixture innovation model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9). If π_1 is assumed unknown, the hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_1 = 1$ and $\underline{b}_1 = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. If the break probability is assumed known, it is fixed at the prior mean, such that $\pi_1 = 1/60$ and $\pi_2 = \pi_3 = 1$. TVP-VAR is the model with $\pi_1 = \pi_2 = \pi_3 = 1$, and static is the model where $\pi_1 = \pi_2 = \pi_3 = 0$. See Section 3.2.3 for the other prior specifications.

if the break probability is 0.05, the break risk for the one period investor is reasonably small. For an investor holding a portfolio for 120 periods, i.e. ten years, the expected number of breaks in this period is 5 and this risk can affect the allocation to stocks.

First, we introduce the asset allocation problem. Second, we discuss the term structure of risk, which is the annualized per period standard deviation of the predictive density, and third, we evaluate the economic performance of the models.

Figure 3.7: Three posterior draws of predictability coefficient under unknown π



The figure presents three random posterior draws of β_{2t}, the coefficient of dp on r, from the mixture innovation model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9), where the break probability is assumed to be unknown. The hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_1 = 1$ and $\underline{b}_1 = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. See Section 3.2.3 for the other prior specifications..

Table 3.1: In-sample fit

Model	Loglikelihood (mean)	$\hat{k}$	WAIC
MI (π unknown)	4,376.8	64.6	$-8{,}682.4$
TVP-VAR	4,393.1	48.5	$-8{,}731.9$
MI ($\pi_1 = 1/60$)	4,351.7	61.7	$-8{,}635.4$
MI ($\pi_1 = 1/120$)	4,341.6	61.9	$-8{,}615.2$
MI ($\pi_1 = 1/600$)	4,322.5	58.5	$-8{,}580.2$
MI-b ($\pi_1 = 0$)	4,315.6	50.0	$-8{,}575.3$
Static	4,288.6	5.4	$-8{,}571.7$

The table presents the loglikelihood mean over the posterior draws, the effective number of parameters $\hat{k}$, and the Watanabe-Akaike Information Criterion (WAIC) for the MI model presented in Section 3.2 with different prior break probabilities. MI is the mixture innovation model, TVP-VAR is the time-varying parameter VAR model, static is no time-variation and the addition of -b means that the intercepts and loadings are restricted to be constant. The number of observations is 1,680.

3.5.1 Asset allocation problem

We consider a power utility (constant relative risk aversion) maximizing buy-and-hold investor with an investment horizon h up to 240 months, i.e. 20 years. The investor allocates his wealth now, and receives the return at the end of the investment period, without rebalancing in between. (S)he faces the following problem,

$$\max_w U(W_{T+h}) = \max_w \frac{W_{T+h}^{1-\gamma} - 1}{1-\gamma}, \tag{3.1}$$

with risk aversion $\gamma > 0$ and wealth at time $T + h$ is defined as

$$W_{T+h} = W_T \left(w \exp\left(\sum_{t=T+1}^{T+h} (r_t + r_f) \right) + (1-w) \exp\left(\sum_{t=T+1}^{T+h} r_f \right) \right) \tag{3.2}$$

$$= W_T \left(w \exp\left(hr_f + \sum_{t=T+1}^{T+h} r_t \right) + (1-w) \exp(hr_f) \right), \tag{3.3}$$

where w is the fraction of wealth allocated to stocks and r_f is the log risk-free rate, which we set to the historical average of the one-month T-bill rate. Wealth at time T is normalized to one. The investor optimizes the utility by changing the allocation to stocks w, with short selling constraints, such that $0 \leq w \leq 0.99$.[16]

To solve the optimization problem in Equation (3.1), the investor needs to specify the distribution of excess stock returns h periods ahead. We use draws from the posterior to sample from the predictive density of stock returns. Hence, the predictive density incorporates estimation uncertainty, parameter instability, and uncertainty regarding the risk of breaks. Excess returns are simulated up to the investment horizon h, conditional on the model, a time period, and the posterior draws from the m-th MCMC iteration. This is repeated ten times per posterior draw, to increase precision on the moments of the predictive density. Using these draws, we compute the expected utility for a given weight w. The optimal weight is then found by a one dimensional adaptive grid search over w, where we decrease the step size to a minimum of 10^{-5}.

The maximum horizon h is 240 months, so we want to make sure that the predictive density is stationary. Otherwise, the mean and the variance will be inflated, due to a few extreme draws, and vary from one simulation to the other. It would be better to impose this in the estimation procedure, but this would make the estimation procedure substantially more difficult, as we need to impose this every period. Then, we cannot easily rewrite the model to the Gaussian state space model and the state space methods we employ are no longer valid. An alternative is to perform a Metropolis-Hastings step,

[16]If $w = 1$, the utility is unbounded, see e.g. Barberis (2000).

introducing more autocorrelation to the MCMC chain, reducing the sampler's efficiency. Instead, we opt to impose stationarity by discarding the non-stationary draws, when drawing from the predictive density. This implies imposing that $\beta_{4t} < 0.999$ and applying the same restriction to the autoregressive parameters on the diagonal of $\boldsymbol{\Phi}$.[17]

The portfolio is evaluated using the certainty equivalent return (CER). It is the return that makes the investor indifferent between holding a riskless asset with that return and investing in the (risky) portfolio. In other words, CER is the return the investor is willing to pay to hold the specific portfolio instead of being fully invested in the risk free asset. It is the rate that yields the same utility as the portfolio of model i, calculated as $\text{CER}_{ih} = (1/h)\big((U_{ih}(1-\gamma)+1)^{\frac{1}{1-\gamma}} - 1\big)$, with utility U_{ih} and investment horizon h.

We do not evaluate the portfolios out-of-sample for two reasons.[18] First, evaluating long-term portfolios out-of-sample is difficult because the sample consists of few non-overlapping periods. Second, an out-of-sample analysis is time-consuming. It requires re-estimating the model each period, because we use smoothed estimates. For example Johannes et al. (2014) use a filter, which alleviates this issue.

Instead, we opt for a different approach. We are interested in the performance of each model, under various types of parameter stability. We assume one model to be the data generating process (DGP) and use that as the predictive density. We take the weights from each of the models, and calculate the CER under the assumed DGP. This is similar to the analysis of Ang and Bekaert (2002) and Pettenuzzo and Timmermann (2011), among others. The model that we assumed to be the DGP will always have the highest CER, because we optimized the weights over it for those draws. Therefore, we repeat the exercise, assuming each of the models as DGP. Finally, we can evaluate the loss compared to the optimal CER (for correctly specified time-variation) as $\Delta\text{CER}_{ijh} = \text{CER}_{jjh} - \text{CER}_{ijh}$, the loss for model i at horizon h with model j as the true model. This provides insight into the costs of misspecifying the break process.

[17] We impose $\beta_{4t} < 0.999$ instead of $\beta_{4t} < 1$ to avoid numerical non-stationarity.

[18] See Diris et al. (2014) for an example on how to do this.

3.5.2 Term structure of risk

Figure 3.1 shows the term structure of risk for the MI (with unknown break probability and fixed probability at the expected value of the prior), TVP-VAR, a static VAR model, and MI-b, the MI model where the intercepts and loadings are restricted to be constant. The term structure of risk is the annualized per period standard deviation, such that we can compare results across investment horizons. We conditioned on 1987M11, one month after Black Monday, a period with relatively high conditional variance.[19] This explains the relatively high short-run variance for the MI and TVP-VAR models. In the medium run, the per period standard deviation decreases due to predictability and mean reversion. This is strongest for MI-b and the constant VAR due to stable intercepts and loadings. This also indicates that even though the MI-b includes time-varying volatility, the effect at the long horizons is mostly driven by the (in)stability of the mean.

In the long run, the MI model displays the largest risk. We have three possible explanations for this. First, there is added uncertainty regarding the break probability. The MI model with unknown break probability has a larger variance than the MI model with known break probability. This also translates to wider posterior distributions for the AR parameters and long run mean. This alone does not explain the difference, as the MI model with known break probability also has higher variance than the TVP-VAR.

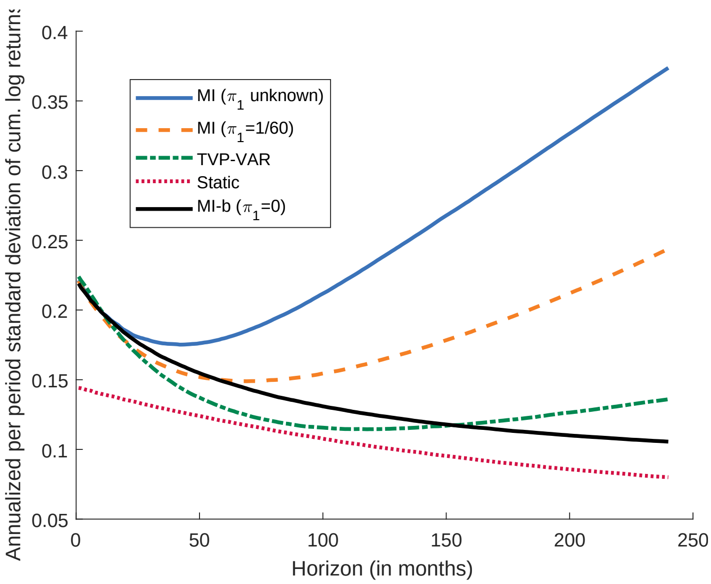
Second, there might be stronger mean-reversion in the time-varying parameters for the TVP-VAR model than the MI model. As shown in Figure 3.7, there are some posterior draws where the intercepts and loadings are constant for longer periods. If the parameters are in an extreme regime, it does not revert back to its mean, which can inflate the long run variance.

Third, there is a difference in the effect of predictability of the dividend price ratio. Figure 3.2 shows the effect of imposing stationarity. The parameters in the multivariate normal distribution are truncated. The high correlation between the coefficients implies that by truncating the AR coefficient for the dividend price ratio, we 'softly truncate'

[19] The long-run results are similar for different conditioning periods. For brevity, we only present the results of 1987M11 here, where we select this period to highlight possible differences from time-varying volatility.

the coefficient of the dividend price ratio on the excess returns, hence increasing the average predictability coefficient. If the break probability is high, the dip in the variance is larger, because each period with a break nudges the predictability upwards. This is in favor of the TVP-VAR model and partly explains the smaller per period annualized standard deviation for the TVP-VAR model compared to the MI model.

Figure 3.1: Term structure of risk



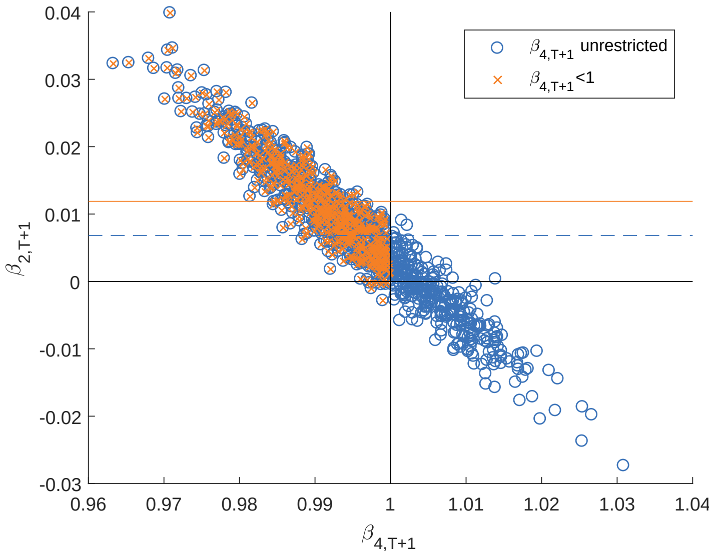
The figure presents the annualized per period standard deviation of the predictive density of stock returns, for an investment horizon up to $h = 240$ months. Results are conditional on time period 1987M11.

3.5.3 Weights and certainty equivalent return

Figure 3.3 presents the fraction of wealth allocated to stocks from optimizing the asset allocation problem in Equation (3.1) with risk aversion of $\gamma = 5$.[20] The allocation can be extreme, with the allocations based on the static VAR and MI-b hitting boundary solutions. This is because of the relatively low standard deviation at the long horizons. When using the static VAR model to optimize the allocation problem, the allocation is at least 70% and quickly rises to 100% for longer horizons. This is a result of the relatively small estimated risk of the returns for those models. When comparing the MI and TVP-VAR, we see that at the longer horizon, MI has a smaller allocation than

[20]Results are qualitatively similar for a risk aversion of 2 and 8.

Figure 3.2: Effect of imposing stationarity



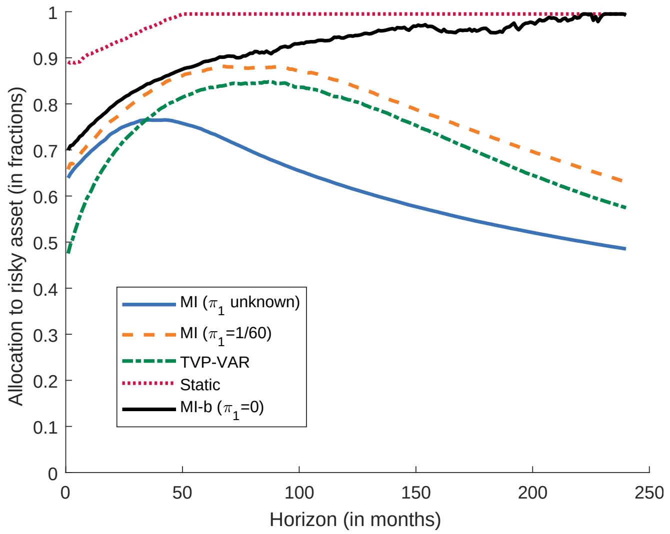
The figures present a 1,000 draws from the predictive density for the intercepts and loadings. The blue circles are from the multivariate normal distribution given by the posterior mean and variance for $\boldsymbol{\mu}_1$ (from the static model) and $\boldsymbol{Q}_1$ (from the TVP-VAR model). The orange crosses are in the subset where $|\beta_{4,T+1}| < 1$. The dashed blue line is at the height of the sample mean of $\beta_{2,T+1}$ for the full sample, and the orange solid line is at the height of the sample mean of $\beta_{2,T+1}$ for the subset where $|\beta_{4,T+1}| < 1$.

TVP-VAR. This is due to the large risk at the longer horizon according to the MI model.

The weights show that when allowing for unstable parameters, stocks are less interesting for the (long-term) investor than when using a static VAR. However, it does not tell how much we would lose if we were to assume an incorrect instability of the relationship between the dividend price ratio and excess stock returns. Therefore, we consider the difference in CERs, presented in Table 3.1, constructed as explained before. The CERs are annualized and therefore comparable across horizons. The exact values should be taken with a grain of salt, as they are based on assuming one of the models as a DGP. However, the DGPs are based on actual data, and the values indicate where and when it pays to assume instabilities.

The CERs provide some interesting results. First, assuming constant intercepts and loadings can be costly, especially at the longer horizons. At the short horizon, the losses are limited, but increase as the horizon becomes 5 year or more, up to 17% in CER at the 20 year horizon, see panel D in Table 3.1. Second, vice versa, the MI

Figure 3.3: Allocation to stocks (conditioning period is 1987M11)



The figure presents the fraction of wealth allocated to stocks for a buy-and-hold investor with power utility with risk aversion parameter γ is 5. Results are conditional on time period 1987M11.

models and TVP-VAR show a large loss in CER when the DGP has a constant mean. This is because the per period standard deviation is relatively small in the DGP in that case and the investor misses these extra profits. However, it is not that bad for the investor because the absolute CER is still positive. Third, it is important to take parameter stability into account, but the difference between the MI model and the TVP-VAR is limited (0.2%). Fourth, the influence of time-varying volatility is clear at the short horizon. The difference in CER in panel A in Table 3.1 for the static model is much larger than for any other model, even the MI-b. At the long horizon though, instability in the mean seems to dominate, as the allocation and hence CERs for the MI-b and static model are equivalent. Fifth, from assuming a DGP with instability uncertainty and considering the mixture ΔCER for the mixture innovation models without instability uncertainty, we find that the utility costs of instability uncertainty is about 0.7% in annualized returns, which may be small but non-negligible.

Finally, from the bottom row in Table 3.1, we can identify the model choice that is most robust to the time-variation misspecification, consistent with a min-max utility investor (Hansen and Sargent, 2001). It is the model that maximizes the minimum CER over the various model DGPs, where the consideration set is defined by the mixture innovation model we specified, with the break probability prior as choice set.[21] We find that assuming a TVP-VAR is most beneficial in the long run for the long-term investor, although a reasonable alternative would be to allow for any type of instability, with a preference for a high break probability.[22] This might be because even though the TVP-VAR might adjust more quickly to breaks, and providing a suitable estimate for the size of the instability. In the long run, this is perhaps more important than the type of instability.

3.6 Prior sensitivity analysis

This section discusses the sensitivity of estimates to the break probability prior. The hyperparameters for the different priors are chosen such that we have the same level of informativeness as the baseline prior ($\underline{a}_1 = 1$ and $\underline{b}_1 = 59$), but match the mean to each of the considered break probabilities in Section 3.4.1. For example, to get an expected value of 1/120, we take $\underline{a}_1 = 0.5$ and $\underline{b}_1 = 59.5$.

Figure 3.1a shows the differences in the prior distribution, and how more probability mass is pushed towards zero as a gets smaller compared to b, and implies more stable parameters. When we consider the posterior break probability for the intercepts and loadings in Figure 3.1b, we notice some differences. First, the modes of the models are ordered as we would expect from with the smallest value for very few and the highest for the many breaks prior. Because the duration of a regime is inversely related to the break probability, especially differences in break probability close to zero can translate to large differences in regime durations. Second, the many breaks case ($E[\pi] = 1/12$) has fewer probability mass in the 0.2 to 0.5 interval than the other models and very little near zero. The posterior for the other models is slightly wider, including both very stable draws, and more volatile draws. The posterior densities are all quite wide, and reflect the uncertainty regarding the break probability. Figure 3.2 shows that the general pattern of the possibly time-varying parameters is quite robust to the

[21] Alternatively, we could minimize over the maximum ΔCER, but then investor would only be interested in the utility relative to the maximum, even though this could imply negative CERs.

[22] The minimum CER is the same as $\pi_1 = 1/60$ in Table 3.1. When considering the next digit, TVP-VAR's minimum CER is slightly higher: 0.759 versus 0.755.

Table 3.1: Difference in certainty equivalent return

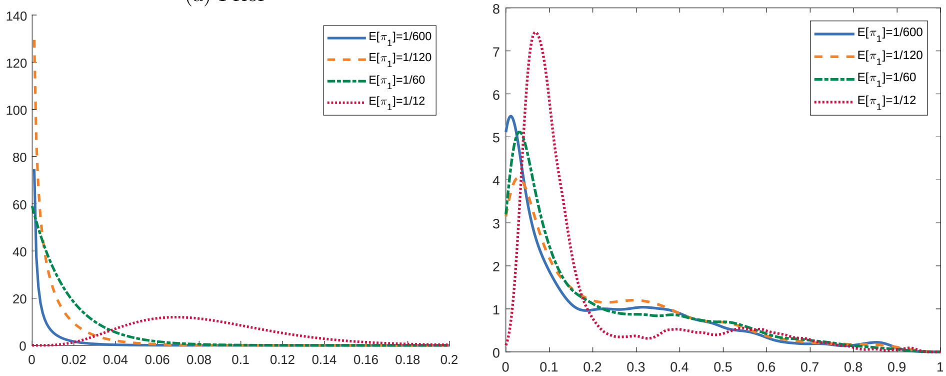
		ΔCER						
DGP	CER_{DGP}	MI π_1 unknown	TVP-VAR	MI $\pi_1 = 1/60$	MI $\pi_1 = 1/120$	MI $\pi_1 = 1/600$	MI -b ($\pi_1 = 0$)	Static
				Panel A: $h = 1$				
MI (π_1 unknown)	0.062		0.004	0.000	0.000	0.000	0.001	0.018
TVP-VAR	0.048	0.004		0.005	0.007	0.007	0.009	0.040
MI ($\pi_1 = 1/60$)	0.064	0.000	0.005		0.000	0.000	0.000	0.016
MI ($\pi_1 = 1/120$)	0.067	0.000	0.007	0.000		0.000	0.000	0.014
MI ($\pi_1 = 1/600$)	0.068	0.000	0.007	0.000	0.000		0.000	0.013
MI-b ($\pi_1 = 0$)	0.070	0.001	0.008	0.000	0.000	0.000		0.011
Static	0.074	0.008	0.017	0.007	0.006	0.006	0.005	
				Panel B: $h = 12$				
MI (π_1 unknown)	0.067		0.001	0.000	0.000	0.000	0.001	0.012
TVP-VAR	0.057	0.001		0.002	0.002	0.003	0.003	0.019
MI ($\pi_1 = 1/60$)	0.069	0.000	0.002		0.000	0.000	0.000	0.009
MI ($\pi_1 = 1/120$)	0.070	0.000	0.002	0.000		0.000	0.000	0.009
MI ($\pi_1 = 1/600$)	0.073	0.000	0.003	0.000	0.000		0.000	0.008
MI-b ($\pi_1 = 0$)	0.076	0.001	0.003	0.000	0.000	0.000		0.007
Static	0.078	0.006	0.010	0.005	0.005	0.004	0.003	
				Panel C: $h = 60$				
MI (π_1 unknown)	0.078		0.003	0.008	0.007	0.009	0.011	0.150
TVP-VAR	0.073	0.001		0.000	0.000	0.000	0.001	0.007
MI ($\pi_1 = 1/60$)	0.086	0.003	0.000		0.000	0.000	0.000	0.005
MI ($\pi_1 = 1/120$)	0.088	0.003	0.000	0.000		0.000	0.000	0.008
MI ($\pi_1 = 1/600$)	0.095	0.005	0.001	0.000	0.000		0.000	0.008
MI-b ($\pi_1 = 0$)	0.101	0.006	0.001	0.000	0.000	0.000		0.006
Static	0.099	0.012	0.006	0.004	0.004	0.003	0.003	
				Panel D: $h = 240$				
MI (π_1 unknown)	0.093		0.002	0.007	0.009	0.016	0.136	0.138
TVP-VAR	0.076	0.001		0.000	0.001	0.002	0.075	0.080
MI ($\pi_1 = 1/60$)	0.104	0.005	0.001		0.000	0.001	0.135	0.141
MI ($\pi_1 = 1/120$)	0.106	0.006	0.001	0.000		0.001	0.139	0.143
MI ($\pi_1 = 1/600$)	0.128	0.017	0.007	0.002	0.001		0.162	0.167
MI-b ($\pi_1 = 0$)	0.233	0.103	0.083	0.070	0.066	0.054		0.000
Static	0.251	0.128	0.109	0.096	0.092	0.080	0.001	
min(CER)		0.075	0.076	0.076	0.075	0.074	−0.043	−0.045

The table presents the difference in annualized certainty equivalent return (ΔCER, in fractions) between the between assuming a model to choose portfolio weights (columns) and the DGP (rows), and the CER for the DGP, for investment horizon $h \in \{1, 12, 60, 240\}$ in months. Panel D also presents the minimum CER at the 20 year horizon over the DGPs for each of the models. All CERs are for a buy-and-hold investor with power utility and risk aversion of $\gamma = 5$. The conditioning period is 1987M11. Results are based on 100,000 draws.

prior values, and the variance in Figure 3.2b is basically the same for all models, and time-variation in the intercepts and loadings is of a similar magnitude.

Figure 3.1: Distribution of π_1

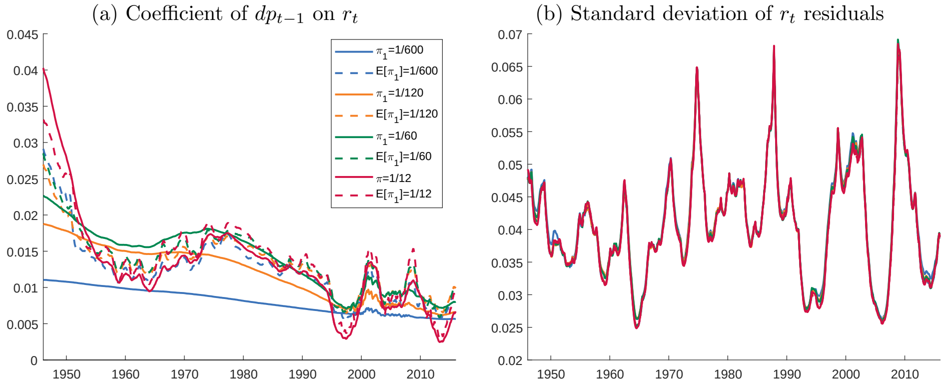
(a) Prior (b) Posterior



The figures present the prior and posterior density for the break probability of the intercepts and loadings for mixture innovation models with prior expected values for the break probability equal to $\mathrm{E}[\pi_1] \in \{1/600, 1/120, 1/60, 1/12\}$, with hyperparameters $\underline{a}_1 + \underline{b}_1 = 60$. The hyperparameters for the break probability of the variances and the covariance term are set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$.

Figure 3.2: Posterior mean of time-varying parameters – prior sensitivity

(a) Coefficient of dp_{t-1} on r_t (b) Standard deviation of r_t residuals



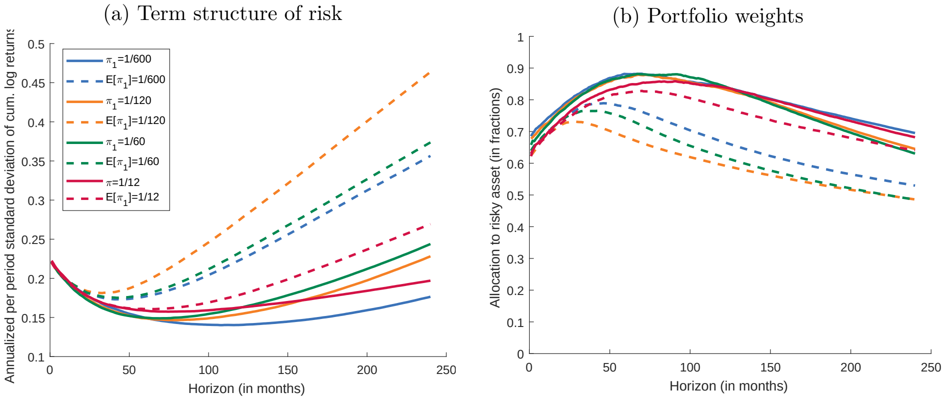
The figures present the posterior mean the predictability coefficient and the residual standard deviation of r from the mixture innovation model in Equations (3.1)–(3.4) with the time-varying parameters described by Equations (3.6)–(3.9). If π_1 is assumed unknown, the priors are set such that the expected value equal is to $\mathrm{E}[\pi_1] \in \{1/600, 1/120, 1/60, 1/12\}$, with hyperparameters $\underline{a}_1 + \underline{b}_1 = 60$. The hyperparameters for the break probability of the variances and the covariance term set at $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. If the break probability is assumed known, it is fixed at the prior mean, such that $\pi_1 \in \{1/600, 1/120, 1/60, 1/12\}$ and $\pi_2 = \pi_3 = 1$. See Section 3.2.3 for the other prior specifications.

The priors can alternatively be interpreted as the investor's views on the stability of the relationship between the dividend price ratio and stock returns. Here, we analyze the consequences of possibly incorrect views regarding the instability. The term structure of risk and the allocation in Figure 3.3 show that the per period standard deviation varies depending on the break probability. It seems a non-linear relationship. The variance at $h = 240$ is small for no breaks, then quickly increases for very few breaks. Then, for some value of $\mathrm{E}[\pi_1]$, the standard deviation decreases. to the TVP-VAR, still higher than the no breaks case. The break size is very similar, so this is not the reason for the difference. There are probably multiple effects driving this. First, a higher break probability implies a higher risk of breaks and a higher per period standard deviation. Second, mean-reversion of the time-varying parameters limits the effects of breaks, if they are stationary. Although the ARMI process has the same long run variance as the AR model, this is not true in finite samples. For the few breaks cases, there are periods where there is no break and can stay in an extreme regime. The process can mean-revert more quickly for the TVP-VAR or many breaks case as there are multiple breaks.

