Essays in Economic Forecasting

Florens Odendahl

TESI DOCTORAL UPF / 2018

DIRECTOR DE LA TESI
Professor Barbara Rossi
Departament d'Economia i Empresa

Für meine Familie.

Acknowledgments

While this thesis represents a collection of my search and research over the last years, this work would not have been possible without the faculty, my colleagues and my family and friends. I would like to take the opportunity to express my gratitude for this support.

First of all, I want to express my sincere gratefulness to my advisor Barbara, for her guidance, the interest in my work and ideas and the plenty of time she invested in me. Throughout the years I could count on her insights and, most notably, on her encouragement whenever I got stuck or frustrated. Thank you for everything!

Among the faculty and professors, I would like to thank in particular Christian, Geert, Majid, and Tatevik, who have always been available when I had questions or needed advice, and whose numerous comments considerably improved my work. The participants of the internal Econometrics Seminar at UPF, of the Time Series Workshop in Zaragoza and of the job market talks gave valuable feedback that helped writing this thesis.

For the many great days on campus, I would like to thank my fellow Ph.D. students: Adrian, Alex, Christian, Christi and Giulia, Donghai, Flavio, Greg, Paul, and Stefan, to name only a few. Special thanks go to my officemates Chris, Shengliang, and Shangyu. I want to thank Paolo, Timo, Chrisi, Benni, Lukas, Mirko and all my friends from Essen for their regular visits, for some of the most memorable evenings in Barcelona and for making my visits to Germany all the more worthwhile. I am thankful to my friends from Bonn, Dirk, Knauth, Mark, and Sven, for the regular contact, which endured the long distance and hopefully keeps enduring it. I would like to thank Friedrich for encouraging me to study economics and for his enthusiasm for research. Furthermore, and as we all know, nobody would actually get his Ph.D. without Marta, Laura and Mariona - thank you!

I have had the luck to spent the last two years in Barcelona with my wonderful girlfriend Ana. Thank you for your support, your patience and your humor that turned my thoughts away from my worries during stressful times. Los dos años contigo han sido increíbles.

I dedicate this thesis to the people I consider a privilege to call my family, above all Christa, Gerd, Tobi, and my grandparents. At all times, I can rely on your unconditional love, your encouragement and, most importantly, your faith in me. It is obvious that without you I would not have completed this chapter of my life and I remain deeply indebted to you.

Abstract

This thesis consists of three chapters on forecasting techniques in economics. In chapter 1, I use copulas to estimate multivariate density forecasts based on univariate densities from survey data. Survey-based predictions are often competitive to time series models in their forecasting performance but have a univariate focus and my estimation strategy exploits the information in the surveys' marginal densities. I subsequently demonstrate the importance of the multivariate aspect for forecasters. In chapter 2, we propose novel tests for forecast rationality, which are robust under the presence of Markov switching. Existing tests focus on constant out-of-sample performances or use non-parametric techniques; consequently, they may lack power against the alternative of discrete switches. Investigating the Blue Chip Financial Forecasts, we find evidence against forecast unbiasedness during periods of monetary easing. Chapter 3 provides an empirical investigation of the real-time forecasting performance of quantile regressions for predicting different vintages of real US GDP growth. My results indicate that quantile regressions are competitive to current benchmark models and that the in-sample estimation strategy matters for the performance concerning different data vintages.

Resumen

Esta tesis consta de tres capítulos sobre métodos predictivos en economía. El primer capítulo propone el uso de cópulas para la elaboración de previsiones de distribuciones multivariantes utilizando datos de encuestas sobre distribuciones univariantes. Las previsiones basadas en sondeos son, a menudo, equiparables a las obtenidas por modelos de series temporales, pero sólo hay datos disponibles para distribuciones univariantes. La estrategia de estimación propuesta utiliza la información de las distribuciones univariantes de los sondeos. Posteriormente queda demostrada la importancia de la perspectiva multivariante en la elaboración de previsiones. El segundo capítulo propone nuevos tests para evaluar la racionalidad de las previsiones, los cuales, resultan sólidos bajo la presencia de Markov switching. En comparación, los tests existentes se centran en probar la prueba entera o usan técnicas no-paramétricas y tienen menos poder contra la alternativa de cambios discretos. Mediante la investigación empírica de la racionalidad del las previsiones del Blue Chip Financial Forecasts, se encuentra evidencia a favor de la hipótesis de un sesgo con Markov switching durante los periodos de relajación monetaria. El tercer capítulo es una investigación empírica de la eficacia del modelo de regresión de cuantiles para prever en tiempo real el crecimiento del PIB estadounidense. Los resultados obtenidos indican que dicho modelo es comparable a los modelos de referencia actuales y que la estrategia de estimación aplicada con differentes muestras de datos influye los resultados.

Preface

This thesis consists of three self-containing chapters, all of which are related to economic forecasting and contain both, theoretical and empirical results. Chapter one and three present and evaluate models for multivariate and univariate density forecasts. Chapter two provides a framework to test for time variation in the forecasting performance of competing models.

In chapter 1, I present a methodology to estimate multivariate density forecasts based on univariate densities from survey data. Survey-based predictions are often competitive to time series models in terms of their forecasting performance but have a univariate focus. My methodology exploits the information in the surveys' marginal densities for the estimation of the multivariate densities. I demonstrate the importance of the multivariate aspect for new measures of the state of the economy and a novel measure of joint macroeconomic uncertainty. A stronger distributional dependence between the variables has different implications for the two types of measures. It tends to increase the probability of "recession-type" events and reduces uncertainty. Empirical results based on SPF data from the euro area and the U.S. show that the survey-based joint density forecasts are competitive to current benchmark econometric models. When considering joint macroeconomic uncertainty, the dependence has sizeable effects on my uncertainty measure in the aftermath of the Great Recession, a feature of the data that existing measures would not capture.

Chapter 2 proposes novel tests for forecast rationality, which are robust under the presence of Markov switching. Existing tests focus on constant out-of-sample performances or use non-parametric techniques; consequently, they may lack power against the alternative of discrete switches. Monte Carlo results suggest that the tests we propose have better power than existing tests in detecting Markov switching deviations from unbiasedness or efficiency. In an empirical investigation of the forecast rationality of the Blue Chip Financial Forecasts for the federal funds target rate, we find evidence against forecast unbiasedness. During periods of monetary easing, the forecasters tend to overestimate the future interest rate systematically. The size of the systematic bias component is around 25 basis points, a typical interest rate move of the federal reserve.

In chapter 3, I provide an empirical investigation of the real-time forecasting performance of quantile regressions for predicting different vintages of real GDP growth. Given a large number of potential predictors, I pool univariate models using an equal-weighting scheme and bayesian model averaging. My results indicate that equal-weighting outperforms bayesian model averaging and that forecasting

first-release GDP growth is best done by also basing the in-sample estimation on first-release data only. When the predictive performance is compared to stochastic volatility models, the pooled bayesian quantile regressions outperform the competitor models at the one-quarter-ahead forecast, in particular for quantiles above the median. For horizons of four quarters-ahead, the results speak in favor of quantile regressions but exhibit substantial time-variation - the multivariate stochastic volatility models tend to perform better during recession periods but worse after the great recession.

Contents

List of Figures

List of Tables

Chapter 1

Survey-Based Multivariate Density Forecasts

1.1 Introduction

Expectations about future events are of central concern for policymakers, the private sector and academia alike. Not only are accurate probabilistic forecasts critical to good decision making but can also be an important determinant of the actual realisations of present and future outcomes. For instance, firms and households take the expected future inflation rate into account when negotiating wages or setting prices and thereby influence the realised inflation rate today. In 2004, Alan Greenspan emphasised *...a central bank needs to consider not only the most likely future path for the economy, but also the distribution of possible outcomes about that path* (Greenspan, 2004).[1] For an economy, which is a multivariate system of variables, *multivariate* density forecasts provide the most comprehensive information about future events. In light of the importance of predictions, the present paper seeks to make the following contributions.

We propose to estimate joint density forecasts based on the marginal density forecasts of survey data through copula functions. The joint densities will be characterised by marginal densities from a survey, a copula function and the thereby estimated distributional dependence of the variables.[2] Survey forecasts are known to be competitive to time series model predictions (Faust and Wright, 2013) and our methodology allows to exploit the information provided by survey marginals

[1] Manski (2017) most recently highlighted the importance of including predictive densities in the decision process of policy institutions.

[2] It is important to note that, in the estimation, the copula is evaluated using the cumulative distribution function of the survey implied marginals, which in turn are evaluated at the realised values of the variables.

for the estimation of joint density forecasts. Existing macroeconomic surveys ask participants only for the univariate marginal distributions and hence, studies have been limited to univariate analyses with survey data. We show the importance of the *multivariate aspect* for the two following novel applications.

We demonstrate how the joint density forecasts can be used to extract information about the state of the economy. Specifically, we propose a probabilistic *economic downturn* indicator (for the euro area), a probability of the risk of *rare events*, such as negative output growth accompanied by low inflation, and the probability of remaining in a secular stagnation region. Importantly, the joint density forecasts allow to base these indicators on the *joint behaviour of several variables*, which reflects the dating procedures of the NBER Business Cycle Dating Committee and the Centre for Economic Policy Research (CEPR).

Further, we consider macroeconomic uncertainty through the lens of the *joint* distribution function of several variables. Existing measures of uncertainty are typically a functional of univariate density forecasts. In contrast, we propose to use a multivariate version of Shannon's entropy and our *joint uncertainty* accommodates the computation of uncertainty in a multivariate system of variables. Cover and Thomas (1991) provide a rigorous introduction to entropy from an information theoretical point of view. While we focus on the importance of distributional dependence for a multivariate measure of dispersion, joint density forecasts, more generally, allow to compute uncertainty for loss functions of several variables, and can therefore be of use beyond the specific example considered in this paper.

We apply the methodology to data from the U.S. and euro area Survey of Professional Forecasters (SPF) and estimate joint density forecasts based on the marginal density forecasts of the surveys. We find that the univariate as well as the multivariate density forecasts, estimated using the SPF, are competitive to econometric model predictions.[3] Further, for both data sets, we find substantial movements in the time-varying correlations between key macroeconomic variables, in particular for the time period after the Great Recession. Moreover, the estimated multivariate densities imply non-linear conditional expectations. The strength of the out-of-sample dependence varies over different parts of the distribution at a given point in time, and the non-linearity follows directly from the non-Gaussian shape of the SPF marginal forecast densities.

Using the estimated multivariate density, we produce one-year ahead out-of-

[3]We choose an AR(1) with stochastic volatility (AR-SV) and a time-varying parameter VARs with stochastic volatility (BVAR) (Primiceri, 2005; Del Negro and Primiceri, 2015). It is well known that stochastic volatility models perform well for density forecasts (see e.g. Clark (2011); D'Agostino et al. (2013); Clark and Ravazzolo (2015)).

sample forecasts for different indicators of the state of the economy. We find that the empirical performance of the SPF-based economic downturn forecasts is competitive with the BVAR. The SPF-based and the BVAR-based indicator perform similarly in predicting the end of the Great Recession, while the BVAR performs better in predicting its onset. However, using our indicator, the SPF-based joint density predicts markedly well the downturn during euro area debt recession and outperforms the BVAR. Still, in 2017, almost a decade after the financial crisis, the probability of remaining in a secular stagnation region, defined by low growth and low inflation (and high unemployment for the eurozone), is high. For the *joint uncertainty* measure, our empirical results show that dependence can have a sizeable effect on macroeconomic uncertainty for the euro area. We compare the measure to the results of Abel et al. (2016) and find that, when taking distributional dependence into account, our results differ.

Our paper builds on several strands of the literature. Faust and Wright (2013), among others, stress the good performance of survey data point forecasts and Kenny et al. (2015a) discuss the performance of euro area SPF density forecasts. D'Agostino et al. (2013), Clark and Ravazzolo (2015) and Smith and Vahey (2016), among others, explore the performance of predictive densities estimated by different econometric models. We contribute to this literature, firstly, by analysing the usefulness of survey marginal density forecasts for estimating joint density forecasts and, secondly, by comparing the performance of the resulting multivariate density to existing benchmark econometric models.

Our work is further related to the literature on Business Cycle (BC) dating, recently surveyed by Hamilton (2011). Recent work on (nowcast) BC dating, that is based on the joint behaviour of several highly aggregated series, includes Chauvet and Piger (2008) and Camacho et al. (2014).[4] Chauvet and Piger (2008) use a Markov-switching dynamic factor model and Camacho et al. (2014) use a mixed-frequency Markov-switching factor model in order to determine the state of the BC in real time. While we build on the idea of using several highly aggregated variables, we differ from the existing approaches in the following ways. First, our ex-ante probabilities are, to a large extent, based on agents' subjective probabilities, i.e. they reflect beliefs of economic decision makers. Second, joint density forecasts allow obtaining not only recession probabilities but also probabilities for crisis events of different economic magnitudes, for example, a recession accompanied by deflation. In addition, we are particularly interested in the out-of-sample predictive perfor-

[4] Another recent example includes Billio et al. (2016). The authors use a panel Markov-switching VAR and, for the determination of the state of the Business Cycle, take the interconnectedness of euro area countries as well as the US into account.

mance, while for example Chauvet and Piger (2008) and Camacho et al. (2014) focus on the real-time ability to call a recession.

Starting with the seminal paper by Bloom (2009), the task of quantifying uncertainty has attracted considerable attention over the last decade. Our measure of uncertainty is related to existing work by Lahiri and Sheng (2010), D'Amico and Orphanides (2014), Abel et al. (2016), Rossi et al. (2017) and Jo and Sekkel (forthcoming), who used survey data to estimate ex-ante uncertainty measures. All of these papers' measures are functionals of univariate distributions, and we contribute to this literature by providing a measure, multivariate Shannon's entropy, that is a functional of a multivariate distribution. Further, we build on the work of Adrian et al. (2016). The authors, among other things, apply a univariate version of conditional Shannon entropy to conditional GDP growth to calculate downside and upside uncertainty. In comparison, we use a multivariate version of entropy in order to compute the joint *simultaneous* uncertainty of a system of macroeconomic variables. Lastly, we show how our framework relates to the influential uncertainty measure proposed by Jurado et al. (2015).

On the technical side, we estimate the joint distribution function using copulas, based on Sklar's Theorem (Sklar, 1959). The theorem states that every joint distribution function can be decomposed into its marginals and a copula function that characterises the dependence of the random variables.[5] It, therefore, allows modeling the dependence between variables and the marginals separately. Our estimation is hence based on two key ingredients: marginals from a survey and the choice of a copula family, which allow us to estimate a joint density forecast using past realisations. Patton (2006) used copulas to model exchange rate dependence. Smith and Vahey (2016) use copulas for modeling both cross-sectional *and* serial correlation in four macroeconomic variables and show that their out-of-sample forecasts are superior to a Bayesian VAR, albeit not to a specification that includes stochastic volatility. However, the authors focus on model-based estimation of the joint forecast density and do not base their predictions on survey data. Our approach can also be interpreted from the angle of combining (subjective) survey information with an econometric model. In a recent paper, Krueger et al. (2017) develop a technique for point forecasts, that allows, inter alia, to incorporate survey information into a BVAR. We differ because we estimate joint density forecasts using survey data, where the joint density is characterised by the marginals from the survey, a copula function, and the thereby estimated distributional dependence.

Finally, we would like to highlight that, although the analysis of mean and

[5] Joe (1997) and Nelsen (2006) provide rigorous introductions to copula theory.

variance survey forecasts is standard in economics, little is known about forecasters' expectations on the joint behaviour of variables.[6] For a review of the literature on univariate survey data in time series econometrics, see for example Pesaran and Weale (2006).

A few remarks on notation. *Origin* denotes the date *on* which the forecast was made. *Target* denotes the date *for* which the forecast was made. Throughout the paper, the variable Y_t will denote GDP growth, π_t will denote the inflation rate and U_t will denote the unemployment rate.

The remainder of this paper is organised as follows. Section 1.2 motivates the importance of multivariate density forecasts. Here we introduce the ex-ante probabilistic measures of the state of the economy and *joint uncertainty*. Section 1.3 explains the methodology used for constructing the joint density forecasts and describes in detail the empirical estimation strategy. Section 1.4 describes the two survey data sets used. Section 1.5 and 1.6 contain the empirical results for the multivariate density for the euro area and the U.S. Section 1.7 concludes.

1.2 Why Joint Densities?

This section motivates why joint densities can be an essential tool in economics. Further, we introduce the measures for the ex-ante probabilities and the *joint uncertainty*. Using these two examples, we will illustrate that, on the one hand, stronger distributional dependence can imply higher probabilities of recessions/economic downturns and *disastrous events*. On the other hand, stronger dependencies can dampen the uncertainty about future outcomes. Their implications, therefore, differ and depend on the specific application.

By $F_t(\mathbf{x})$ we denote the cumulative distribution function (cdf) of a (vector-valued) random variable $\mathbf{X}_t$ of dimension d.[7] The conditional distribution function $F_{t+h|t}(\mathbf{x})$ denotes the forecasts with origin t for the target period $t+h$.

1.2.1 Probabilistic Indicators for the State of the Economy

Forecasting the state of the economy, in particular, the Business Cycle is an important task for policymakers. In general, there are two different approaches: a

[6] See for example Manski (2004, 2017) for a discussion of the importance of measuring agents' probabilistic expectations through survey data.

[7] With a slight abuse of notation, $F_{t+h|t}(\mathbf{x})$ will denote true distribution function or an estimated, possibly misspecified distribution function.

model-based type[8] and a survey-based type, such as the *anxious index*[9]. We focus on the latter, which has been shown to be competitive to a model-based approach.

In particular, we propose a way to construct an ex-ante Business Cycle indicator, based, to a large extent, on the expectations of economic decision makers collected via surveys. Three important features emerge. First, the multivariate density allows to estimate an *economic downturn* probability based on the joint behaviour of several variables, which reflects how the majority of Business Cycle indicators is estimated. The *economic downturn* indicator aims at resembling recession-like events. Second, we can construct an index for the euro area in real time, for which no direct equivalent of the *Anxious Index* exists.[10] Third, the joint density allows to obtain probabilities for specific events, i.e., we can additionally construct an *anxious-type index* for different magnitudes of economic downturns.

In Example 1, we motivate why it is the joint density forecasts that matters for estimating probabilistic ex-ante Business Cycle indicators and highlight the perils of relying on univariate indicators.

Example 1 *Suppose the Business Cycle is determined by the following bivariate system of GDP growth, denoted by* Y_t*, and the unemployment rate, denoted by* U_t*. The joint distribution is a bivariate normal* $[Y_t, U_t]' \sim N(\mu_t, \Sigma_t)$*. For simplicity, suppose further, that values of* Y_t *smaller than zero and* U_t *bigger than 10 indicate an economic downturn. For period* t*, the forecaster estimates a conditional mean forecast* $\mu_{t|t-1} = [0, 10]'$ *and a conditional variance-covariance* $\Sigma_{t|t-1} = [1, 0; 0, 1]$*, whereas for period* $t+1$ *the results are* $\mu_{t+1|t} = [0.3, 9.7]'$ *and* $\Sigma_{t+1|t} = [1, -0.9; -0.9, 1]$*. Although the conditional mean vector shifted away from the recession threshold, and the variance stayed constant, the probability of an economic downturn has increased from* $F_{t|t-1}\big(Y_t < 0, U_t > 10\big) = 0.25$ *to* $F_{t+1|t}\big(Y_{t+1} < 0, U_{t+1} > 10\big) = 0.31$*. The counterfactual probability, i.e. assuming a constant zero correlation, is* $\tilde{F}_{t+1|t}\big(Y_{t+1} <$

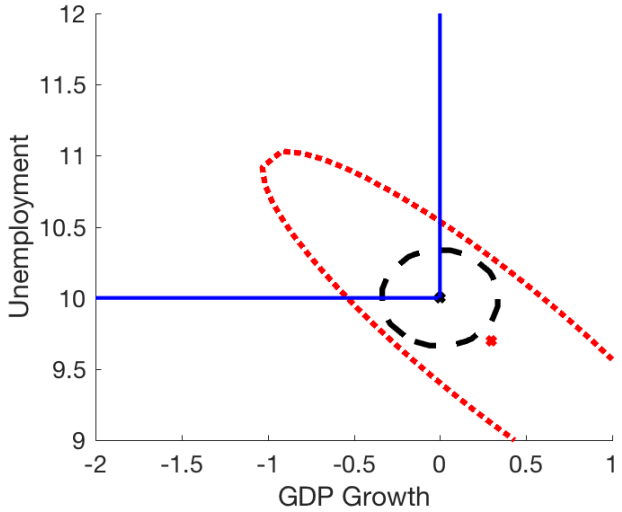
[8] See for example Stock and Watson (1989, 1991, 2014), Koop and Potter (2007), Hamilton (1989), Chauvet and Potter (2005), Chauvet and Piger (2008), Chauvet and Piger (2008), Altissimo et al. (2010) and Camacho et al. (2014), Billio et al. (2016)

[9] *In an article on September 1, 2002, New York Times reporter David Leonhardt used the term anxious index to refer to the probability of a decline in real GDP, as reported in the Survey of Professional Forecasters. The survey asks panelists to estimate the probability that real GDP will decline in the quarter in which the survey is taken and in each of the following four quarters.[...] The index often goes up just before recessions begin. For example, the first quarter survey of 2001 (taken in February) reported a 32 percent anxious index; the National Bureau of Economic Research subsequently declared the start of a recession in March 2001.* taken from https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/anxious-index

[10] Please see A for further elaboration.

$0, U_{t+1} > 10\big) = 0.15$, *where* $\tilde{F}_{t+1|t}$ *is parameterised by* $\mu_{t+1|t} = [0.3, 9.7]'$ *and* $\tilde{\Sigma}_{t+1|t} = [1, 0; 0, 1]$. *Figure 1.1 displays* 0.15 *contour plots for* $f_{t|t-1}$ *(dashed) and* $f_{t+1|t}$ *(dotted). The correlation has spread out the distribution along the diagonal and shifted probability mass to above the downturn threshold. Although the mean* $\mu_{t+1|t} = [0.3, 9.7]'$*, displayed by the cross, shifted down-right, the downturn probability increased.*

Figure 1.1: How Correlation Shifts the Probability Mass



NOTE. – The x and y axis display the values of Y_t and U_t. The dashed and dotted ellipse display 0.15 contour levels of $f_{t|t-1}$ and $f_{t+1|t}$ respectively. Realisations within the upper left-hand side quadrant marked by the solid lines indicate an economic downturn. The crosses denote the mean $\mu_{t|t-1} = [0, 10]'$ and $\mu_{t+1|t} = [0.3, 9.7]'$.

A Probabilistic Ex-Ante Business Cycle Indicator

We define the out-of-sample Business Cycle indicator as follows. A threshold for the target time $t + h$, based on information up to t, for variable i is defined as $\tau^{bc}_{t+h|t,i} \equiv \frac{1}{W}\sum_{w=1}^{W} x_{i,t-w+1}$ where $x_{i,t-w}$ is the realisation of the random variable $X_{i,t-w}$. Then the economic downturn probability $\xi_{t+h|t}$, is defined as

$$\xi_{t+h|t} \equiv F_{t+h|t}(X_{t+h,1} < \tau^{bc}_{t+h|t,1}, ..., X_{t+h,d} < \tau^{bc}_{t+h|t,d}) \tag{1.1}$$

where $F_{t+h|t}$ is the joint density forecast for target period $t+h$ based on information up to time t.[11] Here, the recession probability is determined by the probability of the variables being smaller than some threshold, where the threshold is an average of current and past realisations. While there is no single definition for a recession,

[11] Assuming that for all variables, smaller values indicate less economic activity. A reversion of the operator $<$ straightforwardly complements the usage of variables for which a higher value indicates less economic activity.

the NBER Business Cycle Dating committee, broadly speaking, calls a recession a period of *diminishing economic activity*, rather than *diminished activity*.[12] We want to highlight, that multivariate densities are flexible enough to accommodate a variety of definitions, as the thresholds $\tau^{bc}_{t+h|t,i}$ can be chosen in different of ways, e.g. relative to the latest observations or with respect to values that realised in previous recessions.

Computing the recession probability directly out of the joint distribution forecast takes into account the forecast of the mean, any quantile *and* the dependence of the variables. It is important to note that, using this definition, we can compute an out-of-sample indicator $\xi_{t+h|t}$, as the joint density forecasts and the thresholds are computed only with information up to time t.

Probabilities of Target Inflation, Rare Events and Secular Stagnation

In addition, it is useful to know the probabilities of events of different magnitudes. We propose to estimate the probability of GDP growth being, at least, τ^Y and inflation being within a certain range defined by $[_-\tau^\pi,_+\tau^\pi]$.

$$\delta_{t+h|t} \equiv F_{t+h|t}(Y_{t+h} > \tau^Y,\ {}_-\tau^\pi < \pi_{t+h} <\ {}_+\tau^\pi) \tag{1.2}$$

In addition, we compute the out-of-sample probability of negative growth accompanied by either very low or very high inflation. Both events are rare but have occurred in the past. A negative (positive) pre-subscript indicates low inflation (high inflation). Let ${}_-\zeta_{t+h|t}$ be the probability of negative output growth occurring jointly with low inflation

$${}_-\zeta_{t+h|t} \equiv F_{t+h|t}(Y_{t+h} < 0, \pi_{t+h} <\ {}_-\tau^{VaR}) \tag{1.3}$$

and ${}_+\zeta_{t+h|t}$ be the probability of negative output growth occurring jointly with high inflation

$${}_+\zeta_{t+h|t} \equiv F_{t+h|t}(Y_{t+h} < 0, \pi_{t+h} >\ {}_+\tau^{VaR}) \tag{1.4}$$

where again $F_{t+h|t}$ denotes the joint density forecast and ${}_+\tau^{VaR},_-\tau^{VaR}$ are the respective thresholds for high inflation (low inflation) in a respective currency area.

Further, the joint density forecast allows to compute the probabilities of remaining in a secular stagnation region (SeSt). We will define secular stagnation by low growth, low inflation and high unemployment state of the economy, where *low* and

[12] The website of the NBER provides details; http://www.nber.org/cycles/recessions_faq.html

high will be defined as *below (above) pre-financial crisis averages*, i.e.

$$\eta_{t+h|t} \equiv F_{t+h|t}(Y_{t+h|t} < \tau_y^s, \pi_{t+h|t} < \tau_\pi^s, U_{t+h|t} > \tau_u^s) \tag{1.5}$$

where again $F_{t+h|t}$ denotes the joint density forecast and τ_i^s is the respective secular stagnation threshold.

1.2.2 A Measure of Joint Uncertainty

Starting with the seminal paper of Bloom (2009), the task of quantifying uncertainty has received increasing attention over the last years.[13] Existing work has often used survey data to estimate a ex-ante uncertainty; see Lahiri and Sheng (2010), D'Amico and Orphanides (2014), Abel et al. (2016), Rossi et al. (2017), Jo and Sekkel (forthcoming). While many different definitions of uncertainty have been proposed in the literature, typically, a reduction in the variance of the individual time series is presumed to be a reduction in uncertainty.

We contribute to the literature by proposing an uncertainty measure that provides a systematic way of measuring the uncertainty of *multivariate systems* of variables that takes the dependencies of the variables into account. Here, we think about *joint uncertainty* as measuring the quantity of possible future paths the economy can take, while being agnostic towards the type of possible paths. That means, that it does not measure the risk of a specific event, for example negative growth and deflation, which was discussed in section 1.2.1. Example 2 motivates graphically why the multivariate approach is important and why dependencies can matter.

Example 2 *Consider a bivariate system constituted by GDP growth, Y_t, and inflation π_t, where $Y_t \sim N(0, \sigma^2_{Y,t})$ and $\pi_t \sim N(2, \sigma^2_{\pi,t})$. Additionally, assume that Y_t and π_t are jointly distributed as a bivariate normal $[Y_t, \pi_t]' \sim N([0, 2], \Sigma_t)$ and that $\Sigma_{t|t-1} = [1.3, 0.95; 0.95, 1.3]$ and $\Sigma_{t+1|t} = [1, 0; 0, 1]$. Thus, in absolute terms, both variance and covariance decrease from $t-1$ to t.*

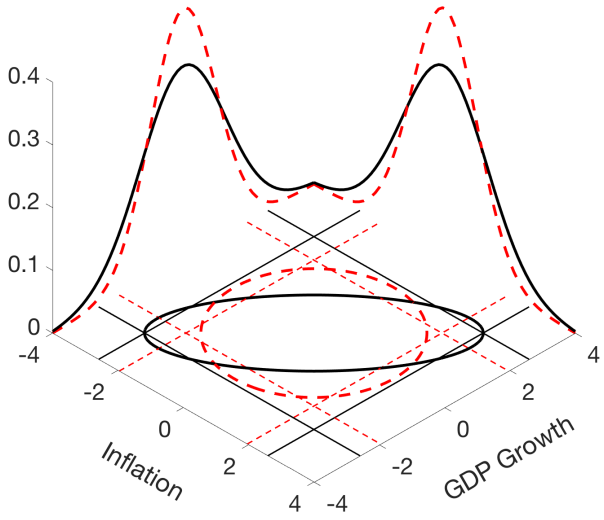
An uncertainty measure based on univariate prediction intervals (or alternatively, on forecast variances or inter-quartile ranges) will therefore indicate that uncertainty decreased. The 95% prediction interval, of the demeaned variables, decreased from $[-2.067, 2.067]$ in $t-1$ to $[-1.96, 1.96]$ in t.

In contrast, a joint uncertainty measure, based on the joint prediction confidence

[13] From a perspective of policy-makers, Alan Greenspan stated that *[uncertainty] is the defining characteristic of the [monetary policy] landscape* (Greenspan, 2004)

region, will indicate that uncertainty increased from $t-1$ to t.[14] *Denote the confidence regions by $CR_{t|t-1}$ and $CR_{t+1|t}$, defined through $F_{t|t-1}\big((Y_t, \pi_t) \in CR_{t|t-1}\big) \geq 1-\alpha$ and $F_{t+1|t}\big((Y_{t+1}, \pi_{t+1}) \in CR_{t+1|t}\big) \geq 1-\alpha$. Then, $CR_{t+1|t}$ increased relative to $CR_{t|t-1}$. Univariate measures fail to take into account the information provided by the covariance of the random variables. Figure 1.2 illustrates the respective confidence region. The solid ellipse displays $CR_{t|t-1}$ and the dashed circle displays $CR_{t+1|t}$.*

Figure 1.2: The Effect of Correlation on Confidence Regions



NOTE. – Solid lines display values for t, dashed lines display values for $t+1$. The solid ellipse is the 95% confidence area $CR_{t|t-1}$ and the dashed ellipse is 95% confidence area $CR_{t+1|t}$ from Example 2. The solid (dashed) densities are the marginal densities of Y_t and the (demeaned) π_t for t (and $t+1$ respectively). The distance between parallel solid (dashed) lines, displays the univariate 95% confidence intervals. Confidence intervals decreased from t to $t+1$, but the confidence area increased from t to $t+1$, i.e. the drop in covariance outweighed the drop in variance.

A Measure of Joint Uncertainty

We define the *joint uncertainty* of the vector of variables $\mathbf{X} = [X_1, ..., X_d]'$ as the differential entropy of $\mathbf{X}$, DE($\mathbf{X}$). Differential entropy is an information theoretical concept and we follow the definitions in Cover and Thomas (1991).[15] DE($\mathbf{X}$) is defined as

$$DE(\mathbf{X}) \equiv -\int_{\mathbb{R}^d} f(\mathbf{x}) \log\big(f(\mathbf{x})\big)\, \mathrm{d}\mathbf{x} \tag{1.6}$$

[14] Under normality, our *joint uncertainty* measure is proportional to the areas of the confidence regions $CR_{t|t-1}$ and $CR_{t+1|t}$.

[15] Arellano-Valle et al. (2012) provide analytical solutions for the differential entropy of multivariate skew-normal and skew-t distributions, which increases computational speed and allows an insight into how different parameters contribute to the entropy measure.

where $f(\mathbf{x})$ denotes the joint density function of $\mathbf{X}$. We can decompose DE into the entropy contributions of the individual variables $X_1, ..., X_d$ and the contribution of their dependencies

$$DE(\mathbf{X}) = H(X_1) + ... + H(X_d) - TC(\mathbf{X}) \tag{1.7}$$

where $H(X_i)$ is the univariate differential entropy of X_i and $TC(\mathbf{X})$, total correlation, is defined as

$$TC(\mathbf{X}) \equiv \int_{\mathbb{R}^d} f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{f_1(x_1) \cdots f_d(x_d)}\right) \mathrm{d}\mathbf{x} \tag{1.8}$$

where f_i denotes the marginal distribution function of X_i and $f(\mathbf{x})$ denotes the joint density. Total correlation is therefore equivalent to a Kullback-Leibler divergence of the joint density and the product density of the marginals and computes distributional dependence between the variables. For given marginals, the upper bound of uncertainty is reached for independent variables because $TC(\mathbf{X})$ is non-negative and equal to zero for independence. Further, the dependence measure $TC(\mathbf{X})$ can be related to instantaneous Granger causality as defined in Taamouti et al. (2014).[16] In their framework, a variable $X_{1,t}$ instantaneously Granger causes a variable $X_{2,t}$ if there is distributional dependence, given the conditioning on the past values of $X_{1,t}, X_{2,t}$. In other words, $X_{1,t}$ instantaneously Granger causes $X_{2,t}$ if $f(X_{1,t}, X_{2,t}|X_{1,t-1}, X_{2,t-1}) \neq f(X_{1,t}|X_{1,t-1}, X_{2,t-1}) f(X_{2,t}|X_{1,t-1}, X_{2,t-1})$. Therefore, if the joint density is modelled as a function of the past values of the variables, $TC(\mathbf{X})$ is equivalent to instantaneous Granger causality. Hence, $TC(\mathbf{X}) = 0$ then implies that there is no instantaneous Granger causality.