The effect of instability uncertainty is consistent across different priors. The extra layer of uncertainty widens the predictive density in Figure 3.3a, and Figure 3.3b shows that this extra risk induces a lower allocation to stocks. The uncertainty accumulates over the horizon, as the differences in Figure 3.3 are larger as the horizon increases. Utility costs of ignoring break probability uncertainty for an investor with a 20 year horizon vary from 0.1% for a prior at many breaks, to 1.2% for lower break probabilities, see the final column in Table 3.1. If the investor does take into account break probability uncertainty, the utility costs of misspecifying the prior are limited. The difference in CER in Table 3.1 is usually around 0.0% to 0.1%, and at most 0.8%.

Figure 3.3: Asset allocation – prior sensitivity

(a) Term structure of risk (b) Portfolio weights



The figures present the annualized per period standard deviation of the predictive density of stock returns (a), and the allocation to stocks for a buy-and-hold investor with power utility with risk aversion parameter γ is 5 (b), for an investment horizon up to $h = 240$ months. Results are conditional on time period 1987M11.

Table 3.1: Difference in certainty equivalent return – prior sensitivity

	ΔCER				
	$\mathrm{E}[\pi_1] = 1/600$	$\mathrm{E}[\pi_1] = 1/120$	$\mathrm{E}[\pi_1] = 1/60$	$\mathrm{E}[\pi_1] = 1/12$	$\pi_1 = \mathrm{E}[\pi_1]$
$\mathrm{E}[\pi_1] = 1/600$		0.001	0.001	0.005	0.012
$\mathrm{E}[\pi_1] = 1/120$	0.001		0.000	0.008	0.009
$\mathrm{E}[\pi_1] = 1/60$	0.001	0.000		0.008	0.007
$\mathrm{E}[\pi_1] = 1/12$	0.003	0.006	0.006		0.001

The table presents the difference in annualized certainty equivalent return (ΔCER, in fractions) between assuming a mixture innovation model to choose portfolio weights (columns) and the DGP (rows), for an investment horizon of $h = 240$ months. The final column presents the difference in CER from assuming a mixture innovation model with known break probability equal to the expected value of the DGP. All CERs are for a buy-and-hold investor with power utility and risk aversion of $\gamma = 5$. The conditioning period is 1987M11.

3.7 Conclusion

We assess the economic costs of misspecifying the type of instability, few (large) or many (small) breaks, in the predictive relationship between the dividend price ratio and stock returns for a long-term investor. We use a mixture innovation model with Bayesian estimation methodology to quantify the effect of uncertainty regarding the break probability.

We find the following results, which are robust to prior specification. First, from estimating the break probability, the instability seems to be subject to many breaks rather than few breaks. At the same time, there is substantial uncertainty on this, probably due to the low signal-to-noise ratio, which shows the challenge in modeling stock returns. The uncertainty adds to the volatility in the predictive density. Second, we show the substantial differences between the different types of instability. These differences affect the predictive density and lead to different asset allocations. Third, from a long-term investor's perspective, the costs of ignoring parameter instability are costly, even if the true process is subject to a limited number of breaks. The costs can run up to about 17% in terms of annualized certainty equivalent return. Conditional on assuming instabilities, the costs of misspecifying the instability are limited. Fourth, the costs of ignoring break probability uncertainty are more modest, but can still be up to 1.2% in certainty equivalent return. These costs are higher if a small break probability is assumed.

Based on the results, we recommend long-term investors to be careful when constructing their view on possible parameter instability, or at least allow for the possibility of it. In fact, a max-min investor would prefer a specification with a relatively high break probability.

Straightforward extensions would be to consider other predictors to see if the results generalize, expand the number of assets, or combine with model uncertainty for example. More interesting would be to let the break probability depend on observable factors such as industrial production or a financial stress indicator. This would make the break probability time-varying and in line with the evidence of countercyclical predictability.

3.A ARMI process properties

Define the autoregressive mixture innovation (ARMI) process X_t as

$$\begin{aligned} X_t &= \mu + f(\kappa_t, \phi)(X_{t-1} - \mu) + \kappa_t \eta_t \\ &= \mu + \big(\phi\kappa_t + (1-\kappa_t)\big)(X_{t-1} - \mu) + \kappa_t \eta_t, \end{aligned}$$

with $\eta_t \sim \mathrm{N}(0, \sigma^2)$ and $\Pr[\kappa_t = 1] = \pi$.

3.A.1 Stationarity conditions

First, we show that the process X_t is stationary if $|\phi| < 1$. Without loss of generality, we take $\mu = 0$ for notational convenience. We rewrite X_t as a sum of the initial condition and its innovations,

$$\begin{aligned} X_t &= X_0 \prod_{s=1}^{t} (\phi\kappa_s + (1-\kappa_s)) + \sum_{s=1}^{t} [\prod_{j=s+1}^{t} (\phi\kappa_j + (1-\kappa_j))]\kappa_s \eta_s \\ &= X_0 \prod_{s=1}^{t} f(\kappa_s, \phi) + \sum_{s=1}^{t} [\prod_{j=s+1}^{t} f(\kappa_j, \phi)]\kappa_s \eta_s. \end{aligned}$$

This sum is non-explosive if $|f(\kappa_t, \phi)| < 1$ and $|[\prod_{j=s+1}^{t} f(\kappa_j, \phi)]\kappa_s| < 1$. This depends on the break process $\boldsymbol{\kappa}$. If we assume, without loss of generality, that a fraction π^* periods for which $\kappa_s = 1$ for $1 \leq s \leq t$ and a fraction $\tilde{\pi}$ periods for which

$\kappa_j = 1$ for $s+1 \leq j \leq t$, then

$$\begin{aligned}
\prod_{s=1}^{t} f(\kappa_s, \phi) &= \prod_{s=1}^{t} (\phi\kappa_s + (1-\kappa_s)) \\
&= (\prod_{1\leq s\leq t|\kappa_s=1} (\phi\kappa_s + (1-\kappa_s)))(\prod_{1\leq s\leq t|\kappa_s=0} (\phi\kappa_s + (1-\kappa_s))) \\
&= (\prod_{1\leq s\leq t|\kappa_s=1} \phi)(\prod_{1\leq s\leq t|\kappa_s=0} 1) \\
&= \phi^{\pi^* t},
\end{aligned}$$

and

$$\begin{aligned}
[\prod_{j=s+1}^{t} f(\kappa_j, \phi)]\kappa_s &= [\prod_{j=s+1}^{t} (\phi\kappa_j + (1-\kappa_j))]\kappa_s \\
&= \phi^{\tilde{\pi}(t-s)}\kappa_s.
\end{aligned}$$

Since $0 \leq \pi^* \leq 1$, it follows that $|\phi^{\pi^* t}| < 1$ and $|\phi^{\tilde{\pi}(t-s)}\kappa_s| \leq 1$ if $|\phi| < 1$. Strict inequality holds for both, unless $\tilde{\pi} = 0$. Then the first part becomes one and we are left with κ_s, which is either zero or one. However, $\kappa_s\eta_s$ is not explosive if $\kappa_s = 1$, as it is only a single normally distributed random variable, not a sum. Moreover, the realization of η_s will then be discounted for future values X_j where $j > s$ if $|\phi| < 1$. This shows that the ARMI process X_t is stationary if $|\phi| < 1$.

3.A.2 Unconditional distribution

We show that the unconditional distribution of the ARMI process given by Equation (3.6), under stationarity, i.e. $|\phi| < 1$, and $\Pr[\kappa_t = 1] = \pi > 0$,[23] is the same as for a stationary autoregressive process, i.e. if $\pi = 1$. This is equivalent to proving that the unconditional distribution does not depend on the break probability π.

We assume $|\phi| < 1$, $\pi > 0$ and independence between X_{t-1} and κ_t.

[23] If $\pi = 0$, then $X_t = X_0$.

The unconditional mean of X_t is

$$
\begin{aligned}
\mathrm{E}[X_t] &= \mathrm{E}\left[\mu + \kappa_t\phi(X_{t-1}-\mu) + (1-\kappa_t)(X_{t-1}-\mu) + \kappa_t\eta_t\right] \\
&= \mu + \mathrm{E}\left[\kappa_t\phi(X_{t-1}-\mu)\right] + \mathrm{E}\left[(1-\kappa_t)(X_{t-1}-\mu)\right] \\
\mathrm{E}[X_t-\mu] &= \mathrm{E}[\kappa_t]\phi\,\mathrm{E}[X_{t-1}-\mu] + \mathrm{E}[1-\kappa_t]\,\mathrm{E}[X_{t-1}-\mu] \\
&= \pi\phi\,\mathrm{E}[X_t-\mu] + (1-\pi)\,\mathrm{E}[X_t-\mu] \\
\big(1-\pi\phi-(1-\pi)\big)\,\mathrm{E}[X_t-\mu] &= 0 \\
\mathrm{E}[X_t-\mu] &= 0 \\
\mathrm{E}[X_t] &= \mu.
\end{aligned}
$$

The unconditional variance of X_t is

$$
\begin{aligned}
\mathrm{var}(X_t-\mu) &= \mathrm{var}\left(\mu + \kappa_t\phi(X_{t-1}-\mu) + (1-\kappa_t)(X_{t-1}-\mu) + \kappa_t\eta_t\right) \\
&= \phi^2\,\mathrm{var}\left(\kappa_t(X_{t-1}-\mu)\right) + \mathrm{var}\left((1-\kappa_t)(X_{t-1}-\mu)\right) \\
&\qquad + \mathrm{var}(\kappa_t\eta_t) \\
&= \phi^2\,\mathrm{E}[\kappa_t]\,\mathrm{var}(X_{t-1}-\mu) + \mathrm{E}[1-\kappa_t]\,\mathrm{var}(X_{t-1}-\mu) \\
&\qquad + \mathrm{E}[\kappa_t]\,\mathrm{var}(\eta_t) \\
&= \phi^2\pi\,\mathrm{var}(X_t-\mu) + (1-\pi)\,\mathrm{var}(X_t-\mu) + \pi\sigma^2 \\
\big(1-\phi^2\pi-(1-\pi)\big)\,\mathrm{var}(X_t-\mu) &= \pi\sigma^2 \\
(\pi-\phi^2\pi)\,\mathrm{var}(X_t-\mu) &= \pi\sigma^2 \\
\mathrm{var}(X_t-\mu) &= \frac{\pi\sigma^2}{\pi-\phi^2\pi} = \frac{\sigma^2}{1-\phi^2}.
\end{aligned}
$$

Hence, the unconditional distribution of X_t is $\mathrm{N}\big(\mu,\sigma^2/(1-\phi^2)\big)$ and does not depend on the break probability π.

3.B Gibbs sampler

The Gibbs sampler is closely related to the one used by Koop et al. (2009). The difference is the addition of autoregressive parameters and long run means. The sampler is based on the algorithm for time-varying parameter VAR (TVP-VAR) models, see

Primiceri (2005), with the difference that we additionally need to draw the break processes $\boldsymbol{\kappa}_t = (\kappa_{1t}, \kappa_{2t}, \kappa_{3t})'$ and the break probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)'$. The step of drawing the break process is also explained by Groen et al.'s (2013) online appendix for a univariate model. We explain it here for the multivariate model.

We can split the Gibbs sampler into four steps, which we iterate over until we have a sufficient number of posterior draws,

1. Draw the coefficients $\boldsymbol{B} = \{\boldsymbol{\beta}_t\}_{t=1}^T$ and their break process $\boldsymbol{\kappa}_1 = (\kappa_{1,1}, \ldots, \kappa_{1,T})$.
2. Draw the volatilities $\boldsymbol{S} = \{\log \boldsymbol{\sigma}_t^2\}_{t=1}^T$ and their break process $\boldsymbol{\kappa}_2 = (\kappa_{2,1}, \ldots, \kappa_{2,T})$.
3. Draw the covariance terms $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$ and its break process $\boldsymbol{\kappa}_3 = (\kappa_{3,1}, \ldots, \kappa_{3,T})$.
4. Draw the parameters in the state equations $\boldsymbol{\theta} = (\boldsymbol{\Phi}, \boldsymbol{Q}, \boldsymbol{\pi})$.

Each of the first three steps consists of writing (parts of) the VAR model into state space form such that we can draw the state variables in two steps,

a. Draw the break process $\boldsymbol{\kappa}_k$ from the full conditional posterior, with the state variables integrated out, using the algorithm of Gerlach et al. (2000).

b. Draw the state variables, i.e. the long-run mean and the time-varying part, using the simulation smoother of Durbin and Koopman (2002a).

3.B.1 Step 1: Drawing the coefficients

Define $\boldsymbol{y}_t = (r_t, z_t)'$ and $\boldsymbol{x}_t = (1, z_{t-1})'$, the left hand side and right hand side variables, and $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t - \boldsymbol{\mu}_1$, the coefficients' deviation from the mean at time t. Then, we can write the state space model,

$$\boldsymbol{y}_t = (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)(\boldsymbol{\mu}_1 + \tilde{\boldsymbol{\beta}}_t) + \boldsymbol{\varepsilon}_t, \tag{3.B.1}$$

$$\tilde{\boldsymbol{\beta}}_t = f(\kappa_{1t}, \boldsymbol{\Phi}_1)\tilde{\boldsymbol{\beta}}_{t-1} + \kappa_{1t}\boldsymbol{\eta}_t, \tag{3.B.2}$$

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1, \tag{3.B.3}$$

where $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$ and $\boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_1)$ are independent, $\otimes$ is the Kronecker product and $\boldsymbol{I}_n$ the $n \times n$ identity matrix, with $n = 2$ the number of left hand side variables.

Now, we can first apply the algorithm of Gerlach et al. (2000) to efficiently draw the break process $\boldsymbol{\kappa}_1$ from the full conditional posterior where the state variables have been integrated out. Then, conditional on the break process $\boldsymbol{\kappa}_1$, we use the simulation smoother of Durbin and Koopman (2002a) to draw the state variables $\tilde{\boldsymbol{\beta}}_t$ and $\boldsymbol{\mu}_1$.

3.B.2 Step 2: Drawing the volatilities

Define

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{A}_t\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{A}_t\big(\boldsymbol{y}_t - (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)\boldsymbol{\beta}_t\big) = \begin{bmatrix} 1 & 0 \\ \alpha_t & 1 \end{bmatrix} \begin{pmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{1t} \\ \alpha_t\hat{\varepsilon}_{1t} + \hat{\varepsilon}_{2t} \end{pmatrix}. \tag{3.B.4}$$

Next, transform $\boldsymbol{\varepsilon}_t^*$ into $\boldsymbol{\varepsilon}_t^{**} = \log\big((\boldsymbol{\varepsilon}_t^*)^2 + \bar{c}\big)$, with $\bar{c} = 0.0001$ the off-set constant to avoid numerical issues.

Further, define $\widetilde{\log \boldsymbol{\sigma}_t^2} = \log \boldsymbol{\sigma}_t^2 - \boldsymbol{\mu}_2$, the volatilities' deviation from the mean at time t.

Then, we can write the state space model,

$$\boldsymbol{\varepsilon}_t^{**} = \boldsymbol{\mu}_2 + \widetilde{\log \boldsymbol{\sigma}_t^2} + \boldsymbol{e}_t, \tag{3.B.5}$$

$$\widetilde{\log \boldsymbol{\sigma}_t^2} = f(\kappa_{2t}, \boldsymbol{\Phi}_2)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2} + \kappa_{2t}\boldsymbol{\zeta}_t, \tag{3.B.6}$$

$$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2, \tag{3.B.7}$$

where $\boldsymbol{\zeta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_2)$ independent from $\boldsymbol{e}_t$.

The state space model is non-Gaussian, because the disturbances $\boldsymbol{e}_t$ follow a χ^2 distribution with one degree of freedom. Carter and Kohn (1997) and Kim et al. (1998) show that we can accurately approximate this using a mixture of normals. We use the mixture of seven normals used by Kim et al. (1998), and can then consecutively draw the break process $\boldsymbol{\kappa}_2$, and the time-varying part of $\widetilde{\log \boldsymbol{\sigma}^2}$ and the long-run mean $\boldsymbol{\mu}_2$.

Stroud et al. (2011) suggest a Metropolis-Hastings step to correct for the approximation error of using a mixture of normals. Kim et al. (1998) show that the approximation error is negligible, though, and Del Negro and Primiceri (2015) confirm this with a TVP-VAR model. Therefore, we skip the Metropolis-Hastings step.

3.B.3 Step 3: Drawing the covariance term

Define $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)\boldsymbol{\beta}_t$, and $\boldsymbol{A}_t\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{u}_t$, with $\boldsymbol{u}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t')$, then we can use the lower triangular structure of $\boldsymbol{A}_t$ to rewrite $\hat{\boldsymbol{\varepsilon}}_t$ as

$$\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{C}_t\alpha_t + \boldsymbol{u}_t = \begin{pmatrix} 0 \\ -\hat{\varepsilon}_{1t} \end{pmatrix} \alpha_t + \boldsymbol{u}_t. \tag{3.B.8}$$

Further, define $\tilde{\alpha}_t = \alpha_t - \mu_3$ the covariance term's deviation from the long run mean.

Then, we can write the state space model,

$$\hat{\varepsilon}_{2t} = -\hat{\varepsilon}_{1t}(\mu_3 + \tilde{\alpha}_t) + u_{2t}, \tag{3.B.9}$$

$$\tilde{\alpha}_t = f(\kappa_{3t}, \phi_3)\tilde{\alpha}_{t-1} + \kappa_{3t}\xi_t, \tag{3.B.10}$$

$$\mu_3 = \mu_3, \tag{3.B.11}$$

where $u_{2t} \sim \mathrm{N}(0, \sigma_{2t}^2)$ and $\xi_t \sim \mathrm{N}(0, q_3^2)$ are independent.

We again apply steps a and b to draw the break process $\boldsymbol{\kappa}_3$, and then the time-varying part of the covariance term $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha_T})$ and the long-run mean μ_3.

3.B.4 Step 4: Drawing the state equation parameters

The full conditional posterior distribution for the diagonal elements of the autoregressive parameter $\boldsymbol{\Phi}_k$ is a truncated normal distribution,

$$\mathrm{diag}(\boldsymbol{\Phi}_k)|\ldots \sim \mathrm{N}(\overline{\boldsymbol{m}}_k, \overline{\boldsymbol{V}}_k)I(|\phi_{ki}| < 1, \forall i = 1, \ldots, n_k), \tag{3.B.12}$$

where $I(A)$ is one if A holds and zero otherwise, n_k the number of diagonal elements in $\boldsymbol{\Phi}_k$, with variance

$$\overline{\boldsymbol{V}}_1 = \left(\underline{\boldsymbol{V}}_1^{-1} + \sum_{t=2}^{T} \kappa_{1t}(\tilde{\boldsymbol{\beta}}_{t-1}' \boldsymbol{Q}_1^{-1} \tilde{\boldsymbol{\beta}}_{t-1})\right)^{-1}, \tag{3.B.13}$$

$$\overline{\boldsymbol{V}}_2 = \left(\underline{\boldsymbol{V}}_2^{-1} + \sum_{t=2}^{T} \kappa_{2t}\big((\widetilde{\log \boldsymbol{\sigma}_{t-1}^2})' \boldsymbol{Q}_2^{-1} \widetilde{\log \boldsymbol{\sigma}_{t-1}^2}\big)\right)^{-1}, \tag{3.B.14}$$

$$\overline{v}_3^2 = \left(\underline{v}_3^{-2} + q_3^{-2} \sum_{t=2}^{T} \kappa_{3t} \tilde{\alpha}_{t-1}^2\right)^{-1}, \tag{3.B.15}$$

and mean

$$\overline{\boldsymbol{m}}_1 = \overline{\boldsymbol{V}}_1\left(\underline{\boldsymbol{V}}_1^{-1}\underline{\boldsymbol{m}}_1 + \sum_{t=2}^{T} \kappa_{1t}(\tilde{\boldsymbol{\beta}}_{t-1}' \boldsymbol{Q}_1^{-1} \tilde{\boldsymbol{\beta}}_t)\right), \tag{3.B.16}$$

$$\overline{\boldsymbol{m}}_2 = \overline{\boldsymbol{V}}_2\left(\underline{\boldsymbol{V}}_2^{-1}\underline{\boldsymbol{m}}_2 + \sum_{t=2}^{T} \kappa_{2t}\big((\widetilde{\log \boldsymbol{\sigma}_{t-1}^2})' \boldsymbol{Q}_2^{-1} \widetilde{\log \boldsymbol{\sigma}_t^2}\big)\right), \tag{3.B.17}$$

$$\overline{m}_3 = \overline{v}_3^2\left(\underline{v}_3^{-2}\underline{m}_3 + q_3^{-2} \sum_{t=2}^{T} \kappa_{3t}(\tilde{\alpha}_{t-1}\tilde{\alpha}_t)\right), \tag{3.B.18}$$

and the off-diagonal elements of $\boldsymbol{\Phi}_k$ are zero.

We draw from the truncated normal distribution using an accept-reject algorithm with a normal distribution with the same mean and variance to reduce computational costs. If after a 1,000 tries no draw has been accepted, we draw directly from the truncated normal distribution.

The full conditional posterior distribution for the break probability π_k is a beta distribution,

$$\pi_k | \ldots \sim \text{Beta}(\overline{a}_k, \overline{b}_k), \tag{3.B.19}$$

where

$$\overline{a}_k = \underline{a}_k + \sum_{t=2}^{T} \kappa_{kt}, \tag{3.B.20}$$

$$\overline{b}_k = \underline{b}_k + \sum_{t=2}^{T} (1 - \kappa_{kt}), \tag{3.B.21}$$

for $k = 1, 2, 3$.

The full conditional posterior distribution of the break size $\boldsymbol{Q}_k$ is an inverse Wishart distribution,

$$\boldsymbol{Q}_k | \ldots \sim \mathrm{IW}\big(\overline{\boldsymbol{W}}_k, \overline{\nu}_k\big), \tag{3.B.22}$$

with location parameter

$$\overline{\boldsymbol{W}}_1 = \underline{\boldsymbol{W}}_1 \underline{\nu}_1 + \sum_{t=2}^{T} \kappa_{1t} \big(\tilde{\boldsymbol{\beta}}_t - f(\kappa_{1t}, \boldsymbol{\Phi}_1)\tilde{\boldsymbol{\beta}}_{t-1}\big)\big(\tilde{\boldsymbol{\beta}}_t - f(\kappa_{1t}, \boldsymbol{\Phi}_1)\tilde{\boldsymbol{\beta}}_{t-1}\big)', \tag{3.B.23}$$

$$\overline{\boldsymbol{W}}_2 = \underline{\boldsymbol{W}}_2 \underline{\nu}_2 + \sum_{t=2}^{T} \kappa_{2t} \left(\widetilde{\log \boldsymbol{\sigma}_t^2} - f(\kappa_{2t}, \boldsymbol{\Phi}_2)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2}\right)\left(\widetilde{\log \boldsymbol{\sigma}_t^2} - f(\kappa_{2t}, \boldsymbol{\Phi}_2)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2}\right)', \tag{3.B.24}$$

$$\overline{w}_3 = \underline{w}_3 \underline{\nu}_3 + \sum_{t=2}^{T} \kappa_{3t} \big(\tilde{\alpha}_t - f(\kappa_{3t}, \phi_3)\tilde{\alpha}_{t-1}\big)^2, \tag{3.B.25}$$

and degrees of freedom

$$\overline{\nu}_k = \underline{\nu}_k + \sum_{t=2}^{T} \kappa_{kt}, \tag{3.B.26}$$

for $k = 1, 2, 3$.

3.C MCMC convergence analysis

We follow the convergence analysis by Groen et al. (2013) in their appendix. That is, we compute inefficiency factors and analyze the convergence of the MCMC chain using the Geweke (1992) test, for our baseline model.

The inefficiency factors in table Table 3.C.1 are calculated as $\mathrm{IF} = 1 + 2\sum_{i=1}^{\infty} \rho_i$, with ρ_i the i-th order autocorrelation of the posterior draws of a parameter. The autocorrelation is computed using the Newey and West (1987) estimator with a Bartlett kernel and 4% bandwidth. A guideline is that you need approximately 100 times the inefficiency factor such that at most 1% of the variation is due to the data, see Kim et al. (1998) and Groen et al. (2013) (appendix). Most parameters are reasonably converged for 10,000 draws, the number of retained draws we use. The results do suggest that we require extra draws to get a more accurate estimate of the parameter in the covariance term's state equation. The inefficiency factor for the long run mean is small, so it only concerns the parameters describing the dynamics of α_t.

Table 3.C.1: Inefficiency factors

Parameter	Number	Median	Mean	Min	Max	5% quantile	95% quantile
Panel A: Intercepts and loadings							
$\tilde{\boldsymbol{\beta}}$	3,360	15.44	16.92	2.60	49.39	5.06	37.62
$\boldsymbol{\mu}_1$	4	52.25	52.78	48.18	58.46		
$\boldsymbol{\kappa}_1$	840	19.00	20.72	8.01	65.09	16.05	30.33
π_1	1	64.06	64.06	64.06	64.06		
$\boldsymbol{\Phi}_1$	4	95.07	101.69	51.97	164.66		
$\boldsymbol{Q}_1$	16	53.85	53.63	51.22	55.46		
Panel B: Variances							
$\widetilde{\log \boldsymbol{\sigma}^2}$	1,680	5.23	9.89	1.61	83.15	2.69	27.90
$\boldsymbol{\mu}_2$	2	16.06	16.06	6.66	25.46		
$\boldsymbol{\kappa}_2$	840	1.23	1.17	0.52	5.24	0.75	1.84
$\boldsymbol{\pi}_2$	1	20.49	20.49	20.49	20.49		
$\boldsymbol{\Phi}_2$	2	96.47	96.47	28.74	164.19		
$\boldsymbol{Q}_2$	4	61.21	62.62	42.29	85.77		
Panel C: Covariance term							
$\boldsymbol{\alpha}$	840	1.35	1.44	0.72	6.04	0.92	2.24
μ_3	1	6.12	6.12	6.12	6.12		
$\boldsymbol{\kappa}_3$	840	81.85	81.85	74.23	88.31	77.79	85.93
π_3	1	289.44	289.44	289.44	289.44		
ϕ_3	1	95.08	95.08	95.08	95.08		
q_3	1	203.38	203.38	203.38	203.38		
$\log p(\boldsymbol{Y}\vert\boldsymbol{\kappa}_1)$	1	4.40	4.40	4.40	4.40		

This table presents a summary of the inefficiency factors for the parameters of the mixture innovation model, where the break probability is unknown, and the hyperparameters are set as $\underline{a}_1 = 1, \underline{b}_1 = 59$, and $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$. The inefficiency factors are estimated using Newey and West (1987) estimator.

The Geweke (1992) tests whether the first 20% and the last 40% of the draws have an equal mean. The rejection rates in table Table 3.C.2 show that in general the MCMC chain is converged. However, the null hypothesis is strongly rejected for the parameters governing the covariance's break process. We experimented with the burn-in, and test results are mostly due to sample selection rather than a lack of convergence. Due to the high autocorrelation, it depends on the selected subsamples whether the null of equal means is rejected.

Combining the inefficiency factors and Geweke (1992) test results, only the parameters in the covariance term's state equation are not well identified. It is not

Table 3.C.2: Geweke test results

Parameter	Number	10% rejection rate	5% rejection rate	1% rejection rate
Panel A: Intercept and loadings				
$\tilde{\boldsymbol{\beta}}$	3,360	0.001	0.000	0.000
$\boldsymbol{\mu}_1$	4	0.000	0.000	0.000
$\boldsymbol{\kappa}_1$	840	0.004	0.000	0.000
π_1	1	0.000	0.000	0.000
$\boldsymbol{\Phi}_1$	4	0.000	0.000	0.000
$\boldsymbol{Q}_1$	16	0.000	0.000	0.000
Panel B: Variances				
$\widetilde{\log \boldsymbol{\sigma}^2}$	1,680	0.080	0.039	0.007
$\boldsymbol{\mu}_2$	2	0.000	0.000	0.000
$\boldsymbol{\kappa}_2$	840	0.092	0.061	0.010
$\boldsymbol{\pi}_2$	1	0.000	0.000	0.000
$\boldsymbol{\Phi}_2$	2	0.000	0.000	0.000
$\boldsymbol{Q}_2$	4	0.000	0.000	0.000
Panel C: Covariance term				
$\boldsymbol{\alpha}$	840	0.096	0.056	0.018
μ_3	1	0.000	0.000	0.000
$\boldsymbol{\kappa}_3$	840	1.000	1.000	1.000
π_3	1	1.000	1.000	1.000
ϕ_3	1	1.000	1.000	1.000
q_3	1	1.000	1.000	1.000
$\log p(\boldsymbol{Y}\|\boldsymbol{\kappa}_1)$	1	0.000	0.000	0.000

This table presents the rejection rates of the Geweke (1992) test, testing whether the first 20% and the last 40% of the draws have an equal mean, for the parameters of the mixture innovation model, where the break probability is unknown, and the hyperparameters are set as $\underline{a}_1 = 1, \underline{b}_1 = 59$, and $\underline{a}_2 = \underline{b}_2 = \underline{a}_3 = \underline{b}_3 = 1$.

surprising, because there is little instability found in this the covariance term. We are not too concerned about the effect on the overall convergence, given that the test results for the other parameters are fine. Concluding, we find that the overall convergence of the sampler for the current number of draws and burn-in is satisfactory.

Chapter 4

Uncertainty and the macroeconomy: A real-time out-of-sample evaluation

Joint work with Dick van Dijk

4.1 Introduction

Understanding the fundamental causes of business cycles has intrigued macroeconomists for decades, if not centuries. According to real option theory (Bernanke, 1983; Dixit and Pindyck, 1994), uncertainty is one of the key drivers of such cyclical fluctuations: as uncertainty increases, businesses hold off on investment and consumers hold off large purchases, thus reducing economic activity, such as output and employment. Bloom (2009) sparked a new line of research, on empirically measuring economic uncertainty and assessing its relationship with real macroeconomic variables.[1] This is not a straightforward exercise, because uncertainty is a latent concept and its exact definition can be debated. Not surprisingly then, a multitude of measures of economic uncertainty has been proposed over the last decade. Examples include financial volatility (Bloom, 2009), news based indices (Baker et al., 2016),

[1] See Bloom (2014) for a recent overview.

dispersion in micro data (Bloom, 2009), and disagreement or errors of professional forecasters (Rossi et al., 2016).

On the introduction of a new measure of economic uncertainty, it is usually added to a vector autoregressive model to assess the impact on macroeconomic variables. The comparison to other measures is usually limited to simple correlation, a visual comparison of extremes, and of impulse response functions. All uncertainty measures are proxies of a latent entity, which makes it difficult to assess their validity. This partly explains why a thorough (statistical) comparison of the proposed measures is lacking. Furthermore, inference thus far is based on in-sample analysis only, while it seems important for the validity to test whether the relationship holds out-of-sample.

In this paper we address both open issues identified above. First, we collect a comprehensive set of different uncertainty measures and conduct a factor analysis. This allows us to examine the similarities and differences between the various measures. Furthermore, the resulting factors, essentially combining the information in the different measures, might provide more comprehensive and accurate proxies of (different aspects of) the underlying notion of 'uncertainty.' Second, we conduct a real-time out-of-sample forecasting analysis to assess whether a forecaster is able to take advantage of the implied relationship between uncertainty and economic activity. This is important to gain insight into the practical usefulness of the various uncertainty measures. Though in-sample tests have more power (Inoue and Kilian, 2005), a forecasting analysis is relevant for policy makers. They might be able to use this extra source of information to improve decision making.

For the first part of our analysis, we identify 15 monthly uncertainty measures, that comply with a number of restrictions such as being freely and directly available for a substantial time period. An important additional restriction is that the data should be available ex ante, such that we can exploit it in our forecasting exercise. We show that the various measures can be categorized into five categories, based on their source: (i) volatility, (ii) cross-sectional dispersion, (iii) news, (iv) surveys, and (v) forecast errors. The collected measures are spread quite evenly across these categories.

Our factor analysis shows that there is indeed a relatively strong common component. The first principal component explains about 40% of total variation. It can be interpreted as general economic uncertainty, because it loads positively on all measures, though it loads slightly more on financial information. Interestingly, the importance of the factor increases during the financial crisis. Additionally, we identify a second factor, which can be interpreted as media/consumer uncertainty. It loads most heavily on news based and consumer survey based uncertainty measures. This factor remains elevated after recessions, reflecting that the media and consumers need more time to become confident about the recovery than reflected by the fundamentals.

For the second part of our analysis, we set up an extensive real-time out-of-sample analysis to forecast industrial production, employment, manufacturing and trade sales, and personal income excluding transfer payments. These variables are taken into account by the NBER business cycle dating committee[2] and are the constituents of the Conference Board's US Coincident Economic Index.[3] In contrast with pseudo out-of-sample analyses, we use different vintages to take into account that publications of macroeconomic variables are revised multiple times after the first release. Using the vintages, allows us to assess whether a forecaster is able to gain from using the values that are available at that point in time. We are not only interested in mean forecasts, but also in other parts of the predictive density. For the latter purpose, we perform a quantile forecasting analysis. This provides insight into possible asymmetries in the relationship between uncertainty and macroeconomic variables. Forecasts are produced for the period 1999 to 2016, based on an expanding window starting in 1986. We consider multiple forecasting horizons, from one month up to 24 months ahead.

We find that there is limited predictive ability in the uncertainty measures for the mean of the various macroeconomic variables. At the lower quantiles however, the uncertainty measures do have forecasting power for all but personal income. The results are especially strong for employment.[4] From the individual uncertainty measures, the Jurado et al. (2015) measures, financial volatility, and business forecast dispersion (Bachmann et al., 2013) perform best and most consistent. The results are consistent when we consider a one or two factor model rather than the individual measures. The second factor helps with forecasting at longer horizons, even though the performance of individual media and news based measures is disappointing. Results are robust to

[2]The dates of peaks and troughs can be found at `http://www.nber.org/cycles/recessions.html`.

[3]See `https://www.conference-board.org/data/bcicountry.cfm?cid=1`.

[4]Interestingly, Bloom (2009) considers the effect of uncertainty on the labor market.

using a smaller panel of uncertainty measures over a longer time period and expanding the set of measures by allowing for an unbalanced panel, as well as using a rolling instead of an expanding estimation window.

Our results indicate that the relationship between economic uncertainty and activity is nonlinear. This is in line with the regime-switching model by Jones and Enders (2016). Further, it corresponds with the finding by Bloom et al. (2018) that a combination of negative shocks to the mean and positive shocks to the variance are required to induce recessions.

Our paper provides three main contributions. First, we contribute to the empirical uncertainty literature by showing how the different uncertainty measures are related and that they can be summarized by two common factors. Second, we add to the literature by conducting a real-time forecasting exercise.[5] This is relevant for example to policy makers, as they have to make decisions in real-time. Third, the extension to quantile forecasts provides insight into possible nonlinearities in the relationship between economic uncertainty and economic output.

Multiple papers are interested in forecasting economic output using financial volatility or conditions. For example, Chauvet et al. (2015) performs a real-time forecasting exercise linking multiple financial volatility measures in a Markov-switching dynamic factor model and find evidence for nonlinearity in the form of Markov-switching regimes. Further, Adrian et al. (2018) consider the impact of financial conditions on the density of economic output. They allow for asymmetry across the density based on quantile forecasts and find that especially the left tail is affected by financial conditions. Most closely related to this paper in spirit is the work by Giglio et al. (2016), who do a similar analysis for systemic risk measures. They gather multiple systemic risk measures and conduct a quantile forecasting exercise. They find that a single common factor improves their forecast accuracy, and that predictive power for the mean is limited. In that light, our results are comparable too. The main difference is – other than using uncertainty measures instead of systemic risk measures – that we conduct a real-time rather than a pseudo out-of-sample forecasting exercise, taking into account revisions. Systemic risk and financial conditions are close in concept to economic uncertainty. Hence, the findings by Giglio et al. (2016) and

[5]Other work that performs quantile forecasts using real-time data includes Korobilis (2017).

Adrian et al. (2018) are consistent with our finding that economic uncertainty is useful mostly in forecasting the lower quantiles of economic output rather than the mean.

The paper is structured as follows. Section 4.2 describes the uncertainty measures, the selection criteria and the different categories, followed by the factor analysis in Section 4.3. Section 4.4 provides the methodology and implementation details of the real-time forecasting analysis. The forecasting results for the mean are discussed in Section 4.5, and those for the quantiles in Section 4.6. Section 4.7 presents robustness checks and Section 4.8 concludes.

4.2 Uncertainty measures

Our selection of uncertainty measures is based on a number of criteria. First, we restrict to US data such that all measures aim to capture the same entity. By far the largest number of measures is available for the US and it makes the results better comparable to the existing literature. Second, to match the frequency of the output variables used in the second part of our analysis we focus on monthly data. Measures available at a higher frequency are transformed to monthly frequency appropriately, by using either the average or the end of month value. Measures reported at a lower frequency are excluded. They can be included using mixed frequency methods, see e.g. Carriero et al. (2017), but we choose to focus the analysis for now. Third, the data should be available to the forecaster in real-time, because we are interested in whether she had been able to take advantage of the extra information. This excludes measures that are estimated using ex-post data, such as forecast error distributions and many other decompositions. This does not exclude filtered time series though. Fourth, we require a sufficient time series length such that we have reasonable power for the forecast evaluation. Fifth, on a more practical note, the data should be relatively easy to collect or compute.

Table 4.A.1 lists the selected uncertainty measures, including a brief description, the source and sample size.[6] It is a reasonably sized set of 15 measures, that includes most of the popular ones that have been proposed thus far.

[6]To increase the sample period, the uncertainty measures VXO and OVX are merged with realized volatility measures, TYVIX is merged with the implied volatility measure of Choi et al. (2017), and EPU is merged with the historical EPU measure that is only based on six to ten large newspapers. The backfilled data is normalized using the overlapping period. Correlations between the merged series in the overlapping periods range from 0.87 to 0.99.

Notable exclusions are cross-sectional dispersion of firm level profit growth (Bloom, 2009), total factor productivity growth (Bloom, 2009; Kehrig, 2015), Livingstone survey GDP forecasts (Bloom, 2009), price changes (Vavra, 2013), and employment growth (Bachmann and Bayer, 2014). Conditional volatility from decomposing financial volatility into risk aversion and uncertainty (Bekaert et al., 2013), shocks from political turmoil, natural disasters or terrorist attacks (Baker and Bloom, 2013), Fama-French factor residual variance (Gilchrist et al., 2014), and fiscal volatility shocks (Fernández-Villaverde et al., 2015) are excluded as well, either because they need to be computed ex post, or because they are only available at a lower frequency. Furthermore, we ignore measures based solely on the Survey of Professional Forecasters (SPF) because these are of quarterly frequency (Lahiri and Sheng, 2010; Rossi et al., 2016).

In Section 4.7.2 we do add Scotti's (2016) US uncertainty index and a Google trends news uncertainty index (Castelnuovo and Tran, 2017), which are available starting in 2003 and 2004. Adding them does not impact the results.

The descriptions in Table 4.A.1 show that economic uncertainty can be proxied in a variety of ways and from multiple sources. We identify five categories related to how economic uncertainty is measured. The uncertainty measure can be based on (A) volatility, (B) cross-sectional dispersion, (C) news, (D) surveys, and/or (E) forecast errors. Combinations are possible, but most measures fit into one of these categories. Table 4.1 shows our data set is well balanced as none of the categories is overrepresented. The following subsections discuss each type and the associated uncertainty measures.

Table 4.1: Uncertainty categories

Code	Type	Count
A	Conditional volatility	4
B	Cross-sectional dispersion	3
C	News	3
D	Surveys	4
E	Forecast errors	4

The table presents the different types of uncertainty measures and the number of associated uncertainty measures.

4.2.1 Conditional volatility

The first type of uncertainty measure is a volatility estimate of some underlying, typically a financial asset. Times of high conditional volatility are assumed to be related to times of high uncertainty. Probably the best known uncertainty measure is the VXO index (VXO), proposed by Bloom (2009). VXO is the implied volatility on the S&P100 index, which is an estimate of conditional volatility using option prices traded on the Chicago Board Options Exchange (CBOE).

Implied volatilities can be similarly computed for other assets, such as 10 year bonds (TYVIX) and the WTI oil price (OVX). These are also obtained from the CBOE. We include TYVIX to diversify the underlying assets and OVX because of the importance of the oil price for the economy, see e.g. Hamilton (1983). The negative impact of oil price volatility or uncertainty on economic activity has been shown empirically by for example Ferderer (1996), Jo (2014) and Kellogg (2014). Kellogg (2014) uses implied volatility – similar to OVX – and finds that this provides a better fit than GARCH or historical volatility.[7]

The conditional volatility of financial assets is a combination of uncertainty and risk aversion (Bekaert et al., 2013). To obtain a purer estimate of uncertainty, Chuliá et al. (2017) suggest to remove the time-varying predictable component of the VIX (CGU).[8] The predictable component is estimated using a filter and hence available to the forecaster at the time of prediction.

4.2.2 Cross-sectional dispersion

The second type of uncertainty measure utilizes micro data by estimating cross-sectional dispersion in each time period for a set of individuals, forecasters or firms. More dispersed outcomes suggest higher economic uncertainty.

Bloom (2009) proposes the cross-sectional standard deviation of stock returns (CSDR) and, to adjust for industry effects, a version that is the mean over industry

[7]The measures are backfilled to increase the sample size. For the TYVIX, we use the implied volatility estimate of (Choi et al., 2017), available on the websites of Philippe Mueller (`https://sites.google.com/site/philippebmueller/`) and Andrea Vedolin (`https://sites.google.com/site/andreavedolin/`). For the VXO and OVX, we use the underlying's realized volatility calculated using daily data from CRSP for the S&P500 stock returns and from St. Louis Fed's FRED for WTI oil returns, in line with Bloom (2009).

[8]Available at `http://www.ub.edu/rfa/uncertainty-index/`.

specific cross-sectional standard deviations using SIC3 (CSDRsic). These measures are computed with data from CRSP, for firms with at least 500+ monthly observations.

Bachmann et al. (2013) use forecast disagreement (FDISP) between respondents of the Philadelphia Fed's Manufacturing Business Outlook Survey (MBOS). The MBOS is a monthly survey among large manufacturing companies in the Third Federal Reserve District (Delaware, Pennsylvania and New Jersey) that dates back to May 1968. Though the survey is regional, the results have been found to be relevant for predicting national activity (Trebing, 1998; Schiller and Trebing, 2003). An advantage is that the questions and composition of respondents is stable over time (Bachmann et al., 2013). The measure is computed using responses to the question regarding the expectations of general business conditions six months from now (the time of the survey). Respondents can reply that they expect an "increase", "no change" or "decrease." Now define Frac^+ and Frac^- as the fractions of respondents that expect an increase (+) or decrease (-). Then FDISP is computed as

$$\text{FDISP} = \sqrt{\text{Frac}^+ - \text{Frac}^- - (\text{Frac}^+ - \text{Frac}^-)^2}. \tag{4.1}$$

Missing responses are treated the same as a "no change" reply, in line with how the Philadelphia Fed computes their diffusion index: $\text{Frac}^+ - \text{Frac}^-$.

The MBOS survey responses are subject to an annual revision to adjust for seasonality. Bachmann et al. (2013) propose to use the seasonally adjusted (sa) responses. Instead, we use the non-seasonally adjusted (nsa) responses to compute FDISP, which ensures that it is available to the forecaster in real-time and not subject to revisions. The nsa FDISP version is very similar to the sa version, with a correlation of 0.88. Moreover, the R^2 from regressing the nsa FDISP on monthly dummies is only 0.043 and none of the dummies are significant at a 5% level, suggesting that seasonal patterns are only of minor importance.

4.2.3 News

The third type of uncertainty measure is based on news, such as newspaper articles or Bloomberg announcements. In uncertain times, newspapers will publish (more) articles to report on the uncertainty and Bloomberg announcements of data will deviate more from expectations.

Counting news articles as a way to measure uncertainty goes back to Alexopoulos and Cohen (2009).[9] They count New York Times articles that contain words related to uncertainty and economic activity. The uncertainty measure of Baker et al. (2016) also uses text based analysis and spans multiple newspapers. They perform a thorough and labor intensive validity check of the methodology. This has become one of the most popular uncertainty measures, partly because it is freely available and continuously updated on a companion website.[10]

We include three versions of Baker et al.'s (2016) newspaper-based uncertainty measure. First, the original newspaper-based economic policy uncertainty (EPU) index, which is based only on the normalized number of articles in ten large newspapers[11] that contain a set of words related to economy, policy and uncertainty. Second, the main version presented on the economic policy uncertainty website (EPU+), which is a combination of the newspaper-based EPU index, tax code provisions and forecaster disagreement in the SPF. Third, we include the monetary policy uncertainty (MPU) index, a sub-index of EPU. The MPU index only counts the articles that include additional terms related to monetary policy.[12]

The news based measures of Baker et al. (2016) are subject to light revisions because there is a delay in posting all the articles online for some newspapers. The revisions do not seem to be severe,[13] and given that it is such a popular uncertainty measure, we include it in our selection. The measures are included with a one month lag, which should further alleviate this.

[9] Alexopoulos and Cohen (2015) is an updated version.

[10] Economic Policy Uncertainty website: `http://policyuncertainty.com/index.html`.

[11] USA Today, the Miami Herald, the Chicago Tribune, the Washington Post, the Los Angeles Times, the Boston Globe, the San Francisco Chronicle, the Dallas Morning News, the New York Times, and the Wall Street Journal.

[12] The additional terms are: federal reserve, the fed, money supply, open market operations, quantitative easing, monetary policy, fed funds rate, overnight lending rate, Bernanke, Volcker, Greenspan, central bank, interest rates, fed chairman, fed chair, lender of last resort, discount window, European Central Bank, ECB, Bank of England, Bank of Japan, BOJ, Bank of China, Bundesbank, Bank of France, Bank of Italy. See Baker et al.'s (2016) online appendix or `http://policyuncertainty.com/categorical_terms.html`.

[13] This is suggested by the wording used on the companion website: "With each monthly update, data from the preceding two months may be revised slightly, as well. This is driven by the fact that some online newspapers do not immediately update their online archives with all articles, leading to slightly changing totals for the previous 1-2 months."

Newspapers article counts have the advantage of going back far in time. However, subscriptions to newspapers have dropped recently and more information is consumed on alternative (online) platforms. This makes the Google trends version (GTU)[14] of Castelnuovo and Tran (2017) relevant. The monthly average of Scotti's (2016) uncertainty index (Scotti)[15] is the final news type measure. It is a weighted average of surprises from Bloomberg announcements that is computed in real-time. We use the monthly average of daily observations. Unfortunately, GTU and Scotti are only available from 2004M1 and 2003M5 onwards. Our main analysis employs a balanced panel, but we experiment with including GTU and Scotti after five or ten years of data is observed in Section 4.7.2. It does not impact our results or factor analysis.

4.2.4 Surveys

The fourth type of uncertainty measure is based on survey data. These are outcomes of polls or surveys taken among consumers, professional forecasters or firms to gauge their expectations for the coming period. It is a direct way of measuring the uncertainty perceived by economic agents. FDISP and EPU+ fall into this category. FDISP is a measure that depends on a survey among firms, and EPU+ partly depends on a survey among professional forecasters.

Further, two of the selected measures use data from the Reuters/University of Michigan Survey of Consumers. One of the questions asks consumers whether they are planning to buy a vehicle in the next 12 months, and a follow-up question asks for their motivation. Leduc and Liu (2016) propose to use the number of respondents that replied "uncertain times" as reason for not buying a vehicle as measure of uncertainty (LLv). Fajgelbaum et al. (2017) similarly define a measure that uses the question on buying large household goods (LLh).

[14] Available at Efrem Castelnuovo's website: `https://sites.google.com/site/efremcastelnuovo/home/publications`

[15] Available at Chiara Scotti's website: `https://sites.google.com/site/chiarascottifrb/data-and-other-materials`.

4.2.5 Forecast errors

The fifth type of uncertainty measure is based on the idea that uncertainty can be inferred from the part of the data that is not forecastable ex ante. This is to distinguish the uncertainty measure from 'forecastable' time-varying volatility. CGU can be interpreted as this type of uncertainty measure. Chuliá et al. (2017) remove the forecastable part from the VIX, thus cleaning it from the risk aversion and leaving the uncertainty.

Jurado et al. (2015) construct some of the most well known uncertainty measures, by removing the forecastable part of the data. They extract factors from a large panel of macroeconomic and financial time series. For multiple horizons, they estimate a forecasting model with the factors and their squares included as predictors, and a stochastic volatility component. The uncertainty measure JLNm is then the common variance of the errors across the panel of macroeconomic time series for a particular forecasting horizon. Ludvigson et al. (2015) extend this to measures based on the forecast error variance for the financial time series (JLNf), and a selection of real activity variables (JLNr).[16] We select the 12 month forecasting horizon version, which is slightly smoother than the other horizons. This is also the version used by for example Rossi et al. (2016).

Though the JLN measures are conditional on the information set at time t, Jurado et al. (2015) do not use real-time data in their forecasting exercise to construct the measures. It is therefore not, strictly speaking, available to the forecaster in real-time. Jurado et al. (2015) argue that using real-time data is not possible, because they need a large panel of time series and this would be limited in real-time. Further, they argue that the revisions are updates on past events rather than news, and that the financial time series in the panel (which are not subject to revision) capture timely events sufficiently well. Also, the measures are aggregates over many variables. Therefore, the effect of revisions should be small. Due to these arguments and because of the unique category and popularity, we include the JLN measures.

[16] The JLN uncertainty measures are available from Sydney Ludvigson's website: `https://www.sydneyludvigson.com/`.

4.3 Factor analysis

The comparison of uncertainty measures is thus far limited to comparing the pattern of the different time series or computing correlations. Furthermore, it is usually limited to a set of about four uncertainty measures. We analyze the commonalities for our more extensive set of measures, and assess the underlying factor structure.

Figure 4.A.1 presents the time series of all 15 selected uncertainty measures. They are largely similar in that all uncertainty measures peak around the time of recessions. The correlation matrix in Figure 4.A.2 indicates multiple blocks of quite strongly correlated measures. The survey based measures (excluding FDISP) are highly correlated, as well as measures based on stock market data, such as the VXO, CGU and JLNf. The correlations for FDISP and CSDRsic are small in general. They are actually (weakly) negatively correlated with the consumer survey based measures. The average correlation level seems quite modest and might be less than expected given that they all aim to measure US economic uncertainty. A possible explanation for the moderate correlations is that the different measures capture different aspects of economic uncertainty.

To determine the commonality between the uncertainty measures more formally, we extract a number of factors using principal components analysis. Table 4.1 presents the factor loadings, and explained fraction of total variance for the first five principal components. The first factor represents the average (economic) uncertainty, with a slight emphasis on financial uncertainty. The loadings are all positive, and it explains about 40% of the total variance, see Table 4.1. The factor level spikes during recessions and periods of financial turmoil, such as Black Monday in October 1987, the Russian financial crisis in 1998, and the Greek government debt crisis in 2012, see Figure 4.1. Figure 4.2 shows the explanatory power for the first three recursively estimated factors. It is interesting to observe that the explanatory power of the first factor increases during recessions. Most measures increase during recessions, which is captured by the first factor.

The second factor loads most heavily and positively on the consumer confidence measures LLv and LLh and news based uncertainty measures EPU and EPU+, see Table 4.1. This factor can be interpreted as a consumer/media uncertainty factor.

Consumers rely on media outlets for economic news, which explains why they are linked. It is interesting to see in Figure 4.1 that its value remains relatively high after the recession has ended. Apparently, while fundamentals are recovering, the public is still uncertain. This can be because the recovery still has to feed back to consumers, e.g. in the form of new jobs – the unemployment rate typically lags other output variables. Further, consumer spending probably lags as well, as their savings might be depleted or at least diminished so they might want to save before spending again. Alternatively, consumers and the media are simply not confident whether the recovery has fully started or if it is simply just a coincidental good output number. This is plausible, given that even the NBER's Business Cycle Dating Committee has a delay of several months in 'officially' calling the end of recessions.

The other factors lack a clear interpretation or explain only a single measure, see Table 4.1. The third factor explains quite some variance of MPU and JLNr, but the loadings are in opposite directions. The fourth factor explains most of FDISP, and the fifth loads heavily on TYVIX and OVX.

There is a clear commonality between the measures. FDISP is somewhat different from most measures though. This is clear from Table 4.1. The loading is small for the first factor compared to the others, and less than 20% of FDISP's variance is explained by the first three common factors. The time series in Figure 4.A.1g confirms its idiosyncratic behavior. FDISP displays less of a jigsaw pattern compared to the others. The level does not suddenly increase before recessions, and it drops quickly after recessions. It is hard to pinpoint the reason for FDISP's difference, but it could either be due its regional focus, because business surveys capture a unique part of uncertainty, or – more mechanically – because it is restricted between zero and one, which makes it harder to get the jigsaw pattern.

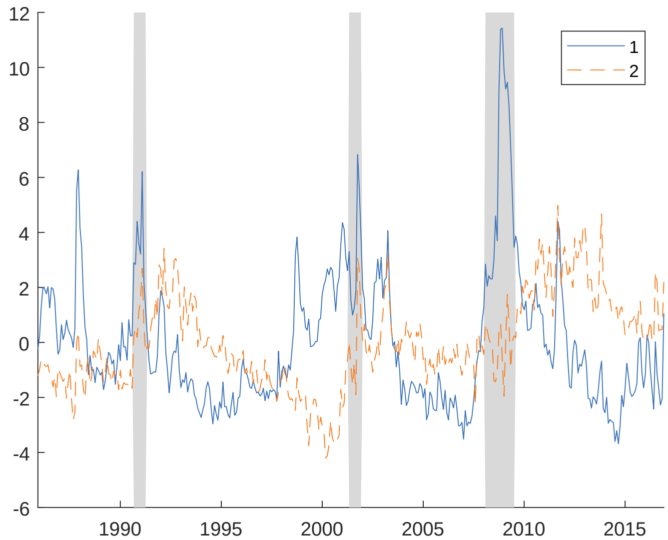
The principal component analysis results suggest the presence of two common factors. Together, they explain almost 60% of the total variance and both have a clear interpretation. We refrain from estimating the number of common factors k using the large panel methods such as Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). The results are sensitive to the choice of $kmax$, the maximum number of factors considered. The number of selected factors k also fluctuates heavily depending on whether or not all measures were included. This is likely because our

Table 4.1: Factor loadings and eigenvalues

k	1	2	3	4	5
VXO	0.340	−0.084	0.081	−0.052	0.211
TYVIX	0.232	−0.065	0.129	0.249	0.691
OVX	0.233	−0.074	−0.064	0.266	−0.612
CGU	0.305	−0.192	−0.075	−0.330	−0.029
CSDR	0.307	−0.207	0.099	−0.350	−0.188
CSDRsic	0.201	−0.301	0.346	−0.196	−0.115
FDISP	0.094	−0.207	0.128	0.658	−0.086
LLv	0.166	0.464	−0.194	−0.080	−0.029
LLh	0.199	0.423	−0.269	−0.117	0.022
EPU+	0.237	0.403	0.158	0.113	0.000
EPU	0.246	0.372	0.270	0.039	−0.110
MPU	0.231	0.158	0.511	0.123	−0.081
JLNm	0.321	−0.114	−0.381	0.178	0.061
JLNf	0.343	−0.147	−0.047	−0.185	0.141
JLNr	0.286	−0.123	−0.453	0.214	−0.048
R^2	0.404	0.190	0.089	0.077	0.052

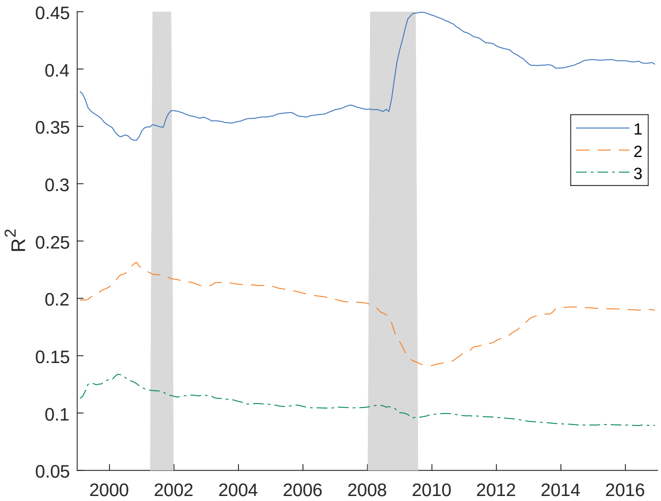
The table presents the loadings and the marginal R^2, the fraction of total variance explained by the x-th principal component, for the first five principal components on the sample 1986M4–2016M12. See Table 4.A.1 for an explanation of the abbreviations.