Example 2 continued *Under joint normality, DE(**X**) is available in closed form and gives the following results in the forecaster's problem in Example 2*

$$DE(Y_{t|t-1}, \pi_{t|t-1}) = 1 + log(2) + \frac{1}{2} log(|\Sigma_{t|t-1}|) = 2.72$$

and

$$DE(Y_{t+1|t}, \pi_{t+1|t}) = 1 + log(2) + \frac{1}{2} log(|\Sigma_{t+1|t}|) = 2.84$$

The uncertainty, of GDP growth and inflation jointly, increased from a perspective of a forecaster from $t-1$ to t using joint uncertainty, while the univariate forecast

[16] See Wiener (1956) and Granger (1969) for the original work about Wiener-Granger causality and Geweke (1984) for the introduction of mean-squared forecast errors as a measure for Wiener-Granger causality.

variance decreased from 1.3 to 1.

We emphasise four important features of our measure of uncertainty that contribute to the literature. First, it allows for a decomposition into the contributions of individual variables and a contribution of the joint behaviour of the variables. In the context of *Example 2*, it would, therefore, capture both the changes in forecast variance as well as the changes in distributional dependencies. Second, the measure of dependence, denoted by $TC(\mathbf{X})$, is not restricted to linear types of dependencies but is a measure of distributional dependence because it is based on a Kullback-Leibler divergence. Third, for given marginals, the researcher can infer lower and upper bounds of entropy, which will be determined by their dependencies. The upper bound is reached for independence and reduces to a sum of individual variables' uncertainty. The lower bound is determined by the dependencies and is directly related to copula entropy. A fourth important feature is that the measure allows computing conditional entropy. For example, if a researcher estimates the multivariate density of GDP growth, inflation, and the policy interest rate, the measure would allow computing the change in uncertainty, conditioned on a specific policy rate in the future. In other words, in this example, it would allow insights into the central bank's influence on *joint uncertainty*.

An additional feature of joint densities is that they naturally enable the researcher to learn something about the uncertainty of a general function z_t that depends on the random variables $X_{1,t}, ..., X_{d,t}$, i.e. about $z_t \equiv g(X_{1,t}, ..., X_{d,t})$. For example, if $g(\cdot)$ takes the form of a square-loss function, $g(X_{1,t}, ..., X_{d,t}) = -(X_{1,t} + ... + X_{d,t})^2$, the joint density straightforwardly enables the researcher to compute the uncertainty of z_t. Given $f(\mathbf{x})$, we can sample values of the random vector $\mathbf{X}$ and obtain the density of z_t, denoted by $h_t(z)$. The density h_t enables the researcher to compute $DE(z_t) = \int_{\mathbb{R}} h_t(z) log\big(h_t(z)\big) \mathrm{d}z$.

Relationship to existing work:

In a simple example we can show how our measure of uncertainty relates to the influential framework of Jurado et al. (2015) (and therefore similarly to Jo and Sekkel (forthcoming)). We adopt their notation in the following paragraph in order to facilitate legibility with respect to their paper. The authors define uncertainty as the *conditional volatility of the unforecastable component of the future*. Their measure of uncertainty is based on $\mathcal{U}_t(h) = \frac{1}{N_y} \sum_{j=1}^{N_y} \Omega_{jt}^y(h)$, where $\Omega_{jt}^y(h)$denotes the $h-step$ forecast error variance of the forecast of variable y_{jt+h}. The authors provide

a stylised example[17] in order to explain the variance decomposition. Following their example, assume that y_{jt+h} is governed by a factor structure such that

$$y_{jt+1} = \phi_j^y y_{jt} + \gamma_j^F \mathbf{F}_t + v_{jt+1}^y$$

where $\mathbf{F}_t$ is a single factor estimated on a large panel of macroeconomic time series, and $v_{jt+1}^y = \sigma_{jt+1}^y \epsilon_{tj+1}^y$ with $\epsilon_{tj+1}^y \sim_{\text{iid}} N(0,1)$, where iid stands for independently and identically distributed, where σ_{jt+1}^y follows a stochastic volatility equation and is depends only on its own past values. Importantly, the factor $\mathbf{F}_t$ is modelled recursively such that

$$\mathbf{F}_t = \Phi^F \mathbf{F}_{t-1} + \mathbf{v}_t^F, \quad \mathbf{v}_t^F = \sigma_t^F \epsilon_t^F, \quad \epsilon_t^F \sim_{\text{iid}} N(0,1)$$

where Φ^F is a vector of coefficients and σ_t^F is characterised by

$$\log(\sigma_t^F) = \alpha^F + \beta^F \log(\sigma_{t-1}^F)^2 + \tau^F \eta_t^F, \quad \eta_t^F \sim_{\text{iid}} N(0,1)$$

The recursive structure of the factor implies a time-varying covariance of the series y_{mt+h}, y_{jt+h}. For $h = 2$ the forecast error for y_{jt+h} is

$$V_{jt+2}^y = v_{jt+2}^y + \phi_j^y V_{jt+1}^y + \gamma_j^F V_{t+1}^F$$

where $V_{t+1}^F = v_{t+1}^F$. It follows therefore that the forecast errors of j and m are characterised by

$$\begin{aligned} \text{cov}(V_{jt+2}^y, V_{mt+2}^y) &= \gamma_j^F \gamma_m^F \text{var}(v_{t+1}^F) = \gamma_j^F \gamma_m^F \text{E}\big[(\sigma_{t+1}^F)^2\big] \\ &= \gamma_j^F \gamma_m^F (\sigma_t^F)^2 \exp\{\alpha^F\} \exp\{\tau^F + \frac{1}{2}\} \end{aligned} \tag{1.9}$$

where $\exp\{\cdot\}$ denotes the Euler number. We observe that the conditional covariance for $h = 2$ of y_{mt+h} and y_{jt+h}, as determined in (1.9), is a function of the factor loadings γ_j^F, γ_m^F, the past realised factor volatility, the level-parameter α^F and the expected mean of a transformation of the factor innovations, i.e. the expected value of a log-normal random variable $\exp\{\eta_t^F\}$. The common component of the forecast errors is therefore $\text{E}\big[(\sigma_{t+1}^F)^2\big]$. Importantly, a factor model directly implies a common component in the forecasting errors and the dynamic factor structure naturally complements the notion of dependence in the future realisations of the variables. The covariance-matrix would further depend on the correlation of v_{jt+2}^y

[17] The example can be found on page 1188 of the published article. Without loss of generality, we drop the hat on $\mathbf{F}_t$.

and v^y_{mt+2}, once the simplifying assumption of independence in ϵ^y_{tj+1} is dropped, i.e. once cross-sectional dependence is allowed for even after conditioning on the factors.

Assuming for simplicity, conditionally, joint normality for a bivariate system of $y_{1,t+h}$ and $y_{2,t+h}$, we can directly compare our measure to Jurado, Ludvigson and Ng (2015). Then, *joint uncertainty* is equal to, up to a constant, the logarithm of the determinant of the variance-covariance matrix, i.e. $DE(\mathbf{y}_{t+h|t}) \propto \log|\Omega^y_t(h)| = \log\big(\Omega^y_{1t}(h)\big) + \log\big(\Omega^y_{2t}(h)\big) + \log(1-\rho^2_{12,h})$ where $\rho_{12,h}$ is the correlation between $y_{1,t+h}$ and $y_{2,t+h}$. *Joint uncertainty* therefore differs from Jurado, Ludvigson and Ng (2015) as it takes all elements of matrix into account, not only the diagonal elements.

1.3 Econometric Framework

In this section, we present the methodology to estimate the survey-based multivariate density forecasts. In addition, we describe the choice of the copula family and elaborate on the specific estimation strategy we use in the application.

1.3.1 Estimation of Survey-Based Multivariate Distributions

Sklar's theorem (Sklar, 1959) states that a joint distribution function $F(\cdot)$ with marginals $F_1(\cdot), ..., F_d(\cdot)$ and domain $\bar{\mathbf{R}}^d$ can be represented by a copula function C such that for all $\mathbf{x}_t = (x_{1,t}, ..., x_{d,t})' \in \bar{\mathbf{R}}^d$

$$F(\mathbf{x}_t) = C(F_1(x_{1,t}), ..., F_d(x_{d,t})) \tag{1.10}$$

For continuously differentiable copulas the joint density of a continuous joint distribution function can be decomposed into

$$f(\mathbf{x}_t) = c(F_1(x_{1,t}), ..., F_d(x_{d,t}))f_1(x_{1,t}) \cdots f_d(x_{d,t}) \tag{1.11}$$

where $c(F_1(x_{1,t}), ..., F_d(x_{d,t}))$ is the copula density, $f(\cdot)$ is the multivariate density and $f_i(x_{i,t})$ denotes the marginal density. Importantly, Sklar's theorem holds both ways, i.e. we can combine arbitrary (continuous) marginals with a copula and get a valid joint distribution function.[18] Copulas allow to model the dependence structure and marginals separately. In particular, any dependence measure that is invariant to a monotone transformation can be represented by the copula alone.

[18] A formal definition of a copula is given in the Appendix C.

One of the contributions of our paper is estimating joint density forecasts based on the marginal densities of survey data. The idea is based on the reversion of Sklar's theorem, which allows to estimate joint density functions given the marginals from a survey. In particular, for given marginals, the copula can be estimated by maximum likelihood

$$\hat{\rho} = \underset{\rho}{\operatorname{argmax}} \sum_{t=1}^{T} c\big(F_{1,t}(x_{1,t}), ..., F_{d,t}(x_{d,t}); \rho\big) \tag{1.12}$$

where $F_{i,t}$ denotes the marginal provided by a survey for variable i at time t, $x_{i,t}$ denotes the realisation of variable i at time t and ρ denotes the copula parameters. $F_{i,t}$ evaluated at $x_{i,t}$ is the well-known Probability Integral Transform (PIT). The estimation is therefore based on the PITs and a choice of a copula family for c.

The estimation procedure is easiest understood through an algorithmic description:

1. Evaluate marginal density forecasts from survey data at the respective realisations
2. Choose a copula family, for instance the Gaussian family
3. Decide on the estimation strategy of the copula parameters, for instance constant copula parameters vs time-varying copula parameters
4. Given (1)-(3), obtain the parameter vector ρ through maximum likelihood

The resulting multivariate density will have univariate marginal densities, which are identical to the survey marginals. The specific shape of the multivariate density will be determined by the marginals, the copula family of c and the parameter (vector) ρ. In turn, the estimation of ρ depends on the marginals, the copula family c, the realised values of the variables and the modeling choice of the potential time-variation in ρ. The choice of the parametric copula family, required in (2), and the estimation scheme, required in (3), are explained in detail in the next section. In general, copula functions are a flexible tool to build new multivariate distributions that allow for variables with bounded and unbounded domains. For the eurozone, we combine two skewed student's marginals, for GDP and inflation, with a truncated normal for the unemployment rate. We are not aware of an existing multivariate distribution that allows for this flexibility. A widely used workaround in economics is to take logs of the variable. However, linear correlation, the most widely used measure of association in economics, is not invariant to this type of monotone-transformations, i.e. in general $corr(X_k, X_j) \neq corr(X_k, log(X_j))$.

1.3.2 Model Choice and Empirical Estimation Strategy

The copula function we choose is a Gaussian copula defined as

$$C_{R_t}(\mathbf{u}_t) = \Phi_{R_t}\big(\Phi^{-1}(u_{1,t}), .., \Phi^{-1}(u_{d,t})\big) \tag{1.13}$$

where Φ_{R_t} is the cumulative distribution function of a normal with mean-zero and correlation matrix R_t with elements $\rho_{kj,t}$ and Φ^{-1} is the inverse cdf of a univariate standard normal. The subscripts k and j of $\rho_{kj,t}$ denote respectively the variable k and j. We would like to highlight four important features of the Gaussian copula. First, although the copula functions is based on Gaussian distributions, the resulting joint distribution function would only be a multivariate Gaussian if the marginals were Gaussian as well. Second, although the Gaussian copula is determined by the correlation matrix R, it allows for a non-linear mean dependence between $X_1, ..., X_d$. Dropping the time subscript for simplicity, Spearman's rho, denoted by ρ^S_{kj}, has a closed form functional relationship with ρ_{kj}, $\rho^S_{kj}(X_i, X_j) = \frac{6}{\pi}\arcsin\big(\frac{\rho_{kj}}{2}\big)$. Therefore, the Gaussian copula is able to capture monotone, linear and non-linear, dependencies. Third, many of the existing copulas are best suited for the dimension $d = 2$. In contrast, the Gaussian copula is easily scalable to dimensions $d \geq 2$. Other examples of scalable distributions include the Student's t-copula and, including for very high-dimensions, the factor copula proposed by Oh and Patton (2017).

As the marginals are time-varying, we will specify the copula as having a time-varying parameters. In order to allow the parameters $\rho_{kj,t}$ to change over time and at the same time ensure that it stays within the range of $(-1, 1)$, the time-varying parameter will be estimated by a transformed auxiliary parameter $\gamma_{kj,t}$. The specification is based on Patton (2006).[19] Let $\gamma_{kj,t}$ evolve according to

$$\gamma_{kj,t} = \beta_1 + \beta_2\gamma_{kj,t-1} + \beta_3\frac{1}{\tau}\sum_{i=1}^{\tau}\Phi^{-1}(z_{t-i,j})\Phi^{-1}(z_{t-i,k}) \tag{1.14}$$

where β_2 captures the persistence in $\gamma_{kj,t}$ and $z_{t,k}$ is the PIT values based on the realisations of variable k at time t. The PIT is computed using the corresponding survey data marginal density. The correlation parameters are then obtained through

$$\rho_{kj,t} = \Lambda(\gamma_{kj,t}) \tag{1.15}$$

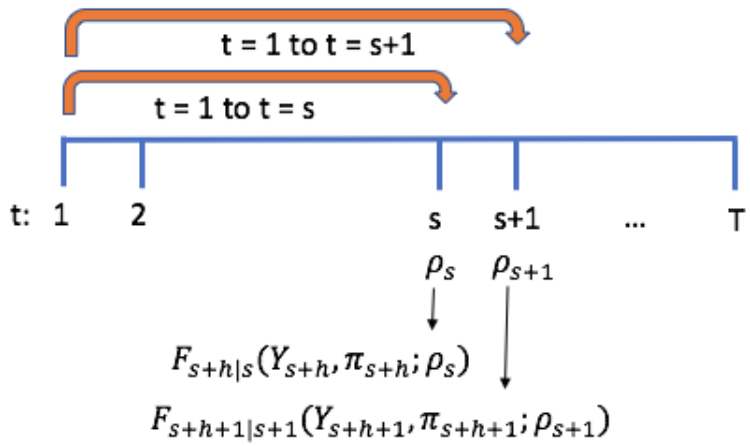
[19] Manner and Reznikova (2012) provide a survey of alternative models for time-varying copulas.

where the transformation function Λ is a modified logistic function

$$\Lambda(y) = \frac{1 - e^y}{1 + e^y} \tag{1.16}$$

The intuition for equation (1.14) is similar to that of a GARCH model, where today's correlation evolves according to a constant, last periods correlation and last periods' realised correlations. The time-variation in the dependence is therefore observation driven. As $\rho_{kj,t}$ is correlation measure of two random variables, it is symmetric, i.e. $\rho_{kj,t} = \rho_{jk,t}$. The sum $\frac{1}{\tau}\sum_{i=1}^{\tau} \Phi^{-1}(z_{j,t-i})\Phi^{-1}(z_{k,t-i})$ in (1.14) allows to take into account the smoothed co-movements of the PITs in the last periods. In the estimation, the value of τ is set to 8 in order to balance the trade-off of a quick update of the correlation versus a smooth estimate of the coefficient. Results are robust for different values for τ. Note that the estimation based on (1.14) does not ensure a positive definite correlation matrix. However, we do not encounter this problem with our estimated parameter values. Similar to time-varying parameter VARs with stochastic volatility, the multivariate density forecasts using the SPF marginals exhibits time-variation in the mean, variance, skewness, kurtosis *and* the cross-sectional dependence of the variables.

Figure 1.3: Estimation Scheme



NOTE. – The recursive window estimation is performed with a starting window size of s for the coefficients $\beta_1, \beta_2, \beta_3$ of (1.14). The baseline value for τ is equal to 8. The parameter estimated at $s + k$ is estimated using information up to time $t = s + k$. This parameter is then used in constructing the joint density forecasts, $F_{s+k+h|s+k}$, where h is the forecasting horizon.

The estimation procedure only uses information available to the survey participants at the time they produced their marginal forecast densities. A parameter $\rho_{t,kj}$, used to construct joint density forecasts with a target period $t + h$, where h is the forecast horizon, is estimated only with information available up to time t. Figure

1.3 illustrates the estimation scheme. We use a recursive window estimation scheme. We chose the recursive window scheme on top of the time-varying parameter model to construct out-of-sample forecasts. We define *out-of-sample* as forecasts that would have been available in real time. More details on how we achieve to obtain actual real-time forecasts can be found in Section 1.4. It is important to note that only the copula parameter estimation would be affected by the data revisions, as the survey forecasts are true out-of-sample forecasts. For both surveys, the starting size of the recursive window, denoted by s, is equal to 50 periods.

1.4 Data

In this section, we provide a brief discussion of the two SPF data sets. Appendix A provides a more thorough description of the surveys' procedures. In addition, we plot the smoothed marginals obtained from the SPF histograms.

1.4.1 Euro Area Data

The ECB Survey of Professional Forecasters started at the beginning of 1999 and is a quarterly survey that asks experts, working in the financial and non-financial industry, for their forecasts of eurozone output, inflation, and unemployment.[20] The output measure to be forecasted is the real GDP growth of the eurozone, the inflation measure is the Harmonised Index of Consumer Prices (HICP)[21] and unemployment is the seasonally adjusted level of the eurozone unemployment rate. Our dataset ranges from 1999:Q1 to 2017:Q2. The total number of realisations and forecasts we match is 70.

Besides point forecasts, the survey reports marginal density forecasts by asking the participants to provide a discrete number of percentiles of the forecast density in the form of a histogram. Depending on the year and quarter, the number of percentiles provided in the questionnaire ranges from 8 to 22. Our estimation is based on *the consensus* of these marginal distributions, which is also the measure directly published on the website of the ECB.[22] The consensus is obtained by an equal-weighted averaging over all answers for a given percentile. In total, the survey provides data on up to six different types of forecast horizons: end-of-current year (fixed-event), end-of-next year (fixed-event), end-of-the-year after next (fixed-event),

[20] The ECB website regarding the survey can be found under https://www.ecb.europa.eu/stats/ecb_surveys/html/index.en.html

[21] Appendix A describes in more detail the properties of the HICP.

[22] Genre et al. (2013) show that the simple equal-weighted average is competitive when compared to more refined techniques.

one year ahead (fixed-horizon), two years ahead (fixed-horizon) and five years ahead (fixed-horizon). Our focus will be the one year ahead forecast for a fixed-horizon as it provides the longest time series of density forecasts that we can evaluate at the realisations of the variables. The fixed-horizon forecast ensures that the horizon of the forecast of the dependence parameter is fixed. However, the proposed methodology to estimate joint density forecasts is equally applicable for fixed-event horizons. The timing is such that one-year-ahead forecasts with the origin (beginning of) Q1 has the target Q3. For fixed-horizons (f-h) the growth rates for real output growth and inflation are year-on-year (cumulative) growth rates, i.e. Let Y_{t,q_i}, denotes the level in year t at quarter q_i and h the forecast horizon. Then the year-on-year growth rate is defined as $\frac{Y_{t+h,q_i}}{Y_{t,q_i}} - 1$. For the unemployment rate, the forecasters predict the unemployment rate, i.e., the level of the rate directly. The timing reflects the lag in the official release of the euro area GDP growth data and Appendix A provides details on the timing of the surveys. For the euro area, the joint forecast densities with origin 2012:Q2 and after that are out-of-sample forecast, as we use real-time data.[23] To achieve out-of-sample forecasts, we align each SPF round with the latest official publication of GDP growth, HICP, and unemployment by the ECB, that took place before the submission deadline.

1.4.2 U.S. Data

The U.S. Survey of Professional Forecasters is currently conducted by the Federal Reserve Bank of Philadelphia and started in 1968.[24] It is conducted at a quarterly frequency and provides, besides point forecasts, marginal density forecasts for output growth and inflation.[25] The survey's timing is such that for a survey published at quarter t, information up to $t-1$ is available. Again, for the analysis of the copulas, we will focus on the consensus density forecasts, i.e., the average for a given percentile taken over all respondents. At 1981:Q3 the output measure the survey asks for changed from nominal GNP to real GNP. Therefore, the sample we use starts at 1981:Q3.[26] The inflation measure considered in the survey is the GDP deflator. The total number of realisations and targets that we match for each variable are 139, from 1982:Q2 to 2016:Q4. The timing is such that a Q1 origin has the target Q4. The forecasts are originally fixed-event forecasts. To provide comparability to the

[23] The real-time is taken from the real-time database of the ECB.

[24] Details about the survey can be found at the website of the Federal Reserve Bank of Philadelphia https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/

[25] Again, the data we use is the *final release*.

[26] In the 1992:Q1, the survey changed from real GNP to real GDP. As the two measures of output are closely related, our estimation proceeds without taking the break into account.

euro area survey, we use the methodology proposed for point-forecast by Dovern et al. (2012a) to transform the densities into fixed-horizon forecasts. The forecasts are then annual-average over annual-average predictions, which is different from the year-on-year euro area predictions. Due to the transformation, the density forecasts are not true out-of-sample forecasts. Appendix A provides more details.

1.4.3 SPF Smoothed Marginal Densities

The SPF forecast densities for the U.S. and the eurozone are reported in histograms over a range of values $m^i_{1,t}, ..., m^i_{M,t}$ for variable i at time t. We decide to work with smoothed histograms. Although the theorem of copulas extends to the discrete case, it does so with some pitfalls. For example, an estimated copula that implies an independence copula is in general, not a sufficient condition for independence of the random variables.[27] An interesting alternative could be to model the dependence through intraclass correlations as in Jondeau and Rockinger (2006). The authors split the unit square into different sub-squares and calculate the dependence as a function of joint occurrences in the sub-squares. However, existing work on density forecasts uses almost exclusively continuous distributions, and to guarantee comparability, we stay within the framework of continuous distributions.

In order to smooth the distribution function based on the histogram data, we use a parametric distribution function. The minimisation takes the form of

$$\min_{\mu_{t,i},\alpha_{t,i},\sigma_{t,i},\nu_{t,i}} \quad \sum_{j=1}^{M_t} \left(F_{SPF_{t,i}}(m^i_{j,t}) - T^s_i(m^i_{j,t}; \mu_{t,i}, \alpha_{t,i}, \sigma_{t,i}, \nu_{t,i})\right)^2 \tag{1.17}$$

where $F_{SPF_{t,i}}(m^i_{j,t})$ is the probability $Pr(X_i < m_{j,t})$ implied by the SPF histograms for point $m^i_{j,t}$ at time t for variable i and $T^s_i(m^i_{j,t}; \mu_{t,i}, \alpha_{t,i}, \sigma_{t,i}, \nu_{t,i})$ (hereafter denoted by $T^s_{i,t}$) is the distribution function, evaluated at $m^i_{j,t}$, of the skewed student's distribution proposed by Azzalini and Capitano (2003).[28] Estimation of the parameters $\mu_{t,i}, \alpha_{t,i}, \sigma_{t,i}, \nu_{t,i}$ is performed independently for every time step and will result in a set of parameters $\left\{\mu_{t,i}, \alpha_{t,i}, \sigma_{t,i}, \nu_{t,i}\right\}_{t=1}^{T}$, where $\mu_{t,i}$ measures the centrality of the distribution, $\alpha_{t,i}$ the skewness, $\sigma_{t,i}$ the scale and $\nu_{t,i}$ is the degree of freedom of the Student's t. The unemployment rate is naturally truncated at zero whereas the skew-t distribution has the entire real line as its support, i.e. the skew-t allows for negative values. Hence, we use the truncated normal as the parametric smoother for the unemployment rate. A truncated normal is characterised by four parameters

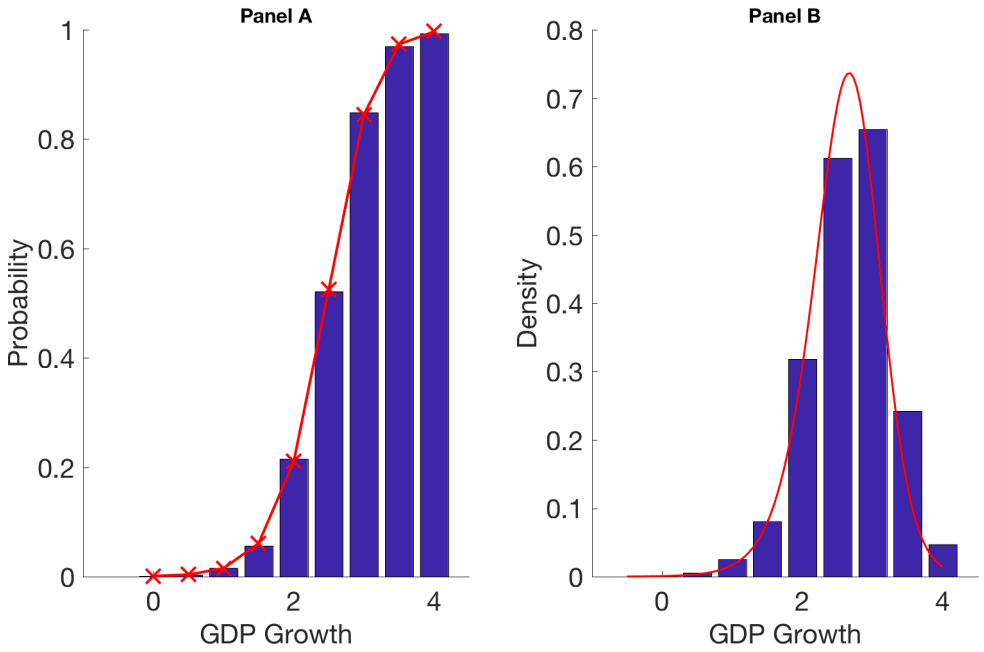
[27] See for example Genest and Nešlehová (2007).

[28] We are not the first to use this distribution function to obtain smooth densities; see for example Adrian et al. (2016).

$(\mu_t, \sigma_t^2, a, b)$ where (μ_t, σ_t^2) are the location and scale of the distribution and a and b are the bounds of the support. We choose $a = 0$ and leave the upper limit of the support unbounded, i.e. $b = \infty$.

The skewed Student's distribution is a flexible function that allows for skewness as well as fat tails. Figure 1.4 displays an example of the matching. Panel (a) displays $T^s_{GDP,2002:Q4}$, i.e. the smoothed distribution function computed for GDP growth for the range of values $[0, 0.5, ..., 3.5, 4]$. If the red crosses are positioned on the top of the respective bar, the resulting matching error is zero.

Figure 1.4: Example of a Smoothed Marginal

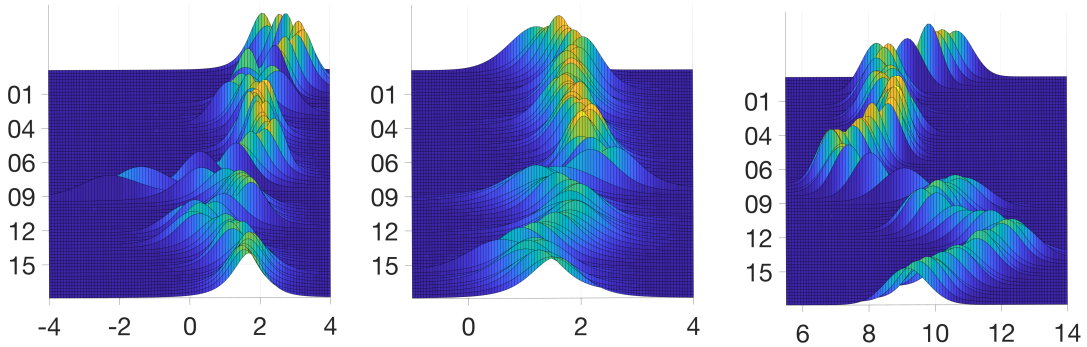


NOTE. – In panel A, the red crossed line displays the estimated smoothed cdf evaluated at the SPF histogram cutoff points. The blue bars displays the cdf computed directly from the SPF histogram. The bars are centred around the mid-point of the respective SPF probability interval. Panel B displays the SPF histogram in blue bars and the smoothed estimated density. The blue bars are centred around the middle of the respective histogram interval.

Figure 1.5 and Figure 1.6 display the smoothed marginal densities forecasts for the euro area, and the U.S. SPF surveys over time. The figures highlight graphically two interesting features of the data set. First, the euro area forecasts have experienced a location shift for both, inflation and the unemployment rate. Although the U.S. inflation measure also experienced a location shift during the 80s, it has remained largely constant over the last two decades. Second, all three euro-area distributions exhibit a strong surge in scale, whereas the scale for the U.S. decreased. The euro-zone distributions remained more spread out after the Great Recession, a fact that will be well reflected in the uncertainty measure.[29] In contrast, the U.S. forecast

[29]Kenny et al. (2015b) use individual-level densities of the euro area SPF and find that the vari-

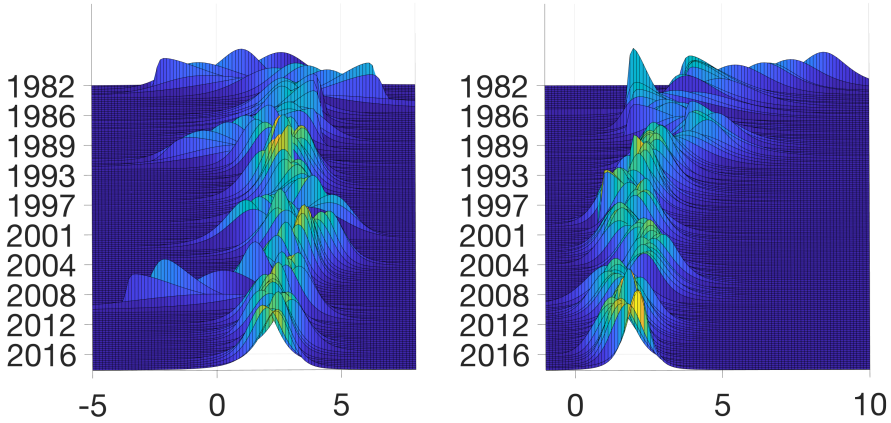
Figure 1.5: Smoothed Marginal Density Forecasts for the Euro Area



(a) Output Growth (b) Inflation (c) Unemployment Rate

NOTE. – The time axis is target aligned, i.e. a forecast displayed at time t was made *for* t and not *in* t. The y-axis displays the values in %. The z-axis displays the density. Panel (a) displays the smoothed marginal density forecasts of real output growth over time. Panel (b) displays the smoothed marginal density forecasts for inflation. Panel (c) displays the smoothed marginal density forecasts for the unemployment rate.

Figure 1.6: Smoothed Marginal Density Forecasts for the U.S.