Figure 4.1: Uncertainty factors



The figure presents the time series of the first (solid blue line) and second factor (dashed orange line) from the full sample principal components analysis. The gray bars are recessions as determined by NBER's Business Cycle Dating Committee.

Figure 4.2: Explanatory power over time



The figure presents the marginal explanatory power of the first (solid blue line), second (dashed orange line) and third principal component (dash-dotted green line) of the uncertainty factor model, estimated recursively.

cross-section is modest compared to the large N and T asymptotics underlying these methods. Indeed, Bai and Ng (2002) find that their information criterion does not work well in the case of $\min(N, T) = 10$, close to our setting. Additionally, the rules of thumb for selecting $kmax$ suggested by Bai and Ng (2002) and Ahn and Horenstein (2013) result in $kmax$ of zero to two, suggesting that the methods are not suitable for our data set. Alternatively, testing the rank of a covariance matrix using rank tests such as Cragg and Donald (1997) or Kleibergen and Paap (2006) relies on large N or T asymptotics. But these rank tests are not suitable for covariance matrices, see Donald et al. (2007) for more details. Instead, we select the number of factors ex ante in our forecasting exercise. An advantage from an interpretation perspective is that by keeping k fixed over time, we test against the same model – or number of factors – across the sample.

4.4 Methodology

This section describes the target variables, the real-time forecasting setup, and explains the quantile regression as well as the forecast evaluation.

4.4.1 Coincident variables

Theory suggests a link of economic uncertainty with business cycle variables (Bernanke, 1983). Therefore, we are interested in forecasting output variables that are considered by the NBER's Business Cycle Dating Committee for dating peaks and troughs. The four monthly coincident variables linked to the business cycle are industrial production (IP), nonfarm payroll employment (EMP), manufacturing and trade industries sales (MTS), and personal income excluding current transfer receipts (PIX).

In order to asses whether a forecaster is able to improve the accuracy of her predictions, real-time data should be used. That is, the vintages with values that were available to the forecaster at that time period. It is relevant, because macroeconomic variables are reported with a lag and are subject to revisions. Relying on final vintage data would misrepresent the forecaster's information set, see e.g. Croushore (2006) for more information on the importance of real-time data forecasting. Real-time data for the four coincident variables is obtained from the data set of Chauvet and Piger (2008).[17] The data set is updated using the Philadelphia Fed's Real-Time Data Set for Macroeconomists (Croushore and Stark, 2001) for industrial production and employment. The most recent vintages for sales and personal income are from St. Louis Fed's ALFRED.[18]

The target variables are defined as follows. Industrial production, employment, sales, and personal income are all modeled as I(1) and we transform them into

[17] To be exact, it is an updated version of the Giusto and Piger (2017) data set, which updates the Chauvet and Piger (2008) data set to 2013. Thanks to Jeremy Piger for uploading the raw data set on his website: `https://pages.uoregon.edu/jpiger/research/published-papers/`.

[18] In particular, we use CMRMTSPL for manufacturing and trade industries sales. Three vintages of MTS are missing (2013M10, 2014M01, and 2015M09), and we use vintages from the Conference Board to fill post-1996 observations. For PIX, we follow Giusto and Piger (2017) by computing the real personal income excluding transfer receipts as the difference between personal income (PI) and personal current transfer receipts (PCTR), and dividing by the ratio of nominal (DSPI) to real disposable income (DSPIC96). Three vintages are missing of PIX due to a large (NIPA) revision at the end of 1995. Imputing the values as Chauvet and Piger (2008) is not possible because it requires observations before and after the missing sample, from the same vintage. In fact, Chauvet and Piger (2008) and Giusto and Piger (2017) skip the 1995M11–1996M1 vintages. Similarly, we delete the rows corresponding to the missing values before estimation. Since it involves only three vintages, it costs at most six observations. Additionally, there are five (additive) outliers in the level of PIX (1992M12, 1993M12, 2004M12, 2005M08, and 2012M12). For now, we simply ignore this and include them in the estimation sample – it is probably difficult for the forecaster to identify outliers in real-time. However, due to the AR model, the outliers also affect other forecasts. As an alternative, we impute the outliers by the final vintage's unconditional median growth rate. In both cases, the periods at which the outliers are observed are excluded in the evaluation.

annualized growth rates:

$$y_{t+h}^{h,t+h+1} = (1200/h)\log(\mathrm{Y}_{t+h}^{t+h+1}/\mathrm{Y}_{t}^{t+h+1}), \tag{4.1}$$

where Y_t^v is the original variable at time t from vintage v. If $h = 1$, we drop the h notation, so $y_t^{1,v} = y_t^v$.

4.4.2 FRED-MD

As benchmark, we include factors from a large set of macroeconomic variables to capture what is already in the forecaster's information set. In particular, we consider the FRED-MD data set (McCracken and Ng, 2016).[19] The FRED-MD covers a large number of macroeconomic variables that closely resembles the so-called Stock-Watson data set, which is typically used to represent a big data setting in macro forecasting. Further, the factors from the set have been shown to hold predictive power for the output variables, see e.g. Stock and Watson (2002a).

An advantage of the FRED-MD data set is that it is freely available, regularly updated, and maintained by the St. Louis Fed, whereas the Stock-Watson data set contains proprietary data. A downside is that the data vintages are available only from 1999M08, leaving us with a limited sample. Instead, we use the latest vintage at the time of accessing the data set.[20] The results can loosely be thought of as providing an upper bound on using real-time FRED-MD data.

4.4.3 Real-time forecasting design

The aim is to forecast output in real-time, emulating reality as close as possible. Therefore, we use the 'real-time vintage' approach (Koenig et al., 2003; Clements and Galvão, 2013), instead of using end of sample data. That is, we use the first release for estimation, matching the release maturity of the leading observations on the left-

[19] Available at Michael McCracken's website: `https://research.stlouisfed.org/econ/mccracken/fred-databases/`.

[20] The 2017M09 vintage.

and right hand side,

$$y_{t+h}^{h,t+h+1} = \beta_0^h + \sum_{j=1}^{p} \beta_j^h y_{t-j+1}^{t+1} + \sum_{j=1}^{m} \boldsymbol{\gamma}_j^h \boldsymbol{f}_{t-j+1} + \varepsilon_{t+h}^{h,t+h+1}, \tag{4.2}$$

for $t = 1, \ldots, T-h$, where $y_t^{h,v}$ is the horizon h observation of the target variable at time t from vintage v, and $\boldsymbol{f}_t$ is the vector of k factors at time t. The lags on the right hand side are from the same vintage as the first lag, and can be lightly revised. For example, the second lag will be the second release of that observation. The factors are not revised, and therefore denoted without a vintage superscript. Financial data is available instantly, but this is not the case for the survey data or the forecast error based measures. To be consistent and to ensure that the information is available, we use lagged values for all uncertainty measures.[21]

4.4.4 Models

We consider the following forecasting models. First, as benchmark, we consider an autoregressive (AR) model,

$$y_{t+h}^{h,t+h+1} = \beta_0^h + \sum_{j=1}^{p} \beta_j^h y_{t-j+1}^{t+1} + \varepsilon_{t+h}^{h,t+h+1}, \tag{4.4}$$

with $0 \leq p \leq 6$ lags. The number of lags p is selected recursively using BIC. This model was usually better than if p is fixed, but results are similar if the lag length is set at three or four for horizons shorther than a year, and zero to one lag for horizons longer than or equal to a year.

Second, the predictive ability of each uncertainty measure is considered individually

$$y_{t+h}^{h,t+h+1} = \beta_0^h + \sum_{j=1}^{p} \beta_j^h y_{t-j+1}^{t+1} + \sum_{j=1}^{m} \delta_{i,j}^h z_{i,t-j+1} + \varepsilon_{t+h}^{h,t+h+1}, \tag{4.5}$$

[21] Equation (4.2) is slightly different for MTS, because there is a two month rather than a one month reporting lag. The equation for MTS becomes

$$y_{t+h}^{h,t+h+2} = \beta_0^h + \sum_{j=1}^{p} \beta_j^h y_{t-j+1}^{t+2} + \sum_{j=1}^{m} \boldsymbol{\gamma}_j^h \boldsymbol{f}_{t-j+1} + \varepsilon_{t+h}^{h,t+h+1}, \tag{4.3}$$

where we keep the same timing for the factors as in (4.2).

where $z_{i,t}$ is the i-th uncertainty measure at time t, and $0 \leq p \leq 6$, and $m = 1$. The maximum number of lags m is in line with literature on diffusion forecasting, see e.g. Stock and Watson (2002a) or McCracken and Ng (2016). The number of lags p is selected recursively using BIC.

Third, we consider the factor model

$$y_{t+h}^{h,t+h+1} = \beta_0^h + \sum_{j=1}^{p} \beta_j^h y_{t-j+1}^{t+1} + \sum_{j=1}^{m} \boldsymbol{\gamma}_j^h \boldsymbol{f}_{t-j+1} + \sum_{j=1}^{m} \boldsymbol{\phi}_j^h \boldsymbol{g}_{t-j+1} + \varepsilon_{t+h}^{h,t+h+1}, \tag{4.6}$$

where $\boldsymbol{f}_t$ is the vector of k uncertainty factors at time t, and $\boldsymbol{g}_t$ is the vector of r FRED-MD factors at time t, $m = 1$, and $0 \leq p \leq 6$, again selected using BIC. The number of FRED-MD factors r is estimated using the Bai and Ng (2002) type 2 information criterion. We consider varieties where either k and/or r are non-zero. This allows us to evaluate the forecasting power of uncertainty factors in isolation, as well as additional to factors from a large set of macroeconomic variables that is known to have been successful in predicting output, see e.g. Stock and Watson (2002a). Following the results in Section 4.3, we consider factor models with a fixed number of one up to three factors. This allows us to assess the relevance of an additional factor. The forecast results presented in the following sections are based on $m = 1$ lag. Results are very similar when the lag of factors/measures $1 \leq m \leq 3$ is selected recursively (in combination with p) using BIC.

Sample

Forecasts are constructed by recursively estimating the models. That is, at each time period t, we first estimate the factors and models – and apply model selection – using data from 1986M04 to time $t - h$. Earlier (initial) observations are included if the lag order is larger than one. We start in 1986M04 because that is the first period where at least three months of data is available for all variables. Recursive estimation is in line with other diffusion forecasting literature, see e.g. Stock and Watson (2002a) and McCracken and Ng (2016). Using all available information improves convergence of the factor estimates. Moreover, results using a rolling window do not indicate the presence of a structural break, while the forecasting results deteriorate in some cases, see Section 4.7.4.

Second, the parameter estimates and time t observations are used to construct the forecast for the $t+h$ value $y_{t+h}^{h,t+h+1}$. We imagine a forecaster, who starts forecasting in January 1991. The first forecast is made for period 1998M12 + h, and the final one for 2016M12, for the horizons of 1 (nowcast), 3, 6, 12, and 24 months. To clarify, the first forecast is constructed using data up to and including 1998M12, so using the vintages available in 1999M01 (1999M02 for sales), to predict the period 1998M12 + h. This yields a sample of 152 initial in-sample and $217-h$ out-of-sample observations.

4.4.5 Quantile forecast

Economic uncertainty is expected to mainly affect the left tail of the distribution of the coincident variables. Additional to forecasting the mean, we therefore also forecast quantiles, with a focus on the lower tail quantiles. Following Giglio et al. (2016), we are interested in the quantiles on the shocks rather than the growth rates. The shocks are computed based on the AR(p) model (see Equation (4.4)), using real-time vintage data as in Equation (4.2). Then we define the first release residuals $\hat{u}_{t+h}^{h,t+h+1}$ as the shocks, which are used in estimation of the quantile forecasting models. When FRED-MD factors are included as predictors, the results change quantitatively, but there is little difference qualitatively.

To evaluate the quantile forecast we use the prebenchmark residuals as actuals, which we define as

$$\hat{u}_{t+h}^{h,PBM} = y_{t+h}^{h,PBM} - (\hat{\beta}_0^h + \sum_{j=1}^{p} \hat{\beta}_j^h y_{t-j+1}^{t+1}), \tag{4.7}$$

with p the same value as in Equation (4.4) and $\hat{\beta}_j^h$ estimated from the model in Equation (4.4). This is consistent with how we define the actuals for the point forecasts in Section 4.4.3. Because the estimated (mean) model using first vintage data is an efficient estimate of the actual mean (Koenig et al., 2003), we can use it do 'demean' the actuals too to get the actual shocks. So the mean is consistent across estimation and evaluation.

Quantile regression is a semiparametric method that dates back to the seminal work by Koenker and Bassett (1978). The estimate of α-quantile q for data set y is the solution to the optimization

$$Q_\alpha(y) = \arg\inf_q \mathrm{E}[\rho_\alpha(y-q)], \tag{4.8}$$

where $\rho_\alpha(x) = \big(\alpha - \mathbf{1}(x \leq 0)\big)x$ the tick loss function, and q can be a linear function of (exogenous) regressors.

We consider the α-quantile forecast

$$Q_\alpha(\hat{u}_{t+h}^{h,t+h+1}|\Omega_t) = \psi_{\alpha,0}^h + \sum_{j=1}^{m} \boldsymbol{\psi}_{\alpha,j}^h \boldsymbol{w}_{t-j+1}, \tag{4.9}$$

with Ω_t the information set at time t, and $\boldsymbol{w}_t$ the regressors at time t with $1 \leq m \leq 3$. As regressors, we consider the uncertainty measures individually, the uncertainty factors, and FRED-MD factors, equivalent to regressors in the models for the mean (Section 4.4.4). The historical quantile estimate $\hat{q}_{\alpha,t}$, using data up to and including time t, is the benchmark. For values of α, we focus on 0.2, but we also check results for 0.1, 0.5 – the median – and 0.8. The in-sample analysis does cover the full range of quantiles, from 0.05 up to 0.95. The parameters $\boldsymbol{\psi}_{\alpha,j}^h$ are estimated using the interior point algorithm. For a review on quantile forecasting, see Komunjer (2013).

Adrian et al. (2018) present an easy and intuitive way to move from quantile forecasts to a density forecast, by matching the quantiles with those of a skewed Student's t, or skew-t, density at each time period. It is a natural way to allow for nonlinearity and asymmetry. We are looking into extending our analysis to include this.

4.4.6 Evaluation

A question with real-time data is which values to use as 'actuals' to evaluate the forecasts. Preferably, these are the true values that are no longer revised. This is impossible however, because of benchmark revisions. For example, due to a change of the index year, annual updating following the consensus numbers, and redefinitions or measurement changes. Broadly, there are three alternatives. One option is to use the x-th release observations $y_{t+h}^{h,t+h+x}$ for some $x \geq 1$. Many empirical studies use x-th release data to evaluate their forecasts, see e.g. Romer and Romer (2000), Groen et al. (2013) and D'Agostino et al. (2013). Selecting x requires some knowledge on the revision process. For quarterly data the second revision (third release) is often used because this is usually the 'final' revision from the statistical agency. A second option is to use the final vintage observations $y^{h,T+1}_{t+h}$. The final vintage is the most recent publication of the numbers. For example Koenig et al. (2003) and Clements and Galvão (2013) use the vintage published about a year and a half after the end of their sample. An advantage is that it incorporates the latest available information and are currently closest to the true values as a single time series. The third option is to use the prebenchmark observations $y^{h,\mathrm{PBM}}_{t+h}$ as actuals. Prebenchmark values are the final observation before the first benchmark after a first value for a given date has been reported. Some prebenchmark observations are subject to regular revisions, while x-th release and final vintage observations are subject to benchmark revisions. In contrast to regular non-benchmark revisions, benchmark revisions can and should not be predictable to the forecaster (Croushore, 2006). The actuals should represent the forecasters' target, rather than be closest to the current truth. Since the prebenchmark values are most consistent with what the forecaster aims to predict, we opt to use those as actuals.[22]

We select the following measures to evaluate the performance. The relative forecast accuracy of the mean forecasts is evaluated using the mean squared prediction error (MSPE). Similarly, the quantile forecasts are evaluated using the mean tick loss (MTL). Statistical significance is tested using one-sided Diebold and Mariano (1995) tests, where we test the null of equal predictive accuracy, versus the alternative of smaller loss compared to the benchmark model. Since the Diebold-Mariano test is defined for a general loss function, it can be used to test the significance for both the mean and quantile forecasts, with quadratic loss for the mean and tick loss for the quantiles.

[22]Benchmark revision dates are from the documentation of Philadelphia Fed's Real-Time Data Set for Macroeconomists (Croushore and Stark, 2001), from the Federal Reserve Board of Governors (`https://www.federalreserve.gov/releases/g17/`), from the Bureau of Labor Statistics (`https://www.bls.gov/web/empsit/cestn.htm#section7`), and from the Bureau of Economic Analysis (Page 1-10, note 22, of the November 2017 edition of the NIPA handbook, `https://www.bea.gov/resources/methodologies/nipa-handbook`). Further, we check for revisions in the data by looking at non-zero revisions of the sixth up to the twelfth release per vintage to identify remaining revisions. Though the reporting of revisions is quite accurate, we do identify some additional ones, mostly in the in-sample period.

Quite some econometric difficulties arise because of our setup of comparing nested models, estimated using an expanding window with real-time data, see Clark and McCracken (2013) for an overview. Clark and McCracken (2009) derive the limiting distribution of tests of equal predictive accuracy when data is subject to revisions. Their setting ignores benchmark revisions, which is in line with our data as we use pre-benchmark observations as evaluation values. However, their test is for comparing predictive accuracy in population, while we are interested in the finite sample performance. We follow the arguments by Faust and Wright (2013), as do Groen et al. (2013), by relying on the Monte Carlo evidence presented by Clark and McCracken (2013). Their simulation study shows that the Diebold-Mariano test statistic with standard normal critical values and the corrections by Harvey et al. (1997) yields satisfactory size, even for nested models.[23] Clark and McCracken's (2013) DGP ignores revisions, so we should still be careful with making strong statements.

For the quantile forecasts, we are also interested in testing absolute performance: whether the coverage is close what is expected from the quantile level. The most popular tests of absolute accuracy of quantiles are the Christoffersen (1998) tests. They test the unconditional coverage, dependence and the conditional coverage of the quantile forecasts by computing the first order Markov probabilities of a violation.

Additionally, we consider Engle and Manganelli's (2004) dynamic quantile (DQ) test. It tests the coverage conditional on Ω_{t-1}, the information set at time $t-1$. Define $e_t = \mathbf{1}(y_t \leq q_t) - \alpha$ the 'demeaned' hits, and the vector of k instruments $\boldsymbol{x}_t$, which are in the information set at time $t-1$. It may contain q_t or its lags and lags of e_t for example. The null hypothesis is $\mathrm{E}[u_t \boldsymbol{x}_t] = 0$. The out-of-sample DQ test statistic is given by

$$DQ_{OOS} = \big[(\sum_t e_t \boldsymbol{x}_t')(\sum_t \boldsymbol{x}_t \boldsymbol{x}_t')^{-1}(\sum_t e_t \boldsymbol{x}_t)\big]/[\alpha(1-\alpha)], \tag{4.10}$$

[23] Harvey et al. (1997) propose two corrections to the Diebold-Mariano test statistic. First, they use the knowledge that the forecast errors from h period ahead forecasts follow an MA(h) structure, to argue the use of a rectangular window in calculating the long-run variance of the loss difference. This has the downside that the resulting variance estimate is not guaranteed to be positive semi-definite. In those few cases that it occurs, we use the Newey-West variance estimator with a window of 1.5 times the forecast horizon. Second, to correct the size at longer horizons, the test statistic is $t_{\mathrm{HLN}} = \big(1 + T_p^{-1}(h-1) + T_p^{-2}h(h-1)\big)^{1/2} t_{\mathrm{DM}}$, where T_p is the number of out-of-sample observations and t_{DM} the Diebold-Mariano test statistic.

with α the quantile level. DQ_{OOS} follows a χ^2 distribution, with k degrees of freedom.[24] We use $\boldsymbol{x}_t = (1, q_t)$. This is equivalent to a Wald test on a quantile version of the Mincer-Zarnowitz regression.

4.5 Mean forecasting results

This section presents the mean forecasting results. The MSPEs relative to the benchmark AR(p) model (RMSPE) are presented for various horizons in Figure 4.1 and for subsamples of the three month forecasting horizon in Figure 4.2.[25] Filled symbols indicate significant predictive power according to a one-sided Diebold-Mariano test against the AR(p) benchmark at the 5% level.

First, consider the performance of the individual uncertainty measures – the gray circles. The uncertainty measures do not hold consistent significant predictive power for the coincident variables across the horizons. There is some promise in forecasting employment at the short horizon, with significant gains from 6% (FDISP, $h = 1$) to 28% (JLNm, $h = 3$) compared to the benchmark, and modestly so for sales at the short horizon, with gains up to a 5% significance level. Also, the majority beats the benchmark at the medium range for personal income, though not significantly so at the 5% level. Industrial production is hardest to predict using the uncertainty measures. Going into more detail, Table 4.B.1 shows that including JLN measures yields significant gains at shorter horizons for all variables but industrial production. The financial measures VXO, and OVX and CGU perform reasonably well for employment and personal income forecasting up to a year ahead. Other measures with some modest success are FDISP for sales and employment, and the measures based on the University of Michigan Survey on Consumers in forecasting personal income at the medium range.

Including uncertainty factors rather than uncertainty measures helps in the sense that the RMSPEs are smaller compared to most models with a single uncertainty

[24] The equation for DQ_{OOS} in Engle and Manganelli (2004) contains a small typo. The right hand side is incorrectly divided by the forecasting sample size. Manganelli's code implements the correct version, available at `http://www.simonemanganelli.org/Simone/Research.html`. Oddly, it seems like it has not been noticed, because Komunjer (2013) copies the mistake.

[25] To preserve space, we plot the RMSPEs rather than include tables. The tables can be found in Section 4.B.

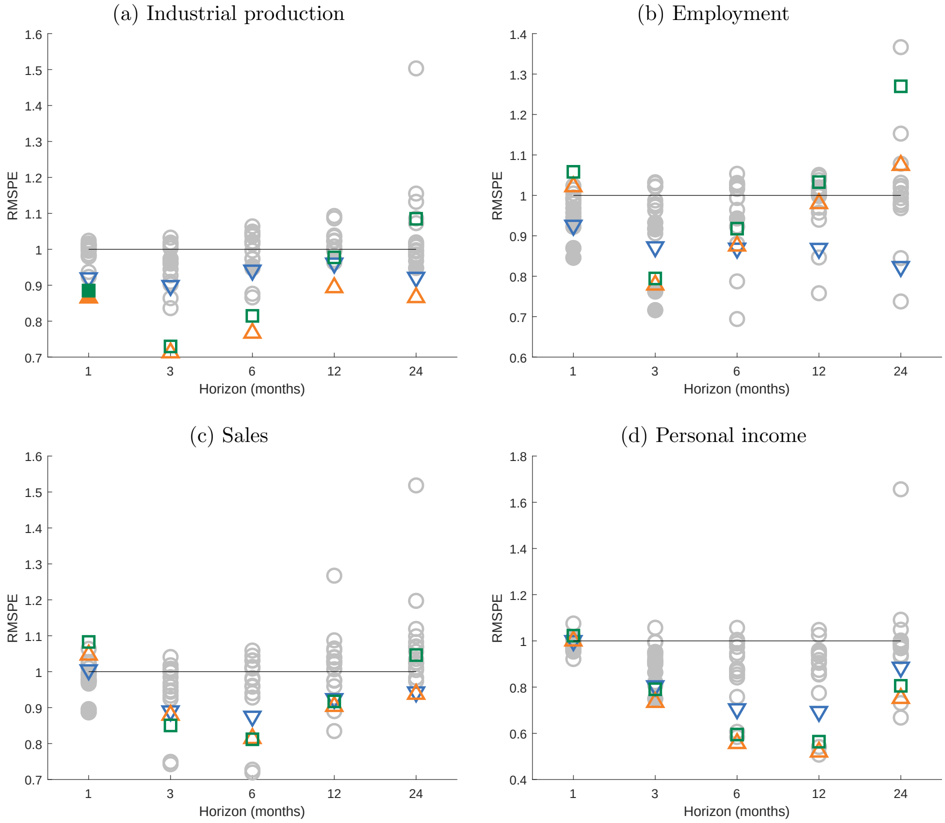
measure across the forecasting horizons, but it does not yield significant gains. The RMSPEs are close to those measures that it loads most heavily on: the JLN measures and the financial measures. Adding a second uncertainty factor decreases performance at the short horizon, increasing the RMSPE by about 3 percentage points on average, but increases the performance at the longer horizons reducing the RMSPE by 5 to 10 percentage points in many cases. It yields modest significance gains (at the 10% level) in forecasting employment and personal income at multiple horizons. Interestingly, the coefficient of the second factor is not as expected. It is not significantly different from zero at the shorter horizons, and significantly positive at longer horizons. This implies that if the consumer/media uncertainty increases, economic output is projected to increase. It does seem to help mainly at the longer horizons, suggesting a reversal effect. Alternatively, because the factor remains elevated after recessions (see Figure 4.1), it might be used statistically to capture the recovery period after the trough, the turning point at the end of a recession. An additional third factor leads to bad performance at longer horizons and no significant gains at the shorter horizon. Including at most two factors is consistent with results in the Section 4.3.

Turning to the FRED-MD factor model, Figure 4.1 shows that it is a tough benchmark to beat. The RMSPEs are lower than most or all models with uncertainty measures, with values ranging from 0.52 (for PIX at $h = 12$) to 1.074 (for EMP at $h = 24$). The FRED-MD factors hold predictive power over the sample period, mostly for industrial production at the shorter horizon, see Figure 4.1a. There is also significance at the 10% level for a three month forecasting horizon. For other variables and horizons, the gains are mostly at the medium term horizon. The uncertainty factor models do outperform the FRED-MD factor models for employment, especially at the one month and longer horizons. The FRED-MD models cannot beat the benchmark here, with RMPSEs of 1.022 ($h = 1$) and 1.074 ($h = 24$). Combining the FRED-MD factors and the first two uncertainty factors does not seem to lead to superior forecasts over the other factor models, but is close to the best model. Sometimes it is much worse, such as at the 24 month horizon, see Figure 4.1.

For all considered models, gains in RMSPE come largely from beating the benchmark model during recessions, see Figure 4.2. The MSPE in those periods is in many cases less than 40% to 50% of the benchmark's loss. This is in line with previous literature using FRED-MD factors, see e.g. McCracken and Ng (2016). It is also visible from inspecting forecasts during recession periods in Figure 4.B.1. Models with FRED-MD and uncertainty factors are able to match the depth of the recession. This translates to large gains during these periods compared to the other models, in particular for the FRED-MD factor models, see the cumulative sum of squared errors differential in Figure 4.B.2.

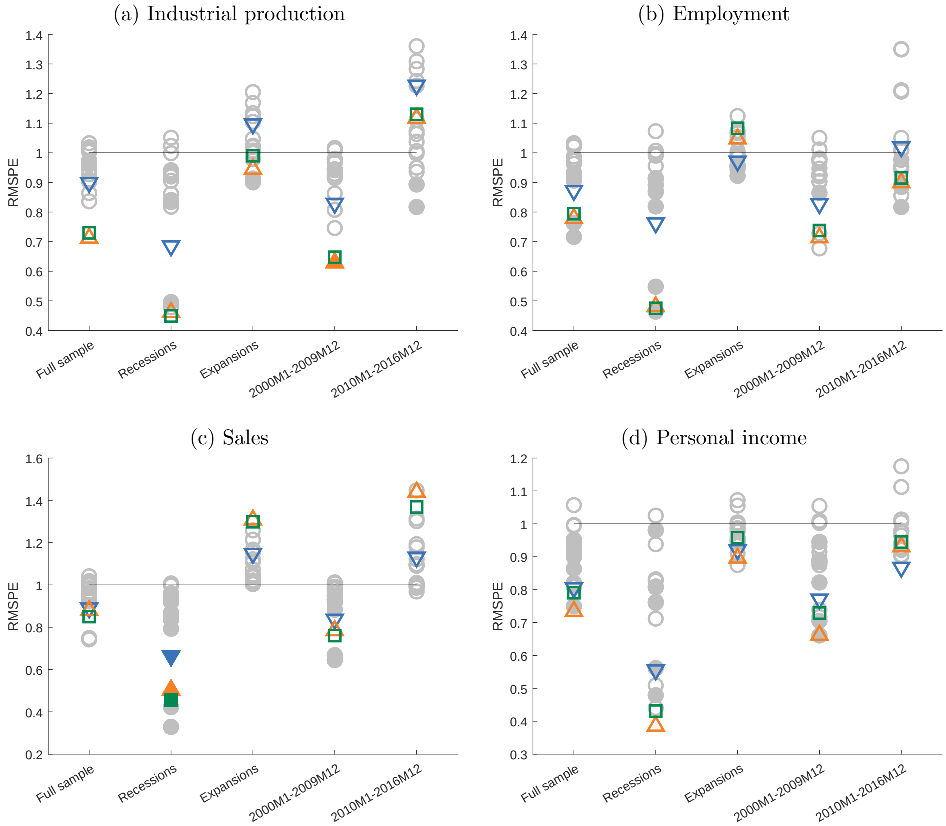
Overall, the results show that the forecasting power of uncertainty measures for the mean of the coincident variables is limited – though there is some promise in forecasting employment, and perhaps personal income.

Figure 4.1: RMSPE by forecast horizon



The figures present the relative mean squared prediction error (RMSPE), with the AR(p) model as benchmark, for multiple forecast horizons (in months). Gray circles are models with a single uncertainty measure, blue down-pointing triangles are models with two uncertainty factors, orange up-pointing triangles are models with FRED-MD factors, and green squares are models with both two uncertainty factors and FRED-MD factors. Filled symbols indicate significance of the one-sided DM test against the AR(p) model at the 5% significance level.

Figure 4.2: RMSPE by evaluation period, $h = 3$



The figures present the relative mean squared prediction error (RMSPE), with the AR(p) model as benchmark, for multiple evaluation periods. The forecast horizon is three months. Gray circles are models with a single uncertainty measure, blue down-pointing triangles are models with two uncertainty factors, orange up-pointing triangles are models with FRED-MD factors, and green squares are models with both two uncertainty factors and FRED-MD factors. Filled symbols indicate significance of the one-sided DM test against the AR(p) model at the 5% significance level.

4.6 Quantile forecasting results

The previous section shows that the out-of-sample forecasting power is limited when we consider a linear relationship for the mean. Most studies consider a linear relationship between economic uncertainty and output using a simple VAR, though some look at a nonlinear relationship. For example, a smooth transition model where the state depends on uncertainty (Jones and Enders, 2016), or only extreme values of the uncertainty index (Bloom, 2009).[26] Still, these studies focus on an impact on the mean rather than other parts of the density of output, such as particular quantiles.

Our in-sample quantile estimates show that there is substantial evidence that a linear relationship does not describe the relationship well, see Figure 4.1.[27] In particular, the impact of the first uncertainty factor is stronger at the lower quantiles and is close to zero at higher quantiles. It holds for all coincident variables, except personal income. The nonlinear effect is smaller at the one month horizon, but clear at the twelve month horizon. This is in line with how systemic risk affects quantiles of output shocks (Giglio et al., 2016). Therefore, we investigate the real-time out-of-sample predictive power of uncertainty measures for various quantiles of industrial production and employment shocks, with a focus on the lower quantiles.

The predictive performance is measured in relative MTL (RMTL), where values smaller than one indicate a smaller loss than the benchmark historical quantile. The results are presented for multiple horizons in Figure 4.2, for the three month horizon across subsamples in Figure 4.4, both for the quantile $\alpha = 0.2$. Figure 4.3 further presents full sample results for the three month horizon across different quantile levels.[28]

All individual uncertainty measures reduce the mean tick loss for employment. In particular the JLN measures, the financial measures, FDISP and MPU yield strongly significant gains – the same measures that provided gains in mean forecasting in the previous section. The gains are mostly in the range of 1% to 9% reduction in mean tick loss compared to the historical quantile. The improvements are mostly for the short horizon, though it is difficult to discern a clear consistent pattern across forecast horizons. FDISP and the JLN measures consistently improve significantly upon the benchmark, with gains from 7% at the shorter horizon up to 35% at $h = 12$ for JLN in predicting employment, but it is somewhat mixed for other measures.

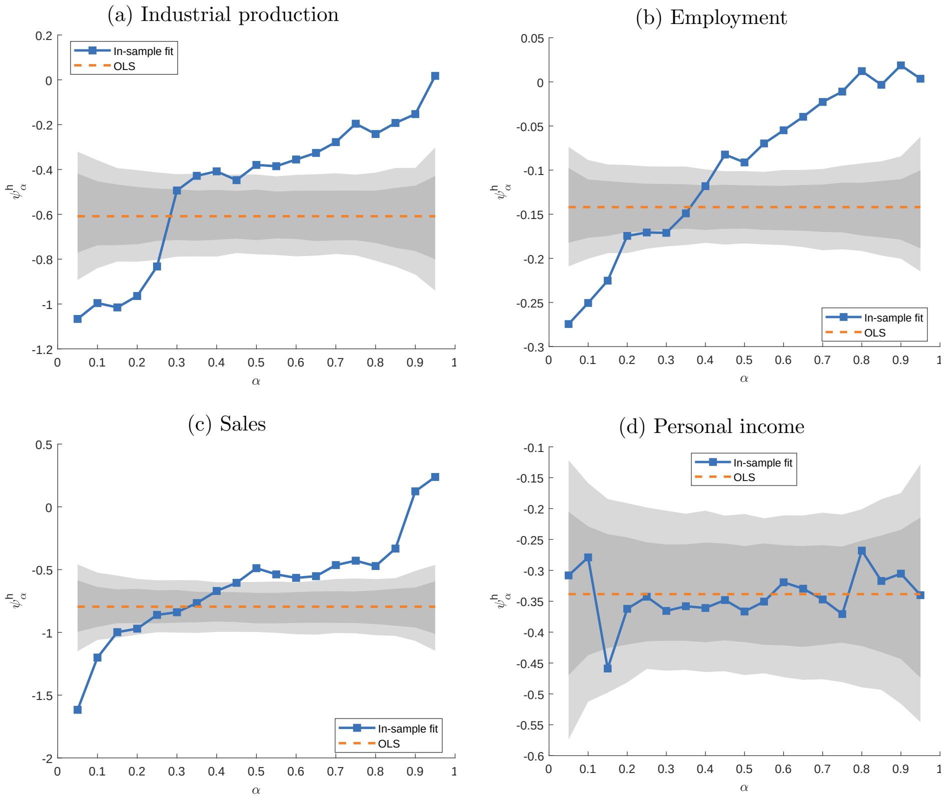
Neither the consumer confidence measures LLv and LLh, nor the newspaper based EPU(+) measure seem to provide much useful information in forecasting quantiles of

[26] Bloom (2009) defines uncertainty shocks as observations that are more than four standard deviations larger than the mean.

[27] Full sample estimates, for the period 1986M04–2016M12. Following Adrian et al. (2018), the confidence intervals are based on bootstraps under the null hypothesis of a linear model, using 1000 bootstrap samples.

[28] As with the mean forecasting results, we plot the RMTL to preserve space. The tables can be found in Appendix 4.C.

Figure 4.1: In-sample quantile estimates for first factor



The figure presents the in-sample quantile estimates for the full sample of the first uncertainty factor on industrial production, employment, sales, and personal income, at the three month forecasting horizon. The shaded areas are bootstrapped confidence bounds at the 90% and 95% level for a linear model based on 1000 bootstrap samples.

output shocks, see Table 4.C.1. RMTLs range from 0.835 to 1.085 and fail to beat the benchmark 41 out of a 100 times. Still, adding the second consumer/media confidence factor provides significant gains at the longer horizons for employment, and sizable gains of about 5% ($h = 12$) to 10% ($h = 24$) for other variables.

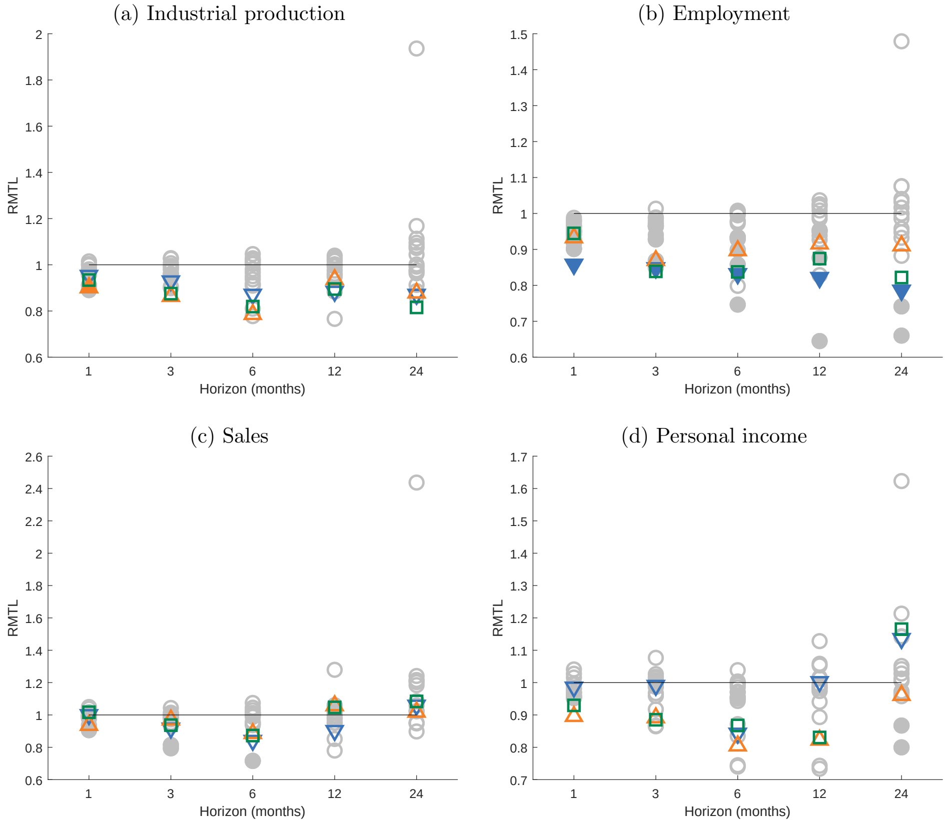
The results in forecasting sales are disappointing at the $\alpha = 0.2$ quantile level, except for JLNm and JLNf. If we move further in the left tail, the results significantly improve for forecasting sales, see Figure 4.3c. All but one beat the benchmark, many by more than 10%. There are some gains in personal income but it is hard to distinguish a pattern across the horizons. TYVIX, OVX and CSDRsic yield strong significant gains when forecasting personal income, but only at the one month horizon.

The uncertainty factor models perform quite well, in line with the uncertainty measures. They outperform the historical quantile consistently, except at the long horizon for sales and personal income, see Table 4.C.1. The two factor model yields strongly significant gains for employment across the horizons, with gains up to 21%. For other variables the RMTL are smallest in magnitude at the three to twelve month horizon. Most of the gains are from the recession periods, which are not captured by the historical quantile, see Figure 4.4, Figure 4.C.1 and Figure 4.C.2. Furthermore, the forecasting power of the uncertainty factor models is better than most of the individual uncertainty measures. Only the JLN measures match or outperform the factor models' performance. The one factor model significantly reduces the mean tick loss for shorter horizons up to a 12 month horizon. Adding a second factor further increases the performance. Separate DM tests comparing the quantile forecasts of the uncertainty factor model with two factors versus one factor show that there are significant gains in tick loss of adding the second factor. The gains are mostly at the longer horizons of 12 and 24 months. There are also significant gains for employment at the one month horizon.

Comparing across quantile levels, we observe in Figure 4.3 that adding uncertainty measures or factors yields smaller average tick losses relative to the benchmark as you move further into the left tail. Even quantile forecasts of industrial production improve when $\alpha = 0.1$. The RMTLs for $\alpha = 0.1$ are 0.84 (IP), 0.75 (EMP), 0.84 (MTS) and 0.94 (PIX). These reductions in tick loss are significant for all but personal income, and strongly significant for employment. Vice versa, forecasting the upper quantile using uncertainty factors is not recommended, with the uncertainty factor models and most uncertainty measures resulting in RMTLs larger than one. The results for PIX are mixed, which is not surprising given that we did not find evidence of non-linearity in our forecasting results. Some measures yield sizable tick loss gains at the lower quantiles. It is a bit puzzling that more measures yield significant gains at the upper quantile ($\alpha = 0.8$) though. Also, the relatively high RMTL in Figure 4.4 indicate a lack of predictive power of economic uncertainty for downward movements of personal income.

The results provide substantial evidence that the uncertainty factors contain information that is not in the FRED-MD factors that is useful for predicting quantiles of output. In particular gains for employment are large and strongly significant. This is in line with the in-sample estimate of a more negative coefficient of the first factor at the lower quantiles in Figure 4.1. Changing the perspective, adding FRED-MD factors to uncertainty factors deteriorates the forecasting performance at most horizons, perhaps due to extra parameter uncertainty. This suggests that the FRED-MD factors capture level movements – mostly at the shorter horizon, while uncertainty factors are related to downside risk.

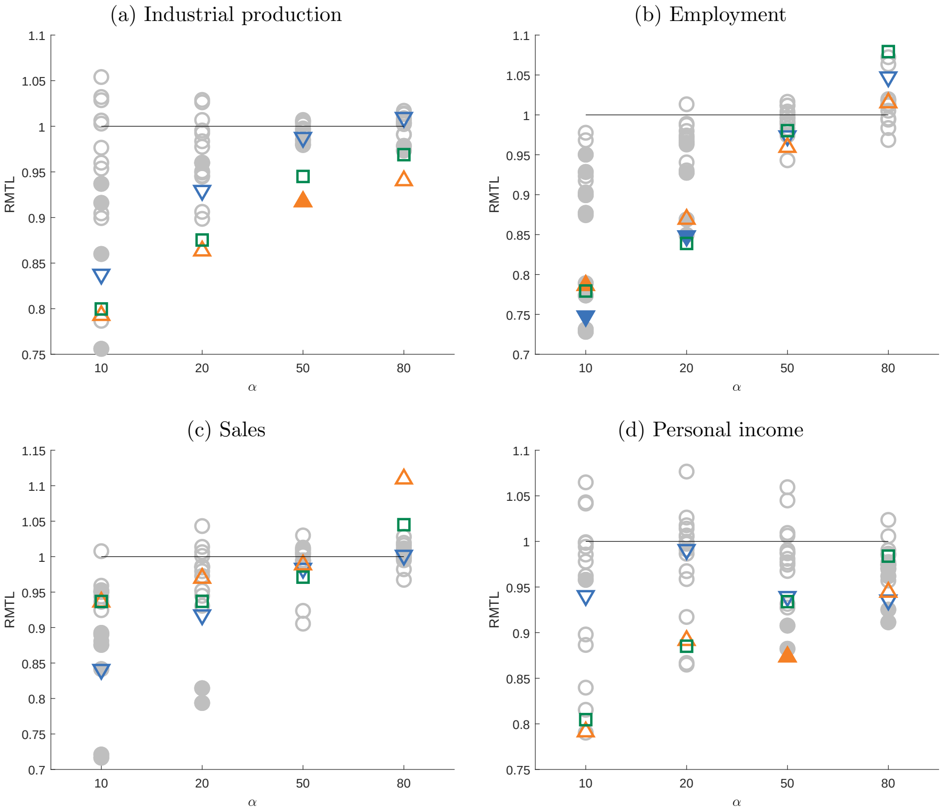
Figure 4.2: RMTL by forecasting horizon, $\alpha = 0.2$



The figures present the relative mean tick loss (RMTL), with the historical quantile as benchmark, for multiple forecast horizons. The quantile level is $\alpha = 0.2$. Gray circles are models with a single uncertainty measure, blue down-pointing triangles are models with two uncertainty factors, orange up-pointing triangles are models with FRED-MD factors, and green squares are models with both two uncertainty factors and FRED-MD factors. Filled symbols indicate significance of the one-sided DM test against the historical quantile at the 5% significance level.

Although our main goal is to improve upon the benchmark, the quantile forecasts ought to have coverage close to the expected value. The (unconditional) coverage is

Figure 4.3: RMTL by quantile level, $h = 3$

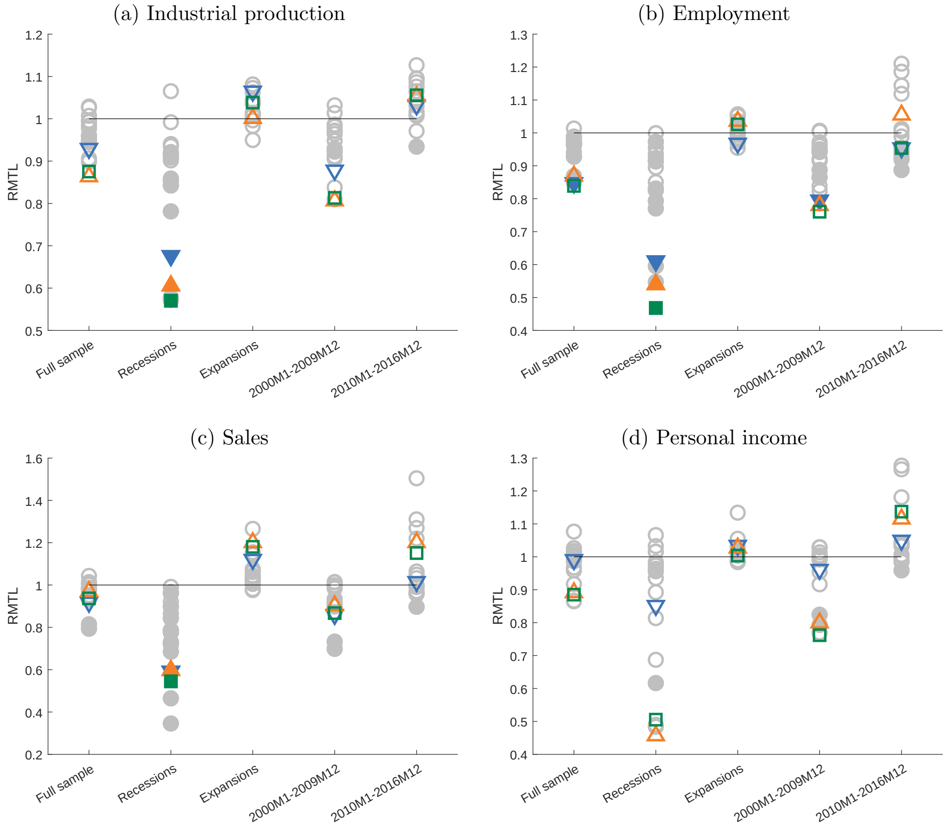


The figures present the relative mean tick loss (RMTL), with the historical quantile as benchmark, for multiple quantile levels α. The forecast horizon is three months. Gray circles are models with a single uncertainty measure, blue down-pointing triangles are models with two uncertainty factors, orange up-pointing triangles are models with FRED-MD factors, and green squares are models with both two uncertainty factors and FRED-MD factors. Filled symbols indicate significance of the one-sided DM test against the historical quantile at the 5% significance level.

good at shorter horizons, but deteriorates at longer horizons, see Table 4.1. The null of correct coverage is more frequently rejected – at the 5% level – at longer horizons as indicated by the markers in the table. This is consistent across target variables and quantile levels, see Table 4.C.4.

For forecasting horizons shorter than a year, the unconditional coverage is good, and so is the efficiency of the forecasts, as given by the DQ test results in Table 4.1. We would expect dependency in the violations for the historical quantile, but the FRED-MD factors and uncertainty measures or factors do not provide much better results. The null of correct conditional coverage is rejected at the 5% level at a one

Figure 4.4: RMTL by evaluation period, $h = 3$ and $\alpha = 0.2$



The figures present the relative mean tick loss (RMTL), with the historical quantile as benchmark, for multiple evaluation periods. The forecast horizon is three months and the quantile level is $\alpha = 0.2$. Gray circles are models with a single uncertainty measure, blue down-pointing triangles are models with two uncertainty factors, orange up-pointing triangles are models with FRED-MD factors, and green squares are models with both two uncertainty factors and FRED-MD factors. Filled symbols indicate significance of the one-sided DM test against the historical quantile at the 5% significance level.

year horizon for all models, due to dependence in the quantile violations. They are clustered around recessions, with hit rates up to 0.7 and 0.8 during those periods. Figure 4.C.1 shows that the quantile forecasts are late for the recessions, but also miss the (size of the) recovery period at the 12 month horizons. This recovery peak is much smaller for the three month horizon, see Figure 4.C.1.

Not even the historical quantile is able to capture the unconditional coverage well at longer horizons, with hit rates often higher than 30%. This is partly due to that the data is not described well by the AR(p) model to get the shocks at the longer horizons. The smoothness at these longer horizons (due to the overlapping periods in the growth) is hard to capture with higher frequency variables and a limited lag structure. A possible solution might be to allow for more lags. Additionally, the in-sample period contains only one recession, while the out-of-sample period contains two recession periods, including the financial crisis which is more severe compared to the previous ones. This makes it very challenging for quantile forecasting models to perform well. Though expanding the information set with economic uncertainty measures reduces the tick loss, the results show that it is difficult to accurately forecast quantiles.

4.7 Robustness checks

To confirm the robustness of our results, we repeat the factor analysis and real-time forecasting exercise using (i) a longer sample but smaller set of uncertainty measures, (ii) an unbalanced panel to include Scotti and GTU, (iii) a sample excluding the JLN and EPU measures such that only variables strictly available in real-time are included, and (iv) using a rolling rather than an expanding window. Detailed results are not reported, but are available upon request.

4.7.1 Longer sample

The current sample is relatively modest with the first forecast made in 1999M01. This choice is a trade-off between number of measures, and sample length. If we favor the sample length, we can start the in-sample period in 1976M11, adding almost 10 years of data. It is possible to go back even further in time for industrial production and employment – starting in 1964M11. The first forecast is made in 1988M01 for sales and personal income. For industrial production and employment, we start in 1977M01. The longer sample does reduce the available number of measures to nine: the volatility measures VXO and CGU, the cross-sectional measures CSRD and CSDRsic, the business forecast dispersion FDISP, the news paper based EPU, and the three JLN measures. Starting in 1964M11 excludes FDISP from this list. There is still a reasonable balance across the measure types.

The principal component analysis yields one common factor, which explains 51.5% of the total variation. It can be interpreted as average economic uncertainty and the dynamics are similar to the one from the shorter panel. The explanatory power is stable over time, but slightly increases during the financial crisis. The second factor of media/consumer uncertainty is not found in this data set. This is not surprising, because EPU is the only media or consumer uncertainty measure in the long panel.

The forecasting results for the mean are worse than for the short panel. We find some gains at the one month horizon, but the RMSPEs are above one for horizons of six months and longer. The results are also worse for personal income. The gains are still positive, but no longer significantly so. The gains come mostly from the financial crisis period.

The quantile forecasts indicate that the gains for the one factor model are comparable to those for the short panel. Further, if we split the performance in ten year periods, the gains are consistent across these periods. The magnitude of gains varies though, as the quantile forecasts are mainly improved during recessions. For industrial production, the historical quantile is actually better than the one factor model, again in line with the results for the shorter panel. For sales, the forecast accuracy for the lower quantiles and a short horizon improves upon the shorter sample in terms of statistical significance. Overall, the results for the long panel confirm the general findings for the shorter but wider panel.

4.7.2 Unbalanced panel

We can expand the set of uncertainty measures if we allow for an unbalanced panel as input for the factor analysis. By treating the unobserved values as missing, it is straightforward to use an EM algorithm to estimate the factors (Stock and Watson, 2002a). The Scotti uncertainty measure and the Google trends based GTU index are included in the data set after five years of data is available.

The factor structure is largely unchanged, and the interpretation is not affected. Similarly, the forecasting results do not change much. Quantitatively, the results are very close, though using the balanced panel seems to yield slightly smaller losses. The robustness in this unbalanced panel setting is perhaps not surprising, given that

Table 4.1: Hit rates by forecast horizon

Horizon	1	3	12	1	3	12
	Industrial production			Sales		
HQ	0.236†§‡	0.252†	0.302	0.167† *	0.215† *	0.224†
Factor models						
$r = 1$	0.259 §‡*	0.266 *	0.327	0.181†§‡*	0.201† *	0.244†
$r = 2$	0.255 §‡*	0.285	0.395	0.185†§‡*	0.224† *	0.298
$r = 3$	0.241†§‡*	0.224† *	0.361	0.167†§‡*	0.192† *	0.283
FRED-MD	0.236†§‡*	0.234† *	0.327	0.162†§‡*	0.234†	0.259
FRED-MD and $r = 2$	0.227†§‡	0.229† *	0.459	0.171†§‡*	0.243†	0.342
Uncertainty measures						
VXO	0.259 §‡*	0.276	0.346	0.181†§‡*	0.229† *	0.249†
TYVIX	0.236†§‡	0.238†	0.298	0.190†§‡*	0.252† *	0.263
OVX	0.255 §‡	0.234† *	0.283	0.153†§‡*	0.206† *	0.254† *
CGU	0.250†§‡*	0.262 *	0.302	0.190†§‡*	0.262	0.283
CSDR	0.264 §	0.257 *	0.351	0.190† ‡*	0.262 *	0.293
CSDRsic	0.232† ‡	0.234†	0.356	0.185†§‡*	0.243† *	0.361
FDISP	0.255 §‡*	0.238† *	0.332	0.185†§‡*	0.243† *	0.263 *
LLv	0.222†§‡*	0.229† *	0.298	0.153†§‡*	0.196† *	0.259
LLh	0.232†§‡*	0.229† *	0.298	0.153†§‡*	0.192† *	0.244†
EPU+	0.264 §‡	0.243† *	0.327	0.157†§‡	0.201†	0.288
EPU	0.236†§‡*	0.206† *	0.302	0.181†§‡*	0.201† *	0.244†
MPU	0.269 §	0.248† *	0.332	0.213†§‡*	0.229† *	0.234†
JLNm	0.250†§‡*	0.243† *	0.249† *	0.111 §	0.168† *	0.229†
JLNf	0.241†§‡*	0.276	0.317	0.185†§‡*	0.220† *	0.244†
JLNr	0.255 §‡*	0.234† *	0.298	0.153†§‡*	0.196† *	0.224† *
	Employment			Personal income		
HQ	0.181† *	0.187† *	0.263	0.148† *	0.202† *	0.256
Factor models						
$r = 1$	0.148†§‡*	0.196† *	0.278	0.148†§ *	0.226† *	0.236† *
$r = 2$	0.157†§‡*	0.201† *	0.273	0.124	0.202† *	0.292
$r = 3$	0.162†§‡*	0.173† *	0.298	0.148†§	0.188†	0.261
FRED-MD	0.167†§‡*	0.206† *	0.254†	0.143 §	0.178†	0.281
FRED-MD and $r = 2$	0.204†§‡*	0.220† *	0.337	0.148†§‡	0.178† *	0.286
Uncertainty measures						
VXO	0.144	0.196† *	0.273	0.143 *	0.231† *	0.281
TYVIX	0.176† *	0.196† *	0.278	0.162† *	0.226† *	0.271
OVX	0.157†	0.182† *	0.254† *	0.143 §	0.197† *	0.256 *
CGU	0.139	0.201† *	0.302	0.162†	0.245†	0.327
CSDR	0.171† *	0.215† *	0.293	0.157†§‡*	0.236† *	0.271
CSDRsic	0.227† *	0.196† *	0.298	0.181†§‡*	0.226† *	0.251†
FDISP	0.176† *	0.196† *	0.288	0.148† *	0.221† *	0.276
LLv	0.153† *	0.182† *	0.283	0.143 § *	0.173† *	0.246† *
LLh	0.162† *	0.168† *	0.283	0.138 *	0.168† *	0.226† *
EPU+	0.171† *	0.173† *	0.283	0.143	0.216†	0.256
EPU	0.153† *	0.178† *	0.268	0.143	0.192† *	0.236† *
MPU	0.176† *	0.187† *	0.278	0.148† *	0.212† *	0.256
JLNm	0.120	0.159†	0.205† *	0.138 *	0.173† *	0.196†
JLNf	0.157†	0.182† *	0.244† *	0.148†§ *	0.212† *	0.271
JLNr	0.144	0.178† *	0.234† *	0.143 *	0.173† *	0.231† *

The table presents hit rates for various forecasting horizons, for the full sample and a quantile level $\alpha = 0.2$. † denotes no rejection of the null hypothesis of correct unconditional coverage, § denotes no rejection of the null hypothesis of independence, ‡ denotes no rejection of the null hypothesis of correct conditional coverage, and * denotes no rejection of the null hypothesis of correct coverage conditional on an intercept and the quantile estimates q_t, all at a 5% significance level. See Table 4.A.1 for an explanation of the abbreviations.

these are only two extra uncertainty measures. Still, it is a confirmation of the factor structure that we find in the balanced panel.