(a) Output Growth (b) Inflation

NOTE. – The time axis is target aligned, i.e. a forecast displayed at time t was made for t and not in t. The y-axis displays the values in %. The z-axis displays the density. Panel (a) displays the smoothed marginal density forecasts of real output growth over time. Panel (b) displays the smoothed marginal density forecasts for inflation.

distributions exhibit a rise in scale during the Great Recession but return to the pre-crisis level after a few quarters. In comparison to the conditional distribution of GDP growth of Adrian et al. (2016), the U.S. SPF data exhibits more shifts in location whereas the distribution in Adrian, Boyarchenko and Giannone (2017) shows

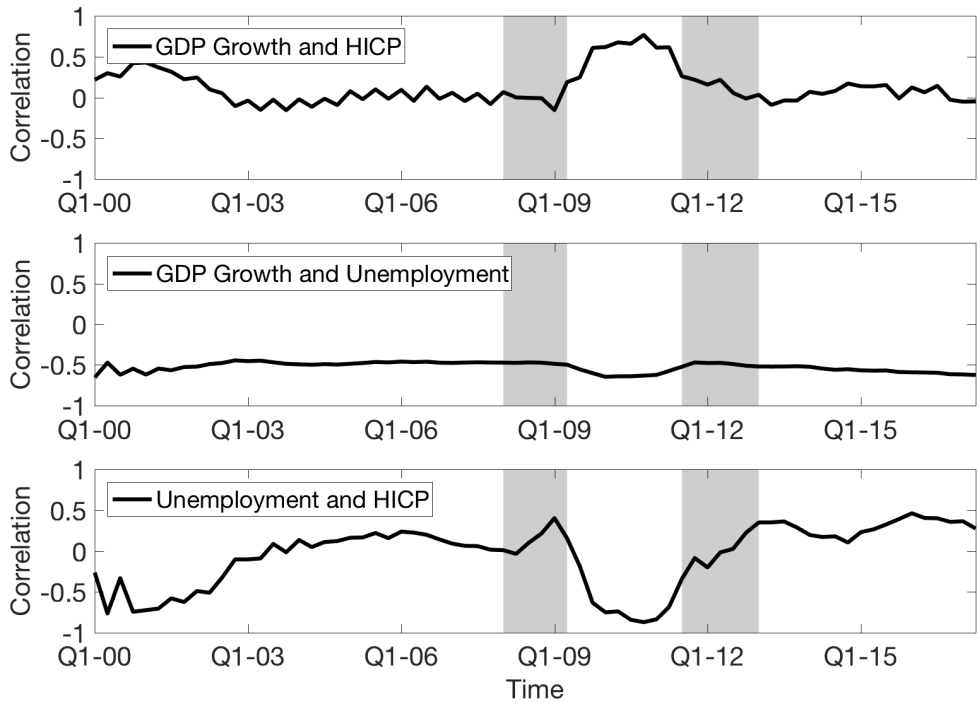
ance increased after the Great Recession but that the majority of forecasters exhibits a downward bias in their variances.

a stronger downside asymmetry with fatter tails.

1.5 Eurozone Results on the Joint Densities

Estimation results show large movements in the expected dependence between GDP and HICP and HICP and the unemployment rate. Figure 1.7 displays the estimated copula parameters. The time axis is origin aligned, i.e. the parameter value is plotted against the date on which the forecast of the joint density based on the respective correlation parameter would have been made. The shaded areas are the period from peak to trough as determined by the Business Cycle committee of the Centre for Economic Policy Research (CEPR).

Figure 1.7: Time-Varying Copula Parameters for the Eurozone



NOTE. – The lines display the evolution of the time-varying copula correlations for the eurozone. The dates on the time axis are origin based. For example, the parameter value plotted against 2009:Q1 is based on information available at the beginning of 2009:Q1. Shaded areas mark CEPR recession periods.

With the end of the Great Recession, dependence between GDP and inflation increased from about zero to around 0.65, negative dependence between GDP and the unemployment rate is constant with only minor fluctuations around the time after the Great Recession. For the dependence of HICP and the unemployment rate we observe a drop, from slightly positive for origins right before the financial crisis,

to -0.9 for the period during and after the Great Recession. Starting with 2012 the dependencies moved back to approximately pre-crisis level. The variation of the dependence will affect the *joint uncertainty* measure and the probabilistic indicators for the state of the economy because these measures are functions of the marginals *and* the dependence. We compare the density forecast performance of the SPF to two models, an AR(1) (AR-SV) with stochastic volatility and a time-varying parameter VAR(2) (BVAR) with stochastic volatility.[30] Using real-time data, when we test for statistically significant differences, the SPF overall outperforms the two competing models for an out-of-sample period from 2008:Q1 to 2016:Q4. This result holds for both univariate as well as multivariate density forecasts. Appendix **??** provides a detailed table of the univariate and multivariate density forecast comparisons.

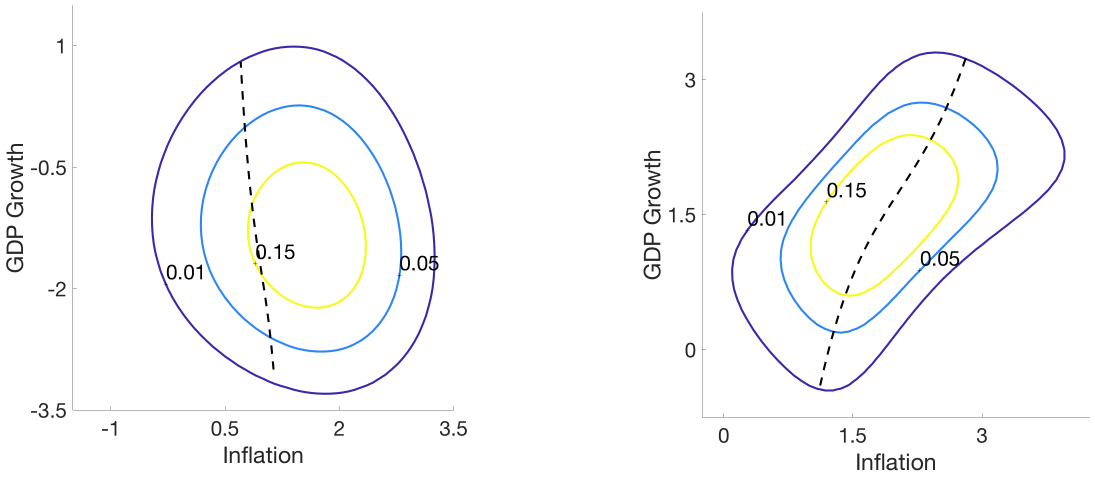
Figure 1.8 displays two contour plots for the joint distribution of output growth and inflation at density levels of $0.01, 0.05$ and 0.15 that show several features of the estimated joint forecast densities. The dashed black line depicts the conditional mean of inflation, conditioned on the respective values of output growth. It provides a straightforward graphical way to assess whether the conditional mean dependence is linear. Panel (a) displays the contour plot for output growth and inflation for the origin period of 2009:Q1, i.e. a target of 2009:Q3. Panel (a) of Figure 1.7 shows that for 2009:Q1, the expected dependence between output and inflation was close to zero, $\rho_t = -0.03$. In the midst of the Great Recession, the joint density predicts a negative growth and a low to moderate inflation. The contour plot does not resemble an ellipse because of the non-gaussian marginals, which allow for both asymmetry and fat-tails. For output growth, the skewness parameter, $\alpha_{t,gdp}$, is positive, which is reflected in the relatively longer positive tail along the horizontal axis for the 0.01 contour isoline. Further, the distributions does not exhibit excessive tails, which shapes the contour into a circular instead of squared-edged region. The distribution for inflation is negatively skewed, which is reflected in the larger distances between different isolines for smaller values on the x-axis.

In contrast to panel (a), the dependence for the time period displayed in Panel (b) is positive and strong with $\rho_t = 0.64$. Further, the distributions exhibit fatter tails and positive asymmetry for inflation. The fat-tails let the isoline appear more rectangular than elliptical. For both cases, the distributions deviate largely from the contour plot of a bivariate normal. The multivariate density predicts the euro area economy to exhibit by positive output growth and positive inflation. In addition, the conditional mean exhibits some non-linearities as depicted by the dashed black line. It is steeper for values around the mean of output growth and less steep in the

[30] A detailed description of the models can be found in Appendix D and E.

tails, which implies a stronger dependence between GDP and inflation for values around the centre of the distribution.

Figure 1.8: Contour for Output Growth and Inflation for the Eurozone



(a) 2009:Q1 (b) 2011:Q1

NOTE. – The date is origin aligned, i.e. it displays the forecast made in 2009:Q1 and 2011:Q1. The x-axis displays inflation values in %. The y-axis displays real output growth values in %. The isolines are for density levels of 0.01, 0.05 and 0.15 and are based on the joint densities estimated from the smoothed SPF marginals and the Gaussian copula. The dashed black line displays the conditional mean of inflation, conditioned on output growth.

1.5.1 SPF Relative Forecasting Performance

We compare the univariate density forecasts of the SPF to two popular time series models. The aim of this exercise is twofold. First, we want to evaluate the forecasting performance of the (smoothed) univariate density forecasts of the SPF. Second, we want to evaluate the forecasting performance of our estimated *survey-based* multivariate density forecasts.

The competing models are an AR(p_{ar}) (AR-SV) with stochastic volatility and a bayesian time-varying parameter VAR(p_{var}) with stochastic volatility (BVAR).[31] These models are known for models performing well for density forecasts (D'Agostino et al., 2013; Clark and Ravazzolo, 2015). In addition, we compare the forecasting performance of the joint densities estimated based on the SPF marginals and the BVAR.

[31]The lag length $p_{\text{ar}} = 1$ and $p_{\text{var}} = 2$ are chosen according to the BIC.

Table 1.1: Density Forecast Comparison - Euro Area

Panel A: Log Scores

	2008:Q1 - 2016:Q4				2009:Q3 - 2016:Q4			
Model	Y	π	U	(Y, π, U)	Y	π	U	(Y, π, U)
AR-SV	17.41	0.79	0.26	-	0.63‡	1.32*	0.76†	-
	(13.90)	(0.69)	(0.37)	-	(0.20)	(0.76)	(0.27)	-
BVAR	18.97*	7.46†	9.12†	10.01†	4.75‡	9.26†	9.82	5.40‡
	(10.10)	(3.50)	(3.20)	(4.80)	(1.54)	(4.25)	(6.01)	(1.36)

Panel B: CRPS Scores

	2008:Q1 - 2016:Q4				2009:Q3 - 2016:Q4			
Model	Y	π	U	(Y, π, U)	Y	π	U	(Y, π, U)
AR-SV	0.52‡	0.32	0.25†	-	0.56‡	0.53†	0.34‡	-
	(0.17)	(0.22)	(0.10)	-	(0.20)	(0.23)	(0.10)	-
BVAR	1.58‡	0.54	0.40‡	2.32‡	1.66‡	0.78‡	0.31‡	2.51‡
	(0.40)	(0.30)	(0.14)	(0.53)	(0.37)	(0.31)	(0.11)	(0.55)

NOTE. – Panel A displays the log-scores of the SPF minus the log-score of the AR-SV and the BVAR model respectively. Panel B displays the negative orientation of CRPS difference of SPF and the AR-SV and the BVAR model respectively. In both cases, positive numbers indicate a superior performance of the SPF. The number of out-of-sample observations is 36 and 30 for the sample that excludes the financial crisis respectively. Y denotes year-on-year GDP growth, π denotes the year-on-year HICP inflation and U denotes the unemployment rate. (Y, π, U) denotes the multivariate density forecast. HAC standard errors are in parenthesis and *, † and ‡ denote significance at the 10%, 5% and 1% level.

Both, the AR-SV and the BVAR, are estimated over a rolling window, with a window size of 40.[32] To imitate realistic forecasting conditions we use real time data for both models to produce the forecasts. The out-of-sample period starts at 2008:Q1. We test for differences in log-scores and continuous ranked probability scores (CRPS) of the two econometric models to the SPF based forecasts, using the out-of-sample from 2008:Q1 to 2016:Q4. In addition, we compute the difference for a sample that excludes the financial crisis for two reasons. First, the SPF density forecasts after the Great Recession are less over-confident in general, i.e. the density forecasts improved. Secondly, the Great Recession is a rare and volatile period and potentially not representative for the forecasting performance of the rest of the sample. The following table compares the log-scores (CRPS) of the two econometrics models relative to the eurozone SPF, such that positive numbers indicate a better performance of the SPF density forecasts. Based on the log score, the SPF density forecasts significantly outperform the BVAR density forecasts for both GDP growth and unemployment, and for unemployment for the period that includes the Great Recession. For the AR-SV model, the log-score difference remains positive throughout but the SPF's performance is only significantly better for the period that excludes the Great Recession. Using the CRPS, all but the inflation forecasts

[32]For a description of the models please see D and E.

perform statistically significantly better.

1.5.2 Economic Downturns, Target Inflation and Rare Events

Let Y_t be the GDP growth rate and U_t be the unemployment rate of the eurozone, then we define the ex-ante *economic downturn* probability of the euro area as

$$\xi_{t+4|t} \equiv F_{t+4|t}(Y_{t+h} < \tau^{bc}_{t+4|t,Y}, U_{t+4} > \tau^{bc}_{t+4|t,u}) \tag{1.18}$$

The baseline specification uses $W = 2$, i.e. the average of the two quarters preceding the forecast origin. The threshold then takes the form of $\tau^{bc}_{t+4,i} \equiv \frac{1}{2}\sum_{w=1}^{2} x_{i,t-w+1}$.[33]

We start by comparing the bivariate measure that we propose to the univariate analogues, i.e. for GDP growth

$$\xi^{Y}_{t+4|t} \equiv F^{\mathrm{Y}}_{t+4|t}(Y_{t+4} < \tau^{bc}_{t+4|t}) \tag{1.19}$$

and for unemployment

$$\xi^{\mathrm{U}}_{t+4} \equiv F^{\mathrm{U}}_{t+4|t}(U_{t+4} > \tau^{bc}_{t+4|t}) \tag{1.20}$$

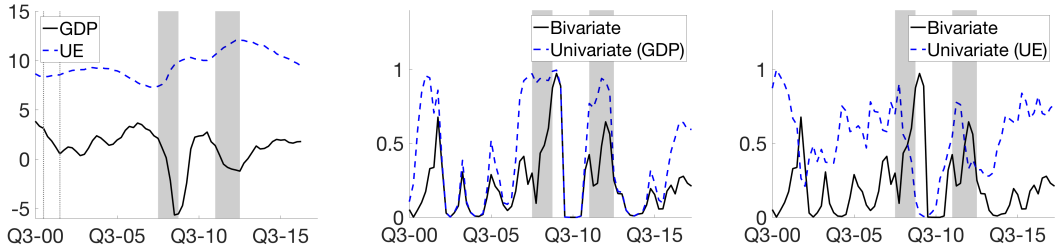
where $F^{\mathrm{Y}}_{t+4|t}$ and F^{U}_{t+4} are the respective marginal forecast densities of GDP growth and unemployment. Figure 1.9 plots in Panel (a) the realised values of GDP and unemployment over time and in Panel (b) and (c) we show the comparison of the *economic downturn* measure based on the joint density, $\xi_{t+4|t}$, to ξ^{U}_{t+4} and ξ^{Y}_{t+4}. In Panel (a), the beginning of the 2000s emphasises, why looking at more than one variable can be important in order to determine the state of the Business Cycle. From the end of 1999 until the beginning of 2002, the GDP growth rate decreased by 3.8 percentage points, but the unemployment rate only increased by about half a percentage point during the same time period. In comparison, during the eurozone debt-crisis recession, GDP growth decreased by 3.9 percentage points but unemployment increased by over 2 percentage points. The CEPR did not call a recession for the period of the early 2000s, although looking only at GDP growth rates could suggest to do so. Panel (b) and (c) display the univariate versus the bivariate *economic downturn* probabilities, ξ^{Y}_{t+4} vs ξ_{t+4} and ξ^{U}_{t+4} vs ξ_{t+4}, and the bivariate probability is clearly preferable.

The competitor model we use is the BVAR evaluated in 1.5.1. As the BVAR produces joint density forecasts, we can define the recession probability identically

[33] Appendix F shows results for an alternative, more conservative, definition of an *economic downturn*.

to the definition in (F.9). The model is estimated using a rolling window and the first out-of-sample density forecast of the model is made for the target 2008:Q1.[34]

Figure 1.9: Bivariate vs Univariate *economic downturn* Probabilities



(a) Realised Values (b) $\xi_{t+4|t,\mathrm{Y}}$ vs $\xi_{t+4|t}$ (c) $\xi_{t+4|t,\mathrm{U}}$ vs $\xi_{t+4|t}$

NOTE. – The dates on the time axis are target based. In Panel (a), the solid line displays realised, cumulative, year-on-year GDP growth rates and the dashed line displays the unemployment rate. In Panel (b) and (c), the solid line denotes the one year-ahead probability of an *economic downturn* based on the joint density forecast. In Panel (b) the dashed line denotes the *economic downturn* probability based on the univariate measure of GDP growth, as defined in (1.19). In Panel (c) the dashed line denotes the *economic downturn* probability based on the univariate measure of the *economic downturn*, as defined in (1.20). Shaded areas mark CEPR recession periods. The grey, dotted vertical lines denote the early 2000s U.S. recession.

Figure 1.10 displays the one year-ahead *economic downturn* probabilities as defined by $\xi_{t+4|t}$ above. The solid line is the *economic downturn* probability calculated using the joint density forecasts estimated based on the SPF data. As discussed in the estimation section, from 2012:Q2 onwards, the values are out-of-samples estimates. The dashed line denotes the respective probability computed using the joint density forecast from the BVAR. The crossed markers display the actual realisations of the event *lower GDP growth and higher unemployment relative to today*.

Overall, the SPF did somewhat worse for the Great Recession than the BVAR model but somewhat better for the euro debt crisis and the aftermath. The beginning of the sample is marked by a spike of $\xi_{t+4|t}$, with a peak at the target date 2002:Q2 (i.e. origin 2001:Q4) of about 70%. While the euro area as a whole did not experience a recession during that period, following the CEPR dating, the strong surge of the *economic downturn* probability can be rationalised by at least two important events that occurred. First, forecasters could have been wary of the eurozone reaction to the ongoing U.S. economic turmoil.[35] The U.S. recession had its trough at mid 2001:Q4 and the forecast origins for the highest values of $\xi_{t+4|t}$ are

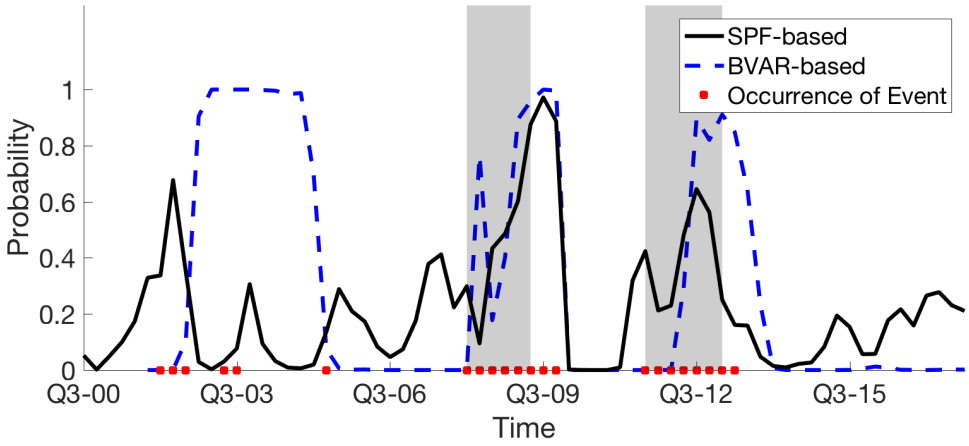
[34] An alternative could be to compare the performance to the Eurocoin index, which is computed by the Bank of Italy and published on the website of the Centre for Economic Policy Research (CEPR). Similar to the Eurocoin index, the Federal Reserve Bank of Philadelphia publishes the Aruoba-Diebold-Scotti Business Conditions Index based on Arouba et al. (2009).

[35] The NBER recession dates for the early 2000s assign the peak to March 2001 and the trough to November 2001.

dated earlier than the U.S. trough. Second, Germany entered a recession during mid to end 2001 and France a near-recession during the second half of 2001. In other words, two major member countries already experienced an economic downturn. The solid, crossed line displays euro area GDP growth, which experienced a strong decrease although it stayed positive. Overall, the period of the early 2000s therefore fulfils the criteria of a time of *diminishing activity* in GDP. The BVAR picks up that movement of GDP calls strongly *economic downturn*. In addition, it is important to note that the early 2000s are the beginning of the data set over which the BVAR is estimated, and the prediction is entirely in-sample.

For the Great Recession, the joint density based on the SPF is late in the forecast of the *economic downturn* and determines the trough three quarters after the date of CEPR. In other words, forecasters overestimated the length of the *economic downturn*. Again, it is important to note that, in contrast to the joint density forecast, the CEPR dating procedure is in-sample and the final decision on the date was made at the end of 2010. The BVAR is quite precise out-of-sample, as the model recognises early on the probability of an economic downturn. Similar to the SPF-based forecast, the BVAR overestimates the length of the *economic downturn* using the indicator proposed in (F.9).

Figure 1.10: Probability of Economic Downturns

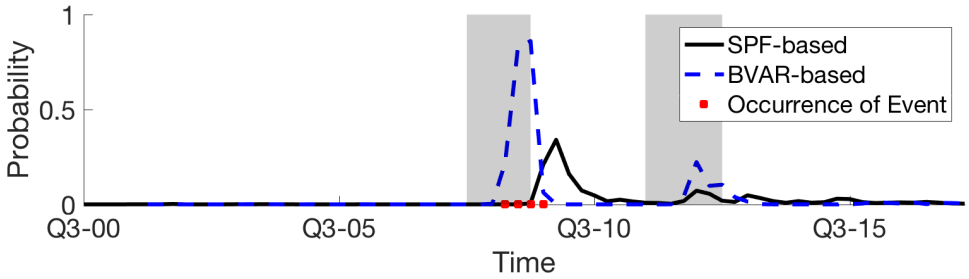


NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the beginning of 2009:Q2. The solid line denotes the one year-ahead probability of lower GDP growth and higher unemployment based on the joint density forecast using the SPF data. The dashed line denotes the respective probability based on the BVAR. The crossed markers denote the realisation of the event of lower GDP growth and an increase in the unemployment rate. Shaded areas mark CEPR recession periods.

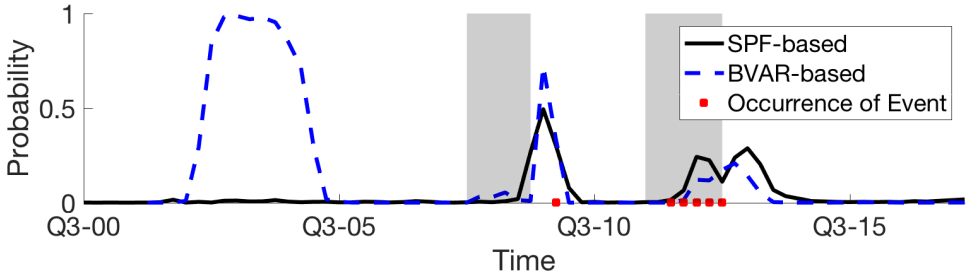
On the other hand, the euro area debt crisis that led to the second *economic downturn* is estimated quite accurately out-of-sample by the SPF and the plotted probability is in real time as the sample is for the period after 2012:Q2. The BVAR picks

up the second dip of the Business Cycle somewhat late when compared to the CEPR recession dates. The increase in a *economic downturn* probability during the years 2015 to mid 2017 is in line with IMF reports (IMF, 2014, 2016) on a remaining risk of a triple-dip economic performance in the euro area. Here again, the performance of the SPF forecasts are superior to the BVAR model. Summarising, the forecasters overestimated the length of the Great Recession but improved timing for the euro area debt crisis. However, we conclude that joint density forecast could be an interesting tool to predict *economic downturns*, in particular a combination of model and survey-based densities.

Figure 1.11: Probabilities of Rare Events



(a) Low Inflation & Negative GDP Growth



(b) High Inflation & Negative GDP Growth

NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the beginning of 2009:Q2. In panel (a), the black solid line denotes the one year-ahead probability of negative output growth and deflation. In panel (b), the black solid line denotes the one year-ahead probability of negative output growth and high inflation. The dashed line is the respective BVAR model implied probability. Shaded areas mark CEPR recession periods.

Panel (a) of Figure 1.11 displays ${}_{-}\zeta_{t+4|t}$, the one-year ahead probability of negative output growth accompanied by a low inflation in the euro area. The BVAR predicts rather well the event of negative output growth and low inflation, which is driven by low mean forecasts as well as the high volatility of the model during that period. For the euro area debt crisis and the remaining crisis the probabilities of a low inflation recession increase slightly but remain very low. In contrast, the SPF-based predictions are lagged with respect to the realisations of the actual events.

Panel (b) of Figure 1.11 displays ${}_{+}\zeta_{t+4|t}$, the one-year ahead probability of a recession accompanied by high inflation in the euro area. The graph remains flat, according to the SPF-based model there was no risk of negative output growth and very high inflation since the establishment of the euro currency area. The BVAR spikes somewhat during the Great Recession and calls a high risk during the beginning of the 2003 period. Both spikes are largely driven by the high variance of the BVAR model density during that period.

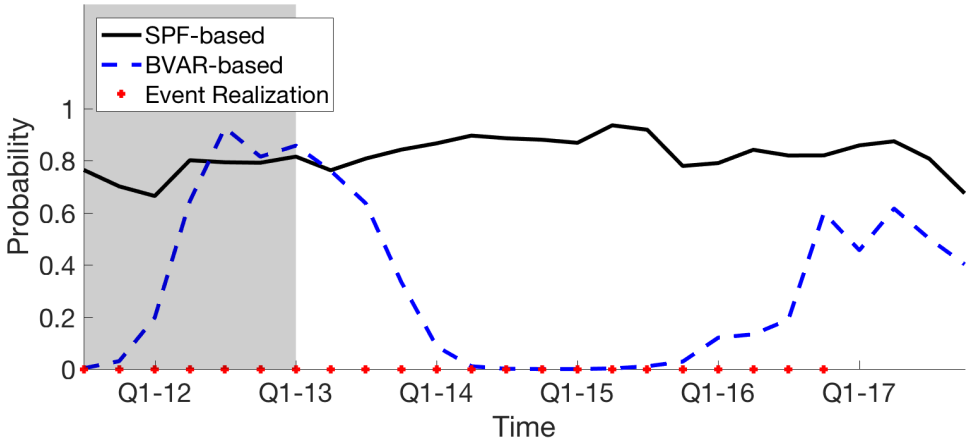
1.5.3 Probability of Remaining in a Secular Stagnation

We define the secular stagnation threshold for variable i, as the pre-financial crisis averages, i.e. $\tau_i^s = \frac{1}{t_f}\sum_{w=1}^{t_f} x_{i,w}$, where t_f denotes 2008:Q1 and $x_{i,w}$ denotes the realisation of the respective values of GDP growth, inflation and of the unemployment rate. Then the probability of remaining the secular stagnation region can be computed using

$$\eta_{t+4|t} \equiv F_{t+4|t}(Y_{t+4|t} < \tau_y^s, \pi_{t+4|t} < \tau_\pi^s, U_{t+4|t} > \tau_u^s) \tag{1.21}$$

where again $F_{t+h|t}$ denotes the joint density forecast.

Figure 1.12: Eurozone Probability to Remain in the Secular Stagnation Region



NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the beginning of 2009:Q2. The black solid line denotes the one year-ahead probability for remaining in a secular stagnation region. Shaded areas mark CEPR recession periods.

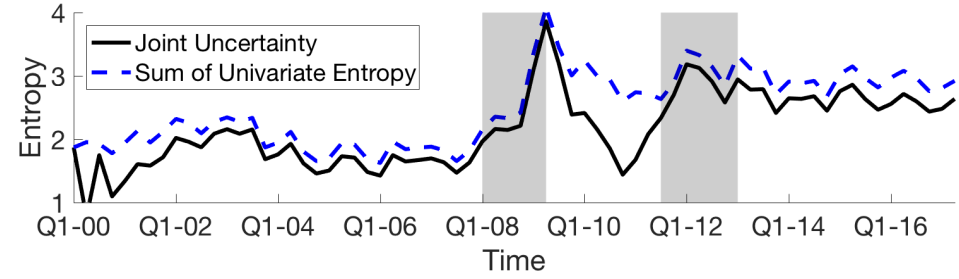
Figure 1.12 displays the one year-ahead probability of remaining in a secular stagnation region, i.e. both output growth and inflation remaining below and unemployment remaining above a pre-crisis level. We can observe a drop for the time before 2012:Q1 and a sharp increase of almost 15% during the euro area debt crisis. With

the end of the second euro area recession, the probability decreases slightly only to surge an additional 12% a quarter later. The outlook on future development in the euro area remains pessimistic.

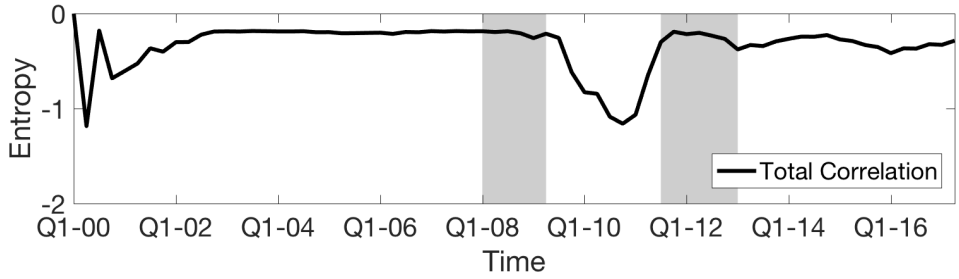
1.5.4 Joint Uncertainty

Panel (a) of Figure 1.13 displays the *joint uncertainty* for the eurozone. Note that the graph is origin based in order to display uncertainty at the time of the forecast origin.The graph reveals three important aspects. First, uncertainty surges with the onset of the Great Recession with a peak for the forecast origin of 2009:Q2. Second, although the macroeconomic uncertainty spikes during the euro area debt crisis recession, it does much less so relative to the Great Recession. Third, *joint uncertainty* has experienced a level shift over time. In particular, euro area uncertainty did only briefly converge back to the pre-Great Recession level for the period of 2009:Q3 to 2011:Q3 and remains currently still higher than prior to the 2009 economic downturn.

Figure 1.13: Entropy Measures for the Eurozone



(a) Multivariate vs Univariate Entropy



(b) Total Correlation

NOTE. – The dates on the time axis are origin based. For example, the uncertainty plotted against 2010:Q1 is based on information available at the beginning of 2010:Q1. The solid line in panel (a) displays the entropy based on the joint density forecasts. The dashed line displays the sum of the univariate entropies. The solid black line in panel (b) displays the negative value of total correlation based on the joint density forecasts. Shaded areas mark CEPR recession periods.

To understand the different behaviour of the *joint uncertainty* and the univariate un-

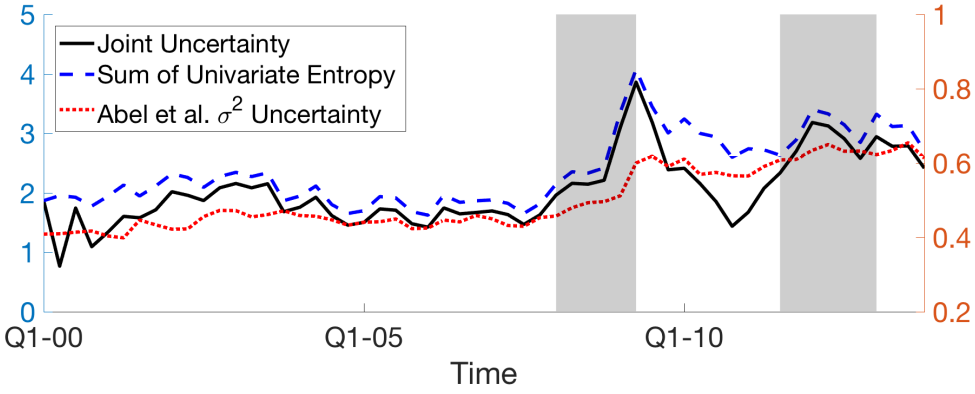
certainty around 2009:Q3 to 2011:Q3 we need to consider Panel (b), which displays the negative value of total correlation. As total correlation is a Kullback-Leibler distance and therefore bounded below by zero, the graph displays directly the contribution to uncertainty. It contributed strongly to a reduction in *joint uncertainty* during the first years of the existence of the euro. Moreover, in the aftermath of the financial crisis, starting during mid 2009 until the beginning of the euro area debt crisis, total correlation contributed an average reduction of 32% to the entropy measure, i.e. the increase in dependencies of the macroeconomic variables reduced macroeconomic uncertainty by about 32%.[36] The importance of $TC(\mathbf{X})$ highlights the fact that *uncertainty* of a multivariate system of variables behaved differently than uncertainty of any of the univariate variables.

Panel (a) and Panel (b) of Figure 1.14 compare our entropy measure to the results of Abel et al. (2016), which are based as well on the euro area SPF data set and from an empirical perspective most closely related to us.[37] Note the different scaling on the right-hand y-axis. The authors provide two measures of uncertainty, one that is based on the median of the variance and one that is based on the median of the interquartile range (iqr) of the individual forecast densities. The dashed lines show the evolution of their measure compared to us. With the onset of the Great Recession, both the variance and the iqr based measure increase markedly. Importantly, the measures do not show a return to pre-crisis levels for the period in between the two recessions, which for *joint uncertainty* is based on the increase in cross-sectional dependence.

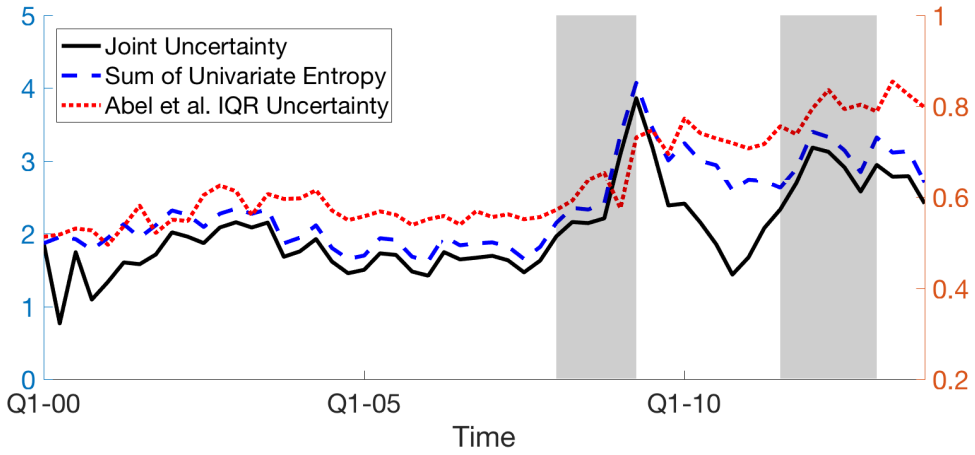
[36] Computed as the average of the percentage difference between *joint uncertainty* and the sum of the univariate uncertainty, i.e. it is the negative value of $\frac{TC(\mathbf{X})}{DE(\mathbf{X})-TC(\mathbf{X})}$.

[37] Abel et al. (2016) provide estimates for each individual SPF variable, i.e. GDP growth, inflation and unemployment. We display the mean of the three measures.

Figure 1.14: Entropy Measures for the Eurozone



(a) Abel et al. Variance Uncertainty



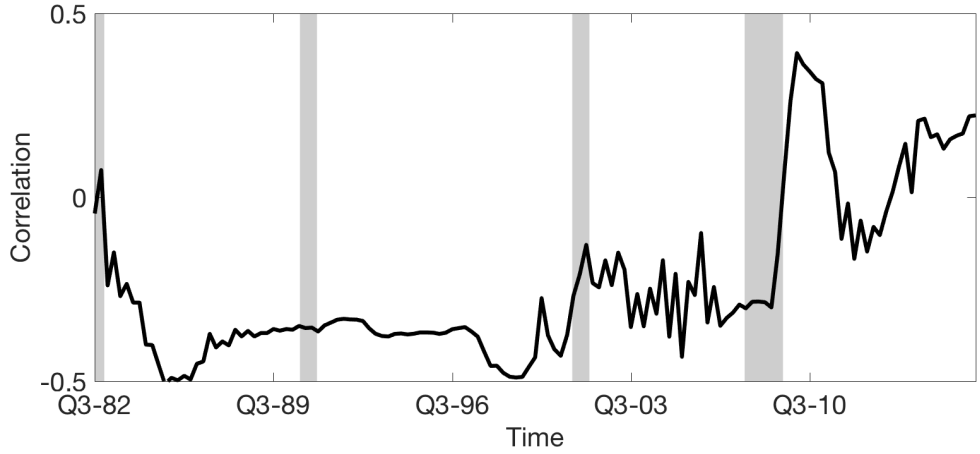
(b) Abel et al. IQR Uncertainty

NOTE. – The left-hand side y-axis denotes the entropy measure proposed in this paper. The right-hand side axis denotes the entropy measure of Abel et al. The dates on the time axis are origin based. For example, the uncertainty plotted against 2010:Q1 is based on information available at the beginning of 2010:Q1. The solid black lines display the entropy based on the joint density forecasts and the dashed lines display the sum of univariate entropies. The dotted lines display the uncertainty measure computed by Abel et al. (2016). Shaded areas mark CEPR recession periods.

1.6 U.S. Results on the Joint Densities

Figure 1.15 shows that the predicted dependence between real output and inflation increases sharply, from around $-.25$ up to 0.4 during and in the aftermath of the Great Recession and the increase is qualitatively comparable to the eurozone results. In 2011 it starts to drop but remains above the pre-crisis level. Again, the increase in predicted co-movement during and after the Great Recession will dampen the increase in *joint uncertainty* relative to a measure that is limited to only consider univariate variances, although less so than in the euro area. Further, probabilities of recessions and disastrous events will be affected by the evolution of the association of GDP and CPI. As well as the euro area data, we compare the forecasting performance of the US SPF to an AR an AR(1) (AR-SV) with stochastic volatility and a time-varying parameter VAR(2) (BVAR) with stochastic volatility.[38] For inflation, the econometric models tend to outperform the two competing models for an out-of-sample period from 1999:Q4 to 2016:Q4. Appendix 1.6.1 provides a detailed table of the univariate and multivariate density forecast comparisons.

Figure 1.15: Time-Varying Parameter Copula for the U.S.

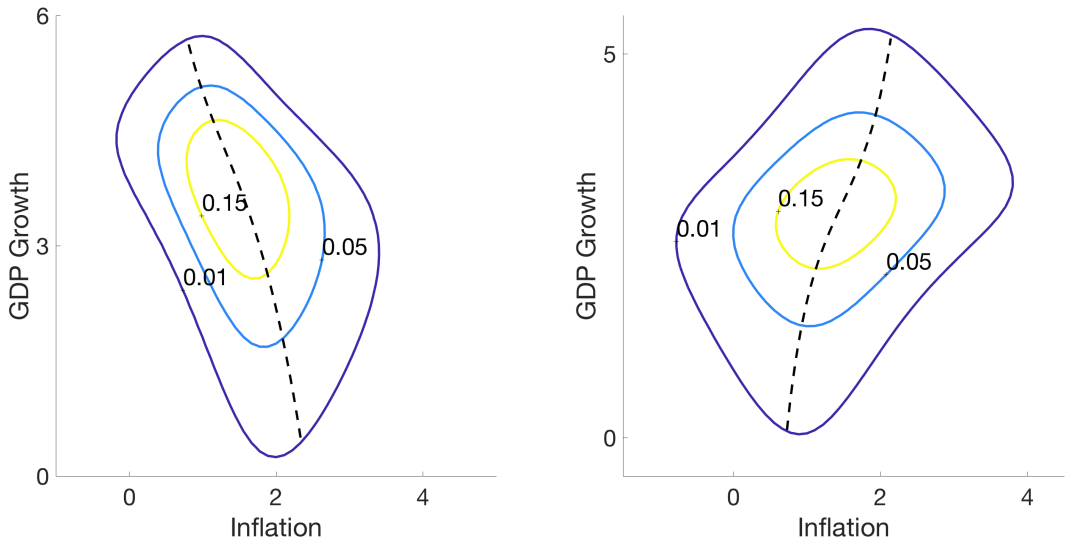


NOTE. – The figure displays the evolution of the time-varying copula correlation for the U.S. The dates on the x-axis are origin based. For example, the parameter value plotted against 2010:Q1 is based on information available at the beginning of 2010:Q1. Shaded areas mark NBER recession periods.

Figure 1.16 shows two example contour plots of the estimated joint forecast density for the origin period 1999:Q2 (target 2000:Q1) and 2010:Q1 (target 2010:Q4), i.e. Panel (a) shows a forecast density with origin during the boom year of 1999 and the second for a forecast origin after the Great Recession. The contour plots allow an insight into several interesting features of both, the marginals as well as the joint distribution function.

[38] A detailed description of the models can be found in Appendix D and E.

Figure 1.16: Contour for Output Growth and Inflation for the U.S.



(a) 1999:Q2 (b) 2010:Q1

NOTE. – The date is origin aligned, i.e. it displays the forecast for 2000:Q1 and 2010:Q3 made in 1999:Q2 and 2010:Q1. The x-axis displays inflation values in %. The y-axis displays real output growth values in %. The isolines are for density levels of 0.01,0.05 and 0.15 and are based on the joint densities estimated from the smoothed SPF marginals and the Gaussian copula. The dashed black line displays the conditional mean of inflation, conditioned on output growth.

Although the forecasts are optimistic for the output growth rate and inflation for the 2000:Q1 period, we observe strong asymmetry along the horizontal axis of output growth in panel (a). The asymmetry suggests that the model not yet assigns a lot of weight to an economic downturn, reflected by the strong positive mean, but at the same time remained wary of a risk of a recession with a relatively high inflation, reflected by the long negative tail along the horizontal axis. Moreover, there conditional mean (solid line) of inflation is not a linear function of GDP growth as the curvature somewhat flattens for smaller growth values. Panel (b) of Figure 1.16 exhibits three interesting features. First, in the beginning of 2010, the model generally implies a positive outlook for the year of 2010 with expectations on positive growth and a moderate inflation. However, although the centre of the distribution is in the positive region of the outcome space, there is a risk of a returning economic downturn for real output growth, as reflected in the long tail. Third, although the relationship between GDP growth and inflation is monotone, it is not from being linear. The dashed black line depicts the conditional mean of inflation, conditioned on output growth. Its shape implies that dependence is weaker for large (small) values of output growth as the line is less steep for large (small) values on the y-axis.

1.6.1 Relative SPF Forecasting Performance

Again, the AR-SV and the BVAR, are estimated over a rolling window, with a window size of 70. The out-of-sample period starts at 1999:Q1.[39] We test for differences in log-scores of the two econometric models to the SPF based forecasts, using the full out-of-sample from 1999:Q1 to 2016:Q4. The econometric models perform better in forecasting the annual-average over annual-average GDP deflator, but only the forecasts of the AR(1) are significantly better. For GDP growth the SPF is slightly better, but the difference is not statistically significant. When comparing the log-scores of the multivariate densities, the BVAR clearly outperforms the SPF based density forecast.

Table 1.2: Density Forecast Comparison - US

	1999:Q1 - 2016:Q4			Excluding Financial Crisis		
Model	Y	π	(Y, π)	Y	π	(Y, π)
AR-SV	0.16	-0.22‡	-	0.21*	-0.17†	-
	(0.13)	(0.08)	-	(0.12)	(0.08)	-
BVAR	0.15	-0.17	-2.60	0.16	-0.25	-2.29‡
	(0.21)	(0.20)	(0.28)	(0.18)	(0.21)	(0.13)

NOTE. – The table displays the log-scores of the SPF minus the log-score of the AR-SV and the BVAR model respectively. The number of out-of-sample observations is 72 and 64 for the sample that excludes the Great Recession respectively. Y denotes annual-average over annual-average one year-head GDP growth, π denotes the annual-average over annual-average one year-head HICP inflation. (Y, π) denotes the multivariate density forecast. HAC standard errors are in parenthesis and *, † and ‡ denote significance at the 10%, 5% and 1% level.

1.6.2 Probabilities of Target Inflation and Rare Events

As we do not have a long time series for US SPF marginal density forecasts of unemployment, we will restrict our analysis to, firstly, predicting, at least, moderate GDP growth and target inflation and, secondly, predicting the joint probability of negative output growth accompanied by low inflation and negative output growth accompanied by high inflation.

Starting with the analysis moderate GDP growth and target inflation, let Y_t denote GDP growth and π_t inflation. We set the value for the target inflation to be in a range of 1.8% to 2.2% and define moderate GDP growth to be, at least, 1%.

$$\delta_{t+4|t} \equiv F_{t+4|t}(Y_{t+4} > 1, 1.8 < \pi_{t+4} < 2.2) \tag{1.22}$$

Figure 1.17 displays the probability $\delta_{t+4|t}$. The models overall attribute higher prob-

[39] The out-of-sample for the joint density forecasts starts as well in 1999:Q1, as the periods before are used for the estimation of the copula parameter.

abilities to time periods where more of the events realised. In comparison to the SPF-based probability, the BVAR based probability is more volatile and predicts particularly well the events in the period around 1995, 1999 and 2003.

Turning to the analysis of negative output growth accompanied by low inflation and negative output growth accompanied by high inflation, we set the value of τ^{VaR} in

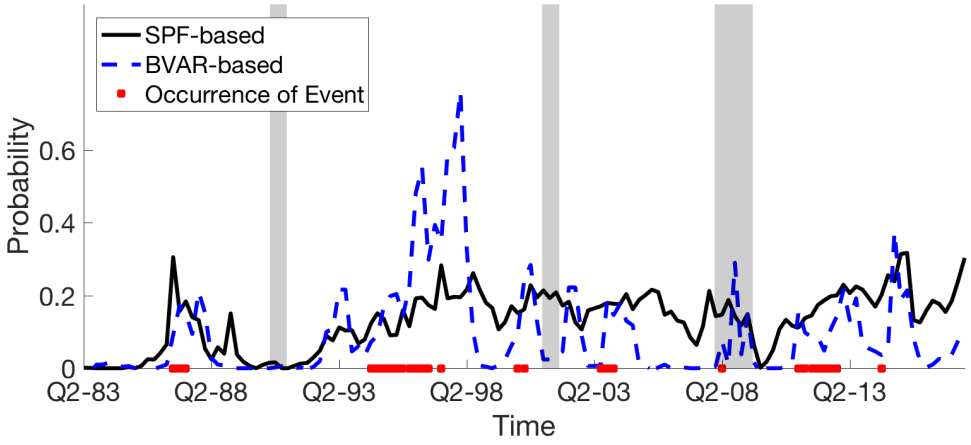
$$_{+}\zeta_{t+4|t} \equiv F_{t+4|t}(Y_{t+4} < 0, \pi_{t+4} > 3) \tag{1.23}$$

to 3%, i.e. an inflation above 3% is considered a high inflation. The value for *low inflation* is set to 1, such that

$$_{-}\zeta_{t+4|t} \equiv F_{t+4|t}(Y_{t+4} < 0, \pi_{t+4} < 1) \tag{1.24}$$

Similar to the euro area results, we compare the results to a time-varying parameter VAR with stochastic volatility (BVAR).

Figure 1.17: Probability of Moderate GDP Growth and Target Inflation



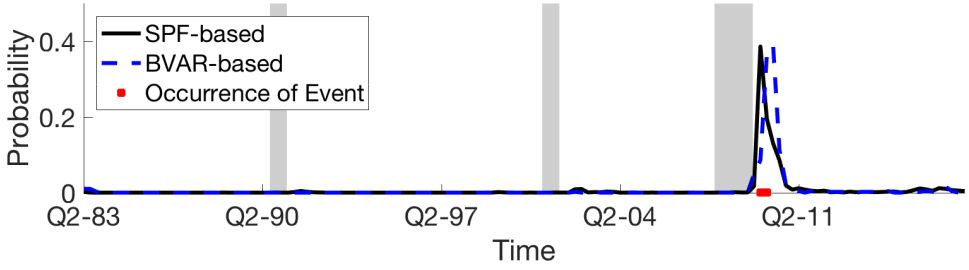
NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the beginning of 2009:Q2. The solid line denotes the one year-ahead probability of, at least, moderate GDP growth and target inflation based on the joint density forecast using the SPF data. The dashed line denotes the respective probability based on the BVAR. The crossed markers denote the realisation of the event of, at least, moderate GDP growth and inflation being within a target range. Shaded areas mark NBER recession periods.

Panel (a) of Figure 1.18 displays $_{-}\zeta_{t+4|t}$, the one-year ahead probability of a recession accompanied by a low inflation. The peak of 11% is reached in quarter 4 of 2009, which is around 6 months after the trough established by the NBER. However, for the annual-average over annual-average rates, the realisation of the event negative GDP growth and low GDP deflator occurs in 2009Q:3 and 2009:Q4. Both models perform rather well in the prediction and the SPF-based outperforms the BVAR

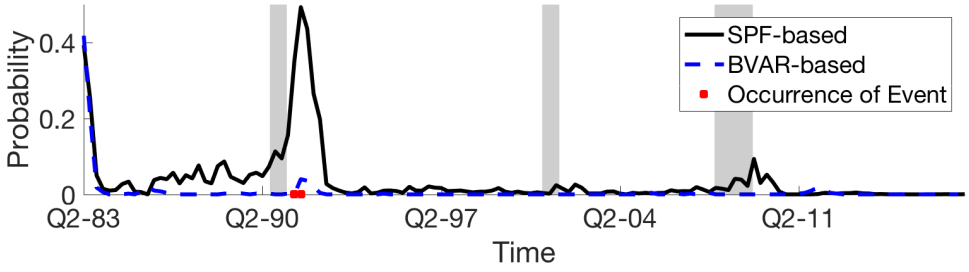
model. Interestingly, the pseudo out-of-sample probabilities remain low for all other U.S. recessions that occurred in the data sample.

Panel (b) of Figure 1.18 displays ${}_{+}\zeta_{t+4|t}$, the one-year ahead probability of a recession accompanied by high inflation. The SPF-based model predicts a high risked of a recession and high inflation for the period after the 1982 U.S. recession that had its trough in November 1982, following the NBER Business Cycle Dating Committee. For the annual-average over annual-average growth rates, the event of negative output growth and high inflation occurs somewhat later than the NBER recession dates. The first target date plotted on the time axis is 1983:Q2. The next U.S. recession occurred in the beginning of the 90s and experienced its trough in March 1991. Output growth remained positive throughout the recession, however, the joint density computes a probability of a recession accompanied by high inflation of around 35%.

Figure 1.18: Probabilities of Rare Events



(a) Low Inflation & Negative GDP Growth



(b) High Inflation & Negative GDP Growth

NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the end of 2009:Q1. In panel (a), the black solid line denotes the one year-ahead probability of negative output growth and deflation. In panel (b), the black solid line denotes the one year-ahead probability of negative output growth and high inflation. The dashed line depicts the BVAR based estimation. Shaded areas mark NBER recession periods.

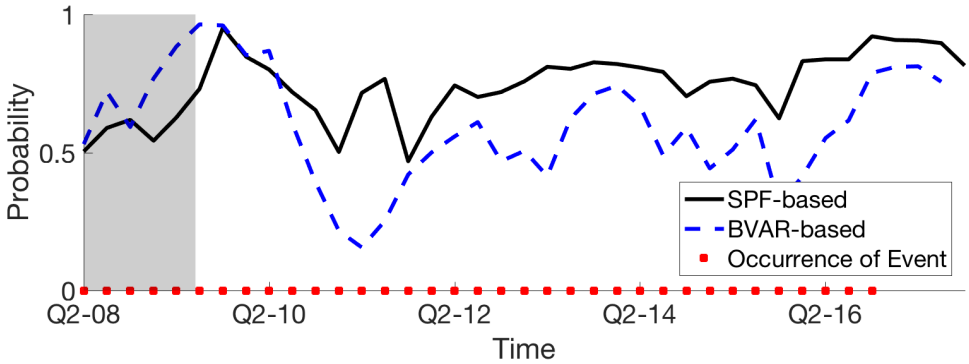
Similar to the plot in Panel (a) and the analogue euro area result, there is a two to three quarter mismatch between the highest values of ${}_{+}\zeta_{t+4|t}$ and the end of the recession due the use of annual-average over annual-average growth rates. Importantly, as the GDP deflator includes energy prices, the Iraqi invasion of Kuwait in

mid 1990 contributed to a severe oil-price shock, which potentially contributed to the risk of a high inflation.

1.6.3 Probability of Remaining in a Section Stagnation

Similar to the eurozone results, we compute the probability of remaining in a secular stagnation region. Due to the lack of unemployment forecasts, the probability is computed based on output growth and inflation only, otherwise the definition is the same.[40] While the probability lowered after the financial crisis, it increased during the years of 2013 to 2015 back to about 85%. Federal Reserve outlooks during that time expected a moderate output growth and low inflation. While unemployment decreased over the last years in the U.S. and nearly reached its pre-crisis level in 2017, the models' predictions for jointly, inflation and output growth, remains low throughout the years of 2016 and 2017. Interestingly, for both the eurozone and the U.S., the SeSt probability increased for the target year 2016, but experienced a slight decrease with the beginning of 2017.

Figure 1.19: U.S. Probability to Remain in the Secular Stagnation Region



NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the end of 2009:Q1. The black solid line denotes the one year-ahead probability for remaining in a secular stagnation region. The dashed line depicts the BVAR based estimation. Shaded areas mark NBER recession periods.

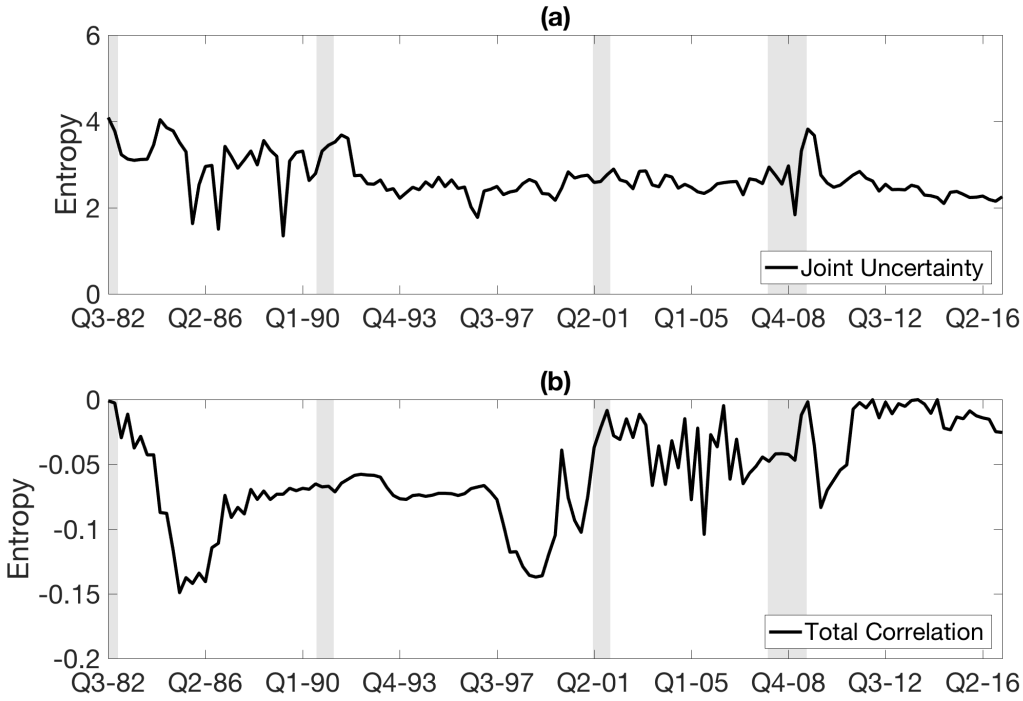
1.6.4 Joint Uncertainty

Panel (a) of Figure 1.20 displays the *joint uncertainty* based on the one year-ahead joint density forecasts. First, the level of joint entropy has declined comparing the beginning and the end of the sample which stands in contract to the euro area results. Second, the volatility has declined as well. Starting in the beginning of

[40] The computation of the thresholds is performed in an analogous way using the respective U.S. data dating back to 1982.

2009, uncertainty spikes for three periods and drops back to the pre-crisis level. Uncertainty of the forecasts surged during the height of the Great Recession and in the immediate aftermath. Panel (b) of Figure 1.20 displays again the negative value of total correlation. We observe that during the recession at the beginning of the new century and in the aftermath of the financial crisis, the dependence of the two variables contribute only slightly to reduction of *joint uncertainty* by about 3%, a pattern that is similar to the eurozone. The contribution to a reduction in uncertainty was somewhat larger for the periods of the early to mid 80s and in the 90s, i.e. to periods of high positive growth. Overall, $TC(\mathbf{X})$ reduces U.S. uncertainty by less than the equivalent measure for the euro area.

Figure 1.20: Entropy Measures for the U.S.



NOTE. – The dates on the x-axis are origin based. For example, the uncertainty plotted against 2010:Q1 is based on information available at the beginning of 2010:Q1. The solid black line in panel (a) displays the entropy based on the joint density forecasts. The solid black line in panel (b) displays the negative value of total correlation based on the joint density forecasts. Shaded areas mark NBER recession periods.

1.7 Conclusion

In this paper, we present a methodology to estimate joint densities based on univariate densities obtained through survey data. Existing analyses have been limited by the univariate focus of the surveys. We demonstrate the usefulness of joint den-

sity forecasts for a variety of applications such as new indicators of the state of the economy and a novel *joint uncertainty* measure. The survey-based joint densities are better or at least competitive to density forecasts from a bayesian time-varying parameter VAR with stochastic volatility and an AR(1) with stochastic volatility. While the BVAR and SPF perform equally well in predicting a recession following the Great Recession, the survey-based forecast is better for the euro area debt crisis and its aftermath. In turn, the BVAR performs better in predicting the intensity of the economic downturn during the Great Recession. Also, we introduce a novel *joint uncertainty* measure that takes dependencies between variables into account. We show that, due to an increase in the expected co-movement in the eurozone after the Great Recession, the rise in uncertainty has been dampened relative to existing univariate uncertainty measures.

In the present paper, we only modeled cross-sectional dependence between different variables through copulas. In general, the methodology readily extends to modeling time-dependence as well, i.e., we can estimate multivariate density forecasts of GDP growth, inflation, and unemployment one-year ahead *jointly with* two-years ahead. The resulting joint densities could be used to answer questions such as: how much does the two-year ahead inflation depend on next year's inflation? This extension is left for future research.

Appendices

A Euro Area Data

The HICP includes all goods and services purchased by household money, i.e. energy and food prices are included, with the exception of expenditures on housing by homeowners. In addition, in european countries health care and education are commonly state-provided and are therefore excluded to the extent to which they are not directly paid by household.[41]

The survey is usually sent out within the first twenty days of the first month of the respective quarter. Participants normally have to reply within a week to ten days to the questionnaire. This timing implies that the forecasts of a survey dated at t is based on information up to $t-1$.[42] In particular, the timing is such that the fixed-horizon is fixed with respect to the last published information regarding that variable. For unemployment and inflation, data becomes available at a higher frequency than for output. In particular, the unemployment data becomes available such that for the first, second, third and fourth quarter survey publication the target month are November, February, May and August. For inflation, the first, second, third and fourth quarter survey publication are aligned with the targets December, March, June and September. Publication of output data is delayed relative to unemployment and inflation. For the surveys published in the first, second, third and fourth quarter the respective fixed-horizon targets for output are the third, fourth, first and second quarter.

These features of the survey have two relevant implications for our analysis. First, one can only estimate the expected dependencies of the variables with the slight misalignment in target dates. Second, while output forecasts are quarterly, inflation and unemployment are monthly. A possible procedure to mitigate the effect could be based on interpolating between different horizons in order to achieve the same target date for all three variables. Further, although monthly realisations of unemployment and inflation vary within a given quarter, the variations are rather small. More importantly, variations in model based forecasts for relatively large horizons tend to be even smaller across neighbouring months. In our baseline estimation, we do not control for this potential confounding factor.

Starting in 2013:Q1 the survey asks participants for the *on the point in the forecast horizon in which according to your baseline outlook the level of real GDP*

[41] More information can be found on https://www.ecb.europa.eu/stats/macroeconomic_and_sectoral/hicp/html/index.en.html

[42] Details about the dates involving the forecast of all published surveys can be found under the tab 'Background' on the ECB's main page for the SPF survey.

will increase again in a sustained manner (i.e.witness successive quarter-on-quarter growth rates larger than 0.0%).[43]

B U.S. Data

Let $f^{FE}_{t+k|t}$ denote the fixed-event forecast k-periods ahead. In the U.S. SPF dataset we observe k-period and $k+4$-period ahead forecasts. The fixed-horizon forecasts are formed through a weighted sum of the k and $k+4$-period ahead forecast where the weights are proportional to the horizon overlap. This means that the FH forecast with origin in t, four periods ahead is denoted by $\hat{f}^{FH}_{t+4+k|t}$ and constructed from

$$\hat{f}^{FH}_{t+4|t} = \frac{k}{4} f^{FE}_{t+k|t} + \frac{4-k}{4} f^{FE}_{t+k+4|t} \tag{B.1}$$

For example, a forecast with origin in quarter 1, 4 periods ahead will be a forecast for quarter 4 of the same year. Therefore, the US density forecasts will be in-sample as todays one year-ahead forecast is constructed using information from future density forecasts. For U.S. and the euro area, final release data is taken from the database of the Federal Reserve Bank of St. Louis.

C Formal Definition of Copulas

A copula is a function that maps from the d-dimensional $[0,1]^d$ unit cube on to an interval on the real line: $C : [0,1]^d \to [0,1]$. Following the definition of Nelsen (2006), a function C that maps from the unit d-cube $\mathbf{I}^d$ to $\mathbf{I} = [0,1]$ is called a copula if

1. For every $\mathbf{u}$ in $\mathbf{I}^d$ it follows that

$$C(u) = 0 \tag{C.2}$$

if at least one coordinate of $\mathbf{u}$ is 0
and
if all coordinates of $\mathbf{u}$ are 1 except u_k, then $C(u) = u_k$.

2. For every $\mathbf{a}, \mathbf{b} \in \mathbf{I}^d$ such that $\mathbf{a} \leq \mathbf{b}$ it follows that

$$V_C([\mathbf{a}, \mathbf{b}]) \geq 0 \tag{C.3}$$

[43] Taken from a sample questionnaire published on
https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html

where $V_C(\cdot,\cdot)$ is the C-volume of the hyperrectangles formed by $[\mathbf{a},\mathbf{b}]$.

D Time-Varying Parameter VAR with Stochastic Volatility

The model is based on Primiceri (2005). The estimation takes into account the correction of the algorithm proposed in Del Negro and Primiceri (2015). We describe the model briefly, for more details, we refer the reader to the original papers. Let y_t be a vector of variables at time t, then

$$y_t = B_{0,t} + B_{1,t}y_{t-h} + .. + B_{p,t}y_{t-p-h} + A_t^{-1}\Sigma_t e_t$$

where

$$B_t = B_{t-1} + v_t \tag{D.4}$$

$$\alpha_t = \alpha_{t-1} + \xi_t \tag{D.5}$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t \tag{D.6}$$

where B_t is matrix of stacked $B_{i,t}$, the α_t are the stacked non-zero elements of A_t and $\log(\sigma_t)$ is the log of the vector of diagonal elements of Σ_t. The variable h denotes the horizon, which here is four quarters-ahead. The errors $(e_t, v_t, \xi_t, \eta_t)$ are assumed to be normal and mutually independent. Let $\Theta^T = [\Theta_1, ..., \Theta_T]$ be the collection of all model parameters, let Φ denote the parameters that are constant and let $y^T = [y_1, ..., y_T]$. The predictive density is computed as

$$p(y_{T+h}|y^T) = \int p(y_{T+h}|y^T, \Theta^T, \Phi)p(\Theta^T, \Phi|y^T)\mathrm{d}\Theta^T \tag{D.7}$$

where $p(y_{T+h}|y^T, \Theta^T, \Phi)$ is a Normal due to normality of e_t and $p(\Theta^T, \Phi|y^T)$ is approximated using a Gibbs sampler as proposed by Primiceri (2005), using the correction of Del Negro and Primiceri (2015). A natural extension, as used for example by D'Agostino et al. (2013), would be to simulate future draws Θ_{T+h} using the posterior distribution of Θ^T and the equations in (D.4),(D.5),(E.8), and then draw from $p(y_{T+h}|y^T, \Theta^{T+1}, \Phi)$.

For the euro area, the data used is real-time data and consists of the three-variables GDP growth, unemployment and inflation. For the U.S. we estimate a bivariate VAR using GDP growth and the corresponding GDP deflator. We adopt the exact same timing as for SPF forecasts. For example, for the target 2008:Q1 we use data available until the end of 2007:Q1. We choose the number of lags equal to two and we choose a direct forecasting scheme for the one-year ahead forecast

instead of an iterative scheme.[44]

We use the prior specification of Primiceri (2005), D'Agostino et al. (2013) and Clark and Ravazzolo (2015). Hyperparameters for the prior of B_0, A_0 and $\log(\sigma_0^2)$ are based on the respective OLS estimate using the first T_0 observations. The prior for the initial states is a Normal, such that, for example, $B_0 \sim N\big(\hat{B}_{ols}, 4V(\hat{B}_{ols})\big)$. For the euro area, T_0 is equal to 25 and due to the short data set, we do not drop these estimations in the posterior estimation. For the U.S. we drop these observations for the estimation of the posterior. The priors for the variances of v_t, ξ_t and η_t are inverse Wishart with degrees of freedom 40, 4 and [2,3] respectively and the same scale as in Primiceri (2005) and Del Negro and Primiceri (2015). We set the total number of draws from the Gibbs sampler for $(D.7)$ to 15000 and drop the first 3000 draws. In addition, we thin out the results by a factor of 10.