4.7.3 Exclude EPU and JLN

The JLN and EPU measures are not measured in real-time, strictly speaking. Here, we check their impact, by removing them from the sample. Because the revisions in the EPU measures are supposed to be small, whereas it is unclear how revisions affect JLN, we also consider only removing the JLN measures from the sample of uncertainty measures.

Without JLN and EPU measures, the factor structure remains strong. The first factor is still identified as average economic uncertainty and explains about 39%. Moreover, the correlation with the factor when all measures are included is 94%. Though the second factor cannot load on the EPU measures, it does load heavily on the consumer confidence measures LLv and LLh, and has an 87% correlation with the second factor from Section 4.3. The main difference with Section 4.3 is that the third factor drops out. But it was hard to assign a label to it anyway. Instead, the fourth factor (mainly explaining FDISP) becomes the third most important factor.

The mean forecasting results are very similar to those in Section 4.5. The quantile forecasting results are affected quantitatively. The tick loss is larger when JLN (and EPU) measures are excluded. The largest effect occurs when the JLN measures are excluded. In a way this is not too surprising, since they are some of the best performing individual uncertainty measures, see Section 4.6. Still, using the factors based on the set of uncertainty measures that excludes JLN (and EPU) measures yields significant gains at the lower quantiles of employment and sales. Some deterioration of the results could be expected simply because a relatively large part of the cross-section is removed. The main patterns still emerge, and we interpret this is as a sign of robustness.

4.7.4 Rolling window

Because of possible concerns of structural changes in the sample period, we also consider estimation results based a rolling window of 12 years.[29] The general results are similar, though some models' performance deteriorates substantially when estimated using the rolling window, especially at the longer horizons and for the factor models with both uncertainty and FRED-MD factors.

These results show that there is little to suggest a structural change in the factor structure of economic uncertainty measures. Using an expanding window is preferred then, to improve convergence of the factor estimates. We also allow the factors to be estimated using the full available sample (i.e. using a recursive window), while the relationship between the uncertainty measures/factors and the coincident variables is estimated using a rolling window in case of structural changes. This yields similar results, supporting the main findings.

4.8 Conclusion

Many economic uncertainty measures have been proposed over the last decade. We find that they share a factor structure. One common component is probably quite close to the target: economic uncertainty. A second component represents media/consumer uncertainty.

Despite the big interest in economic uncertainty measures and its relationship with the business cycle, they do not have much predictive power for the mean of coincident variables. There is some predictive power in other parts of the density, in particular for employment at lower quantiles. Most gains are during recessions. This suggests a nonlinear relationship between economic uncertainty and economic activity. In that light, Fan et al. (2017) propose a way to estimate a non-linear relationship using factor models that might be interesting to explore.

[29]This corresponds with the range that holds two full business cycles of an average length, according to the NBER business cycle dating committee.

4.A Additional data description

This appendix present the time series plots of the time series, and the correlation matrix for the uncertainty measures.

Table 4.A.1: Uncertainty measures

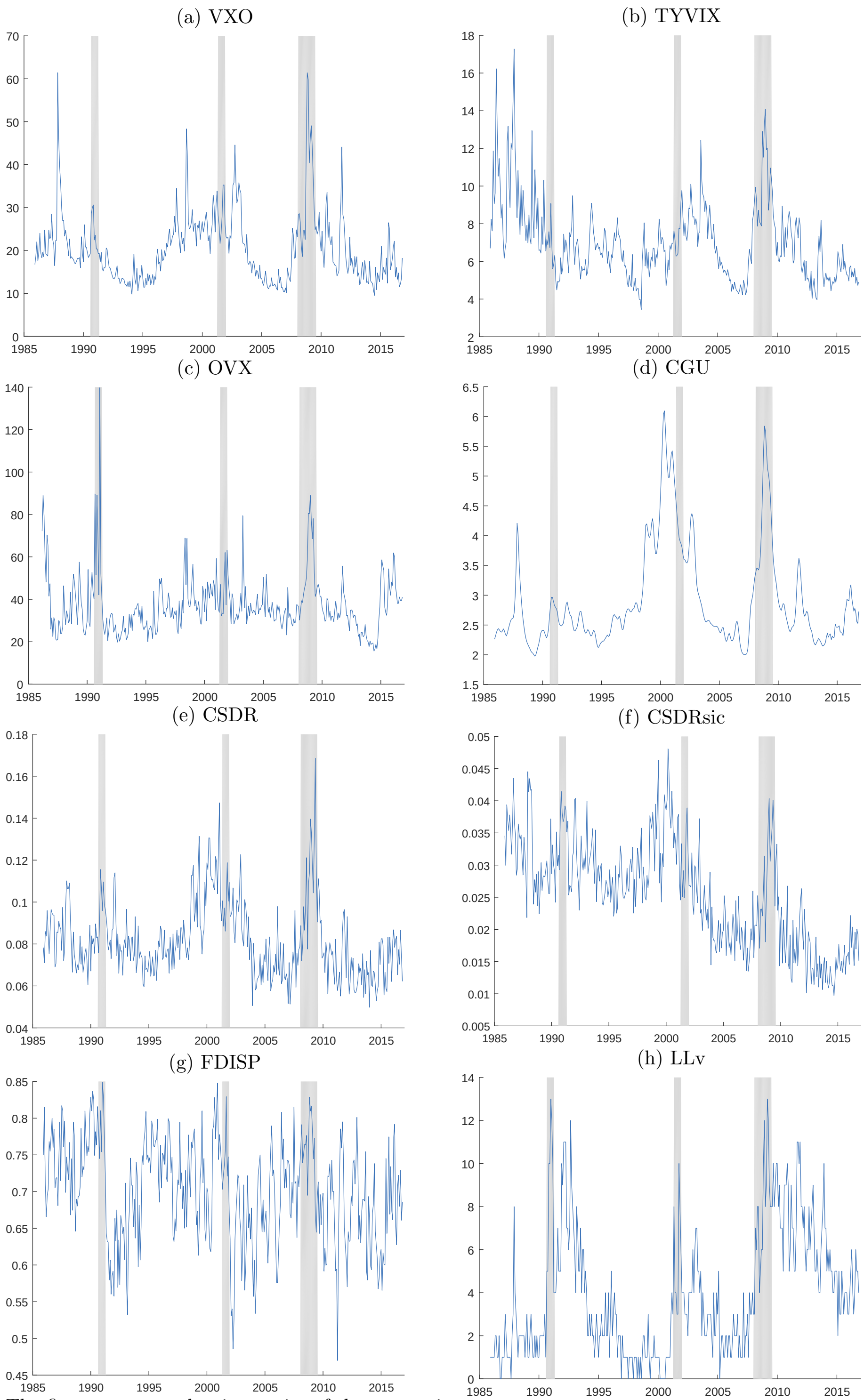
#	Abbreviation	Description	Reference	Source	Type	Start	End	Merge start
1	VXO	(1) Realized volatility of daily S&P500 returns; (2) End of month closing price of VXO	Bloom (2009)	CRSP (daily returns) and CBOE (VXO)	A	1962M7	2017M12	2004M1
2	TYVIX	(1) Implied volatility of 10 year bond; (2) End of month closing price of TYVIX	Choi et al. (2017)	Philippe Mueller/Andrea Vedolin's website (implied volatility) and CBOE (TYVIX)	A	1985M5	2017M11	2003M1
3	OVX	(1) Realized volatility of daily WTI returns; (2) End of month closing price of OVX	Kellogg (2014)	FRED (daily returns) and CBOE (OVX)	A	1986M2	2018M1	2007M5
4	CGU	Time-varying stock market uncertainty – removing common/predictable component; monthly average of daily observations	Chuliá et al. (2017)	University of Barcelona hosted website	A,E	1927M1	2016M11	
5	CSDR	Cross sectional standard deviation of stocks with 500+ month observations	Bloom (2009)	CRSP	B	1926M1	2016M12	
6	CSDRsic	Cross sectional standard deviation of stocks with 500+ month observations; mean over dispersion in SIC3 code	Bloom (2009)	CRSP	B	1926M1	2016M12	
7	FDISP	Ex ante forecast dispersion of general business conditions 6 months ahead; not seasonally adjusted	Bachmann et al. (2013)	Philadelphia Fed's Manufacturing Business Outlook Survey	B,D	1968M5	2018M4	
8	LLv	Consumer confidence (personal vehicle)	Leduc and Liu (2016)	Thomson Reuters/University of Michigan Survey of Consumers	D	1978M2	2017M8	

Table 4.A.1: Uncertainty measures (continued)

#	Abbreviation	Description	Reference	Source	Type	Start	End	Merge start
9	LLh	Consumer confidence (large households)	Fajgelbaum et al. (2017)	Thomson Reuters/University of Michigan Survey of Consumers	D	1978M1	2017M8	
10	EPU+	Economic policy uncertainty; combination of newspaper counts, tax code provisions and forecaster disagreement in the survey of professional forecasters (SPF)	Baker et al. (2016)	Economic Policy Uncertainty website	C,D	1985M1	2017M9	
11	EPU	(1) Historical economic policy uncertainty based on 6 to 10 major US newspapers; (2) Newspaper based economic policy uncertainty (original)	Baker et al. (2016)	Economic Policy Uncertainty website	C	1900M1	2017M9	1985M1
12	MPU	Monetary Policy Uncertainty; category of EPU, counts of articles additionally containing monetary policy related keywords	Baker et al. (2016)	Economic Policy Uncertainty website	C	1985M1	2017M9	
13	JLNm	Macroeconomic variables' forecast error variance based on large factor model, horizon=12	Jurado et al. (2015)	Sydney Ludvigson's website	E	1960M7	2017M6	
14	JLNf	Financial variables' forecast error variance based on large factor model, horizon=12	Jurado et al. (2015)	Sydney Ludvigson's website	E	1960M7	2017M6	
15	JLNr	Real activity variables' forecast error variance based on large factor model, horizon=12	Jurado et al. (2015)	Sydney Ludvigson's website	E	1960M7	2017M6	

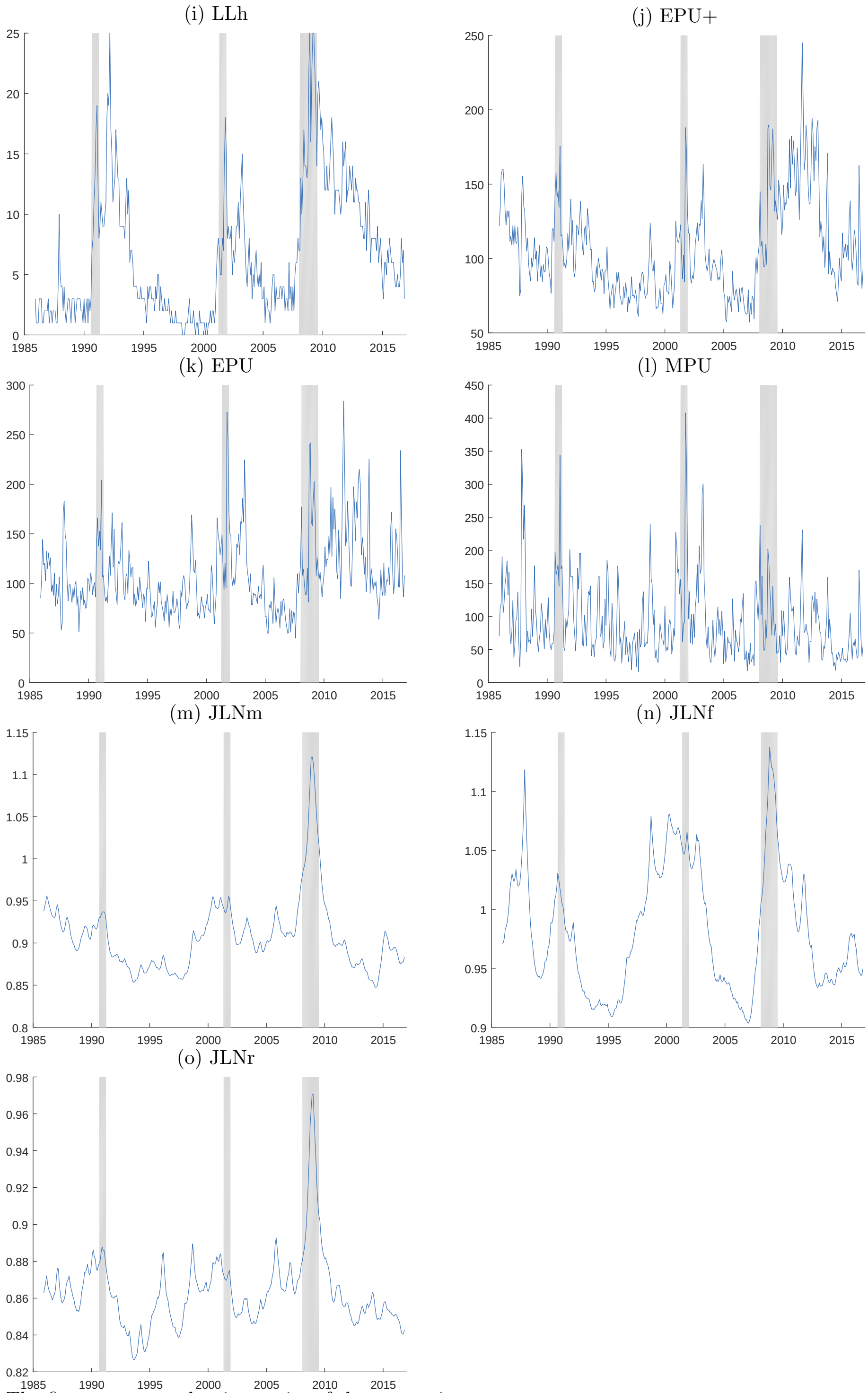
The table presents the uncertainty measures. Sources are University of Barcelona hosted website (Helena Chuliá's and Montserrat Guillén's affiliation, `http://www.ub.edu/rfa/uncertainty-index/`), Philippe Mueller's website (`https://sites.google.com/site/philippebmueller/`), Andrea Vedolin's website (`https://sites.google.com/site/andreavedolin/`), Economic Policy Uncertainty website (`http://www.policyuncertainty.com/index.html`), Sydney Ludvigson's website (`https://www.sydneyludvigson.com/`).

Figure 4.A.1: Uncertainty measure time series



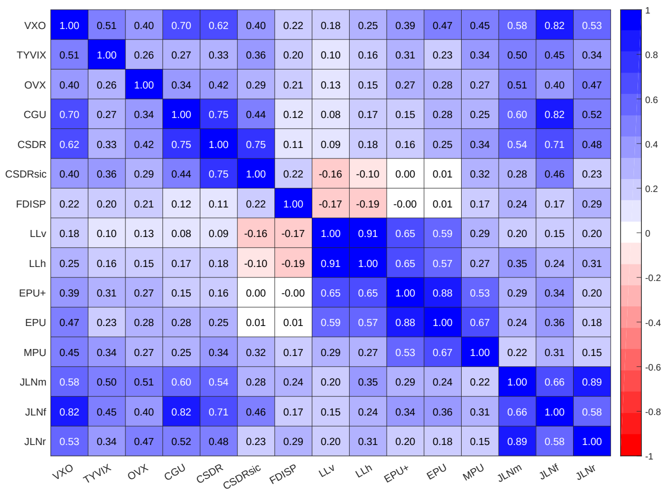
The figure presents the time series of the uncertainty measures.

Figure 4.A.1: Uncertainty measure time series (continued)



The figure presents the time series of the uncertainty measures.

Figure 4.A.2: Correlation matrix

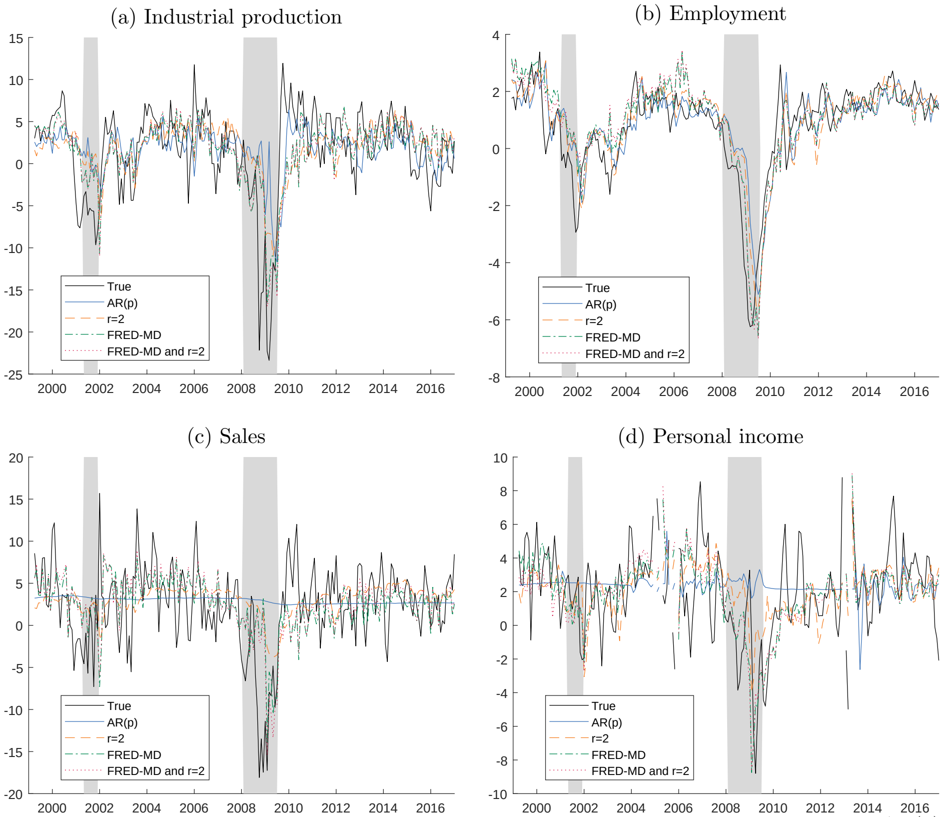


The figure presents the correlation matrix of the uncertainty measures for the full sample.

4.B Additional results on mean forecasts

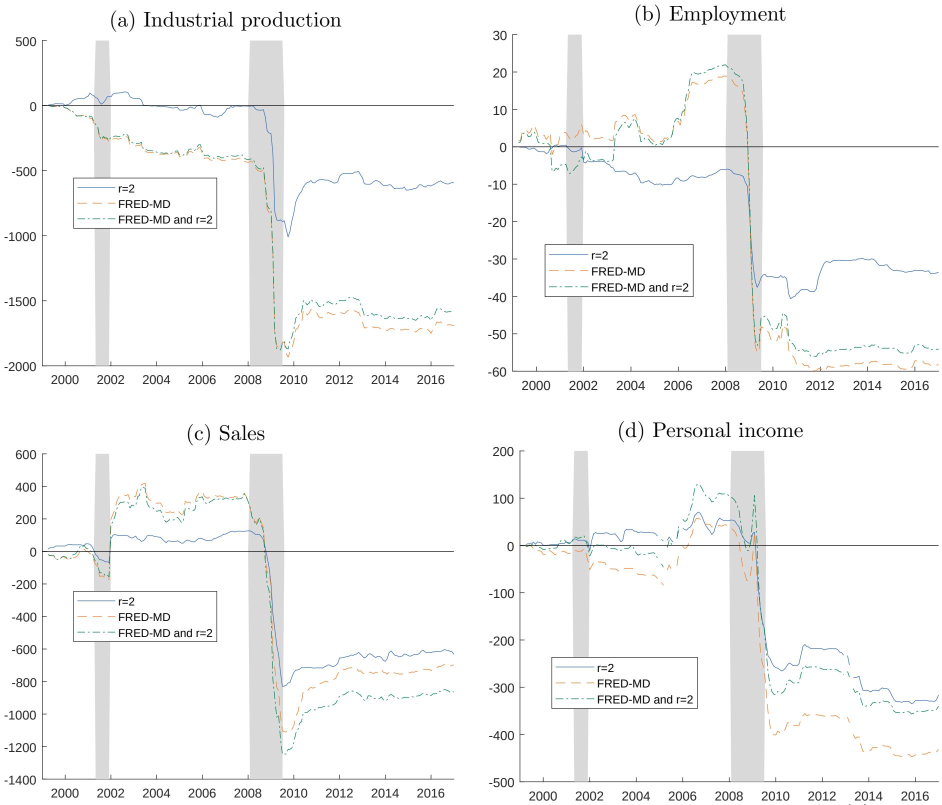
This section present additional tables and figures related to the mean forecasting results.

Figure 4.B.1: Mean forecasts, $h = 3$



The figures present the mean forecasts and the realized values for the factor models and the AR(p) model for a three month forecast horizon. The solid black line are the prebenchmark values, the solid blue line forecasts from the AR(p) model, the dashed orange line forecasts from the factor model with two uncertainty factors, the dash-dotted green line forecasts from the factor model with FRED-MD factors, and the red dashed line are forecasts from the factor model with FRED-MD factors and two uncertainty factors.

Figure 4.B.2: Cumulative sum of squared errors differential, $h = 3$



The figures present the cumulative sum of squared errors differential compared to the AR(p) model. Negative values indicate better forecasts compared to the benchmark. The solid blue line represents the factor model with two uncertainty factors, the dashed orange line the factor model with FRED-MD factors, and the dash-dotted green line represents the factor model with FRED-MD factors and two uncertainty factors.

Table 4.B.1: Relative mean squared prediction error by forecast horizon

Horizon (months)	1	3	6	12	24
			Panel A: Industrial production		
MSPE AR(p)	61.968	27.474	24.065	19.205	11.604
Factor models					
$r = 1$	0.924*	0.894	0.958	0.995	1.039
$r = 2$	0.920*	0.899	0.942	0.961	0.922
$r = 3$	0.901	0.882	0.951	1.174	1.313
FRED-MD	0.865**	0.712*	0.767	0.894	0.866
FRED-MD and $r = 2$	0.885**	0.730*	0.815	0.978	1.085
Uncertainty measures					
VXO	0.982	0.902*	0.953	0.998	1.016
TYVIX	1.002	1.014	1.021	1.024	0.998
OVX	0.980	0.910*	0.947**	0.970	1.011
CGU	0.998	0.927*	0.973	1.093	1.503
CSDR	1.012	0.966	1.022	1.023	1.132
CSDRsic	1.013	1.010	1.015	1.038	1.155
FDISP	1.009	0.960**	0.996	0.995	0.986
LLv	1.010	1.019	1.036	1.035	0.968
LLh	1.006	1.003	1.064	1.086	0.981
EPU+	1.024	1.033	1.044	1.023	0.948***
EPU	1.016	0.974	1.049	1.008	0.994
MPU	0.996	0.973	1.042	0.999	1.007
JLNm	0.923	0.864	0.866	0.963	1.073
JLNf	0.987	0.945	0.974	0.995	1.086
JLNr	0.937	0.836	0.876	0.965	1.019
			Panel B: Nonfarm payroll employment		
MSPE AR(p)	1.425	1.230	1.351	1.708	2.220
Factor models					
$r = 1$	0.890***	0.842**	0.870	0.930	0.995
$r = 2$	0.926*	0.873*	0.869	0.869	0.824*
$r = 3$	0.886*	0.817	0.909	1.141	1.137
FRED-MD	1.022	0.779	0.875	0.980	1.074
FRED-MD and $r = 2$	1.058	0.795	0.918	1.033	1.270
Uncertainty measures					
VXO	0.922***	0.916*	0.923	0.957	0.983
TYVIX	0.995	1.033	1.053	1.038	1.023
OVX	0.946***	0.918**	0.943**	0.969	1.003
CGU	0.957**	0.906*	0.881*	1.003	1.367
CSDR	0.993	0.972	0.964	1.002	1.078
CSDRsic	1.014	1.021	1.018	1.046	1.153
FDISP	0.939***	0.963	1.016	1.007	0.976
LLv	1.024	1.030	1.026	1.011	0.968
LLh	0.987	0.987	1.025	1.051	0.993
EPU+	1.005	1.022	1.031	1.020	0.977
EPU	0.979	0.977	0.994	1.015	1.017
MPU	0.991	0.979	0.994	0.998	1.006
JLNm	0.846***	0.716**	0.694	0.758	0.738
JLNf	0.968***	0.933**	0.923	0.940	1.031
JLNr	0.870***	0.762**	0.787	0.846	0.845

The table presents the relative mean squared prediction error (RMSPE) with the autoregressive model with automatic lag selection (AR(p)) as benchmark for various forecasting horizons for the out-of-sample period 1998M12+h-2016M12. Further, the table presents the mean squared prediction error (MSPE) of the AR(p) model. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.B.1: Relative mean squared prediction error by forecast horizon (continued)

Horizon (months)	1	3	6	12	24
Panel C: Manufacturing and trade sales					
MSPE AR(p)	88.594	27.013	18.289	14.090	9.529
Factor models					
$r = 1$	0.973	0.895	0.912	0.974	1.046
$r = 2$	1.005	0.891	0.876	0.925	0.943
$r = 3$	0.981	0.826	0.847	1.107	1.190
FRED-MD	1.047	0.879	0.814	0.904	0.938
FRED-MD and $r = 2$	1.083	0.851	0.812	0.917	1.046
Uncertainty measures					
VXO	0.967*	0.948	0.958	1.007	1.035
TYVIX	0.973*	0.985	0.983	1.022	1.119
OVX	0.987	0.944	0.941*	0.975	1.011
CGU	0.978*	0.933*	1.013	1.267	1.518
CSDR	0.997	0.968	0.979	1.039	1.070
CSDRsic	1.010	1.006	1.032	1.065	1.197
FDISP	0.978*	0.959*	0.961*	0.959	0.983
LLv	1.023	1.003	1.031	1.059	1.026
LLh	0.993	0.999	1.059	1.087	1.056
EPU+	1.064	1.041	1.044	1.029	1.082
EPU	1.027	1.018	1.011	1.041	1.023
MPU	1.022	1.008	1.009	1.029	1.005
JLNm	0.896**	0.743*	0.719	0.890	1.099
JLNf	0.973**	0.925*	0.928	0.977	1.068
JLNr	0.887**	0.749*	0.728	0.835	0.974
Panel D: Personal income excluding transfer receipts					
MSPE AR(p)	20.475	7.810	4.958	4.892	4.259
Factor models					
$r = 1$	0.955	0.779**	0.662*	0.683*	0.871
$r = 2$	1.001	0.805*	0.706	0.694*	0.886
$r = 3$	1.027	0.796	0.604*	0.642	1.220
FRED-MD	1.000	0.734*	0.556*	0.520*	0.752
FRED-MD and $r = 2$	1.023	0.791	0.595*	0.564	0.806
Uncertainty measures					
VXO	0.972**	0.900**	0.868*	0.917*	0.982
TYVIX	0.954**	0.911**	0.879*	0.937	0.987
OVX	0.979	0.914*	0.854*	0.874*	0.932
CGU	1.003	0.952**	0.934*	1.025	1.656
CSDR	0.984	0.938	0.974	0.952	0.995
CSDRsic	1.013	1.057	1.057	1.048	1.092
FDISP	1.005	0.997	1.001	0.946*	0.940
LLv	0.922*	0.815*	0.876	0.858	0.998
LLh	0.956	0.749**	0.758*	0.775*	0.968
EPU+	1.075	0.946	1.005	0.946	1.049
EPU	1.005	0.924	0.947	0.905	1.002
MPU	0.994	0.994	0.989	0.958*	0.988
JLNm	0.975	0.821	0.607	0.508*	0.668
JLNf	0.961**	0.864**	0.842*	0.855*	0.976
JLNr	0.968	0.808*	0.584*	0.541*	0.731*