E AR with Stochastic Volatility

The model is a standard autoregressive model with stochastic volatility. Let y_t be the variable of interest. We choose an $AR(1)$ and the model takes the form

$$y_t = c + \beta y_{t-h} + \sigma_t e_t$$

where c is the intercept, β is a constant autoregressive parameter and $e_t \sim N(0,1)$. The volatility is modelled following Chan and Grant (2016):

$$\log(\sigma_t) = \mu_\sigma + \gamma_\sigma\big(\log(\sigma_{t-1}) - \mu_\sigma\big) + \eta_t \tag{E.8}$$

where μ_σ is the mean of the volatility, γ_σ is the autoregressive parameter, with $|\gamma_\sigma| < 1$, and $\eta_t \sim N(0, \omega_\sigma^2)$. We assume the same prior specifications as Chan and Grant (2016) and base the estimation on their code, where we set the number of draws of the MCMC to 15000.

F Economic Downturn Probability

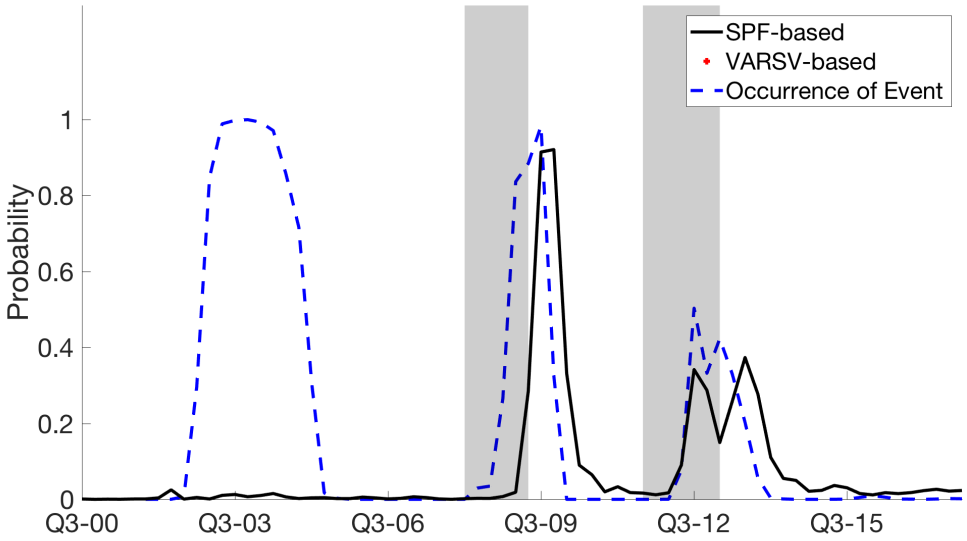
Let Y_t be the GDP growth rate and U_t be the unemployment rate of the eurozone, then we define the *alternative* ex-ante economic downturn probability of the euro area as

$$\xi_{t+h|t} \equiv F_{t+h|t}(Y_{t+h} < 0, U_{t+h} > \tau^{bc}_{t+h|t,u}) \tag{F.9}$$

[44] Teräsvirta (2006) notes that a direct forecasting scheme can be superior in the case of misspecified non-linearities.

The baseline specification uses again $W = 2$, i.e. the average of the two quarters preceding the forecast origin. The threshold then takes the form of $\tau^{bc}_{t+h|t,u} \equiv \frac{1}{2}\sum_{w=1}^{2} U_{i,t-w+1}$. Importantly, a recession is now defined through negative output growth and an increase in unemployment. Using this definition, the BVAR performs somewhat better in terms of predicting actual crises, but it calls wrong events at the beginning of the 2000s.

Figure F.1: Probability of Negative Growth and Higher Unemployment



NOTE. – The dates on the time axis are target based. For example, the probability plotted against 2010:Q1 is based on information available at the beginning of 2009:Q2. The crossed markers denote the realised event of negative GDP growth and higher unemployment. The solid line denotes the one year-ahead probability of a *economic downturn* on the joint density forecast using the SPF data and the dashed line denotes the economic downturn probability based on the BVAR. Shaded areas mark CEPR recession periods.

Chapter 2

Comparing Forecast Performances under Markov Switching

(joint with Barbara Rossi and Tatevik Sekhposyan)

2.1 Introduction

We propose three tests to evaluate the absolute performance of a mean forecast, where the tests are robust under the presence of parametric time-variation in the forecast errors. The time-variation is assumed to take the parametric form of Markov switching, i.e. weakly dependent and potentially discrete changes in the process of the forecast error. Existing tests focus on the constant relative out-of-sample performance or use non-parametric techniques to detect time-varying deviations from an equal performance; both approaches can lack power against the alternative of regime switching.

Producing accurate forecasts for economic variables is an essential task in academic work, central banks, and the private sector. Policymakers and private agents strive to base their decisions on optimal (rational) predictions, i.e., forecasts that are unbiased and efficient.[1] In light of the importance of optimal predictions, we seek to make two contributions.

We are the first to apply the idea of Markov switching directly on the forecast errors and to test for rationality under this specific, but popular, type of parametric time-variation. Second, we modify three existing test such that they satisfy the need of a forecaster to test jointly for both, no instabilities and no constant deviation

[1] See Granger and Newbold (1986) and Diebold and Lopez (1996) for seminal contributions to the definition of forecast optimality in the mean prediction case. See Mincer and Zarnowitz (1969) for a definition of forecast rationality.

from rationality. We label our tests ORS-H, ORS-G and ORS-CHP. Throughout the paper, we focus on forecasts that are model-free or survey-based, i.e. we abstract from parameter estimation error in the forecast errors. Testing for Markov switching requires non-standard statistics as it is subject to two problems. The hyperparameters of the switching process, i.e., the state-to-state transition probabilities are not identified under the null and therefore the standard LR, Wald and LM tests do not have chi-square distribution. The problem of a unidentified nuisance parameters under the null was described by Davies (1977, 1987). It has since generated a substantial literature, for example regarding tests of parameter instability. In the case of Markov switching, a second problem occurs. Under the null, the score of the restricted parameters is identically zero, which violates the regularity conditions that are imposed to derive the asymptotic chi-square distribution of the finite dimensional LR statistic by the usual second-order Taylor expansion.

Our tests are based on a modification of the test against Markov switching parameters proposed by Hansen (1992), García (1998), and Carrasco et al. (2014) (CHP). The main difference to their work is that Hansen, Garcia, and CHP leave the switching parameters unspecified under the null whereas we impose a specific value such that the forecast errors are rational (or unbiased/efficient respectively). Consequently, we can test jointly for rationality in the full out-of-sample as well as against local, regime switching deviations. To summarize our modifications: for Hansen, we partition the parameter vector differently. In the case of Garcia we apply a different restriction matrix, and for CHP we specify the parameters differently under the null.

For several reasons, we are interested in testing for random variation in the forecast error instead of the predictand. First, it allows the forecaster to impose the null hypothesis directly on the object of interest, namely the forecast error. Moreover, we focus on the case when the forecasts are model-free or survey-based, for instance when testing for rationality of the SPF, Greenbook forecasts or the Blue Chip Financial Forecasts.

Importantly, the idea of Markov switching in models' absolute forecasting abilities allows for economic interpretation. For example, considering the case of testing for unbiasedness, our tests can discover recurring periods that exhibit a bias in the predictions, even if the full sample of forecast errors is unbiased; the periods of biased predictions can be times of recessions, financial distress or other, previously unnoticed, circumstances. Our testing procedure treats these periods as an unknown state variable and allows the model under the alternative to recover them. Supporting the idea of changing relations of economic variables, predictors, and

predictands, the St. Louis Fed recently announced (Bullard, 2016) that their new characterization of the economy abandons the view of a single steady-state in favor of a regime-switching world with several steady-states.

In addition to the Markov switching literature, we build on the literature for evaluating absolute predictive abilities (Diebold and Lopez, 1996; West and McCracken, 1998). The closest to our work is Rossi and Sekhposyan (2016). The authors propose a non-parametric test for forecast rationality that is robust under the presence of instabilities. We differ from their work because our tests can detect non-smooth discrete deviations from rationality, whereas their test has better power against smooth and somewhat persistent changes. In other words, relative to Rossi and Sekhposyan (2016), our test can detect deviations from rationality that occur for shorter time periods, as long as the deviations occur repeatedly. Also, the identification of regimes can allow for a meaningful interpretation of the source of the rejection of the null hypothesis, which the Fluctuation test cannot provide.

We demonstrate the usefulness of our approach by investigating the bias in the predictions of the Blue Chip Financial Forecast survey for the federal funds target rate. When we consider the 3-months-ahead forecast error, our test rejects the null hypothesis of unbiasedness and finds evidence in favor of a two-regime model. The estimated regimes indicate that in the majority of times (in the persistent regime) the forecasts are unbiased. However, in regime two the forecasters overestimate the federal funds target rate. The occurrence of regime two is associated with times of monetary easing and not limited to recessionary periods.[2]

The paper is organized as follows. Section 2.2 formalizes our null hypothesis, describes the challenges that arise when testing for Markov switching and gives an overview of the Markov switching test literature. Section 2.3 introduces our test statistics. Section 2.4 provides Monte Carlo results for size and power of our proposed procedure and section 2.5 illustrates the empirical usefulness of our procedure. Section 2.6 discusses an idea for future research. We conclude in section 2.7.

2.2 Literature and the Null and Alternative Hypothesis

2.2.1 Null and Alternative Hypothesis

Assume the researcher has a series $y_{t+h|t}$, of h-step-ahead out-of-sample predictions for the variable y_{t+h}. Let $\epsilon_{t+h|t} = y_{t+h} - y_{t+h|t}$ denote the forecast error. We are

[2]The results are robust when using the 6-months-ahead predictions.

interested in testing for unbiasedness, efficiency and rationality in $\epsilon_{t+h|t}$, robust to the presence of parametric time-variation. Without loss of generality, we consider the leading case of two regimes and we model the forecast errors as

$$\epsilon_{t+h|t} = (1 - S_t)z_t\beta_1 + S_t z_t \beta_2 + \sum_{i=1}^{d} \phi_i \epsilon_{t+h-i|t-i} + e_{t+h} \tag{2.1}$$

where S_t is a stationary Markov chain with $\mathrm{S}_t \in \{0, 1\}$, e_{t+h} is $N(0, \sigma^2)$ and the lag coefficients ϕ_i. Another way of modelling the serial correlation in the forecast error would be to include a MA(h-1) component.[3] However, we stick to the convention in the Markov switching literature to control for autocorrelation by autoregressive components. The vector $z_t = [1, y_{t+h|t}]$ contains the main regressors and the $\beta_r = (\mu_r, \gamma_r)'$ are the parameters of interest, where r denotes regime $r = 1, 2$. The parameters μ_1 and μ_2 denote the intercepts of regime one and two and they are the parameters of interest in our test for unbiasedness. The parameters γ_1 and γ_2 denote the regression coefficients on the forecast itself and denote the parameters of interest in a test for efficiency.[4] For notational simplicity, we drop the the autocorrelation parameters in what follows, i.e. $\phi_i = 0 \; \forall i$.[5] Then (2.1) simplifies to

$$\epsilon_{t+h|t} = (1 - S_t)z_t\beta_1 + S_t z_t \beta_2 + e_{t+h} \tag{2.2}$$

(i) Unbiasedness: In a test for unbiasedness, $z_t = 1$ and (2.3) simplifies to

$$\epsilon_{t+h|t} = (1 - S_t)\mu_1 + S_t \mu_2 + e_{t+h} \tag{2.3}$$

Subsequently, our null and alternative hypothesis are

$$\mathrm{H}_0 : \mu_1 = \mu_2 = 0 \;\; \text{vs.} \;\; \mathrm{H}_A : \mu_1 \neq \mu_2 \text{ or } \mu_1 = \mu_2 \neq 0 \tag{2.4}$$

Under the null, the expected value of ϵ_{t+h} is equal to zero for all t. Existing tests for Markov switching impose a null hypothesis of $\mu_1 = \mu_2$, but not $\mu_1 = \mu_2 = 0$ and the specification of the otherwise unrestricted parameter values under the null will lead to changes in the test statistics. The additionally restriction of $\mu_1 = \mu_2 = 0$ is important in order to have power in cases of the traditional tests for unbiasedness, i.e. where the forecast errors have a constant non-zero mean.

[3] An h period-ahead forecast should at most exhibit a MA(h-1) autocorrelation structure.

[4] In general, the vector z_t can be extended to contain more regressors.

[5] Note that, when we control for autocorrelation, the null hypothesis remains on the parameters of interest in β_i.

(ii) Efficiency: In a test for efficiency, $z_t = [y_{t+h|t}]$ and subsequently

$$\epsilon_{t+h|t} = (1 - S_t)y_{t+h|t}\gamma_1 + S_t y_{t+h|t}\gamma_2 + e_{t+h} \tag{2.5}$$

Our null and alternative hypothesis then take the form

$$\mathrm{H}_0 : \gamma_1 = \gamma_2 = 0 \;\text{ vs. }\; \mathrm{H}_A : \gamma_1 \neq \gamma_2 \text{ or } \gamma_1 = \gamma_2 \neq 0 \tag{2.6}$$

Again, note that in a traditional test for forecast efficiency, the value γ is restricted to be constant.

(iii) Rationality: The test for rationality is a joint test of unbiasedness and efficiency, i.e. $z_t = [1, y_{t+h|t}]$ and

$$\epsilon_{t+h|t} = (1 - S_t)z_t\beta_1 + S_t z_t\beta_2 + e_{t+h} \tag{2.7}$$

where $\beta_r = [\mu_r, \gamma_r]'$ and our null hypothesis takes the form

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \mathbf{0} \;\text{ vs. }\; \mathrm{H}_A : \beta_1 \neq \beta_2 \text{ or } \beta_1 = \beta_2 \neq \mathbf{0} \tag{2.8}$$

2.2.2 Literature on Inference in Markov Switching Models

Testing for the presence of Markov switching in parameters is prone to two problems. First, a nuisance parameters which are only present under the alternative. Consequently, standard likelihood ratio tests do not have an asymptotic chi-squared distribution, a problem first documented by Davies (1977, 1987). The respective nuisance parameters are the hyperparameters that govern the distribution of the random coefficients, i.e. the state-to-state transition probabilities. In the closely related literature of parameter instabilities, the problem is solved by treating unidentified parameters as nuisance parameters and using a supremum type of test. The relevant central theorems rely on the weak convergence of the finite dimensional statistics.[6] However, in Markov switching models, the score with respect to the parameters of the Markov process is identically zero under the null, i.e. the null yields a local optimum and the asymptotic distribution of the finite dimensional statistic cannot be derived by the usual second-order approximation.

The first formal test for Markov switching was developed in the seminal work of

[6] See for example Pollard (1990).

Hansen (1992). The author bounds the likelihood ratio by an empirical process and derives the asymptotic distribution of the bound instead of the likelihood-ratio itself. Main drawbacks of the test by Hansen (1992) are that the bound may decrease power and that the asymptotic distribution is data dependent, i.e. to simulate the asymptotic distribution the alternative model needs to be estimated over a grid of all parameters of interest. On the positive side, the test is robust to weak serial correlation and heterogeneity in the error terms. For very large models, the computation becomes computationally infeasible. García (1998) partitions the parameter vector differently than Hansen (1992), ignores the problem of the identically zero-score, and obtains an asymptotic distribution, using results from Hansen (1996), which is not data dependent. Although the derivation of the asymptotic distribution is invalid from a theoretical point of view, the test exhibits good size properties and a computationally feasible alternative to Hansen (1992). Recently, Qu and Zhuo (2017) derived the exact asymptotic distribution of the likelihood-ratio using higher-order approximations.

Carrasco et al. (2014) propose a test statistic which requires the estimation of the model only under the null and is admissible in the sense of satisfying an optimality condition against local alternatives. Their test relies on the information matrix equality (White, 1982; Chesher, 1984), and further makes use of the auto-covariance of the scores generated by the persistence of the regimes. It is a supremum type of test that factors out the nuisance parameters governing the process of the random coefficient under the alternative.

Whereas Hansen (1996), García (1998), Carrasco et al. (2014) and Qu and Zhuo (2017) focus on the null region of $\mu_s = 0$, Cho and White (2007) extend the null region to contain the boundary value $\pi = 1$ (or $\pi = 0$ respectively), where π is the unconditional probability of regime one. The authors derive a QLR statistic that tests for mixture components, taking into account the extended null region.[7] As we built our test statistic on Hansen (1996), García (1998) and Carrasco et al. (2014), we focus on the null space of μ_s and do not consider the boundary parameter problem of $\pi = 1$ ($\pi = 0$).

[7]Including $\pi = 1$ ($\pi = 0$) can guard against false rejections due to structural breaks or infrequent but extreme values. See Carter and Steigerwald (2013) for an excellent summary of Cho and White (2007) and a detailed description of how to simulate their asymptotic distribution.

2.3 Test Statistics

2.3.1 ORS-H Test Statistic

Hansen (1992) developed the first formal test for Markov switching. Before we describe how we implement our null hypothesis, we summarize the results of Hansen for the simplest case of a switch-in-mean model. To simplify the comparison of our work to Hansen (1992), we adopt the following notation from his work: μ_1 and μ_2 can be equivalently written as $\mu_1 = \mu$ and $\mu_2 = \mu + \mu_s$. The null hypothesis of Hansen (1992) is $\mu_s = 0$ (using Hansen's (1992) notation, our null takes the form of $\mu = \mu_s = 0$).[8] The author considered the likelihood ratio as an empirical process indexed by the parameters of interest, $\alpha = (\mu_s, p, q)$, and additionally depending on the nuisance parameters $\theta = (\mu, \phi_1, ..., \phi_d, \sigma)$. Let $\alpha_0 = (\mu_{s0}, p, q)$ denote the parameter vector under the null hypothesis of Hansen. Let

$$\hat{\theta} = \max_{\theta \in \Theta} \ \mathrm{L}_T\big(\alpha_0, \theta\big)$$

denote the MLE of the nuisance parameters under the null and let

$$\hat{\theta}(\alpha) = \max_{\theta \in \Theta} \ \mathrm{L}_T\big(\alpha, \theta(\alpha)\big)$$

denote the MLE of the nuisance parameters under the alternative α. The likelihood ratio is defined as

$$\widehat{\mathrm{LR}}_T(\alpha) = \mathrm{L}_T\big(\alpha, \hat{\theta}(\alpha)\big) - \mathrm{L}_T(\alpha_0, \hat{\theta}) \tag{2.9}$$

with

$$\mathrm{L}_T\big(\alpha, \hat{\theta}(\alpha)\big) = \sum_{t=1}^{T} \ell_t\big(\alpha, \hat{\theta}(\alpha)\big) \quad \text{and} \quad \mathrm{L}_T\big(\alpha_0, \hat{\theta}\big) = \sum_{t=1}^{T} \ell_t\big(\alpha_0, \hat{\theta}\big)$$

where ℓ_t denotes the log likelihood of observation t. The likelihood ratio can be split into its expected value, $\mathrm{R}_T(\alpha)$, and its deviation from that expectation, $\mathrm{Q}_T(\alpha)$,

$$\widehat{\mathrm{LR}}_T(\alpha) = \mathrm{R}_T(\alpha) + \mathrm{Q}_T(\alpha) + \mathrm{O}_p(1) \tag{2.10}$$

and

$$\mathrm{Q}_T(\alpha) = \sum_{t=1}^{T} q_t(\alpha) = \sum_{t=1}^{T} \Big[\ell_t\big(\alpha, \theta(\alpha)\big) - \ell_t(\alpha_0, \theta) - \mathrm{E}\big[\ell_t\big(\alpha, \theta(\alpha)\big) - \ell_t(\alpha_0, \theta)\big] \Big]$$

[8] Note that we can equivalently write Hansen's (1992) null hypothesis as $\mu_1 = \mu_2$ without the additional zero restriction.

with

$$\mathrm{R}_T(\alpha) = \mathrm{E}\big[\ell_t\big(\alpha, \theta(\alpha)\big) - \ell_t(\alpha_0, \theta)\big]$$

where θ and $\theta(\alpha)$ denote the large sample values of the MLE of $\hat{\theta}$ and $\hat{\theta}(\alpha)$. The parameter estimation error that is present in sample equivalent of $\mathrm{Q}_T(\alpha)$ is captured in the term $\mathrm{O}_p(1)$. Please see appendix C for details. Under the null, $\mathrm{R}_T(\alpha) \leq 0$ and the value of $\mathrm{R}_T(\alpha)$ is maximized at the true parameter α_0 (under the null). It follows that

$$\frac{1}{\sqrt{T}}\widehat{\mathrm{LR}}_T(\alpha) \leq \frac{1}{\sqrt{T}}\mathrm{Q}_T(\alpha) + \mathrm{o}_p(1)$$

Let $\mathrm{V}_T(\alpha)$ denote the variance of the $\mathrm{q}_t(\alpha)$. For a fixed α, the demeaned and standardized $\mathrm{Q}_T(\alpha)$ converges to a standard Normal by a central limit theorem

$$\frac{1}{\sqrt{T}}\frac{\mathrm{Q}_T(\alpha)}{\mathrm{V}_T^{1/2}(\alpha)} = \frac{1}{\sqrt{T}}\mathrm{Q}_T^*(\alpha) \to_d \ N(0,1)$$

Then, by applying a empirical process central limit theory, Hansen derives the asymptotic distribution of the bound $\sup_{\alpha \in A} \mathrm{Q}_T^*(\alpha)$

$$\sup_{\alpha \in A} \ \frac{1}{\sqrt{T}}\widehat{\mathrm{LR}}_T^*\big(\alpha, \theta(\alpha)\big) \leq \sup_{\alpha \in A} \ \frac{1}{\sqrt{T}}\mathrm{Q}_T^*(\alpha) + o_p(1) \to_d \sup_{\alpha \in A} \ \mathrm{Q}^*(\alpha)$$

which is a supremum of standard Normals with covariance matrix $K^*(\alpha_i, \alpha_j)$.

To use the strategy of Hansen for testing our joint null hypothesis, we need to partition the parameters differently. Our null hypothesis includes both μ_1 and μ_2, or more generally speaking β_1 and β_2, i.e. we can not treat β_1 as nuisance parameter but have to add it to the vector of parameters of interest. The three relevant cases for us imply parameter vectors $\tilde{\alpha}_u = (\mu_1, \mu_2, p, q)$, $\tilde{\alpha}_e = (\gamma_1, \gamma_2, p, q)$ and $\tilde{\alpha}_r = (\beta_1, \beta_2, p, q)$, where the subscripts u, e, r denote unbiasedness, efficiency and rationality. For notational ease, we drop the subscript for the remainder of the section. The vector of nuisance parameters reduces to the lag coefficients and the standard deviation, $\theta = (\phi_1, ..., \phi_d, \sigma)$. By partitioning the parameters in this way, we can rely on the same empirical process theory as Hansen did and consider the likelihood ratio as an empirical process indexed by $\tilde{\alpha} = (\beta_1, \beta_2, p, q)$. Let $\tilde{\alpha}_0$ denote the vector of parameters of interest under the null. Analog to Hansen, let

$$\hat{\theta} = \max_{\theta \in \Theta} \ \mathrm{L}_T\big(\tilde{\alpha}_0, \theta\big)$$

denote the MLE of the nuisance parameters under the null and let

$$\hat{\theta}(\tilde{\alpha}) = \max_{\theta \in \Theta} \mathrm{L}_T\big(\tilde{\alpha}, \theta(\tilde{\alpha})\big)$$

denote the MLE of the nuisance parameters under the alternative $\tilde{\alpha}$. Then, we can write

$$\sup_{\tilde{\alpha} \in A} \widehat{\mathrm{LR}}_T\big(\tilde{\alpha}\big) = \mathrm{L}_T\big(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})\big) - \mathrm{L}_T\big(\tilde{\alpha}_0, \hat{\theta}\big)$$

The inequality

$$\frac{1}{\sqrt{T}}\widehat{\mathrm{LR}}_T(\tilde{\alpha}) \leq \frac{1}{\sqrt{T}}\mathrm{Q}_T(\tilde{\alpha}) + \mathrm{o}_p(1) \tag{2.11}$$

follows from the assumptions of Hansen, i.e. that the estimator $\hat{\theta}(\tilde{\alpha})$ converges to its pseudo-true[9] value $\theta(\tilde{\alpha})$, for a given $\tilde{\alpha}$. As the result in *Statement 1* follows directly from Hansen (1992), we summarize his main steps and assumptions in appendix C. The conditions allow for dependence in $\ell_t\big(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})\big)$ or $\ell_t\big(\tilde{\alpha}_0, \hat{\theta}_0(\tilde{\alpha}_0)\big)$ such as mixing or a near epoch dependence and heterogeneity in the error terms.

Statement 1: *Under assumption A1*

$$Pr\big(\widehat{LR}_T^* \geq x\big) \leq Pr(Q_T^* \geq x) + o_p(1) \rightarrow_d Pr(Q^* \geq x) \tag{2.12}$$

where each $Q^*(\tilde{\alpha})$ *converges to a* $N(0,1)$ *variate and the process* $Q^* = \sup_{\tilde{\alpha} \in A} Q^*(\tilde{\alpha})$ *is characterized by its covariance function*

$$K^*(\tilde{\alpha}_1, \tilde{\alpha}_2) = \frac{\sum_{k=-\infty}^{\infty} E\, q_t(\tilde{\alpha}_1)\, q_{t+k}(\tilde{\alpha}_2)}{V(\tilde{\alpha}_1)^{\frac{1}{2}}\, V(\tilde{\alpha}_2)^{\frac{1}{2}}} \tag{2.13}$$

The covariance function now depends on $\mathrm{q}_t(\tilde{\alpha})$ in contrast to $\mathrm{q}_t(\alpha)$ in the original Hansen test. Simulation of the critical values is analog to Hansen and can be achieved by simulating

$$\mathrm{Q}_T^*(\tilde{\alpha}_i) = \frac{\sum_{i=0}^{M} \sum_{t=1}^{T} \hat{q}_t(\tilde{\alpha}_i) v_{t+i}}{\sqrt{1+M} V_T(\tilde{\alpha}_i)^{\frac{1}{2}}} \tag{2.14}$$

where the v_{t+i} are i.i.d. $N(0,1)$ variates and M denotes the bandwidth length for the Bartlett kernel. To obtain a set of $\mathrm{Q}_T^*(\tilde{\alpha}_i)$, the researcher must estimate the model under the alternative over a grid of values for $\tilde{\alpha} = (\beta_1, \beta_2, p, q)$. Table 2.1 displays the average critical values we obtain when we apply the ORS-H and the original Hansen test to the DGP of a standard Normal for the unbiasedness and efficiency

[9] Pseudo-true value means the large-sample value, given a fixed value for $\tilde{\alpha}$.

test with $\tilde{\alpha}_u = (\mu_1, \mu_2, p, q)$ and $\tilde{\alpha}_e = (\gamma_1, \gamma_2, p, q)$. As the approximation of the asymptotic distribution is data dependent, the numbers are obtained by *averaging* the critical value over all Monte Carlo repetitions. The aim of this exercise is not to tabulate critical values, for that would be invalid, but to provide an idea of how the additional parameter restriction changes the empirical critical values we obtain. As expected, the critical values of the ORS-H test are larger on average, reflecting the additional restriction of the ORS-H on the null parameter space.

Table 2.1: Critical Values

	Unbiasedness		Efficiency	
Nominal Size	ORS-H	Hansen	ORS-H	Hansen
1%	3.60	2.80	3.44	3.24
5%	3.01	2.51	2.84	2.64
10%	2.72	2.18	2.53	2.33

NOTE. – The table shows the average critical values based on our simulations for the ORS-H test and the Garcia test for standard Normal DGP and a sample size of $T = 500$ and 500 Monte Carlo repetitions.

2.3.2 ORS-G Test Statistic

García (1998) develops a likelihood ratio test for Markov switching models, which closely follows the seminal work of Hansen (1992).[10] The main difference in their work is that Garcia partitions the parameter vector differently, which allows him to use the theory developed in Hansen (1996). Hansen defines the parameters of interests as $\alpha = (\mu_s, p, q)$, treats $\theta = (\mu, \sigma)$ as nuisance parameters and considers the likelihood as an empirical process indexed by $\alpha = (\mu_s, p, q)$. Garcia, in turn, treats $\alpha = (p, q)$ as nuisance parameters[11] and defines the parameters of interest to be $\theta = (\mu, \mu_s, \sigma)$. By keeping the parameter of interest μ_s in the parameter vector θ and through ignoring the problem of the zero score of the parameter μ_s under the null restriction, Garcia is in the general setting where unidentified nuisance parameters are only present under the null (Hansen, 1996). Importantly, as already noted by Garcia, the score problem violates the regularity conditions imposed in Hansen (1996). However, the Monte Carlo results obtained by Garcia show that the test has good size and power properties and it provides a computationally simple alternative to the statistic of Hansen.

We can equivalently define the parameter vector of interest as $\theta = (\beta_1, \beta_2, \sigma)$, i.e. we do not need to partition the parameters differently and can straightforwardly impose our null hypothesis as $\mathrm{H}_0 : \beta_1 = \beta_2 = 0$ (rationality), $\mathrm{H}_0 : \mu_1 = \mu_2 = 0$

[10] The null hypothesis of Garcia is the same as in Hansen (1992), i.e. $\mathrm{H}_0 : \mu_s = 0$.

[11] Assumed to be in a compact set A.

(unbiasedness) and $H_0 : \gamma_1 = \gamma_2 = 0$ (efficiency). The difference in the test statistics will be a different restriction matrix τ, which we explain below. Although the asymptotic distribution follows directly from results of Garcia, for notational clarity, denote testing the null hypothesis of Garcia by G and denote testing our null hypothesis by ORS-G. Note that the asymptotic distribution will be nuisance parameter free only in special cases, and will depend in general on the autocorrelation parameters as well as one the first two moments of regressors with Markov switching coefficients. For the ease of the reader we summarize Garcia's main arguments, and then show how our restrictions matrix differs.

Let $\mathrm{LR}_T(\alpha) = 2T\big[Q_T\big(\alpha, \hat{\theta}(\alpha)\big) - Q_T(\tilde{\theta})\big]$ denote the likelihood ratio statistic, given a value for α. The parameter vector $\hat{\theta}(\alpha)$ denotes the MLE of the unconstrained model and $\tilde{\theta}$ denotes the MLE of the constrained model. Let

$$\mathrm{LR}_T = \sup_{\alpha \in A} \mathrm{LR}_T(\alpha) \tag{2.15}$$

be the supremum of the likelihood ratio statistics, taken over the nuisance parameter α. Garcia's theorem 1 states that LR_T converges weakly to the supremum of chi-square processes

$$\mathrm{LR}_T = \sup_{\alpha \in A} C(\alpha) \tag{2.16}$$

where $C(\alpha)$ is a chi-square process with covariance matrix $\bar{\mathrm{K}}(\alpha_i, \alpha_j)$, defined as

$$\bar{\mathrm{K}}(\alpha_i, \alpha_j) = \tau V^{-1}(\alpha_i) \mathrm{K}(\alpha_i, \alpha_j) V^{-1}(\alpha_j) \tau' \tag{2.17}$$

The matrices $V(\alpha_i)^{-1}$ and $\mathrm{K}(\alpha_i, \alpha_j)$ are described in Appendix D. The main difference to Garcia is in the matrix of restrictions τ. In the case of testing for Markov switching in the mean, τ simplifies to be vector with a one in the position of the constrained parameter. As we are testing a joint null, τ will be a matrix of size r x k, where r is the number of restrictions and k is the size of the vector θ. For a two-state model, the restrictions matrix τ, for testing for unbiasedness or efficiency[12], takes the form

$$\tau = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \end{pmatrix}$$

[12] Assuming a model without intercept.

For testing for rationality, τ takes the form

$$\tau = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 1 & 0 & ... & 0 \end{pmatrix}$$

Table 2.2 compares the critical values for the unbiasedness test, without autoregressive coefficients, of the original Garcia null to our joint null. Note that for a model without autoregressive parameters, the asymptotic distribution, conveniently, depends on the nuisance parameter $\alpha = (p, q)$ only through $\pi = (1-q)/(2-p-q)$. We follow Garcia in the choice of the grid such that $\pi \in [0.01, 0.99]$ and $\pi \in [0.15, 0.85]$. As expected, the critical values of our null are higher because we are testing an additional constraint. In general, the derivation assumes iid error terms. Conditional heteroskedasticity and serial correlation can be accommodated but will lead to additional nuisance parameters in the asymptotic distribution.

Table 2.2: Comparison of Critical Values

	$\pi \in [0.01, 0.99]$		$\pi \in [0.15, 0.85]$	
Nominal Size	ORS-G	G	ORS-G	G
1%	14.33	13.64	12.66	12.45
5%	10.67	10.18	9.25	8.60
10%	8.94	8.68	7.64	7.08

NOTE. – The table shows the simulated critical values of the ORS-Garcia test and the Garcia test. The grid is $\pi \in [0.01, 0.99]$ and $\pi \in [0.15, 0.85]$ respectively.

2.3.3 ORS-CHP Test Statistic - Test for Unbiasedness

Carrasco et al. (2014) develop an optimal test for Markov switching parameters. We extend their test statistic to satisfy our null hypothesis of unbiasedness, because, different from CHP, we use the test in an absolute forecast evaluation framework. To ease the comparison between their work and ours we adopt, whenever possible, their notation. Before we describe the test statistic, we discuss the relation of their null and alternative to ours. Their null and alternative hypothesis are $H_0 : \theta_t = \theta_0$, for some *unspecified* θ_0, against $H_1 : \theta_t = \theta_0 + \eta_t$. We are interested in the null of $H_0 : \theta_t = \theta_0$ for some parameters of θ_0 being *specified*, i.e. $\mu = 0$. As our null imposes a *specified* value for θ_0, the arguments used by Carrasco et al. (2014) to derive the asymptotic distribution apply only in the special case of a normal likelihood and when testing the intercept for unbiasedness. We leave the general case of testing for

rationality for future research.

We are interested in the alternative of $\eta_t = (cS_t, 0, ..., 0)$, where S_t is a stationary Markov chain and c is the magnitude of the change. This alternative corresponds to Example 3.3 in CHP and allows to describe the process of η_t by the dynamics of S_t alone, which are summarized in α.[13]

Their test statistic builds on two parts: one that exploits the autocorrelation introduced by the Markov chain and one that tests for misspecification in the likelihood and goes back to White (1982). We will directly describe the test statistic under our null θ_0 and derive the asymptotic distribution under the assumption of a Normal likelihood. Let $\ell_t^{(i)}(\hat{\theta}_0)$ denote the i-th derivative of the log likelihood function with respect to θ, evaluated at $\hat{\theta}_0$, where $\hat{\theta}_0$ is the MLE of θ under our null. The statistic in the general form is based on

$$\mu_{2,t}(\alpha, \hat{\theta}_0) = \frac{1}{2} h' \left[\left(\ell_t^{(2)}(\hat{\theta}_0) + \ell_t^{(1)}(\hat{\theta}_0) \ell_t'^{(1)}(\hat{\theta}_0) \right) + 2 \sum_{s=1}^{t-1} \alpha^{(t-s)} \ell_t^{(1)}(\hat{\theta}_0) \ell_s'^{(1)}(\hat{\theta}_0) \right] h \quad (2.18)$$

where the parameter α captures the dynamics of S_t, i.e. $\mathrm{E}(S_t S_{t-i}) = \alpha^i$. The first term is an information matrix type of test, which has power against misspecification of the likelihood, and the second term exploits the autocorrelation introduced by the Markov chain that governs the time-varying coefficients. The vector h selects the parameters to be tested and takes the form $h = (1, 0, ..., 0)'$.[14] Therefore, (2.18) simplifies to

$$\begin{aligned} \mu_{2,t}(\alpha, \hat{\theta}_0) &= \frac{1}{2} \left[\left(\ell_{t,(1,1)}^{(2)}(\hat{\theta}_0) + \{ \ell_t^{(1)}(\hat{\theta}_0) \ell_t^{(1)'}(\hat{\theta}_0) \}_{(1,1)} \right) + 2 \ell_{t,1}^{(1)}(\hat{\theta}_0) \sum_{s=1}^{t-1} \alpha^{(t-s)} \ell_{s,1}^{(1)}(\hat{\theta}_0) \right] \\ &= \frac{1}{2} \left[2 \ell_{t,1}^{(1)}(\hat{\theta}_0) \sum_{s=1}^{t-1} \alpha^{(t-s)} \ell_{s,1}^{(1)}(\hat{\theta}_0) \right] \end{aligned}$$

where the second equality is due to the choice of a Normal likelihood. The subscript (i, j) denotes the (i, j)-element of a matrix and $\ell_{t,1}^{(1)}$ denotes the first element of the score. The restricted MLE implies that the sum of the score vector of the restricted parameters is different from zero. The numerator of our test statistic differs from CHP because it is evaluated at a different null, $\hat{\theta}_0$, where in our case the first element of $\hat{\theta}_0$ is specified to be zero

$$v_T(\hat{\theta}_0, \alpha) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}_0, \alpha) \quad (2.19)$$

[13] Please see appendix E for details.

[14] The time-varying parameter η_t can alternatively be written as $\eta_t = chS_t$.

The denominator differs in the argument $\hat{\theta}_0$ and takes the form

$$u_t(\hat{\theta}_0, \alpha) = \mu_{2,t}(\hat{\theta}_0, \alpha) - \hat{d}_0' \ell_t^{(1)}(\hat{\theta}_0, \alpha) \tag{2.20}$$

where $\hat{d}_0'$ is the regression coefficient from regressing $\mu_{2,t}(\hat{\theta}_0, \alpha)$ on $\ell_t^{(1)}(\hat{\theta}_0, \alpha)$. Further, denote by $u = [u_1, ..., u_T]'$, where we dropped the arguments $(\hat{\theta}_0, \alpha)$ for notational simplicity.

Proposition 1: *Let the test statistic take the form*

$$\text{supTS} = \sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]\}} \frac{1}{2} \left(\max \left(0, \frac{v_T(\hat{\theta}_0, \alpha)}{\sqrt{T^{-1} u' u}} \right) \right)^2 \tag{2.21}$$

Given that Assumption B1 to B3 in appendix E hold and that the likelihood is taken to be a Normal, then it follows that supTS converges weakly under H_0, using results from Theorem 3.1, to the distribution of the supremum of a set of normal random variables

$$\text{supTS} \rightarrow_d \sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} \frac{1}{2} \left(\max \left(0, K(\alpha) \right) \right)^2 \tag{2.22}$$

where $K(\alpha)$ is a linear combination of standard Normal random variables, with a covariance that depends on α.

We impose regularity assumptions on the data analog to CHP. The test of CHP has optimal local power against a variety of random coefficient processes. The subset of feasible random coefficients processes that we are interested in, is discussed in Example 3.3 of CHP and can be described such that $\eta_t = \mu_s h S_t$, where S_t is a stationary Markov chain as described above. This definitions allows for geometric ergodic Markov process with a finite state space, such that $S_t \in \{0, 1, ..., N\}$, as well as for continuous but bounded state spaces such that $S_t \in (a, b)$, where $a, b \in R$ are finite constants. Consequently, we can test against a finite number of regime switches as well as against random variation that can take a smooth form of a continuum of values. As we are interested in testing the intercept we have $h = (1, 0, ..., 0)'$ and analog to CHP we can analytically maximize our test statistic with respect to μ_s. The magnitude of μ_s and the parameter h are thus irrelevant for the computation of the test statistic. We model the forecast errors as

$$\epsilon_{t|t-h} = \mu + \mu_s S_t + \sum_{i=1}^{d} \phi_d \epsilon_{t-i|t-h-i} + e_t \tag{2.23}$$

where $e_t \sim N(0, \sigma^2)$. Again, the parameter μ denotes the unconditional mean and

η_t captures the time-variation in the forecast errors due to Markov switching. For the model in (2.23), under the null, and analog to CHP, define

$$\mu_{2,t}(\hat{\theta}_0, \rho) = \frac{1}{2\hat{\sigma}^4}\left((\hat{e_t}^2 - \hat{\sigma}^2) + 2\sum_{s=1}^{t-1}\alpha^{t-s}\hat{e}_t\hat{e}_s\right) \tag{2.24}$$

where $\hat{e}_t$ are the errors of equation (2.23) estimated under the null of $\mu_s = 0$ and $\hat{\sigma}_0^2 = \frac{1}{T}\sum_t \hat{e}_t^2$. In order to obtain the test statistic one has to compute the statistic in (2.21) over a grid of values for α.

Proposition 1 and equation (2.24) are derived under the assumption of normality of the errors e_t. However, the test is robust to certain misspecification of the likelihood under the null. In particular, as stated before, for the statistic based on a $\mu_{2,t}$ of the form in (2.24) the moment conditions required for Proposition 1 to hold are mild with respect to some properties of the e_t, as they allow for heterogeneity or a random coefficient as long as it is stationary and serially uncorrelated. The important assumption of no serial correlation is motivated by the fact that the test achieves power against weakly dependent random coefficients by explicitly exploiting the autocorrelation generated by these random coefficients. Consequently, any serial correlation that does not arise from the random coefficient will lead to over-rejections under H_0. This feature is inherited from the original CHP test. We do not provide an analytic evaluation of its power against either of the alternatives. In particular we can not assess whether it maintains the property of optimality against local alternatives and we leave this question for future research. A drawback of our test statistic is that it does not provide a strictly formal way to assess the source of rejection. Consequently, we recommend to investigate the sign of the mean of the forecast errors and subsequently estimate a Markov switching model.

Comparison of ORS-CHP to CHP

Consider the model in (2.23) (without the autoregressive component)

$$\epsilon_{t|t-h} = \mu + \eta_t + e_t$$

The numerator takes the form of

$$\begin{aligned}\mu_{2,t}(\hat{\theta}_0, \alpha) &= \frac{1}{2\hat{\sigma}^4}\left((\hat{e_t}^2 - \hat{\sigma}^2) + 2\sum_{s=1}^{t-1}\alpha^{t-s}\hat{e}_t\hat{e}_s\right) \\ &= \frac{1}{2\hat{\sigma}^4}\left((\epsilon_{t|t-h}^2 - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{t|t-h}^2) + 2\sum_{s=1}^{t-1}\alpha^{t-s}\epsilon_{t|t-h}\epsilon_{s|s-h}\right)\end{aligned} \tag{2.25}$$

such that

$$\begin{aligned} v_T(\hat{\theta}_0, \alpha) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}_0, \alpha) \\ &= \frac{1}{\sqrt{T}} \frac{1}{\hat{\sigma}^4} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \rho^{t-s} e_t e_s \end{aligned} \tag{2.26}$$

In comparison, the numerator of Carrasco et al. (2014) takes the form

$$\begin{aligned} v_{T,chp}(\hat{\theta}, \alpha) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t,chp}(\hat{\theta}, \alpha) \\ &= \frac{1}{\sqrt{T}} \frac{1}{\tilde{\sigma}^4} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \alpha^{t-s} (e_t - \tilde{\mu})(e_s - \tilde{\mu}) \end{aligned} \tag{2.27}$$

where $\tilde{\mu} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_{t|t-h}$ and $\tilde{\sigma} = \frac{1}{T} \sum_{t=1}^{T} (\epsilon_{t|t-h} - \tilde{\mu})^2$. For the alternative of $\mu \neq 0$, the sum of errors in (2.26) will not have a mean of zero, i.e. the sum diverges.

2.4 Simulation Results

2.4.1 Monte Carlo Study - Unbiasedness

This section provides Monte Carlo evidence on the empirical size and power of the test for unbiasedness. The DGP that we consider in this section has the form

$$y_t = \phi y_{t-1} + u_t \tag{2.28}$$

The forecasting model is (we abstract from parameter estimation error)

$$y_{t+1|t} = \phi y_t \tag{2.29}$$

and the forecast errors are subsequently

$$\epsilon_{t+1|t} = y_{t+1} - y_{t+1|t}$$

We then model the forecast errors for the ORS-H, ORS-G and ORS-CHP as

$$\epsilon_{t+1|t} = \mu_1(1 - S_{t+1}) + S_{t+1}\mu_2 + e_{t+1} \tag{2.30}$$

where μ_1 and μ_2 are the Markov switching parameters, S_{t+1} is a stationary Markov chain and $e_{t+1} \sim N(0, \sigma^2)$. The null model imposes the restrictions $\mu_1 = \mu_2 = 0$

and takes the form

$$\epsilon_{t+1|t} = e_{t+1} \tag{2.31}$$

Table 2.3 shows the size results for ORS-H, ORS-G and ORS-CHP. In addition, we show size results for the original Hansen, Garcia and CHP test, to provide a comparison of how our null affects the size properties relative to the original tests. As we are going to use the tests of West and McCracken (1998) (WM) and Rossi and Sekhposyan (2016) (Fluctuation) as a benchmark in the power study, their sizes are included as well.[15] Overall, the size results of ORS-CHP and ORS-G are good, although ORS-G performs better for small and medium sized samples than ORS-H. The ORS-H overrejects for small and medium sized samples, however, the size distortions are of a similar magnitude as in the original test of Hansen.

Table 2.3: Size Results - Unbiasedness

Nominal Size	Test	T = 100	T = 200	T = 500
5%	WM	0.047	0.056	0.052
	Fluctuation	0.096	0.059	0.046
	Hansen	0.116	0.100	0.070
	Garcia	0.043	0.035	0.036
	CHP	0.036	0.034	0.040
	ORS-H	0.126	0.090	0.058
	ORS-G	0.054	0.059	0.065
	ORS-CHP	0.091	0.076	0.062
10%	WM	0.100	0.109	0.111
	Fluctuation	0.015	0.109	0.092
	Hansen	0.194	0.174	0.110
	Garcia	0.073	0.055	0.063
	CHP	0.072	0.068	0.075
	ORS-H	0.198	0.156	0.104
	ORS-G	0.107	0.108	0.108
	ORS-CHP	0.146	0.129	0.104

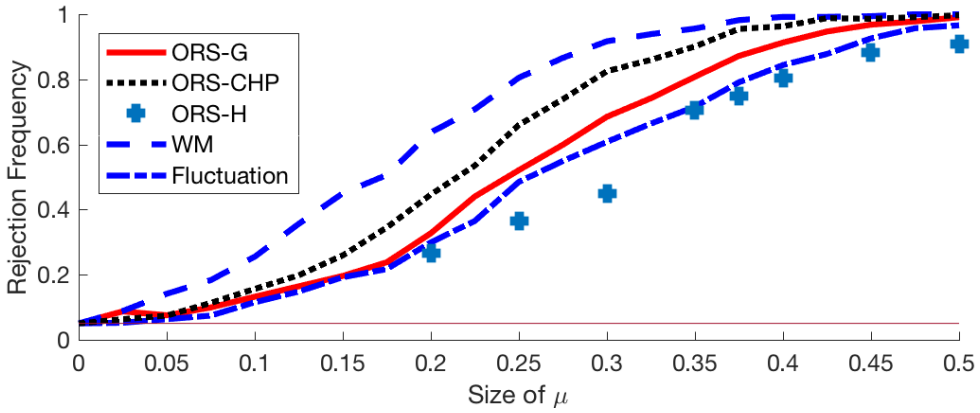
NOTE. – T denotes the sample size. Results are based on 1000 Monte Carlo replications, except for Hansen and ORS-H. Due to the computational burden, the Monte Carlo replications for Hansen and ORS-H are limited to 500. The DGP is a standard Normal. The results for Hansen are based on a 3-tuple of 12 grid points for $(p, q) \in [0.05, 0.95]$ and 20 grid points for $\mu_s \in [0.05, 1]$. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.05, 0.95]$ and 20 grid points for $\mu_1, \mu_2 \in [-1, 1]$. The results for CHP and ORS-CHP are based on 70 grid points for $\alpha \in [0.02, 0.98]$. The results for Garcia and ORS-G are based on $\pi \in [0.01, 0.99]$. The window size m for the Fluctuation test is set to $m = \frac{T}{2}$.

For the alternative of a constant mean we impose $\mu_1 = \mu_2 = \mu$, and estimate the power of the four tests over a grid of values for μ. The grid is set to $[0.05, 2]$ with

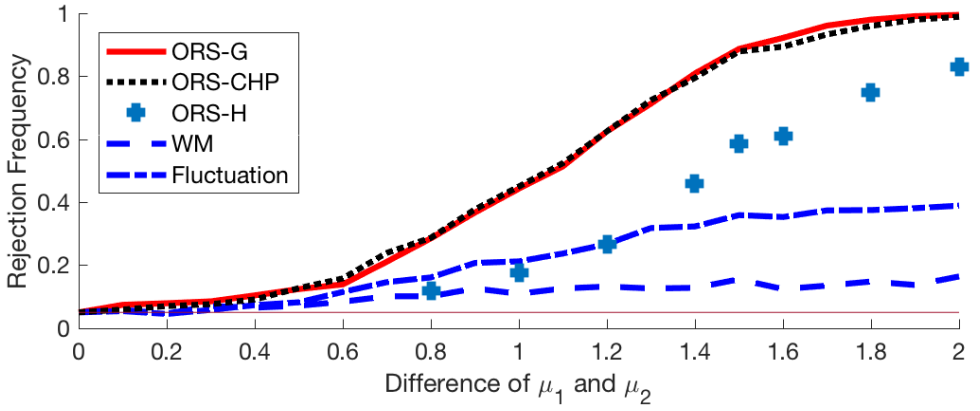
[15] Under our assumption of no parameter estimation error, the West and McCracken (1998) simplifies to a standard t-test.

increments of 0.1. Panel (a) of figure 2.1 shows the size-adjusted power results for a sample size of $T = 100$ at a nominal size of 5%. As expected, the WM test has the highest power against a constant mean. However, the results of the ORS-CHP test are close to the WM rejections frequencies. The ORS-G test exhibits good power against a constant mean and the power increases rapidly with the size of the alternative. For the ORS-H test, we compute the power only at eight grid points due to the computational intensity of the test. Its power is similar, but lower, than that of the ORS-G test.[16]

Figure 2.1: Size-Adjusted Power - Unbiasedness



(a) Constant Mean



(b) Markov Switching in the Mean

NOTE. – The y-axis denotes the rejection frequency. The x-axis denotes the size of $\mu_1 - \mu_2$. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.05, 0.95]$ and 20 grid points for $\mu_1, \mu_2 \in [-1.5, 1.5]$. ORS-CHP is based on 70 grid points for $\alpha \in [0.02, 0.98]$. The results for ORS-G are based on $\pi \in [0.01, 0.99]$

For the alternative of Markov switching, we set the state-to-state transition prob-

[16] An increase in the number of grid points for μ may increase the power of ORS-H.

abilities to be $(p, q) = 0.9$ and impose $\mu_1 = -\mu_2$. These parameter choices ensure that $\mathrm{E}\epsilon_{t+1|t} = 0$, i.e. that we can compute the power against Markov switching only. Again, the rejection frequencies are computed over a grid for μ_1 of $[0.05, 2]$ with increments of 0.1. Panel (b) of figure 2.1 displays the size-adjusted rejections frequencies at a nominal size of 5%.[17] The ORS-CHP and ORS-G exhibit strong power against the alternative of Markov switching. Theoretically, the rejection frequency of the GW test should remain at the nominal level of 5%. However, for small samples, it is quite likely to sample one of the states more often than the other, even if the unconditional state probabilities are 0.5 and this shifts the mean away from zero.[18] Interestingly, the Fluctuation test does not have much power against this type of time variation. This result is driven by the non-parametric approach of the test, i.e. it has less power against parametric discrete switches. Note however, that the power results of the Fluctuation test depend to some extent on the window size - smaller windows potentially improve the tests' power under Markov switching. Analog to the case of a constant mean, we compute the power of ORS-H only at two reference points. ORS-H exhibits a lower power than ORS-G and ORS-CHP, however, we note that the grid size of μ and μ_s can influence the results.

2.4.2 Monte Carlo Study - Efficiency

We now turn to testing efficiency. The DGP, the forecasting model and the forecast errors have the same DGPs as in (2.32), (2.29) and (2.30). We are testing the null hypothesis of $\gamma_1 = \gamma_2 = 0$ in

$$\epsilon_{t+1|t} = \big((1 - S_{t+1})\gamma_1 + S_{t+1}\gamma_2\big)y_{t+1|t} + e_{t+1} \tag{2.32}$$

where γ_1 and γ_2 are the Markov switching parameter, S_{t+1} is a stationary Markov chain and $e_{t+1} \sim N(0, 1)$. Table 2.4 shows the size results for ORS-H, ORS-G, the original Hansen, Garcia, WM and the Fluctuation test. Overall, the size results of ORS-G are good, although ORS-G slightly underrejects. The ORS-H is rather strongly undersized, but again, the size distortions are of a similar magnitude as in the original test of Hansen. The reason for the distortions could be that the critical values are taken from a bound and not from an exact distribution. Another reason for the underrejections could be that the test is more sensible to choice of the grid for p, q and γ_1, γ_2 than in the case of unbiasedness.

[17] Figure and B.5 and B.6 show the raw (not size-adjusted) power.
[18] This the feature should disappear for large samples.

Table 2.4: Size Results - Efficiency

Nominal Size	Test	T = 100	T = 200	T = 500
5%	WM	0.050	0.058	0.055
	Fluctuation	0.119	0.085	0.062
	Hansen	0.000	0.017	0.010
	Garcia	0.018	0.016	0.012
	ORS-H	0.020	0.023	0.010
	ORS-G	0.056	0.034	0.034
10%	WM	0.099	0.113	0.095
	Fluctuation	0.176	0.118	0.115
	Hansen	0.017	0.030	0.017
	Garcia	0.028	0.028	0.034
	ORS-H	0.046	0.033	0.030
	ORS-G	0.096	0.068	0.068

NOTE. – T denotes the sample size. Results are based on 1000 Monte Carlo replications, except for Hansen and ORS-H. Due to the computational burden, the Monte Carlo replications for Hansen and ORS-H are limited to 500. The DGP is a standard Normal. The results for Hansen are based on a 3-tuple of 12 grid points for $(p, q) \in [0.05, 0.95]$ and 20 grid points for $\gamma_s \in [0.05, 1]$. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.05, 0.95]$ and 20 grid points for $\gamma_1, \gamma_2 \in [-1, 1]$. The results for Garcia and ORS-G are based on $\pi \in [0.01, 0.99]$. The window size m for the Fluctuation test is set to $m = \frac{T}{2}$. The ORS-H test is based on a grid for $(\gamma_1, \gamma_2) \in [-1.5, 1.5]x[-1.5, 1.5]$.

To study power, we proceed similarly as in the case of unbiasedness. Under the alternative of a constant, but non-zero efficiency coefficient, we impose $\gamma_1 = \gamma_2$, and estimate the power of three tests over a grid of values for μ.[19]. The grid is set to $[0.05, 1]$ with increments of 0.05. Panel (a) of figure 2.2 shows the size-adjusted power results for a sample size of $T = 100$ at a nominal size of 5%. Again the power of the WM test cleary outperforms the other two tests.[20] However, as we will see in Panel (b) of figure 2.2 the traditional WM test has little power against the the alternative of Markov switching in γ. In the simulation, we set the state-to-state transition probabilities to be $(p, q) = 0.9$ and impose $\gamma_1 = -\gamma_2$.[21] Again, the rejection frequencies are computed over a grid for μ_1 of $[0.05, 2]$ with increments of 0.1. Panel (b) of figure 2.2 displays the size-adjusted rejections frequencies at a nominal size of 5%. The ORS-H and ORS-G exhibit strong power against the alternative of Markov switching, whereas the power of the traditional WM test remains rather flat. The Fluctuation test outperforms the ORS-H and ORS-G test for a constant, but non-zero efficiency parameter. However, in the case of Markov

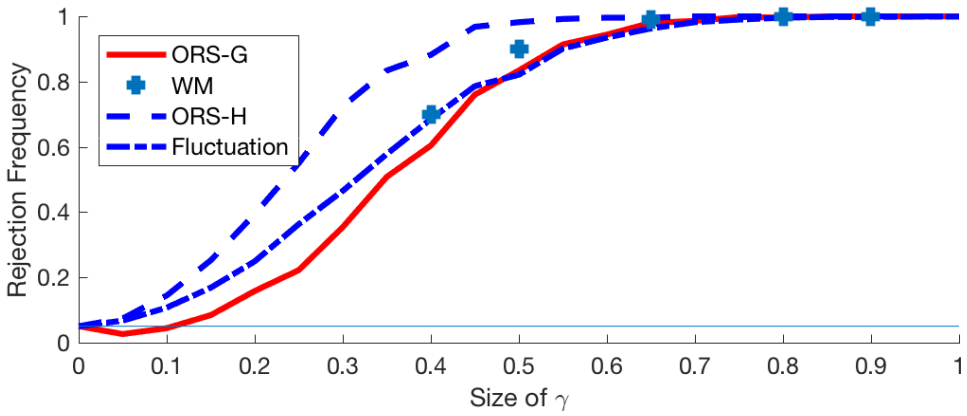
[19] As before, for the ORS-H test, we compute the power at fewer grid points due to the computational intensity of the test.

[20] An increase in the number of grid points for μ may increase the power of ORS-H.

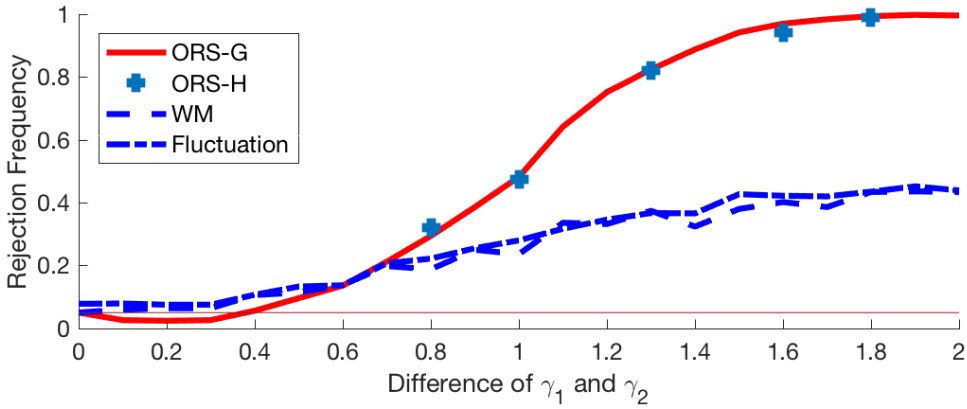
[21] Again, these parameter choices ensure that $\mathrm{E}\epsilon_{t+1|t} = 0$, i.e. that we can compute the power against Markov switching only.

switching efficiency parameters, its power is similar to the WM test. Again, the power of the Fluctuation test will be sensitive to the window size m.

Figure 2.2: Size-Adjusted Power - Efficiency



(a) Constant Efficiency Parameter



(b) Markov Switching Efficiency Parameter

NOTE. –The y-axis denotes the rejection frequency. The x-axis denotes the size of μ. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity.

2.4.3 Discussion of the Tests

Testing for Markov-switching is challenging and all of the proposed tests have advantages and disadvantages. ORS-H and ORS-G allow to test for a specific alternative, for instance a two-state first-order Markov switching model with switching in the mean or an additional regressor. However, ORS-H has the drawback of computational intensity and size distortions for small samples when testing for the intercept and rather strong underrejections in the efficiency test. On the other hand, ORS-H is robust to weakly dependent and heteroscedastic error terms. ORS-G has the disadvantage of violating the regularity conditions that are imposed to derive its

asymptotic distribution. In addition, when testing for efficiency or when including lags of forecast error the asymptotic distribution is not nuisance parameter free. For testing efficiency, the test is slightly undersized. In contrast, the original CHP test is optimal against local alternatives, requires estimation only under the null, is reasonably well sized for medium sized samples and exhibits high power under both alternatives. However, we can only test unbiasedness with it and it does not specify a clear alternative. To the practitioner, we recommend using the test that is most appropriate for the question at hand. When testing for unbiasedness, ORS-CHP can be a powerful test that is robust under the presence of (quite general) parametric time-variation.[22] When the researcher is interested in testing for efficiency and rationality or unbiasedness, with interest on testing against the specific alternative of a Markov switching, ORS-H and ORS-G might be the better choice. When using the ORS-H test, the researcher needs to set the grid of parameters of interest carefully. For the case of unbiasedness, we recommend to plot the forecast errors and decide then on a range for the intercept coefficients. For the test of efficiency, setting a grid around the full sample efficiency parameter could be a sensitive starting point.

2.5 Empirical Evidence

In the empirical section, we investigate the absolute forecasting performance of the federal funds rate forecast of the Blue Chip Financial Forecasts (BCCF) survey. The survey is conducted on a monthly basis and consists of approximately fifty participants in the private financial sector.[23] The predictions are fixed-event forecasts and we use the same methodology as Dovern et al. (2012b) to convert the forecasts to fixed-horizon forecasts. In total, the data ranges from 1983:M4 to 2018:M2. In the analysis, we will focus on the period starting in 1990:M1, as the periods in the 80s, are quite volatile and could potentially confound the results of a Markov switching model.[24] Further, we focus on the 3-month-ahead horizon and provide robustness results, using the 6-month-ahead forecasts, in appendix A.

Our results of testing for unbiasedness are based on a two-regime model, with switches only in the intercept, and we additionally control for lags of the forecast error. Thus the model takes the form

[22] In particular for one period ahead forecast, which should be serially uncorrelated under the assumption of rationality.

[23] We will focus on the consensus forecast of all participants.

[24] Note that our test still rejects but the rejection could be driven by outliers.

$$\epsilon_{t+3|t} = (1 - S_{t+3})\mu_1 + S_{t+3}\mu_2 + \sum_{i=1}^{d} \phi_i \epsilon_{t+3-i|t-i} + e_{t+3} \tag{2.33}$$

Table 2.5 displays the results of the ORS-H test for unbiasedness using (2.33) for $d = 0, 1, 2, 3$.[25] For all lag length, the ORS-H test clearly rejects with pvalues below 0.001 and results are very robust across different values for d. A traditional, West and McCracken (1998) full-sample test for forecast error unbiasedness has a pvalue of around 60% and would not reject the null hypothesis. The Fluctuation test by Rossi and Sekhposyan (2016) rejects at the 5% level for window sizes of $m < 100$. However, it provides less economic interpretation than a Markov switching model. The displayed coefficients are the values that correspond to the maximum of the ORS-H test, where we estimated the test over a 4-tuple of 12 grid points for $(p, q) \in [0.04, 0.96]$ x $[0.04, 0.96]$ and 20 grid points for $(\mu_1, \mu_2) \in [-1, 0.2]$ x $[-1, 0.2]$. The results show the presence of a very persistent regime, with a state-to-state transition probability of 96%, in which the forecasters are unbiased. A t-test on μ_1 does not reject the null hypothesis of $\mu_1 = 0$. However, in the second regime, which is considerably less persistent, the forecasters overestimate the future federal funds rate, and the coefficient μ_2 is significantly different from zero. The difference between the two states, $\mu_1 - \mu_2$, is approximately 25 basis points, a typical interest rate move by the FED. Figure 2.3 plots the smoothed regime probabilities of the MS-AR(1) model against the forecast errors and the federal funds target rate. The probability of regime 2 is associated with monetary easing, i.e., decreases in the federal funds rate and not limited to recessionary periods. In particular, in the time after the 90s recession as well as around 1998 and before the great recession the probability co-moves with movements in the FFR target rate. The regimes are well identified, in a sense that most regime probabilities are close to zero or one. Table A.1 and figure A.1 and A.2 in appendix A show that the results are robust to the exclusion of the zero lower bound period after the great recession. We use the same data set as Dahlhaus and Sekhposyan (2018), and our results relate straightforwardly. The authors find when evaluating forecast unbiasedness in different sub-samples of the data that the bias seems to be mainly driven by periods of monetary easing. We find similar periods of a negative forecast bias, but without having to know the conditioning variable - the periods are identified through the latent states of the Markov switching model.

[25] A rational forecast would exhibit maximum serial correlation length should be h-1, i.e., in this two. We show results for a maximum of three lags to be robust against rejections of the null hypothesis due to other types of misspecifications.

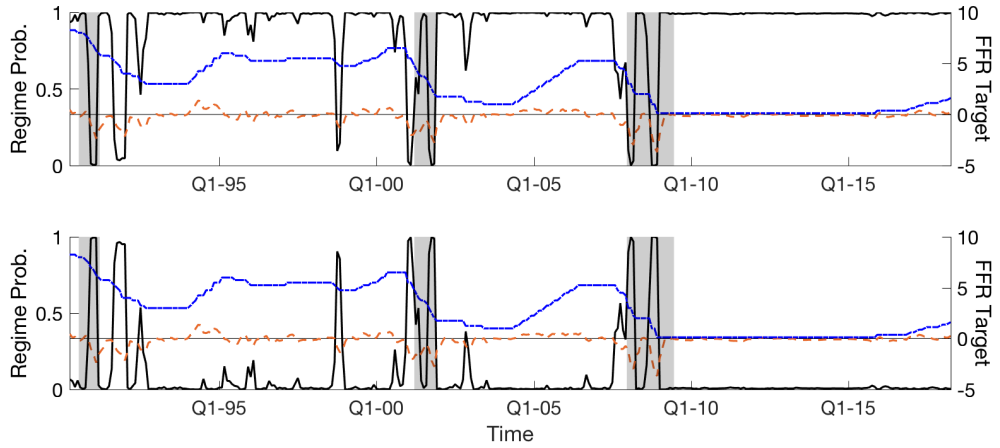
Table 2.5: Estimated Coefficients - 3-Months-Ahead Forecast Error

Model	p	q	μ_1	μ_2	ϕ_1	ϕ_2	ϕ_3	ORS-H	pvalue
AR(0)	0.96	0.50	0.01	-0.50	-	-	-	12.44	< 0.01
	(0.03)	(0.15)	(0.02)	(0.06)	-				
AR(1)	0.96	0.58	0.01	-0.24	0.79	-	-	8.46	< 0.01
	(0.03)	(0.22)	(0.02)	(0.04)	(0.02)				
AR(2)	0.96	0.58	0.01	-0.18	1.08	-0.31	-	8.42	< 0.01
	(0.04)	(0.36)	(0.02)	(0.05)	(0.05)	(0.04)			
AR(3)	0.96	0.63	0.01	-0.24	1.03	-0.26	-0.03	7.95	< 0.01
	(0.03)	(0.21)	(0.02)	(0.04)	(0.05)	(0.06)	(0.04)		

NOTE. – The coefficients displayed correspond to the coefficients of the maximum ORS-H statistic. The sample size is $T = 338$. Numbers in parenthesis denote standard errors. ORS-H denotes the value of the test statistic. pvalue denotes the pvalue obtained from the simulated asymptotic distribution. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.04, 0.96]$ and 20 grid points for $\mu_1, \mu_2 \in [-1, 0.2]$. p denotes the state-to-state transition probability for regime one and q denotes the state-to-state transition probability for regime two.