The table presents the relative mean squared prediction error (RMSPE) with the autoregressive model with automatic lag selection (AR(p)) as benchmark for various forecasting horizons for the out-of-sample period 1998M12+h-2016M12. Further, the table presents the mean squared prediction error (MSPE) of the AR(p) model. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.B.2: Relative mean squared prediction error by evaluation period

Period	1998M12+h–2016M12	Recessions	Expansions	2000M01–2009M12	2010M01–2016M12
		Panel A: Industrial production			
MSPE AR(p)	27.474	108.809	16.226	40.264	12.035
Factor models					
$r = 1$	0.894	0.670*	1.102	0.813	1.256
$r = 2$	0.899	0.685*	1.097	0.829	1.228
$r = 3$	0.882	0.431*	1.299	0.737	1.592
FRED-MD	0.712*	0.461*	0.945	0.629**	1.117
FRED-MD and $r = 2$	0.730*	0.449*	0.990	0.647*	1.130
Uncertainty measures					
VXO	0.902*	0.818*	0.980	0.863*	1.074
TYVIX	1.014	1.023	1.006	1.017	0.998
OVX	0.910*	0.906	0.913**	0.929	0.817**
CGU	0.927*	0.841**	1.008	0.920*	0.936
CSDR	0.966	0.929*	1.000	0.977	0.893**
CSDRsic	1.010	0.999	1.020	1.009	1.006
FDISP	0.960**	0.941*	0.978	0.944**	1.038
LLv	1.019	0.926	1.104	0.966	1.283
LLh	1.003	0.862*	1.134	0.941	1.309
EPU+	1.033	0.934	1.124	0.968	1.360
EPU	0.974	0.920	1.024	0.922	1.227
MPU	0.973	1.051	0.900***	0.980	0.950*
JLNm	0.864	0.496**	1.205	0.807	1.111
JLNf	0.945	0.834**	1.047	0.915*	1.064
JLNr	0.836	0.478*	1.168	0.746	1.243
		Panel B: Nonfarm payroll employment			
MSPE AR(p)	1.230	4.802	0.736	1.642	0.734
Factor models					
$r = 1$	0.842**	0.726*	0.946	0.803**	0.955
$r = 2$	0.873*	0.763*	0.972	0.828*	1.020
$r = 3$	0.817	0.525*	1.080	0.697*	1.166
FRED-MD	0.779	0.481*	1.047	0.714	0.899
FRED-MD and $r = 2$	0.795	0.475*	1.083	0.738	0.915
Uncertainty measures					
VXO	0.916*	0.867	0.960	0.902*	0.952
TYVIX	1.033	1.073	0.996	1.050	0.976**
OVX	0.918**	0.914	0.922***	0.930*	0.885**
CGU	0.906*	0.819**	0.984	0.904*	0.892**
CSDR	0.972	0.956*	0.987	0.981	0.942
CSDRsic	1.021	0.997	1.043	1.011	1.051
FDISP	0.963	0.915**	1.006	0.947*	1.008
LLv	1.030	0.990	1.066	0.978	1.211
LLh	0.987	0.891*	1.074	0.920*	1.207
EPU+	1.022	0.909	1.124	0.923	1.349
EPU	0.977	0.892	1.054	0.865**	1.351
MPU	0.979	1.007	0.953**	0.969*	1.009
JLNm	0.716**	0.464**	0.944	0.677*	0.817**
JLNf	0.933**	0.871**	0.988	0.924*	0.957
JLNr	0.762**	0.548**	0.955	0.728*	0.856*

The table presents the relative mean squared prediction error (RMSPE) with the autoregressive model with automatic lag selection (AR(p)) as benchmark for various evaluation periods, for a forecasting horizon of $h = 3$ months. Recessions and expansions are identified as determined by NBER's Business Cycle Dating Committee. Further, the table presents the mean squared prediction error (MSPE) of the AR(p) model. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.B.2: Relative mean squared prediction error by evaluation period (continued)

Period	1998M12+h–2016M12	Recessions	Expansions	2000M01–2009M12	2010M01–2016M12
		Panel C: Manufacturing and trade sales			
MSPE AR(p)	27.013	118.362	14.380	39.873	10.119
Factor models					
$r = 1$	0.895	0.682***	1.136	0.839*	1.155
$r = 2$	0.891	0.665***	1.148	0.838*	1.131
$r = 3$	0.826	0.440***	1.265	0.745*	1.318
FRED-MD	0.879	0.503***	1.308	0.784	1.439
FRED-MD and $r = 2$	0.851	0.457***	1.299	0.761	1.369
Uncertainty measures					
VXO	0.948	0.859**	1.049	0.916*	1.095
TYVIX	0.985	0.961**	1.012	0.983	1.004
OVX	0.944	0.862**	1.037	0.929*	1.010
CGU	0.933*	0.835***	1.046	0.927	0.994
CSDR	0.968	0.926**	1.017	0.965*	0.969
CSDRsic	1.006	1.007	1.005	1.008	0.987
FDISP	0.959*	0.918***	1.004	0.933***	1.098
LLv	1.003	0.869**	1.155	0.952	1.314
LLh	0.999	0.851**	1.167	0.947	1.302
EPU+	1.041	0.850**	1.258	0.980	1.447
EPU	1.018	0.937**	1.110	0.991	1.178
MPU	1.008	0.997	1.021	1.012	0.998
JLNm	0.743*	0.329***	1.214	0.644**	1.193
JLNf	0.925*	0.793***	1.074	0.883**	1.090
JLNr	0.749*	0.422***	1.121	0.668**	1.121
		Panel D: Personal income excluding transfer receipts			
MSPE AR(p)	7.810	19.811	6.095	9.751	5.346
Factor models					
$r = 1$	0.779**	0.510*	0.904	0.732*	0.825
$r = 2$	0.805*	0.556*	0.921	0.771*	0.868
$r = 3$	0.796	0.527	0.921	0.737	0.973
FRED-MD	0.734*	0.385*	0.896	0.662*	0.931
FRED-MD and $r = 2$	0.791	0.431	0.958	0.729	0.945
Uncertainty measures					
VXO	0.900**	0.807**	0.943	0.883**	0.921
TYVIX	0.911**	0.827*	0.950	0.915*	0.902
OVX	0.914*	0.760*	0.985	0.890*	0.964
CGU	0.952**	0.938*	0.959	0.945**	0.940
CSDR	0.938	0.832*	0.987	0.889*	0.942
CSDRsic	1.057	1.025	1.072	1.054	1.005
FDISP	0.997	0.980**	1.004	1.003	0.979
LLv	0.815*	0.562**	0.933	0.705**	1.112
LLh	0.749**	0.479**	0.875	0.661**	0.980
EPU+	0.946	0.712	1.055	0.874	1.175
EPU	0.924	0.810*	0.976	0.932	0.921
MPU	0.994	0.982	1.000	1.009	0.975
JLNm	0.821	0.440*	0.998	0.717	1.013
JLNf	0.864**	0.763**	0.911*	0.822***	0.905
JLNr	0.808*	0.509*	0.947	0.739*	0.929

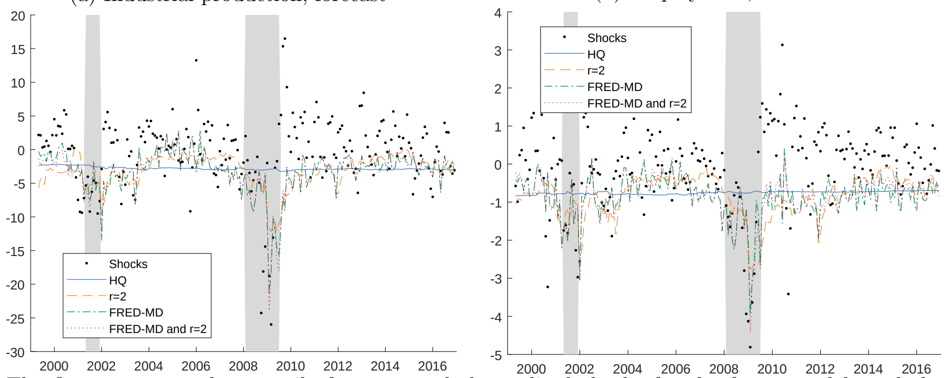
The table presents the relative mean squared prediction error (RMSPE) with the autoregressive model with automatic lag selection (AR(p)) as benchmark for various evaluation periods, for a forecasting horizon of $h = 3$ months. Recessions and expansions are identified as determined by NBER's Business Cycle Dating Committee. Further, the table presents the mean squared prediction error (MSPE) of the AR(p) model. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

4.C Additional results on quantile forecasts

This section present additional tables and figures related to the mean forecasting results.

Figure 4.C.1: Quantile forecasts and hits, $\alpha = 0.2$, $h = 3$

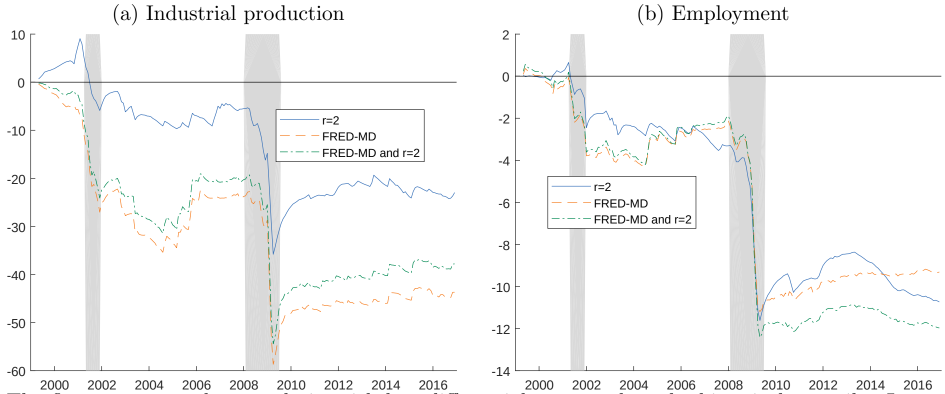
(a) Industrial production, forecast (b) Employment, forecast



The figure presents the quantile forecasts and the realized shocks for the factor models and the historical quantile for a three month forecast horizon.

Figure 4.C.2: Cumulative tick loss differential, $\alpha = 0.2$, $h = 3$

(a) Industrial production (b) Employment



The figure presents the cumulative tick loss differential compared to the historical quantile. Lower values indicate better quantile forecasts.

Table 4.C.1: Relative mean tick loss by forecast horizon, $\alpha = 0.2$

Horizon (months)	1	3	6	12	24
		Panel A: Industrial production			
MTL AR(p)	2.129	1.509	1.533	1.464	1.205
Factor models					
$r = 1$	0.953*	0.912	0.886	0.943	0.999
$r = 2$	0.954*	0.930	0.872	0.885	0.872
$r = 3$	0.894**	0.906	0.816	0.891	0.980
FRED-MD	0.901**	0.863*	0.785*	0.936	0.879
FRED-MD and $r = 2$	0.934*	0.875	0.819	0.895	0.816
Uncertainty measures					
VXO	0.972*	0.946	0.919	0.975	1.073
TYVIX	1.003	1.029	1.020	1.017	1.082
OVX	0.961***	0.978	0.937	0.957*	1.095
CGU	0.979*	0.960	0.910	1.002	1.936
CSDR	0.999	0.984	0.951	0.961	1.113
CSDRsic	1.015	1.026	1.008	1.040	1.168
FDISP	0.979*	0.960**	0.967*	1.002	0.977
LLv	0.973	0.995	1.028	1.001	0.913*
LLh	0.972	1.007	1.046	1.032	0.963
EPU+	0.971	0.993	1.019	1.024	1.045
EPU	0.959*	0.951	1.004	1.002	0.992
MPU	0.959**	0.945*	0.981	0.993	1.001
JLNm	0.912**	0.906	0.778*	0.766	0.864
JLNf	0.980	0.949	0.914	0.949	0.972
JLNr	0.890**	0.899	0.810	0.884	0.886
		Panel B: Nonfarm payroll employment			
MTL AR(p)	0.354	0.332	0.356	0.437	0.523
Factor models					
$r = 1$	0.882***	0.855**	0.827**	0.880*	0.977
$r = 2$	0.858***	0.848**	0.832**	0.821***	0.786**
$r = 3$	0.851***	0.833*	0.760*	0.885	0.780*
FRED-MD	0.933	0.869*	0.897	0.916*	0.910
FRED-MD and $r = 2$	0.945	0.839*	0.837	0.874	0.822
Uncertainty measures					
VXO	0.903***	0.941*	0.899**	0.927*	0.986
TYVIX	0.980**	0.987	0.977	0.985	1.040
OVX	0.951**	0.963	0.928**	0.954*	1.032
CGU	0.901***	0.930*	0.933	1.018	1.479
CSDR	0.948**	0.971	0.934**	0.938*	0.998
CSDRsic	0.950**	0.989	0.995	1.024	1.076
FDISP	0.941***	0.965**	0.979	0.952**	0.882*
LLv	0.988	1.013	1.001	1.008	0.932*
LLh	0.973	0.980	1.008	1.037	0.958
EPU+	0.967	0.974	1.007	1.026	1.075
EPU	0.956	0.968	0.973	0.994	0.999
MPU	0.963***	0.965**	0.933***	0.949*	1.016
JLNm	0.921**	0.850**	0.746**	0.645**	0.660**
JLNf	0.917***	0.928**	0.858***	0.877**	0.948
JLNr	0.929**	0.869**	0.799*	0.828	0.741**

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various forecasting horizons for the out-of-sample period 1998M12+h-2016M12 and quantile level $\alpha = 0.2$. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.1: Relative mean tick loss by forecast horizon, $\alpha = 0.2$ (continued)

Horizon (months)	1	3	6	12	24
	Panel C: Manufacturing and trade sales				
MTL AR(p)	2.652	1.536	1.304	1.160	0.990
Factor models					
$r = 1$	0.974	0.893	0.869*	0.964	1.164
$r = 2$	1.001	0.918	0.842*	0.903	1.059
$r = 3$	0.994	0.879	0.826	0.966	1.018
FRED-MD	0.936	0.970	0.886	1.058	1.018
FRED-MD and $r = 2$	1.017	0.937	0.873	1.047	1.084
Uncertainty measures					
VXO	0.980	0.952	0.937	0.978	1.184
TYVIX	0.996	0.980	1.005	1.036	1.210
OVX	0.993	0.931	0.945	0.950	1.059
CGU	1.006	0.986	1.074	1.279	2.437
CSDR	0.991	0.986	0.988	1.056	1.216
CSDRsic	0.992	0.985	1.002	1.011	1.203
FDISP	0.987	1.001	0.958*	0.953	0.948
LLv	1.017	1.006	1.029	1.021	0.952
LLh	0.998	0.985	1.045	1.046	1.023
EPU+	1.049	1.043	1.012	1.061	1.085
EPU	1.039	1.014	1.011	0.997	0.950*
MPU	1.008	0.973*	0.999	1.015	1.024
JLNm	0.936	0.794**	0.717*	0.851	1.083
JLNf	0.981**	0.945	0.951	0.982	1.241
JLNr	0.906**	0.815**	0.716**	0.780*	0.898
	Panel D: Personal income excluding transfer receipts				
MTL AR(p)	1.266	0.789	0.639	0.705	0.674
Factor models					
$r = 1$	0.946**	0.928	0.799*	0.919	1.138
$r = 2$	0.986	0.991	0.843	1.003	1.136
$r = 3$	0.981	0.959	0.835	0.791	0.988
FRED-MD	0.896*	0.892	0.805	0.822	0.961
FRED-MD and $r = 2$	0.930	0.885	0.868	0.830	1.165
Uncertainty measures					
VXO	0.997	1.007	0.956	0.974	1.028
TYVIX	0.956***	1.000	0.951**	0.995	0.973
OVX	0.969***	0.959	0.955	0.975	0.967
CGU	1.016	1.026	0.999	1.128	1.623
CSDR	0.984*	1.005	0.985	0.989	1.005
CSDRsic	0.958***	1.018	1.039	1.058	1.141
FDISP	0.998	0.997	0.994	1.016	0.867**
LLv	0.962	0.967	0.872	0.941	1.008
LLh	0.952	0.917	0.835*	0.893	1.051
EPU+	1.041	1.077	0.969	0.984	1.143
EPU	1.027	1.014	0.972	0.984	1.042
MPU	0.979	0.987	1.004	1.013	1.012
JLNm	0.969	0.865	0.740*	0.734	0.958
JLNf	0.994	0.991	0.943	1.053	1.213
JLNr	0.954**	0.867	0.745*	0.743	0.800**

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various forecasting horizons for the out-of-sample period 1998M12+h-2016M12 and quantile level $\alpha = 0.2$. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.2: Relative mean tick loss by evaluation period, $h = 3$ and $\alpha = 0.2$

Period	1998M12+h–2016M12	Recessions	Expansions	2000M01–2009M12	2010M01–2016M12
Panel A: Industrial production					
MTL AR(p)	1.509	4.322	1.119	1.962	0.953
Factor models					
$r = 1$	0.912	0.682**	1.034	0.846*	1.049
$r = 2$	0.930	0.676**	1.065	0.878	1.032
$r = 3$	0.906	0.580*	1.080	0.812*	1.176
FRED-MD	0.863*	0.606**	1.001	0.806*	1.053
FRED-MD and $r = 2$	0.875	0.570**	1.038	0.813*	1.055
Uncertainty measures					
VXO	0.946	0.852*	0.996	0.911*	1.031
TYVIX	1.029	1.065	1.010	1.032	1.017
OVX	0.978	0.912*	1.012	0.972	0.971
CGU	0.960	0.859**	1.014	0.948	1.007
CSDR	0.984	0.940*	1.007	0.986	0.934**
CSDRsic	1.026	0.992	1.045	1.014	1.047
FDISP	0.960**	0.917***	0.983	0.929***	1.043
LLv	0.995	0.922*	1.035	0.971	1.076
LLh	1.007	0.902*	1.063	0.981	1.096
EPU+	0.993	0.908*	1.038	0.964	1.095
EPU	0.951	0.842**	1.009	0.905*	1.086
MPU	0.945*	0.936	0.950	0.923**	1.010
JLNm	0.906	0.578**	1.081	0.838*	1.065
JLNf	0.949	0.781***	1.039	0.907*	1.041
JLNr	0.899	0.572**	1.073	0.810*	1.127
Panel B: Nonfarm payroll employment					
MTL AR(p)	0.332	0.916	0.251	0.398	0.254
Factor models					
$r = 1$	0.855**	0.630**	0.969	0.806**	0.947
$r = 2$	0.848**	0.611**	0.968	0.795**	0.954
$r = 3$	0.833*	0.504**	0.998	0.751*	0.990
FRED-MD	0.869*	0.540**	1.036	0.780*	1.055
FRED-MD and $r = 2$	0.839*	0.468**	1.026	0.761*	0.954*
Uncertainty measures					
VXO	0.941*	0.913	0.954*	0.946	0.929
TYVIX	0.987	0.973**	0.994	1.007	0.942***
OVX	0.963	0.895*	0.997	0.951	0.989
CGU	0.930*	0.771**	1.011	0.916	0.940***
CSDR	0.971	0.916**	0.999	0.972	0.956
CSDRsic	0.989	0.999	0.984	1.004	0.954**
FDISP	0.965**	0.924***	0.986	0.942***	1.012
LLv	1.013	0.937	1.052	0.961	1.144
LLh	0.980	0.831**	1.055	0.921*	1.119
EPU+	0.974	0.851*	1.037	0.887**	1.186
EPU	0.968	0.794**	1.057	0.865**	1.210
MPU	0.965**	0.955	0.970*	0.945**	1.006
JLNm	0.850**	0.547**	1.002	0.824*	0.887***
JLNf	0.928**	0.826**	0.979	0.916*	0.945
JLNr	0.869**	0.595**	1.007	0.839*	0.920**

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various evaluation periods for a forecast horizon of three months and quantile level of $\alpha = 0.2$. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.2: Relative mean tick loss by evaluation period, h = 3 and α = 0.2 (continued)

Period	1998M12+h–2016M12	Recessions	Expansions	2000M01–2009M12	2010M01–2016M12
Panel C: Manufacturing and trade sales					
MTL AR(p)	1.536	4.843	1.078	1.947	0.980
Factor models					
$r = 1$	0.893	0.586***	1.084	0.833*	0.988
$r = 2$	0.918	0.592***	1.120	0.860	1.016
$r = 3$	0.879	0.477***	1.129	0.806*	1.123
FRED-MD	0.970	0.597***	1.201	0.902	1.201
FRED-MD and $r = 2$	0.937	0.545***	1.181	0.868	1.151
Uncertainty measures					
VXO	0.952	0.866*	1.005	0.934	0.996
TYVIX	0.980	0.921*	1.016	0.989	0.971
OVX	0.931	0.720**	1.062	0.871*	1.064
CGU	0.986	0.843**	1.075	1.013	0.961
CSDR	0.986	0.898**	1.041	0.999	0.961
CSDRsic	0.985	0.992	0.980	1.014	0.897***
FDISP	1.001	0.961**	1.026	0.993	1.033
LLv	1.006	0.776**	1.149	0.921	1.270
LLh	0.985	0.731**	1.143	0.907	1.220
EPU+	1.043	0.685**	1.266	0.901	1.504
EPU	1.014	0.787**	1.155	0.915	1.310
MPU	0.973*	0.967*	0.976	0.984	0.957*
JLNm	0.794**	0.346***	1.072	0.698**	1.000
JLNf	0.945	0.784**	1.045	0.899*	1.049
JLNr	0.815**	0.465***	1.032	0.733**	0.991
Panel D: Personal income excluding transfer receipts					
MTL AR(p)	0.789	1.501	0.687	0.910	0.613
Factor models					
$r = 1$	0.928	0.726*	0.991	0.881	0.991
$r = 2$	0.991	0.852	1.034	0.961	1.050
$r = 3$	0.959	0.572	1.080	0.866	1.192
FRED-MD	0.892	0.457*	1.027	0.800	1.115
FRED-MD and $r = 2$	0.885	0.505*	1.004	0.762*	1.137
Uncertainty measures					
VXO	1.007	1.016	1.004	0.992	1.034
TYVIX	1.000	1.033	0.989	1.004	0.991
OVX	0.959	0.814*	1.004	0.917*	1.032
CGU	1.026	1.066	1.014	1.030	1.006
CSDR	1.005	0.974	1.015	0.999	0.987
CSDRsic	1.018	1.016	1.018	1.029	0.985
FDISP	0.997	0.957***	1.010	0.993	1.002
LLv	0.967	0.687*	1.055	0.824**	1.277
LLh	0.917	0.617**	1.011	0.793**	1.181
EPU+	1.077	0.892	1.134	1.014	1.266
EPU	1.014	0.964	1.030	1.003	1.042
MPU	0.987	0.984	0.987	0.999	0.959**
JLNm	0.865	0.484*	0.984	0.770*	1.035
JLNf	0.991	0.935*	1.009	0.955	1.042
JLNr	0.867	0.490*	0.985	0.768*	1.035

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various evaluation periods for a forecast horizon of three months and quantile level of α = 0.2. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.3: Relative mean tick loss by quantile level, $h = 3$

α	0.1	0.2	0.5	0.8
Panel A: Industrial production				
MTL AR(p)	1.088	1.509	1.815	1.177
Factor models				
$r = 1$	0.806*	0.912	0.981	1.029
$r = 2$	0.838*	0.930	0.988	1.010
$r = 3$	0.785*	0.906	0.998	1.020
FRED-MD	0.793*	0.863*	0.917**	0.940*
FRED-MD and $r = 2$	0.800*	0.875	0.945	0.969
Uncertainty measures				
VXO	0.860**	0.946	0.990	1.003
TYVIX	1.054	1.029	1.002	1.003
OVX	0.899*	0.978	0.983*	0.991
CGU	0.954	0.960	1.005	1.006
CSDR	0.916**	0.984	1.007	1.009
CSDRsic	1.006	1.026	1.006	1.004
FDISP	0.937**	0.960**	0.998	1.004
LLv	1.032	0.995	0.991	0.979***
LLh	1.029	1.007	0.997	0.973**
EPU+	1.003	0.993	0.995	1.003
EPU	0.977	0.951	0.980	1.004
MPU	0.960	0.945*	0.980***	0.991
JLNm	0.756**	0.906	0.986	1.014
JLNf	0.905*	0.949	1.004	1.017
JLNr	0.787*	0.899	0.985	1.013
Panel B: Nonfarm payroll employment				
MTL AR(p)	0.251	0.332	0.386	0.261
Factor models				
$r = 1$	0.729***	0.855**	1.007	1.024
$r = 2$	0.748**	0.848**	0.974	1.048
$r = 3$	0.756**	0.833*	0.963	1.132
FRED-MD	0.786**	0.869*	0.959	1.015
FRED-MD and $r = 2$	0.779*	0.839*	0.980	1.079
Uncertainty measures				
VXO	0.789**	0.941*	0.992	1.018
TYVIX	0.968*	0.987	1.016	0.969
OVX	0.899**	0.963	0.983	1.012
CGU	0.874*	0.930*	0.988	1.019
CSDR	0.877**	0.971	0.996	0.995
CSDRsic	0.950**	0.989	1.012	1.012
FDISP	0.929**	0.965**	0.997	0.984*
LLv	0.978	1.013	0.991	1.020
LLh	0.923	0.980	0.996	1.013
EPU+	0.917	0.974	1.004	1.014
EPU	0.903*	0.968	1.012	1.012
MPU	0.900***	0.965**	1.001	0.994
JLNm	0.731**	0.850**	0.943	1.072
JLNf	0.774***	0.928**	0.989	1.005
JLNr	0.728**	0.869**	0.974	1.063

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various quantile levels for the out-of-sample period 1998M12+h-2016M12, with a forecast horizon of three months. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.3: **Relative mean tick loss by quantile level, $h = 3$ (continued)**

α	0.1	0.2	0.5	0.8
	Panel C: Manufacturing and trade sales			
MTL AR(p)	1.104	1.536	1.868	1.255
Factor models				
$r = 1$	0.813**	0.893	1.000	1.004
$r = 2$	0.841*	0.918	0.983	1.002
$r = 3$	0.797*	0.879	0.942	1.004
FRED-MD	0.936	0.970	0.989	1.110
FRED-MD and $r = 2$	0.937	0.937	0.971	1.045
Uncertainty measures				
VXO	0.876**	0.952	0.995	1.002
TYVIX	0.890**	0.980	1.002	1.012
OVX	0.881*	0.931	1.005	1.001
CGU	0.936	0.986	0.995	0.995
CSDR	0.893***	0.986	1.006	1.004
CSDRsic	0.944*	0.985	1.009	0.998
FDISP	0.951***	1.001	0.994	1.004
LLv	0.953	1.006	1.010	1.009
LLh	0.925	0.985	1.013	1.020
EPU+	1.008	1.043	1.030	1.016
EPU	0.959	1.014	1.012	1.028
MPU	0.950**	0.973*	1.004	1.006
JLNm	0.721**	0.794**	0.924	0.967
JLNf	0.842**	0.945	0.994	0.998
JLNr	0.717**	0.815**	0.906*	0.982
	Panel D: Personal income excluding transfer receipts			
MTL AR(p)	0.548	0.789	1.061	0.745
Factor models				
$r = 1$	0.911	0.928	0.932	0.937**
$r = 2$	0.941	0.991	0.940	0.936*
$r = 3$	0.914	0.959	0.923	0.949
FRED-MD	0.791*	0.892	0.874**	0.944
FRED-MD and $r = 2$	0.805*	0.885	0.934	0.984
Uncertainty measures				
VXO	0.994	1.007	0.977*	0.975**
TYVIX	0.962	1.000	0.975	0.971**
OVX	0.887*	0.959	0.988	0.974**
CGU	1.042	1.026	1.009	0.985*
CSDR	0.978	1.005	1.045	0.987
CSDRsic	0.998	1.018	1.060	1.024
FDISP	0.958**	0.997	1.006	1.006
LLv	0.898	0.967	0.908**	0.925**
LLh	0.840*	0.917	0.882**	0.911**
EPU+	1.065	1.077	0.987	0.957
EPU	1.043	1.014	0.968	0.969
MPU	0.999	0.987	0.990	0.977*
JLNm	0.815*	0.865	0.928	0.991
JLNf	0.986	0.991	0.981	0.962**
JLNr	0.790*	0.867	0.932	0.986

The table presents the relative mean tick loss (RMTL) with the historical quantile (HQ) as benchmark for various quantile levels for the out-of-sample period 1998M12+h-2016M12, with a forecast horizon of three months. Further, the table presents the mean tick loss (MSPE) of the historical quantile. ***,**, and * denote significance of a one-sided Diebold-Mariano test at the 1%, 5% and 10%, respectively. See Table 4.A.1 for an explanation of the abbreviations.