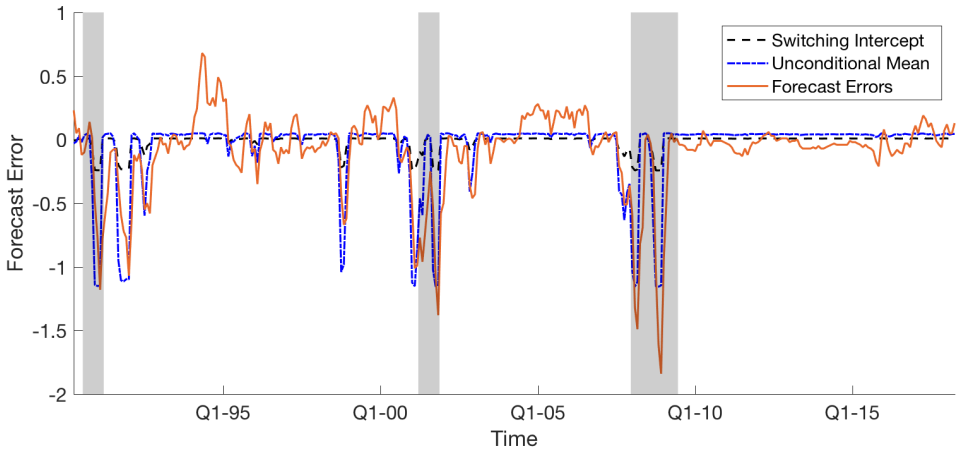
Figure 2.4 plots the forecast errors against the time-varying intercept and the time-varying state-unconditional mean, defined as $\mu_1/(1-\phi)$ and $\mu_2/(1-\phi)$.The switches play a significant role in explaining the recurring periods of negative bias and fit the negative values of the forecast errors rather well.

Figure 2.3: Regime Probabilities



NOTE. – The left y-axis denotes the regime probability. The right y-axis denotes the value of the forecast error and the FFR target. The x-axis denotes time. The solid line displays the smoothed regime probabilities. The dashed line displays the forecast errors. We rescaled the forecast errors by a factor of two, to increase the legibility of the plot. The dotted line displays the FFR target. Grey shaded areas display NBER recession periods.

Figure 2.4: Forecast Errors vs Switching Parameters



NOTE. – The solid line displays the forecasting error. The dashed line displays the time-varying intercept and the dashed line displays the state dependent unconditional mean, i.e. $\mu_1/(1-\phi)$ and $\mu_2/(1-\phi)$. The left y-axis denotes the value of the forecast error. The x-axis denotes time. Grey shaded areas display NBER recession periods.

2.6 Regime Switching in a Forecast Comparison Framework

In this chapter, we focussed on the evaluation of the absolute performance of a forecasting model under the potential presence of Markov switching. However, the idea of Markov switching is not limited to absolute evaluations but can also be interesting in the case of relative model comparisons. Consider the case of two competing models, which produce forecasts for the predictand y and the researcher or policymaker is interested in which model forecasts better, i.e., the object of interest is the loss differential ΔL_t. While the literature has investigated extensively the properties of full sample comparison tests (Diebold and Mariano, 1995; West, 1996; Clark and McCracken, 2001; Clark and West, 2006, 2007; Giacomini and White, 2006), fewer alternatives are available under the presence of instabilities or parametric time-variation. A non-parametric test that is robust to time-variation is provided by Giacomini and Rossi (2010), which is based on computing local Giacomini and White (2006) type of tests. An additional alternative could be the *conditional test* by Giacomini and White (2006), which, however, requires knowledge of a conditional variable. In general, the idea of Markov switching in loss differentials is interesting for three main reasons. First and different from Giacomini and Rossi (2010), it allows for an economic interpretation if the null hypothesis is rejected. Second, it does

not require *a priori* knowledge of the conditioning variable. Third, if a test finds evidence of switching in ΔL_t, a Markov switching model estimated on the loss differential can potential improve forecasts in real-time because the regime probabilities can be considered as a type of model selection or model averaging.

Testing the loss differential is, however, somewhat more challenging than testing the forecast errors. While the normality assumption might be a good approximation to forecast errors for the typical macroeconomic data, the loss differential exhibits strong deviations from normality. For instance, assuming that both of the two forecasts errors are normal and using the MSE as a loss function, ΔL_t will be the difference of two chi-squared variates, i.e., loss differentials typically exhibit fat-tails. As the model under the alternative is a mixture model, strong misspecifications of the likelihood under the null can lead to overrejections, even if no regime switching in the intercepts is present. How to best deal with this problem is an open question that we leave for future research.

2.7 Conclusion

We proposed three tests for evaluating absolute forecast performances and the tests are robust to parametric time-variation in the forecast errors. Overall, the three tests exhibit good size and power properties in a Monte Carlo study for unbiasedness and efficiency, where the ORS-H and ORS-G somewhat underreject in a test for efficiency. We showed that for the alternative of Markov switching, the new tests outperform the available alternative tests for absolute forecast performance. In an empirical investigation of the forecast rationality of the Blue Chip Financial Forecasts, we find that the forecasters exhibit a bias in the 3-months-ahead prediction of the federal funds target rate. While there is no deviation from unbiasedness when we consider the entire sample of forecast errors, we provide evidence that participants tend to systematically overestimate the federal funds rate during times of monetary easing.

Appendices

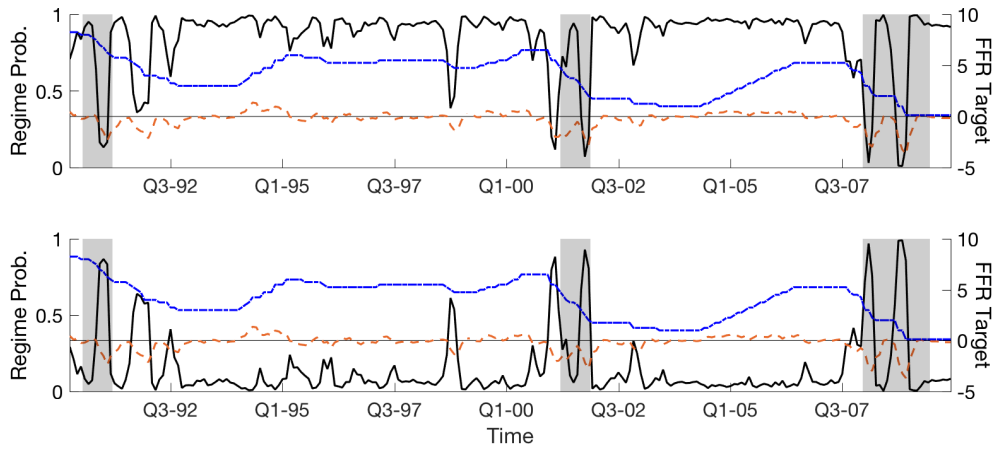
A Empirical Results - Robustness

Table A.1: Robustness - 3-Months-Ahead Forecast Error

Model	p	q	μ_1	μ_2	ϕ_1	ϕ_2	ϕ_3	ORS-H	pvalue
AR(1)	0.83	0.50	0.07	-0.31	-	-	-	7.03	< 0.01
	(0.05)	(0.23)	(0.04)	(0.12)	-				
AR(1)	0.92	0.50	0.01	-0.18	0.83	-	-	7.03	< 0.01
	(0.06)	(0.59)	(0.03)	(0.09)	(0.05)				
AR(2)	0.96	0.63	0.01	-0.18	1.10	-0.35	-	6.61	< 0.01
	(0.07)	(0.56)	(0.03)	(0.09)	(0.06)	(0.29)			
AR(3)	0.96	0.63	0.01	-0.18	1.08	-0.35	-0.29	6.26	< 0.01
	(0.07)	(0.51)	(0.03)	(0.09)	(0.09)	(0.09)	(0.05)		

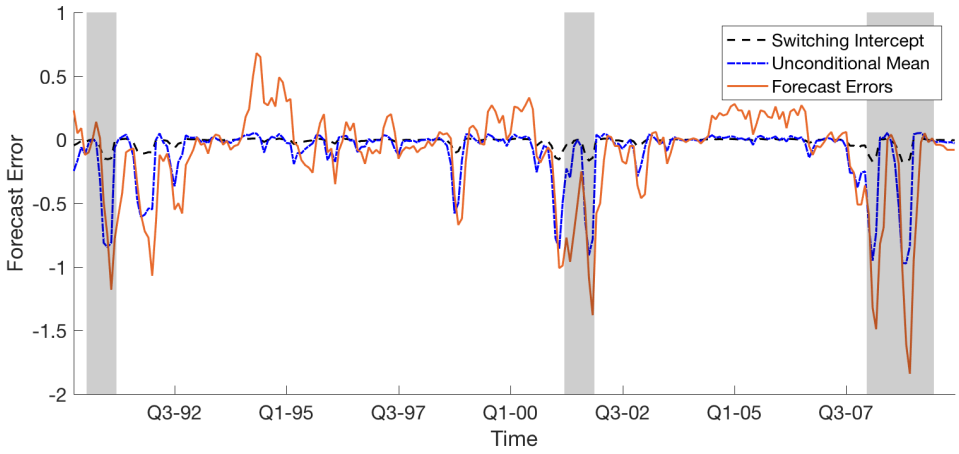
NOTE. – The coefficients displayed correspond to the coefficients of the maximum ORS-H statistic. The sample size is $T = 238$. Numbers in parenthesis denote standard errors. ORS-H denotes the value of the test statistic. pvalue denotes the pvalue obtained from the simulated asymptotic distribution. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.04, 0.96]$ and 20 grid points for $\mu_1, \mu_2 \in [-1, 0.2]$. p denotes the state-to-state transition probability for regime one and q denotes the state-to-state transition probability for regime two.

Figure A.1: 3-Months-Ahead: Regime Probabilities - Robustness



NOTE. – The left y-axis denotes the regime probability. The right y-axis denotes the value of the forecast error and the FFR target. The x-axis denotes time. The solid line displays the smoothed probabilities of regime one (upper panel) and two (lower panel). The dashed line displays the forecast errors. We rescaled the forecast errors by a factor of two, to increase the legibility of the plot. The dotted line displays the FFR target. Grey shaded areas display NBER recession periods.

Figure A.2: 3-Months-Ahead: Forecast Errors vs Switching Parameters



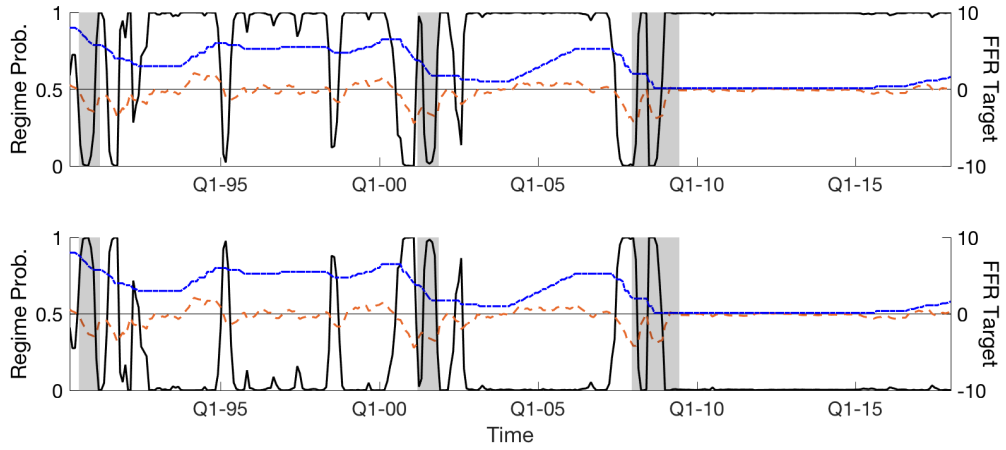
NOTE. – The solid line displays the forecasting error. The dashed line displays the time-varying intercept and the dashed line displays the state dependent unconditional mean, i.e. $\mu_1/(1-\phi)$ and $\mu_2/(1-\phi)$. The left y-axis denotes the value of the forecast error. The x-axis denotes time. Grey shaded areas display NBER recession periods.

Table A.2: Estimated Coefficients - 6-Months-Ahead Forecast Error

Model	p	q	μ_1	μ_2	ϕ_1	ϕ_2	ϕ_3	ORS-H	pvalue
AR(0)	0.79	0.50	0.14	-0.37	-	-	-	6.07	< 0.01
	(0.09)	(0.24)	(0.08)	(0.26)	-				
AR(1)	0.96	0.71	0.01	-0.31	0.84	-	-	6.07	< 0.01
	(0.02)	(0.13)	(0.02)	(0.04)	(0.02)				
AR(2)	0.96	0.75	0.01	-0.12	1.27	-0.38	-	5.29	< 0.01
	(0.11)	(0.56)	(0.05)	(0.10)	(0.10)	(0.06)			
AR(3)	0.96	0.79	0.01	-0.12	1.24	-0.26	-0.09	5.65	< 0.01
	(0.10)	(0.42)	(0.05)	(0.10)	(0.08)	(0.09)	(0.05)		

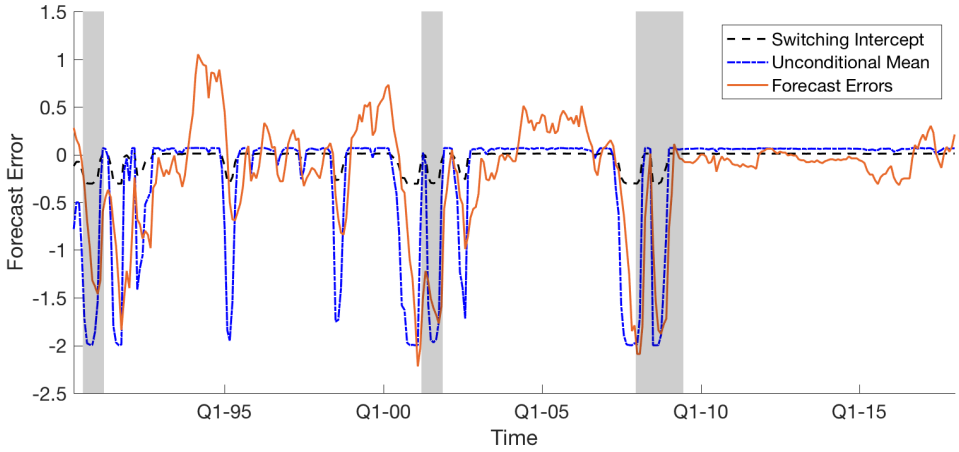
NOTE. – The coefficients displayed correspond to the coefficients of the maximum ORS-H statistic. The sample size is $T = 335$. Numbers in parenthesis denote standard errors. ORS-H denotes the value of the test statistic. pvalue denotes the pvalue obtained from the simulated asymptotic distribution. The results for ORS-H are based on a 4-tuple of 12 grid points for $(p, q) \in [0.04, 0.96]$ and 20 grid points for $\mu_1, \mu_2 \in [-1, 0.2]$. p denotes the state-to-state transition probability for regime one and q denotes the state-to-state transition probability for regime two.

Figure A.3: 6-Months-Ahead: Regime Probabilities



NOTE. – The left y-axis denotes the regime probability. The right y-axis denotes the value of the forecast error and the FFR target. The x-axis denotes time. The solid line displays the smoothed regime probabilities. The dashed line displays the forecast errors. The dotted line displays the FFR target. Grey shaded areas display NBER recession periods.

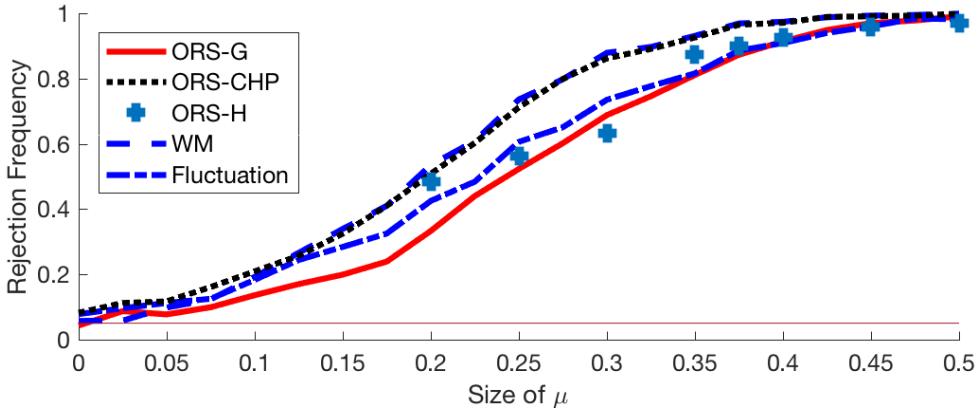
Figure A.4: 6-Months-Ahead: Forecast Errors vs Switching Parameters



NOTE. – The solid line displays the forecasting error. The dashed line displays the time-varying intercept and the dashed line displays the state dependent unconditional mean, i.e. $\mu_1/(1-\phi)$ and $\mu_2/(1-\phi)$. The left y-axis denotes the value of the forecast error. The x-axis denotes time. Grey shaded areas display NBER recession periods.

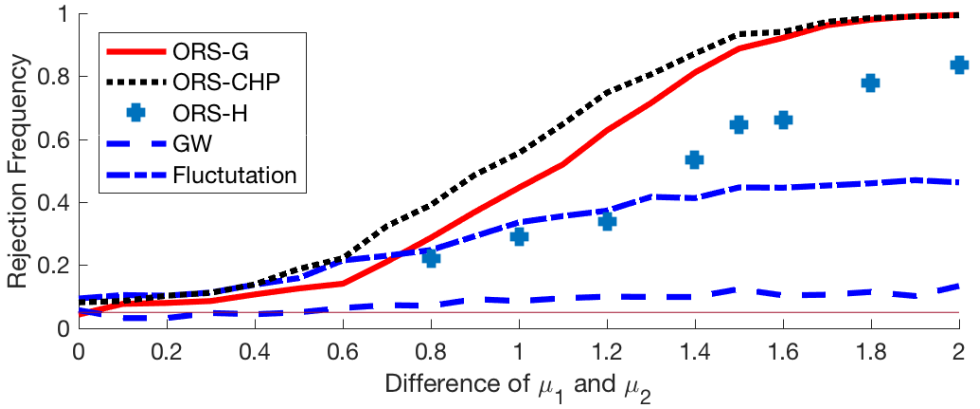
B Power - Not Size-Adjusted

Figure B.5: Power - Constant Mean



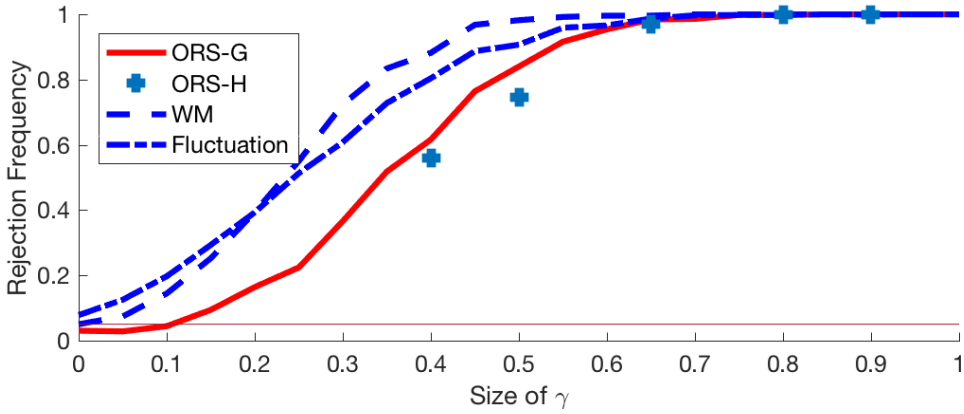
NOTE. – The y-axis denotes the rejection frequency. The x-axis denotes the size of μ. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity.

Figure B.6: Power - Markov Switching in the Mean



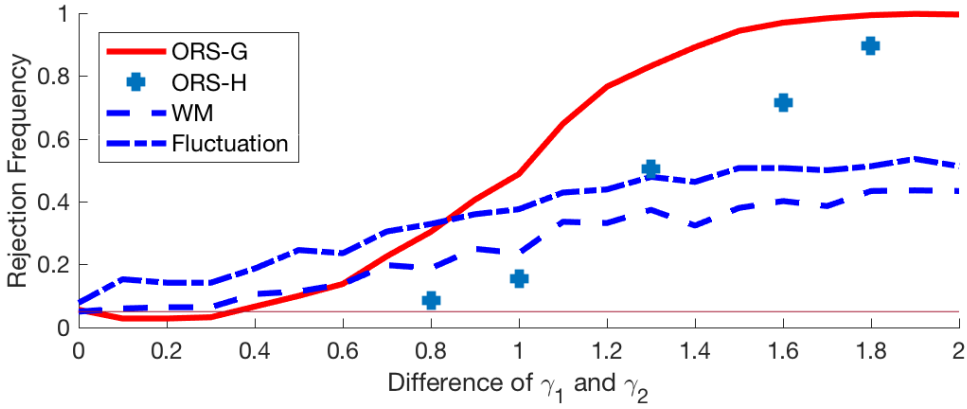
NOTE. – The y-axis denotes the rejection frequency. The x-axis denotes the size of μ. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity.

Figure B.7: Power - Constant Efficiency Parameter



NOTE. –The y-axis denotes the rejection frequency. The x-axis denotes the size of μ. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity.

Figure B.8: Power - Markov Switching Efficiency Parameter



NOTE. – The y-axis denotes the rejection frequency. The x-axis denotes the size of $\mu_1 - \mu_2$. The nominal size is 5% and illustrated by the solid horizontal line. The results are based on 1000 Monte Carlo replications - except for the ORS-H, which is based on 500 replications due to the computational intensity.

C Appendix for ORS-H

The arguments of Hansen (1992) apply in the same way to our partition of the parameter vector as they applied to his partion. In the following, we summarize Hansen's main steps for the convenience of the reader.

Assumption A1: *The first and second derivative of* $L(\tilde{\alpha}, \theta(\tilde{\alpha}))$ *with respect to* $\theta(\tilde{\alpha})$ *is uniformly bounded, i.e.*

$$\sup_{\tilde{\alpha}\in A, \theta\in\Theta} \left|\left|\frac{\partial L(\tilde{\alpha}, \theta(\tilde{\alpha}))}{\partial \theta}\right|\right| = O_p(n) \tag{C.1}$$

$$\sup_{\tilde{\alpha}\in A, \theta\in\Theta} \left|\left|\frac{\partial^2 L(\tilde{\alpha}, \theta(\tilde{\alpha}))}{\partial \theta \partial \theta'}\right|\right| = O_p(n) \tag{C.2}$$

Assumption A2: *Denote as* $D(\tilde{\alpha}) = \hat{\theta}(\tilde{\alpha}) - \theta(\tilde{\alpha})$. *Under the null, assume that*

$$\sup_{\tilde{\alpha}\in A} \left|\left|D(\tilde{\alpha})\right|\right| = o_p(1) \tag{C.3}$$

i.e. $\hat{\theta}(\tilde{\alpha})$ *converges uniformly over* $\tilde{\alpha} \in A$ *to* $\theta(\tilde{\alpha})$.

Denote by $\ell_t(\epsilon_{t|t-h}; \tilde{\alpha}, \theta) = \ell_t(\tilde{\alpha}, \theta) = \ell_t$ the log-likelihood and denote by $\mathrm{L}_T(\tilde{\alpha}, \theta) = \sum_{t=1}^{T} \ell_t(\cdot)$. The notation $\mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha}))$ denotes the same as $\mathrm{L}_T(\tilde{\alpha}, \theta)$ but for a fixed value of $\tilde{\alpha}$. Using a second-order Taylor expansion, it follows that

$$\mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha})) - \mathrm{L}_T(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) = \mathrm{D}(\tilde{\alpha})' \frac{\partial}{\partial \theta} \mathrm{L}(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) + \frac{1}{2} \mathrm{D}(\tilde{\alpha})' \frac{\partial^2}{\partial \theta \partial \theta'} \mathrm{L}(\tilde{\alpha}, \theta^*(\tilde{\alpha})) \mathrm{D}(\tilde{\alpha}) \tag{C.4}$$

where $\theta^*(\tilde{\alpha}) = \lambda \theta_0(\tilde{\alpha}) + (1-\lambda)\hat{\theta}_0(\tilde{\alpha})$ and $0 < \lambda < 1$. Under assumption A1 and A2 it follows that

$$\begin{aligned} &\sup_{\tilde{\alpha}\in A} \left|\left|\mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha})) - \mathrm{L}_T(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha}))\right|\right| \\ &= \sup_{\tilde{\alpha}\in A} \left|\left|\mathrm{D}(\tilde{\alpha})' \frac{\partial}{\partial \theta} \mathrm{L}(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) + \frac{1}{2} \mathrm{D}(\tilde{\alpha})' \frac{\partial^2}{\partial \theta \partial \theta'} \mathrm{L}(\tilde{\alpha}, \theta^*(\tilde{\alpha})) \mathrm{D}(\tilde{\alpha})\right|\right| = O_p(1) \end{aligned} \tag{C.5}$$

Let $\mathrm{R}(\tilde{\alpha})$ denote $\mathrm{R}(\tilde{\alpha}) = \mathrm{ELR}_T(\tilde{\alpha})$. For notational clarity, let the following denote the sample and population counterpart of the variables

$$\widehat{\mathrm{LR}}_T(\tilde{\alpha}) = \widehat{\mathrm{L}}_T(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) - \widehat{\mathrm{L}}_T(\hat{\theta}_0) \tag{C.6}$$

$$\mathrm{LR}_T(\tilde{\alpha}) = \mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha})) - \mathrm{L}_T(\theta_0) \tag{C.7}$$

$$\mathrm{R}(\tilde{\alpha}) = \mathrm{ELR}_T(\tilde{\alpha}) \tag{C.8}$$

$$\widehat{\mathrm{Q}}_T(\tilde{\alpha}) = \widehat{\mathrm{LR}}_T(\tilde{\alpha}) - \mathrm{R}(\tilde{\alpha}) \tag{C.9}$$

$$\mathrm{Q}_T(\tilde{\alpha}) = \mathrm{LR}_T(\tilde{\alpha}) - \mathrm{R}(\tilde{\alpha}) \tag{C.10}$$

Rewriting $\widehat{\mathrm{Q}}_T(\tilde{\alpha})$ and under A1 and A2

$$\begin{aligned}
\widehat{\mathrm{Q}}_T(\tilde{\alpha}) &= \widehat{\mathrm{LR}}_T(\tilde{\alpha}) - \mathrm{R}(\tilde{\alpha}) \\
&= \mathrm{LR}_T(\tilde{\alpha}) - \mathrm{R}(\tilde{\alpha}) + \widehat{\mathrm{LR}}_T(\tilde{\alpha}) - \mathrm{LR}_T(\tilde{\alpha}) \\
&= \mathrm{Q}_T(\tilde{\alpha}) + \left[\widehat{\mathrm{L}}_T(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) - \mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha}))\right] + \left[\mathrm{L}_T(\theta_0) - \widehat{\mathrm{L}}_T, \hat{\theta}_0)\right] \\
&= \mathrm{Q}_T(\tilde{\alpha}) + \left[\widehat{\mathrm{L}}_T(\tilde{\alpha}, \hat{\theta}(\tilde{\alpha})) - \mathrm{L}_T(\tilde{\alpha}, \theta(\tilde{\alpha}))\right] + \left[\mathrm{L}_T(\theta_0) - \widehat{\mathrm{L}}_T(\hat{\theta}_0)\right] \\
&= \mathrm{Q}_T(\tilde{\alpha}) + O_p(1) + O_p(1)
\end{aligned} \tag{C.11}$$

where the last inequality follows from (C.5) and the fact that $\left[\mathrm{L}_T(\theta_0) - \widehat{\mathrm{L}}_T(\hat{\theta}_0)\right]$ is proportional to a likelihood ratio test of the parameters in θ, which has a chi-square limiting distribution. Therefore

$$\frac{1}{\sqrt{T}}\widehat{\mathrm{Q}}_T(\tilde{\alpha}) = \frac{1}{\sqrt{T}}\mathrm{Q}_T(\tilde{\alpha}) + o_p(1) \tag{C.12}$$

As $\mathrm{R}(\tilde{\alpha}) \leq 0$, it follows that

$$\frac{1}{\sqrt{T}}\widehat{\mathrm{LR}}_T(\tilde{\alpha}) \leq \frac{1}{\sqrt{T}}\widehat{\mathrm{Q}}_T(\tilde{\alpha}) = \frac{1}{\sqrt{T}}\mathrm{Q}_T(\tilde{\alpha}) + o_p(1) \tag{C.13}$$

and

$$\frac{1}{\sqrt{T}}\widehat{\mathrm{LR}}^*_T(\tilde{\alpha}) \leq \frac{1}{\sqrt{T}}\widehat{\mathrm{Q}}^*_T(\tilde{\alpha}) = \frac{1}{\sqrt{T}}\mathrm{Q}^*_T(\tilde{\alpha}) + o_p(1) \tag{C.14}$$

Hansen (1992) imposes high-level assumptions on Q^*_T and assumes that Q^*_T obeys an empirical process CLT. In general, there are empirical process CLTs available (Andrews, 1991) that allow for near epoch dependence in the errors as well as heterogeneity.[26]

D Appendix for ORS-G

We derive the variance-covariance matrix of the ORS-Garcia test following García (1998), which is necessary for the simulation of the critical values. The matrix $V(\alpha)$ is defined as

$$V(\alpha) = \lim_{T\to\infty} T\mathrm{E}\left[\mathrm{S}_T(\theta_0, \alpha)\mathrm{S}_T(\theta_0, \alpha)'\right]$$

[26] Which rules out long-memory behaviour in the q_t.

where $\mathrm{S}_T(\theta_0, \alpha)$ is the score with respect to θ evaluated at θ_0, the true parameter values under the null. The matrix $K(\alpha_1, \alpha_2)$ is defined as

$$K(\alpha_1, \alpha_2) = \lim_{T\to\infty} T\mathrm{E}\Big[\mathrm{S}_T(\theta_0, \alpha_1)\mathrm{S}_T(\theta_0, \alpha_2)'\Big]$$

The variance-covariance matrix is

$$\bar{K}(\alpha_1, \alpha_2) = \tau V^{-1}(\alpha_1)^{-1} K(\alpha_1, \alpha_2) V^{-1}(\alpha_2)^{-1}\tau'$$

which is key for the simulation of the critical values, as $\bar{K}(\alpha_1, \alpha_2)$ summarizes the variance-covariance of the chi-square processes indexed by α. García (1998) provides the scores $\mathrm{S}_T(\theta_0, \alpha)$ for a slightly different model specification than we. In the following, we will therefore follow his strategy and derive the scores for our specification. Assume the general model

$$\epsilon_{t|t-h} = \mu + \mu_s + (\gamma + \gamma_s)x_t + \sum_{i=1}^{d} \phi_i \epsilon_{t-i|t-h-i} + e_t \tag{D.15}$$

where x_t denotes the $y_{t|t-h}$ or any additional regressors. The scores for $\mu, \mu_s, \gamma, \gamma_s, \phi_i$ and σ^2, evaluated at θ_0, take the form

$$\mathrm{S}_T(\theta_0, \alpha)_{\mu} = \frac{1}{T}\sum_{t=1}^{T} \frac{e_t}{\sigma_0^2}$$

$$\mathrm{S}_T(\theta_0, \alpha)_{\mu_s} = \frac{1}{T}\sum_{t=1}^{T} \frac{e_t}{\sigma_0^2} \sum_{s_t=0}^{1} p_{s_t}(\alpha) s_t(\alpha)$$

$$\mathrm{S}_T(\theta_0, \alpha)_{\gamma} = \frac{1}{T}\sum_{t=1}^{T} \frac{e_t x_t}{\sigma_0^2}$$

$$\mathrm{S}_T(\theta_0, \alpha)_{\gamma_s} = \frac{1}{T}\sum_{t=1}^{T} \frac{e_t x_t}{\sigma_0^2} \sum_{s_t=0}^{1} p_{s_t}(\alpha) s_t(\alpha)$$

$$\mathrm{S}_T(\theta_0, \alpha)_{\phi_i} = \frac{1}{T}\sum_{t=1}^{T} \frac{e_t \epsilon_{t-i|t-h-i}}{\sigma_0^2}$$

$$\mathrm{S}_T(\theta_0, \alpha)_{\sigma^2} = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\sigma_0^2}\left(\frac{e_t}{\sigma_0^2} - 1\right)$$

where $p_{s_t}(\alpha)$ denotes the probability $\Pr(S_t = s_t | \epsilon, \theta_0, \alpha)$ with $\epsilon = (\epsilon_{1|1-h}, ..., \epsilon_{T|T-h})$. Let $\pi = \frac{1-q}{2-p-q}$, with p, q from the transition matrix $P = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$ and