Table 4.C.4: Hit rates by quantile

α	0.1	0.2	0.5	0.1	0.2	0.5
	Industrial production			Sales		
HQ	0.140†	0.252†	0.519† *	0.117†	0.215† *	0.556†
Factor models						
$r = 1$	0.136† *	0.266 *	0.519† *	0.079† *	0.201† *	0.547† *
$r = 2$	0.131† *	0.285	0.509† *	0.094† *	0.224† *	0.547† *
$r = 3$	0.103† *	0.224† *	0.542† *	0.089† *	0.192† *	0.505† *
FRED-MD	0.117† *	0.234† *	0.533† *	0.126†	0.234†	0.551† *
FRED-MD and $r = 2$	0.126† *	0.229† *	0.533† *	0.131†	0.243†	0.556† *
Uncertainty measures						
VXO	0.122† *	0.276	0.519† *	0.094† *	0.229† *	0.551† *
TYVIX	0.140†	0.238†	0.519† *	0.122† *	0.252† *	0.556† *
OVX	0.112† *	0.234† *	0.509† *	0.103† *	0.206† *	0.551† *
CGU	0.136† *	0.262 *	0.542† *	0.140†	0.262	0.565 *
CSDR	0.112† *	0.257 *	0.528† *	0.122† *	0.262 *	0.570 *
CSDRsic	0.145	0.234†	0.528† *	0.164	0.243† *	0.537† *
FDISP	0.136† *	0.238† *	0.514† *	0.126† *	0.243† *	0.551† *
LLv	0.136†	0.229† *	0.519† *	0.094† *	0.196† *	0.551† *
LLh	0.131† *	0.229† *	0.519† *	0.084† *	0.192† *	0.542†
EPU+	0.122† *	0.243† *	0.495† *	0.112†	0.201†	0.514†
EPU	0.098† *	0.206† *	0.486† *	0.089† *	0.201† *	0.533† *
MPU	0.112† *	0.248† *	0.528†	0.122† *	0.229† *	0.551† *
JLNm	0.108† ‡*	0.243† *	0.500† *	0.065†§ *	0.168† *	0.486† *
JLNf	0.136† *	0.276	0.528† *	0.117† *	0.220† *	0.533† *
JLNr	0.150 *	0.234† *	0.472† *	0.070† *	0.196† *	0.537† *
	Employment			Personal income		
HQ	0.108†	0.187† *	0.500†	0.111† *	0.202† *	0.587
Factor models						
$r = 1$	0.084†	0.196† *	0.523† *	0.106† *	0.226† *	0.514† *
$r = 2$	0.089†	0.201† *	0.561† *	0.130† *	0.202† *	0.495† *
$r = 3$	0.084† *	0.173† *	0.551† *	0.120†	0.188†	0.447†
FRED-MD	0.103† *	0.206† *	0.528† *	0.101†	0.178†	0.476† *
FRED-MD and $r = 2$	0.103† *	0.220† *	0.533† *	0.096† *	0.178† *	0.490† *
Uncertainty measures						
VXO	0.098†	0.196† *	0.533† *	0.115† *	0.231† *	0.591
TYVIX	0.117† *	0.196† *	0.500† *	0.111† *	0.226† *	0.572 *
OVX	0.112† *	0.182† *	0.500† *	0.101† *	0.197† *	0.572 *
CGU	0.089† *	0.201† *	0.542† *	0.115†	0.245†	0.625
CSDR	0.122† *	0.215† *	0.519† *	0.115† *	0.236† *	0.567
CSDRsic	0.131† *	0.196† *	0.486† *	0.130† *	0.226† *	0.563†
FDISP	0.136† *	0.196† *	0.514† *	0.111† *	0.221† *	0.596
LLv	0.108† *	0.182† *	0.509† *	0.101† *	0.173† *	0.438† *
LLh	0.084† *	0.168† *	0.500† *	0.082† ‡*	0.168† *	0.452† *
EPU+	0.108† *	0.173† *	0.523† *	0.130†	0.216†	0.486†
EPU	0.079† *	0.178† *	0.509† *	0.106† *	0.192† *	0.505† *
MPU	0.112† *	0.187† *	0.514† *	0.120† *	0.212† *	0.587
JLNm	0.070†	0.159†	0.519† *	0.082† *	0.173† *	0.505† *
JLNf	0.084†	0.182† *	0.542† *	0.106† *	0.212† *	0.582 *
JLNr	0.094† *	0.178† *	0.523† *	0.067† *	0.173† *	0.548† *

The table presents hit rates for various forecasting horizons, for the full sample and a quantile level $\alpha = 0.2$. † denotes no rejection of the null hypothesis of correct unconditional coverage, § denotes no rejection of the null hypothesis of independence, ‡ denotes no rejection of the null hypothesis of correct conditional coverage, and * denotes no rejection of the null hypothesis of correct coverage conditional on an intercept and the quantile estimates q_t, all at a 5% significance level. See Table 4.A.1 for an explanation of the abbreviations.

Nederlandse Samenvatting (Summary in Dutch)

In de economische wetenschap zijn er veel vraagstukken, met name op macro-economisch en financieel gebied, die moeilijk beantwoord kunnen worden zonder het analyseren van tijdreeksen. Bijvoorbeeld het bepalen van welke factoren gebruikt kunnen worden om aandelenrendementen te voorspellen. Of de vraag of het risico dat bankleningen niet terug worden betaald hoger is tijdens recessies. Antwoorden op deze vragen zijn van belang voor investeerders, banken en beleidsmakers die hier hun portfolio, beleid of regelgeving op kunnen aanpassen. Dit proefschrift is een verzameling van drie essays op het gebied van toegepaste tijdreeksanalyse, waarin een variëteit aan econometrische technieken wordt gebruikt voor het analyseren van belangrijke financiële en economische vraagstukken.

Het eerste essay behandelt de conjunctuurgevoeligheid van verliezen op bankleningen. Het is waarschijnlijk dat bedrijven die geld geleend hebben meer moeite hebben om hun betalingen te voldoen als het economisch slecht gaat dan in een periode waarin het economisch goed gaat. Echter, een bank heeft invloed op dit proces en kan de conjunctuurgevoeligheid van verliezen mogelijk verminderen. De vraag is wat de gevoeligheid in werkelijkheid is. Met behulp van een unieke dataset worden de verliezen tezamen met macro-economische variabelen gemodelleerd. Om rekening te houden met de karakteristieke kenmerken van de distributie van verliezen, wordt deze gemodelleerd als een combinatie van twee normale verdelingen. De ene verdeling staat voor de kleine verliezen, en de andere voor de zware verliezen. De distributie van verliezen verandert over de tijd doordat de fractie van zware verliezen t.o.v. kleine verliezen verschilt over de tijd en hangt mogelijk samen met fluctuaties van macro-economische variabelen. De schattingsresultaten laten zien dat de verliezen op bankleningen hoger zijn tijdens periodes van slechtere economische condities. Uit simulaties blijkt dat een bank hierdoor tijdens recessies ongeveer twee keer zoveel kapitaal zouden moeten aanhouden dan tijdens expansies om kredietwaardig te blijven.

De economie is een dynamisch systeem dat onderhevig is aan veranderingen. Relaties tussen economische variabelen kunnen veranderen over de tijd, bijvoorbeeld door innovaties zoals de computer of de auto. Daarnaast is het gedrag van bedrijven en consumenten anders tijdens recessies dan tijdens expansies. Het tweede essay gaat over het correct detecteren van dit soort mogelijke veranderingen in de relatie over de tijd. Het gaat in op het risico dat een lange termijn investeerder – zoals een pensioenfonds – loopt bij het verkeerd identificeren van de relatie van aandelenrendementen met een voorspeller, in dit geval de dividend-prijs ratio. Eerder onderzoek wijst uit dat de voorspellende waarde van bekende factoren varieert over de tijd. Hier wordt de instabiliteit geschat, in plaats van van te voren aan te nemen dat er of een klein aantal (grotere) wisselingen zijn, of een groot aantal (kleinere) veranderingen. Er blijkt dat de relatie substantieel verschilt over de tijd. De relatie zwakt vanaf de jaren '70 geleidelijk af tot de jaren '90, maar wordt rond 2000 en tijdens de financiële crisis weer iets sterker. Tegelijk is er veel onzekerheid is over hoe vaak de relatie precies verandert. Een analyse van de kosten van het verkeerd bepalen van de instabiliteit geeft aan dat het voor een lange termijn investeerder belangrijk is om toe te staan dat de relatie verandert over de tijd, maar de manier waarop lijkt minder uit te maken.

Het derde essay evalueert of iemand die geïnteresseerd is in het voorspellen van economische output gebruik kan maken van maatstaven van economische onzekerheid. De 'reële optie' theorie geeft economische onzekerheid als verklaring voor conjunctuurbewegingen: als er veel onzekerheid is, stellen bedrijven investeringen uit, wat leidt tot een reductie in economische activiteit. De theorie is lastig te staven, omdat economische onzekerheid niet direct geobserveerd is, en geen exacte definitie heeft. Daarom is er een benadering nodig. De afgelopen tien jaar zijn veel verschillende maatstaven van economische onzekerheid voorgesteld. Er worden 15 maatstaven verzameld en ingedeeld in vijf categorieën op basis van hoe economische onzekerheid gemeten wordt. Uit een factor analyse blijkt dat er een algemene onzekerheidsfactor is, en een media-/cosumentenvertrouwen factor die een groot deel van de fluctuaties van de maatstaven samenvatten. Bij de voorspellingsopzet wordt er rekening mee gehouden dat de informatie beschikbaar moet zijn geweest op het moment van voorspellen. Dit is relevant, omdat macro-economische cijfers gereviseerd worden na publicatie. Daarnaast wordt er niet alleen naar het gemiddelde gekeken, maar ook naar het voorspellen van waarden in het geval van economisch mindere periodes. De resultaten laten zien dat onzekerheid vooral informatief is ten tijde van economische neergang, met name voor werkgelegenheid. Dit suggereert een niet-lineaire relatie tussen economische onzekerheid en output.

Bibliography

Acharya, V. V., Bharath, S. T., and Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics*, 85(3):787–821.

Adrian, T., Boyarchenko, N., and Giannone, D. (2018). Vulnerable growth. *American Economic Review*, forthcoming.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Alexopoulos, M. and Cohen, J. (2009). Uncertain times, uncertain measures. University of Toronto Working Paper No. 352.

Alexopoulos, M. and Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance*, 40:8–28.

Allen, L. and Saunders, A. (2003). A survey of cyclical effects in credit risk measurement models. BIS Working Paper 126, Bank of International Settlements, Basel, Switzerland.

Altman, E., Brady, B., Resti, A., and Sironi, A. (2005). The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications. *Journal of Business*, 78(6):2203–2228.

Altman, E. I. and Kalotay, E. A. (2014). Ultimate recovery mixtures. *Journal of Banking & Finance*, 40:116–129.

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

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458.

Azizpour, S., Giesecke, K., and Schwenkler, G. (2015). Exploring the Sources of Default Clustering. Technical report, Working Paper, Boston University Questrom School of Business.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Bachmann, R. and Bayer, C. (2014). Investment dispersion and the business cycle. *American Economic Review*, 104(4):1392–1416.

Bachmann, R., Elstner, S., and Sims, E. R. (2013). Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics*, 5(2):217–49.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Baker, S. R. and Bloom, N. (2013). Does uncertainty reduce growth? Using disasters as natural experiments. NBER Working Paper No. 19475.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55(1):225–264.

Basel Committee on Banking Supervision (2005). Guidance on Paragraph 468 of the Framework Document. *Bank for International Settlements*.

Bastos, J. A. (2014). Ensemble predictions of recovery rates. *Journal of Financial Services Research*, 46(2):177–193.

Bauwens, L., Carpantier, J.-F., and Dufays, A. (2017). Autoregressive moving average infinite hidden Markov-switching models. *Journal of Business & Economic Statistics*, 35(2):162–182.

Bauwens, L., Koop, G., Korobilis, D., and Rombouts, J. V. (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, 30:596–620.

Bekaert, G., Hoerova, M., and Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788.

Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics*, 98(1):85–106.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic Perspectives*, 28(2):153–175.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, 86(3):1031–1065.

Brandt, M. (2010). Portfolio choice problems. In Aït-Sahalia, Y. and Hansen, L. P., editors, *Handbook of financial econometrics*, volume 1, chapter 5, pages 269–336. North Holland.

Bruche, M. and González-Aguado, C. (2010). Recovery rates, default probabilities, and the credit cycle. *Journal of Banking & Finance*, 34(4):754–764.

Calabrese, R. (2014a). Downturn Loss Given Default: Mixture distribution estimation. *European Journal of Operational Research*, 237(1):271–277.

Calabrese, R. (2014b). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry*, 30(2):99–114.

Calabrese, R. and Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking & Finance*, 34(5):903–911.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic asset allocation: Portfolio choice for long-term investors*. Oxford University Press.

Carriero, A., Clark, T. E., and Marcellino, M. (2017). Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, forthcoming.

Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with Mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):255–268.

Castelnuovo, E. and Tran, T. D. (2017). Google it up! A Google trends-based uncertainty index for the United States and Australia. *Economics Letters*, 161:149–153.

Chauvet, M. and Piger, J. (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, 26(1):42–49.

Chauvet, M., Senyuz, Z., and Yoldas, E. (2015). What does financial volatility tell us about macroeconomic fluctuations? *Journal of Economic Dynamics and Control*, 52:340–360.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241.

Choi, H., Mueller, P., and Vedolin, A. (2017). Bond variance risk premiums. *Review of Finance*, 21(3):987–1022.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–62.

Chuliá, H., Guillén, M., and Uribe, J. M. (2017). Measuring uncertainty in the stock market. *International Review of Economics & Finance*, 48:18–33.

Clark, T. and McCracken, M. (2013). Advances in forecast evaluation. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2, pages 1107–1201. Elsevier.

Clark, T. E. and McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business & Economic Statistics*, 27(4):441–454.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.

Clements, M. P. and Galvão, A. B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28(3):458–477.

Cogley, T. and Sargent, T. J. (2001). Evolving post-World War II US inflation dynamics. *NBER Macroeconomics Annual*, 16:331–373.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Cragg, J. G. and Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, 76(1-2):223–250.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk. *Review of Economics and Statistics*, 96(5):898–915.

Cremers, K. J. M. (2002). Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223–1249.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 1, pages 961–982. Elsevier.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of econometrics*, 105(1):111–130.

D'Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Del Negro, M. and Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: A corrigendum. *Review of Economic Studies*, 82(4):1342–1345.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Diris, B. (2014). The impact of model instability on long-term investors. Available at SSRN: http://ssrn.com/abstract=2480046.

Diris, B., Palm, F., and Schotman, P. (2014). Long-term strategic asset allocation: An out-of-sample evaluation. *Management Science*, 61(9):2185–2202.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under uncertainty.* Princeton, NJ: Princeton University Press.

Donald, S. G., Fortuna, N., and Pipiras, V. (2007). On rank estimation in symmetric matrices: The case of indefinite matrix estimators. *Econometric Theory*, 23(6):1217–1232.

Duffie, D., Horel, G., Saita, L., and Eckner, A. (2009). Frailty correlated default. *Journal of Finance*, 64(5):2089–2123.

Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665.

Düllmann, K. and Gehde-Trapp, M. (2004). Systematic risk in recovery rates: An empirical analysis of US corporate credit exposures. Discussion paper 02/2004, Deutsche Bundesbank, Frankfurt.

Durbin, J. and Koopman, S. J. (2002a). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Durbin, J. and Koopman, S. J. (2002b). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Elliott, G. and Müller, U. K. (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies*, 73(4):907–940.

Emery, K., Cantor, R., Ou, S., Solomon, R., and Stumpp, P. (2004). Recent bank loan research: Implications for Moody's bank loan rating practices. Technical report, Moody's Investor Service.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fajgelbaum, P. D., Schaal, E., and Taschereau-Dumouchel, M. (2017). Uncertainty traps. *Quarterly Journal of Economics*, 132(4):1641–1692.

Fan, J., Xue, L., and Yao, J. (2017). Sufficient forecasting using factor models. *Journal of Econometrics*, 201(2):292–306.

Farmer, L., Schmidt, L., and Timmermann, A. (2018). Pockets of predictability. Available at SSRN: http://ssrn.com/abstract=3152386.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2, pages 2–56. Elsevier.

Ferderer, J. P. (1996). Oil price volatility and the macroeconomy. *Journal of macroeconomics*, 18(1):1–26.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., and Rubio-Ramírez, J. (2015). Fiscal volatility shocks and economic activity. *American Economic Review*, 105(11):3352–84.

Fiorentini, G., Planas, C., and Rossi, A. (2014). Efficient MCMC sampling in dynamic mixture models. *Statistics and Computing*, 24(1):77–89.

Flannery, M. J. (1986). Asymmetric information and risky debt maturity choice. *The Journal of Finance*, 41(1):19–37.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *Annals of Applied Statistics*, 5(2A):1020–1056.

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11(2):317–336.

Frye, J. (2000). Depressing Recoveries. *Risk*, pages 108–111.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gerlach, R., Carter, C., and Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, 95(451):819–828.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian statistics*, volume 4, pages 169–194. Oxford University Press, Oxford.

Giglio, S., Kelly, B., and Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471.

Gilchrist, S., Sim, J. W., and Zakrajšek, E. (2014). Uncertainty, financial frictions, and investment dynamics. NBER Working Paper No. 20038.

Giordani, P. and Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics*, 26(1):66–77.

Giordani, P., Kohn, R., and van Dijk, D. (2007). A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 137(1):112–133.

Giusto, A. and Piger, J. (2017). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*, 33(1):174–184.

Gonzalo, J. and Pitarakis, J.-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics*, 30(2):229–241.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3):199–232.

Groen, J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44.

Grunert, J. and Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance*, 33(3):505–513.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

Hamerle, A., Knapp, M., and Wildenauer, N. (2011). Modelling loss given default: a "point in time"-approach. In Engelmann, B. and Rauhmeier, R., editors, *The Basel II Risk Parameters*, chapter 8, pages 137–150. Springer.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2):228–248.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2):60–66.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.

Hartigan, J. A. and Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, 13(1):70–84.

Hartmann-Wendels, T., Miller, P., and Töws, E. (2014). Loss given default for leasing: Parametric and nonparametric estimations. *Journal of Banking & Finance*, 40:364–375.

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

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Höcht, S. and Zagst, R. (2007). Loan Recovery Determinants - A Pan-European Study. *Working Paper*.

Huber, F., Kastner, G., and Feldkircher, M. (2017). A new approach toward detecting structural breaks in vector autoregressive models. arXiv:1607.04532v3.

Inoue, A. and Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402.

Jo, S. (2014). The effects of oil price uncertainty on global real economic activity. *Journal of Money, Credit and Banking*, 46(6):1113–1135.

Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *Journal of Finance*, 69(2):611–644.

Jones, P. M. and Enders, W. (2016). The asymmetric effects of uncertainty on macroeconomic activity. *Macroeconomic Dynamics*, 20(5):1219–1246.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: An asset-allocation perspective. *Journal of Finance*, 51(2):385–424.

Kehrig, M. (2015). The cyclical nature of the productivity distribution. Available at SSRN: https://ssrn.com/abstract=1854401.

Keijsers, B., Diris, B., and Kole, E. (2018). Cyclicality in losses on bank loans. *Journal of Applied Econometrics*, 33(4):533–552.

Kellogg, R. (2014). The effect of uncertainty on investment: Evidence from Texas oil drilling. *American Economic Review*, 104(6):1698–1734.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–393.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126.

Knaup, M. and Wagner, W. (2012). A Market-Based Measure of Credit Portfolio Quality and Bank's Performance During the Subprime Crisis. *Management Science*, 58(8):1423–1437.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics*, 85(3):618–628.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Komunjer, I. (2013). Quantile prediction. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2, pages 961–994. Elsevier.

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4):997–1017.

Koop, G. and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74(3):763–789.

Koop, G. and Potter, S. M. (2009). Prior elicitation in multiple change-point models. *International Economic Review*, 50(3):751–772.

Koopman, S. J. and Lucas, A. (2008). A non-Gaussian panel time series model for estimating and decomposing default risk. *Journal of Business & Economic Statistics*, 26(4):510–525.

Koopman, S. J., Lucas, A., and Schwaab, B. (2012). Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008. *Journal of Business & Economic Statistics*, 30(4):521–532.

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting*, 33(1):11–20.

Lahiri, K. and Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics*, 25(4):514–538.

Leduc, S. and Liu, Z. (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, 82:20–35.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.

Ludvigson, S. C., Ma, S., and Ng, S. (2015). Uncertainty and business cycles: Exogenous impulse or endogenous response?

Maheu, J. M. and Gordon, S. (2008). Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, 23(5):553–583.

Maheu, J. M. and Song, Y. (2014). A new structural break model, with an application to Canadian inflation forecasting. *International Journal of Forecasting*, 30(1):144–160.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

McCulloch, R. E. and Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, 88(423):968–978.

Miu, P. and Ozdemir, B. (2006). Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD Correlations. *Journal of Credit Risk*, 2(2):43–68.

Nakajima, J. (2012). Bayesian analysis of generalized autoregressive conditional heteroskedasticity and stochastic volatility: Modeling leverage, jumps and heavy-tails for financial time series. *Japanese Economic Review*, 63(1):81–103.

Nakajima, J. and West, M. (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, pages 221–247.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006a). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4):1057–1084.

Pesaran, M. H., Schuermann, T., Treutler, B.-J., and Weiner, S. M. (2006b). Macroeconomic Dynamics and Credit Risk: A Global Perspective. *Journal of Money, Credit and Banking*, 38(5):1211–1261.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4):1201–1228.

Pesaran, M. H. and Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, 9(5):495–510.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

Pettenuzzo, D. and Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, 35(2):183–201.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Pykhtin, M. (2003). Recovery rates: Unexpected recovery risk. *RISK*, 16(8):74–79.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2A, chapter 6, pages 328–383. Elsevier Amsterdam.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate US stock returns. *Journal of Financial Econometrics*, 4(2):238–274.

Ravazzolo, F., Paap, R., van Dijk, D., and Franses, P. H. (2008). Bayesian model averaging in the presence of structural breaks. In Rapach, D. E. and Wohar, M. E., editors, *Forecasting in the presence of structural breaks and model uncertainty*, chapter 15, pages 561–594. Emerald Group Publishing Limited.

Romer, C. D. and Romer, D. H. (2000). Federal Reserve information and the behavior of interest rates. *American Economic Review*, 90(3):429–457.

Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2B, chapter 21, pages 1203–1324. Elsevier Amsterdam.

Rossi, B., Sekhposyan, T., and Soupre, M. (2016). Understanding the sources of macroeconomic uncertainty. Available at SSRN: https://ssrn.com/abstract=2780213.

Schiller, T. and Trebing, M. (2003). Taking the measure of manufacturing. *Federal Reserve Bank of Philadelphia, Business Review*, (Fourth Quarter).

Schuermann, T. (2004). What Do We Know About Loss Given Default? In Shimko, D., editor, *Credit Risk Models and Management*, chapter 9. Risk Books, London, 2nd edition.

Scotti, C. (2016). Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, 82:1–19.

Shleifer, A. and Vishny, R. W. (1992). Liquidation Values and Debt Capacity: A Market Equilibrium Approach. *The Journal of Finance*, 47(4):1343–1366.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41:788–829.

Stroud, J. R., Müller, P., and Polson, N. G. (2011). Nonlinear state-space models with state-dependent variances. *Journal of the American Statistical Association*, 98(462):377–386.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18.

Trebing, M. E. (1998). What's happening in manufacturing: "Survey says…". *Federal Reserve Bank of Philadelphia, Business Review*, (September/October).

Vavra, J. (2013). Inflation dynamics and time-varying volatility: New evidence and an Ss interpretation. *Quarterly Journal of Economics*, 129(1):215–258.

Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. Working paper, Harvard Business School.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Windle, J., Carvalho, C. M., Scott, J. G., and Sun, L. (2013). Efficient data augmentation in dynamic models for binary and count data. Working paper, available at arXiv preprint arXiv:1308.0774.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

681. Y. GAO, *Stability and Adaptivity: Preferences over Time and under Risk*
682. M.J. ZAMOJSKI, *Panta Rhei, Measurement and Discovery of Change in Financial Markets*
683. P.R. DENDERSKI, *Essays on Information and Heterogeneity in Macroeconomics*
684. U. TURMUNKH, *Ambiguity in Social Dilemmas*
685. U. KESKIN, *Essays on Decision Making: Intertemporal Choice and Uncertainty*
686. M. LAMMERS, *Financial Incentives and Job Choice*
687. Z. ZHANG, *Topics in Forecasting Macroeconomic Time Series*
688. X. XIAO, *Options and Higher Order Risk Premiums*
689. D.C. SMERDON, *Everybodys doing it: Essays on Trust, Norms and Integration*
690. S. SINGH, *Three Essays on the Insurance of Income Risk and Monetary Policy*
691. E. SILDE, *The Econometrics of Financial Comovement*
692. G. DE OLIVEIRA, *Coercion and Integration*
693. S. CHAN, *Wake Me up before you CoCo: Implications of Contingent Convertible Capital for Financial Regulation*
694. P. GAL, *Essays on the role of frictions for firms, sectors and the macroeconomy*
695. Z. FAN, *Essays on International Portfolio Choice and Asset Pricing under Financial Contagion*

696. H. ZHANG, *Dealing with Health and Health Care System Challenges in China: Assessing Health Determinants and Health Care Reforms*

697. M. VAN LENT, *Essays on Intrinsic Motivation of Students and Workers*

698. R.W. POLDERMANS, *Accuracy of Method of Moments Based Inference*

699. J.E. LUSTENHOUWER, *Monetary and Fiscal Policy under Bounded Rationality and Heterogeneous Expectations*

700. W. HUANG, *Trading and Clearing in Modern Times*

701. N. DE GROOT, *Evaluating Labor Market Policy in the Netherlands*

702. R.E.F. VAN MAURIK, *The Economics of Pension Reforms*

703. I. AYDOGAN, *Decisions from Experience and from Description: Beliefs and Probability Weighting*

704. T.B. CHILD, *Political Economy of Development, Conflict, and Business Networks*

705. O. HERLEM, *Three Stories on Influence*

706. J.D. ZHENG, *Social Identity and Social Preferences: An Empirical Exploration*

707. B.A. LOERAKKER, *On the Role of Bonding, Emotional Leadership, and Partner Choice in Games of Cooperation and Conflict*

708. L. ZIEGLER, *Social Networks, Marital Sorting and Job Matching. Three Essays in Labor Economics*

709. M.O. HOYER, *Social Preferences and Emotions in Repeated Interactions*

710. N. GHEBRIHIWET, *Multinational Firms, Technology Transfer, and FDI Policy*

711. H. FANG, *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*

712. Y. KANTOR, *Urban Form and the Labor Market*

713. R.M. TEULINGS, *Untangling Gravity*

714. K.J. VAN WILGENBURG, *Beliefs, Preferences and Health Insurance Behavior*

715. L. SWART, *Less Now or More Later? Essays on the Measurement of Time Preferences in Economic Experiments*

716. D. NIBBERING, *The Gains from Dimensionality*

717. V. HOORNWEG, *A Tradeoff in Econometrics*

718. S. KUCINSKAS, *Essays in Financial Economics*

719. O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*

720. E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*

721. M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*

722. N. CIURILĂ, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*

723. N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*

724. S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*

725. S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*

726. Y. ZHU, *On the Effects of CEO Compensation*

727. S. XIA, *Essays on Markets for CEOs and Financial Analysts*

728. I. SAKALAUSKAITE, *Essays on Malpractice in Finance*

729. M.M. GARDBERG, *Financial Integration and Global Imbalances*

730. U. THÜMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change*