$\pi_j = \frac{1-q(\alpha_j)}{2-p(\alpha_j)-q(\alpha_j)}$. Under the assumption of serially uncorrelated and homoskedastic e_t, the matrix $V(\alpha)$ takes the form

$$V(\alpha) = \frac{1}{\sigma_0^2}\begin{pmatrix} 1 & \pi & \mu_x & \mu_x\pi & 0 & 0 & 0 & ... & 0 \\ \pi & \pi & \mu_x\pi & \mu_x\pi & 0 & 0 & 0 & ... & 0 \\ \mu_x & \mu_x\pi & \sigma_x^2 & \sigma_x^2\pi & \sigma^2\rho_{xy_{-1}} & \sigma^2\rho_{xy_{-2}} & \sigma^2\rho_{xy_{-3}} & ... & 0 \\ \mu_x\pi & \mu_x\pi & \sigma_x^2\pi & \sigma_x^2\pi & \sigma^2\rho_{xy_{-1}}\pi & \sigma^2\rho_{xy_{-2}}\pi & \sigma^2\rho_{xy_{-3}}\pi & ... & 0 \\ 0 & 0 & \sigma^2\rho_{xy_{-1}} & \sigma^2\rho_{xy_{-1}}\pi & \sigma^2\rho_0 & \sigma^2\rho_1 & \sigma^2\rho_2 & ... & 0 \\ 0 & 0 & \sigma^2\rho_{xy_{-2}} & \sigma^2\rho_{xy_{-2}}\pi & \sigma^2\rho_1 & \sigma^2\rho_0 & \sigma^2\rho_1 & ... & 0 \\ 0 & 0 & \sigma^2\rho_{xy_{-3}} & \sigma^2\rho_{xy_{-3}}\pi & \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2\rho_0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0\frac{1}{2\sigma_0^2} \end{pmatrix}$$

where $\mathrm{E}(x) = \mu_x$, $\mathrm{Var}(x) = \sigma_x^2$, $\mathrm{E}(\epsilon_{t|t-h}\epsilon_{t-i|t-h-i}) = \rho_i$ and $\mathrm{E}(x_t\epsilon_{t-i|t-h-i}) = \rho_{xy_{-i}}$. The matrix $K(\alpha_1, \alpha_2)$ takes the form

$$K(\alpha_1,\alpha_2) = \frac{1}{\sigma_0^2}\begin{pmatrix} 1 & \pi_2 & \mu_x & \mu_x\pi_2 & 0 & ... & 0 \\ \pi_1 & \min(\pi_1,\pi_2) & \mu_x\pi_1 & \mu_x\min(\pi_1,\pi_2) & 0 & ... & 0 \\ \mu_x & \mu_x\pi_2 & \sigma_x^2 & \sigma_x^2\pi_2 & 0 & ... & 0 \\ \mu_x\pi_1 & \mu_x\min(\pi_1,\pi_2) & \sigma_x^2\pi_1 & \sigma_x^2\min(\pi_1,\pi_2) & \sigma^2\rho_{xy_{-1}}\pi_1 & ... & 0 \\ 0 & 0 & \sigma^2\rho_{xy_{-1}} & \sigma^2\rho_{xy_{-1}}\pi_2 & \sigma^2\rho_0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & 0 & 0 & ... & 0\frac{1}{2\sigma_0^2} \end{pmatrix}$$

To simulate the chi-square process we follow the procedure of Garcia. First, we need to obtain the matrix

$$\Omega = \begin{pmatrix} \bar{K}(\alpha_1,\alpha_1) & \bar{K}(\alpha_1,\alpha_2) & ... & \bar{K}(\alpha_1,\alpha_N) \\ \bar{K}(\alpha_2,\alpha_1) & \bar{K}(\alpha_2,\alpha_2) & ... & \bar{K}(\alpha_2,\alpha_N) \\ \vdots & \vdots & \vdots & \vdots \\ \bar{K}(\alpha_N,\alpha_1) & \bar{K}(\alpha_N,\alpha_2) & ... & \bar{K}(\alpha_N,\alpha_N) \end{pmatrix}$$

and apply a Cholesky decomposition to it, $\Omega = PP'$. Then, let ϵ denote a r x 1 draw of i.i.d. $N(0,1)$ variates. The asymptotic distribution can then be simulated by drawing N times a variate ϵ and compute the distribution according to

$$\sup_{\alpha\in A_N} \left(P(\alpha_i)\epsilon\right)' \bar{K}^{-1}(\alpha_i,\alpha_i)P(\alpha_i)\epsilon$$

Unbiasedness: Let $\mathrm{H}_0 : \tau\beta = 0$ denote the null hypothesis $\mathrm{H}_0 : \mu_1 = \mu_2 = 0$, then

the matrix τ has the form

$$\tau = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \end{pmatrix}$$

and τ has dimension 2 x k, where k is the number of parameters in θ. For the two regime model, with transition matrix $P = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$, define $\pi = \frac{1-q}{2-p-q}$.

No Autocorrelation Under uncorrelated and homoskedastic error terms, the variance-covariance for simulating the asymptotic distribution has elements

$$\bar{K}(\alpha_i, \alpha_j) = \begin{pmatrix} \frac{1}{1-\pi_i} & -\frac{1}{1-\pi_i}\frac{\pi_i}{\pi_j} \\ -\frac{1}{1-\pi_i} & \frac{1}{(1-\pi_i)\pi_j} \end{pmatrix} \tag{D.16}$$

where $\pi_i = \frac{(1-q_i)}{2-q_i-p_i}$.

Efficiency: For the special case of a model without intercept, and/or $\mathrm{E}(x_t) = 0$, the matrix $\bar{K}(\alpha_i, \alpha_j)$ is identical to the case of testing for unbiasedness:

$$\bar{K}(\alpha_i, \alpha_j) = \begin{pmatrix} \frac{1}{1-\pi_i} & -\frac{1}{1-\pi_i}\frac{\pi_i}{\pi_j} \\ -\frac{1}{1-\pi_i} & \frac{1}{(1-\pi_i)\pi_j} \end{pmatrix} \tag{D.17}$$

For $\mathrm{E}(x_t) \neq 0$ and an intercept in the model, the asymptotic distribution is not nuisance parameter free and critical values need to be simulated depending on $\mathrm{E}(x_t)$ and $\mathrm{Var}(x)$. As the quantities are unknown in practice, the researcher has to substitute them with consistent estimates. The restrictions matrix

$$\tau = \begin{pmatrix} 0 & 0 & 1_\rho & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 1_{\rho_s} & 0 & ... & 0 \end{pmatrix}$$

has ones at the position of ρ and ρ_s in the parameter vector and zeros otherwise.

Rationality: For rationality the respective restriction matrix takes the form

$$\tau = \begin{pmatrix} 1_\mu & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 1_{\mu_s} & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 1_\rho & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 1_{\rho_s} & 0 & ... & 0 \end{pmatrix}$$

Again, critical values can depend on $\mathrm{E}(x_t)$ and $\mathrm{Var}(x)$ and can not be tabulated for

a general case.

E Appendix for ORS-CHP

The assumptions we impose are taken from Carrasco et al. (2014):

Assumption B1: *The random coefficient η_t is stationary and its distribution depends on some parameters α, which are not identified under the null. We assume that α belongs to a compact set A. Moreover, η_t is strongly exogenous in the sense that the joint likelihood of $\big(\epsilon_{h|t0}, ..., \epsilon_{T|T-h}, \eta_1, ..., \eta_P\big)$ factorizes to*
$\Pi_t f(\epsilon_{t+h|t}|\epsilon_{t+h-1|t-1}, ...\epsilon h|0, \theta_t) g(\eta_t|\eta_{t-1}, ..., \eta_1; \alpha)$ and the values of θ_t belong to some compact subset of R^p, Θ, containing θ_0.

Assumption B2: *$\epsilon_{t+h|t}$ is stationary under H_0, and the following condition on the conditional log-density of $\epsilon_{t+h|t}$ given $\big(\epsilon_{h|0}, ..., \epsilon_{T|t-1}\big)$, under H_0, are satisfied: $\ell_t(\theta)$ is at least five times differentiable and for $k = 1, .., 5$, it holds that*
$E_{\theta_0} sup_{\theta \in N}\Big(||\ell_t^{(k)}(\theta)||^{20}\Big) < \infty$, where $\ell_t^{(i)}$ denote the i-th derivative of the log likelihood function with respect to θ and N is a neighborhood around θ_0. E_{θ_0} is the expectation with respect to the probability measure corresponding to θ_0 and $||.||$ is the Euclidean norm. Moreover, θ_0 is an interior point of Θ and the information matrix $I(\theta_0) = E_{\theta_0}\big(\ell_t^{(1)}(\theta_0)\ell_t^{(1)'}(\theta_0)\big)$ is non-singular.

Assumption B3: *The latent variable η_t can be represented by $\eta_t = \mu_s h S_t$, where μ_s is a finite scalar constant and h is a vector with $||h|| = 1$. Moreover, S_t is stationary Markov chain, with a finite or a continuous but bounded state space, and is β-mixing with geometric decay such that $var(S_t) = 1$ and $cov(S_t, S_s) = \alpha^{|t-s|}$ for $-1 < \underline{\alpha} \leq \alpha \leq \bar{\alpha} < 1$ and $h = (1, 0, ..., 0)$. Assume that $E\big(\eta_t \eta_k\big)$, for any integer k, is assumed to be continuous in α.*

Let us rewrite the numerator of our test statistic, $v_T(\theta, \alpha)$, as

$$v_T(\theta, \alpha) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \big(\mu_{2,t}(\theta, \alpha) - \mathrm{d}(\alpha)' l_t^{(1)}(\theta)\big) \tag{E.18}$$

where $\mathrm{d}(\alpha) = \mathrm{d}(\alpha, \theta_0) = I^{-1}(\theta_0)\mathrm{cov}\big(\mu_{2,t}(\alpha, \theta_0), \ell_t^1(\theta_0)\big)$, with $I(\theta_0)$ being the information matrix. For the model we consider and under our null, the definition of $v_T(\theta, \alpha)$ is identical to the definition of $v_T(\theta, \alpha)$ in CHP. Note that the first element of $\mathrm{d}(\alpha)$, i.e. the element associated with the score of μ, is equal to zero. Note further that $\sum_{t=1}^{T} d'(\alpha) l_t^1(\hat{\theta}_0) = 0$, because the score with respect to the unre-

stricted parameters is zero when evaluated at the MLE $\hat{\theta}_0$, which implies $v_T(\hat{\theta}_0, \alpha) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}_0, \alpha)$. In other words, $v_T(\theta, \alpha)$ in (E.18) and $v_T(\theta, \alpha)$ in (2.19) are identical when evaluated at $\hat{\theta}_0$. To show that $v_T(\hat{\theta}_0, \alpha)$ has the same limiting distribution as $v_T(\theta_0, \alpha)$, we need to show that $v_T(\hat{\theta}_0, \alpha)$ converges, uniformly in $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, to $v_T(\theta_0, \alpha)$. Following the argument of CHP, we use a second-order Taylor expansion around θ_0, which gives

$$\begin{aligned} v_T(\hat{\theta}_0, \alpha) = v_T(\theta_0, \alpha) &+ \frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} v_T(\theta_0, \alpha) \sqrt{T}(\hat{\theta}_0 - \theta_0) \\ &+ \frac{1}{2}(\hat{\theta}_0 - \theta_0)' \frac{1}{\sqrt{T}} \frac{\partial^2}{\partial \theta \partial \theta'} v_T(\bar{\theta}, \alpha) \sqrt{T}(\hat{\theta}_0 - \theta_0)' \end{aligned} \tag{E.19}$$

where $\bar{\theta} = (1 - \lambda)\theta_0 + \lambda \hat{\theta}_0$ and $\lambda \in (0, 1)$. As our definition of $v_T(\theta, \alpha)$ is identical to the definition of CHP, in what follows we can use the arguments as in their proof. Note that $\sqrt{T}(\hat{\theta}_0 - \hat{\theta}) = O_p(1)$ and $(\hat{\theta}_0 - \hat{\theta}) = o_p(1)$ under the null. Further, we invoke assumption 2, which guarantees the uniform convergence of $\frac{1}{\sqrt{T}} \frac{\partial^2}{\partial \theta \partial \theta'} v_T(\bar{\theta}, \alpha)$ to a constant. Lemma C.1 of CHP is sufficient to establish that $\frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} v_T(\theta_0, \alpha) \to_p 0$ uniformly in α, which gives the required result. The asymptotic distribution of $v_T(\hat{\theta}_0, \alpha)$ can thus be derived from the asymptotic distribution of $v_T(\theta_0, \alpha)$, which is identical to distribution derived in Andrews and Ploberger (1996). Our denominator converges to the same probability limit as in CHP (dropping α for notational convenience)

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{T} \left[\mu_{2,t}(\hat{\theta}_0) - \hat{d}'(\hat{\theta}_0) \ell_t^{(1)}(\hat{\theta}_0) \right]^2 \\ &= \frac{1}{T} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}_0)^2 - \frac{1}{T} \hat{d}'(\hat{\theta}_0) \left(\sum_{t=1}^{T} \ell_t^{(1)}(\hat{\theta}_0) \ell_t^{(1)'}(\hat{\theta}_0) \right) \hat{d}(\hat{\theta}_0) \\ &= \frac{1}{T} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}_0)^2 - \left(\frac{1}{T} \sum_{t=1}^{T} \mu_{2,t} \ell_t^{(1)'} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \ell_t^{(1)}(\hat{\theta}_0) \ell_t^{(1)'}(\hat{\theta}_0) \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \mu_{2,t} \ell_t^{(1)} \right) \\ &\to_p E\left(\mu_{2,t}(\theta_0)^2 \right) - E\left(\mu_{2,t}(\theta_0) \ell_t^{(1)'}(\theta_0) \right) E\left(I(\theta_0) \right)^{-1} E\left(\mu_{2,t}(\theta_0) \ell_t^{(1)}(\theta_0) \right) \end{aligned} \tag{E.20}$$

For $\alpha = 0$, the limit of the test-statistic is indeterminate, a feature we inherit from the original test of CHP. The asymptotic distribution described in (2.22) has the form

$$\text{supTS} \to_d \sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} \frac{1}{2} \left(\max\left(0, K(\alpha) \right) \right)^2 \tag{E.21}$$

For the model without autoregressive components, it follows that

$$K(\alpha) = \text{sign}(\alpha)\sqrt{1-\alpha^2}\sum_{i=1}^{\infty} Z_i \alpha^i$$

where Z_i is an iid standard Normal variate. For a model that includes the autoregressive component the distribution can be derived from the asymptotic distribution of $\sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} v_T(\theta_0, \alpha)$.

Chapter 3

Real-Time Density Forecasts via Pooled Quantile Regressions

3.1 Introduction

Recent developments in the economic forecasting literature aim to give a probabilistic perspective on the predictand, instead of a point prediction only. The probabilistic forecasts are typically estimated using a parametric density assumption, the shape of the distribution is (to some extent) determined by the choice of the parametric family and only the mean is a direct function of observables. In contrast, quantile regressions are a semi-parametric approach for modeling distributions, allow to specify each quantile of the predictive density directly as a function of observables and leave the shape of the distribution unrestricted. Subsequently, the models allow for more flexible, and a growing literature in economics (Gaglianone and Lima, 2012; Manzan, 2015; Korobilis, 2017; Adrian et al., 2016) uses quantile regressions for probabilistic forecasts.

We contribute to this literature in several ways. First, we are the first to compare out-of-sample quantile regression forecasts of real US GDP growth to benchmark stochastic volatility models using *real-time* data. Second, we empirically investigate how two different estimation strategies affect the performance of quantile regressions for forecasting first- and final-release data. For *real-time* out-of-sample forecasts, researchers are confronted with substantial measurement error in first-release data, see Koenig et al. (2003); Croushore and Stark (2003), and need to decide which vintage to use for the estimation of the model. Further, relevant predictands in practice can include both, first-release as well as final-release data. We compare the strategies of *end-of-sample* vintages, the common approach in the literature, and *real-time* vintages for forecasting first-release data. For the prediction of final-release data,

we base the estimation on *end-of-sample* vintages.[1] Third, due to the abundance of candidate predictors, we incorporate the information provided by many predictors by pooling individual regressions.

Within the class of quantile regressions, our results indicate the following. First, forecasting first-release GDP growth is best done by estimating the model with first-release data only. The gain in forecast accuracy is more pronounced, and statistically significant, for one-quarter-ahead forecasts than for four quarter-ahead, and quantiles towards the tails. Importantly, the performance gain from first-release data is more considerable during recessions for both horizons, one and four quarters-ahead. Second, equal-weighting outperforms bayesian model averaging for a horizon of four quarters. Relative to the competitor models, the pooled bayesian quantile regressions outperform (pooled) univariate and multivariate stochastic volatility models at the one-quarter-ahead forecast, where the evidence is stronger for above-median quantiles. For horizons of four quarters-ahead, the results speak in favour of quantile regressions, but no statistically significant differences are observed. As the tick loss function comparison involves multiple testing, we additionally estimate a density based on the quantiles forecasts and compare it to the competitor models based on the continuous ranked probability score (CRPS). The results again are that the quantile models are competitive to stochastic volatility models. However, when we plot the loss differential, we observe that the multivariate stochastic volatility model tends to perform better during recession periods but worse after the great recession.

Our work is mainly related to Gaglianone and Lima (2012), Manzan (2015) and Adrian et al. (2016). Manzan (2015) explores the performance of frequentist quantile regressions with variable selection via Lasso and compares the out-of-sample forecasting performance of the quantile regressions using the tick-loss. The main difference from his paper is that we use *real-time* data only, investigate the performance for *first-* and *final-release* and deal with the problem of many predictors by model pooling instead of variable selection. In addition, we obtain smooth density forecasts based on the predicted quantiles and evaluate the respective forecasting performance using density scoring-rules. Gaglianone and Lima (2012) use SPF mean forecasts as predictors in a quantile regression model and construct density forecasts based on the quantile forecasts but focus on the US unemployment rate and do not use real-time data. Adrian et al. (2016) found that the National Financial Condition Index (NFCI) has strong in-sample and out-of-sample predictive power for lower quantiles of GDP growth. We differ from their paper in several ways. First,

[1] *End-of-sample* vintages are defined as using only the latest data available for model estimation. *Real-time* vintages instead use only first-release data for model estimation. See Croushore (2006) for an overview.

we compare the out-of-sample forecasts not within the class of quantile regressions but to current benchmark competitor density forecast models. Second, we use *real-time* data for both the dependent variable and the predictand, which excludes the use of the NFCI for the largest part of the sample. Third, we consider a broader range of predictor and incorporate the resulting model uncertainty by equal-weight pooling and bayesian model averaging. The paper is further related to the strand of literature that analyzes the effect of data revisions on forecasts (Koenig et al., 2003; Croushore and Stark, 2003). We intend to contribute to that literature through an empirical assessment of the forecasting performance of quantile regressions based on two different strategies, analogue to the procedures used for mean predictions in Koenig et al. (2003), for forecasting first- and final-release data. In a recent paper, Clements and Galvao (2017) investigate the impact of data revision on density forecasts from stochastic volatility models. Different from their work, we focus on the empirical performance of the model class of quantile regressions for real-time forecasts.

The set of predictors we use consists of the main macroeconomic variables and the term spread. To guarantee a true out-of-sample forecast, we rely on the data vintages provided by the Real-Time Data Research Center of the Philadelphia FED for each predictor and the dependent variable.[2] To incorporate the information in a data-rich environment, we choose a model averaging approach in this work, i.e. we estimate a separate univariate model for each predictor and average the resulting pool of forecast by equal-weights and, alternatively, bayesian model averaging.[3]

Quantile regressions in economics started with the work of Koenker and Basset (1978). Bayesian quantile regressions are based on the asymmetric Laplace distribution (ALD) and were introduced by Yu and Moyeed (2001). We will base the estimation of the quantile models on the Gibbs sampler developed by Kozumi and Kobayashi (2011). Recent work of Bernardi et al. (2017) extents the quantile regression to time-varying parameters and dynamic model averaging. Instead of time-varying parameters, we use a rolling-window estimation scheme to deal with potential parameter instabilities. In general, we contribute to the quantile regression literature by an empirical evaluation of the model's forecasting performance for macroeconomic time series.

The paper is organized as follows. Section 3.2 describes the bayesian quantile regression, the competitor models and the empirical estimation strategy with respect

[2]The real-time data set was first compiled by Croushore and Stark (2001).

[3]For mean predictions, the success of equal-weighted averaging is documented in Timmermann (2006); Wright (2009) found that bayesian model averaging improves mean forecasts for inflation and we are evaluating both strategies.

to the use of different vintages. Section 3.3.1 shows some results on the in-sample estimation of the quantile regressions. Section 3.3.2 and 3.3.3 display the main results on the out-of-sample forecasting exercise for first-release and final-release data. Section 3.4 concludes.

3.2 Prediction Models and Empirical Strategy

3.2.1 Bayesian Quantile Regression

Let Y_t be the predictand and X_t a set of K predictor variables. The predictive quantile regressions considered in this paper take the form of

$$Q_\tau(Y_{t+h}|X_t) = X_t\beta(\tau) \tag{3.1}$$

where h denotes the forecast horizon and $Q_\tau(Y_{t+h}|X_t)$ denotes the τ^{th} conditional quantile of Y_{t+h}, with $\tau \in [0,1]$. Yu and Moyeed (2001) showed that quantile regressions can be estimated by maximum likelihood using the asymmetric Laplace distribution (ALD), which takes the form

$$f(e_t|u,\tau) = \tau(1-\tau)\exp\{-\rho_\tau(e_t)\} \tag{3.2}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u<0))$ and we say that $e_t(\tau)$ has the asymmetric Laplace distribution.[4] Then

$$Y_t = X_t\beta(\tau) + e_t(\tau) \tag{3.3}$$

is equivalent to the model by Koenker and Basset (1978). Kozumi and Kobayashi (2011) note that $e_t(\tau)$ can be represented by a mixture of independent exponential and normal distributions and subsequently developed a Gibbs sampler, on which we are going to rely for estimation.

The baseline model takes the form

$$Y_{t+h} = \alpha_k(\tau) + Y_{t,qq}\beta_k(\tau) + X_{t,k}\gamma_k(\tau) + \sigma_k(\tau)e_{t+h}(\tau) \tag{3.4}$$

such that for each predictor, $X_{t,k}$, we run a separate univariate predictive regression and we produce forecasts for h=1 and h=4. $Y_{t,qq}$ denotes the quarter-on-quarter growth rate of real US GDP. In the case of h=1, Y_{t+h} equally denotes the quarter-on-quarter growth rate and for the case of h=4, Y_{t+h} denotes the year-on-year quarterly

[4]The distribution is easily extended to include a scaling coefficient σ. Please see Yu and Moyeed (2001) for details.

growth rate. Let $\theta_k(\tau) = \big(\alpha_k(\tau), \beta_k(\tau), \gamma_k(\tau), \sigma_k(\tau)\big)'$ denote the vector of parameters. In total we obtain for each quantile $K+1$ out-of-sample forecasts $\hat{q}^{(k)}_{t+h|t}(\tau)$, $k = 1, ..., K+1$, which is estimated using information up to time t.[5]

The benchmark forecast will be the *equal-weight* pool of $\hat{q}^{(k)}_{t+h|t}(\tau)$, i.e. $Q^{\text{ew}}_{t+h|t}(\tau) = \frac{1}{K+1}\sum_{k=1}^{K+1} \hat{q}^{(k)}_{t+h|t}(\tau)$. In addition, we use bayesian model averaging, which provides a coherent mechanism to account for model uncertainty that arises from many potential predictors. Given the data, $Y^t = [Y_1, ..., Y_t]$, we assign the following probability to model k

$$\Pr(M_k|Y^t) = \frac{\Pr(Y^t|M_k)\Pr(M_k)}{\sum_{j=1}^{K+1}\Pr(Y^t|M_k)\Pr(M_j)} \tag{3.5}$$

where M_i denotes model i. We will make use of the relationship of the marginal likelihood to the prior and posterior as in Chib (1995)

$$\Pr(Y^t|M_k) = \frac{\Pr(Y^t|\theta_i^*, M_k)\Pr(\theta_i^*|M_k)}{\Pr(\theta_i^*|Y^t, M_i)} \tag{3.6}$$

where θ_k^* is for example the mean of the posterior of θ_k.[6] Each model is assigned the same *a priori* probability $\Pr(M_k)$. The resulting weights $w^{bma}_{t,k} = \Pr(Y^t|M_k)$ are then used to construct the *bma*-pooled forecast $\hat{Q}^{\text{bma}}_{t+h|t}(\tau) = \sum_{k=1}^{K+1} \hat{q}^{(k)}_{t+h|t}(\tau) w^{bma}_{t,k}$.[7]

3.2.2 Quantile Regression Density Forecast

In addition to the quantile predictions, we obtain continuos densities out of the quantile predictions for two reasons.[8] First, the literature typically evaluates probabilistic forecasts on the basis density scoring rules, such as the log-score or the CRPS. Second, the evaluation of the tick-loss function at different quantiles requires multiple testing and, to the best of our knowledge, there is no test available to jointly evaluate the out-of-sample performance for a discrete number of quantiles. Once we obtain density forecasts, we can compare the forecasts of the different models using a scoring rule for densities, which mitigates the problem of multiple testing.[9] To obtain a continuos density out of the quantile predictions, we follow Adrian et al. (2016) and use the skew student's t proposed by Azzalini and Capitano (2003). The

[5] The total $K+1$ quantile forecasts are obtained through K models of type (3.4) and one AR(1) model without any additional predictors.

[6] We suppress the (τ) for notational convenience.

[7] The problem of quantile crossing essentially never occurs in our application.

[8] Similar to Gaglianone and Lima (2012); Adrian et al. (2016)

[9] When we evaluate the scoring rule, the potential additional variation, introduced through the fit of skew student's t, is not accounted for.

minimisation takes the form of

$$\min_{\mu_t,\alpha_t,\sigma_t,\nu_t} \sum_{j=1}^{N} \left(\hat{Q}_{t|t-h}(\tau_j) - T^s(\tau_j;\mu_t,\alpha_t,\sigma_t,\nu_t)\right)^2 \tag{3.7}$$

where $\hat{Q}_{t|t-h}(\tau_{j,t})$ is the predicted value of the τ_j^{th} quantile, N is the number of predicted quantiles and $T^s(\tau_{j,t};\mu_t,\alpha_t,\sigma_t,\nu_t)$ is the distribution function of the skew student's distribution at $\tau_{j,t}$. Estimation of the parameters $(\mu_t,\alpha_t,\sigma_t,\nu_t)$ is performed independently for each time period.

3.2.3 Competitor Models

We compare the forecasts of the bayesian quantile regression models to univariate and multivariate stochastic volatility models.

The univariate stochastic volatility models take the form

$$Y_{t+h} = \alpha_k + Y_{t,qq}\beta_k + X_{t,k}\gamma_k + \sqrt{h_{t+h,k}}\epsilon_{t+h} \tag{3.8}$$

where

$$\log h_{t+h,k} = \log h_{t+h-1,k} + u_{t+h} \tag{3.9}$$

and $\theta_k = (\alpha_k, \beta_k, \gamma_k, \sigma_{u,k}, h_{t,k})$. The errors ϵ_t, u_t are mean zero, iid normally distributed with $\sigma_\epsilon = 1$ and the models are estimated using the sampler of Kim et al. (1998). Following Geweke (1984), we base the predictive density on

$$\text{pr}(Y_{t+h}|Z_{t,k}) = \int\int \text{pr}(Y_{t+h}|\theta^{t+h,k}, Z_{t,k})\text{pr}(\theta^{t+h,k}|\theta_{t,k}, Z_{t,k})\text{pr}(\theta_{t,k}|Z_{t,k})\text{d}\theta_t\text{d}\theta^{t+h}$$

where $Z_{t,k} = [X_t, Y_{t,qq}]$ and $\theta^{t+h,k} = [\theta_{t+h,k}, ..., \theta_{t+1,k}]$, i.e. we take parameter uncertainty and their time variation into account. The quantiles $\hat{q}^{(k)}_{t+h|t}(\tau)$ are then obtained numerically by simulating from $\text{pr}(Y_{t+h}|Z_{t,k})$.

We compute *equal-weighted* pools to deal with the model uncertainty, which are denoted by $\hat{Q}^{ew}_{t+h|t,sv}(\tau)$.

The multivariate competitor model isa standard bayesian VAR with stochastic volatility. The form of the model is

$$Y_{t+h} = A_0 + A_1 Y_t + ... + A_p Y_{t-p+1} + \Sigma^{1/2}_{t+h}\epsilon_{t+h} \tag{3.10}$$

where $\epsilon_{t+h} \sim N(0, \mathbb{I})$ with $\mathbb{I}$ being an identity matrix, $\Sigma_{t+h} = F_{t+h} D_{t+h} F'_{t+h}$ with D_{t+h} being a diagonal matrix and F_{t+h} a lower triangular matrix with ones on the main diagaonal and

$$\log h_{t+h} = \alpha_h + \log h_{t+h-1} + u_{t+h} \tag{3.11}$$

$$\phi_t = \phi_{t-1} + \xi_{t+h} \tag{3.12}$$

where $\log h_{t+h}$ contains the diagonal elements of D_{t+h} and ϕ_t is the vector of elements of F_{t+h}. The errors $u_{t+h}, \xi_{t+h}, \epsilon_{t+h}$ are iid Gaussian. We decided to use a four variable VAR to balance between the problems of noisy parameter estimates in large VARs and a too small VAR that leaves out relevant information. The vector Y_t contains real GDP growth, changes in the unemployment rate, real residential investment growth and the term spread. Including an unemployment indicator in this type of VARs is standard. For two reasons we include real residential investment growth as variable: First, our sample includes the great recession, which was strongly related to the housing market. Further, a recent paper by Aastveit et al. (2018) finds that residential investment is a reliable out-of-sample predictor for economic downturns. Finally, with the term spread, we add an interest rate variable. Analogue to the univariate models, the predictive density $\text{pr}(Y_{t+h}|Y_t)$ is obtained from

$$\text{pr}(Y_{t+h}|Y_t) = \int \int \text{pr}(Y_{t+h}|\theta^{t+1}, Y_t)\text{pr}(\theta^{t+h}|\theta_t, Y_t)\text{pr}(\theta_t|, Y_t)\text{d}\theta_t \text{d}\theta^{t+h}$$

and the quantiles are again obtained numerically by simulating from $\text{pr}(Y_{t+h}|Y_t)$.

3.2.4 Parameter Choices and Priors

We estimate all models using a rolling-window, to mitigate the problem of structural breaks, of size $T = 100$. For the choice of the priors and hyperparameters, we attempt to closely follow the convention in the literature.

Priors for Bayesian Quantile Regressions

Table 3.1: Priors for Bayesian Quantile Regression

Parameter	Prior	Hyperparameters
$\alpha_k(\tau), \beta_k(\tau), \gamma_k(\tau)$	Normal	μ_p, Σ_p
σ_k	Inverse-Gamma	t_0, s_0

The resulting conditional posterior densities are given in Kozumi and Kobayashi (2011). The prior mean of $\big(\alpha_k(\tau), \beta_k(\tau), \gamma_k(\tau)\big)$ is specified as the ordinary least square (ols) estimate and the variance is taken to be an identity matrix with diagonal

elements equal to 30. We chose a very flat prior to allow for the possibility that the coefficients vary over quantiles. The hyperparameters $\sigma_p \sim IG(t_0/2, s_0/2)$, with $t_0 = 3.13$ and $s_0 = 0.1$, which again imply a rather flat prior.

Priors for Univariate Stochastic Volatility Models

Table 3.2: Priors for Univariate SV Model

Parameter	Prior	Hyperparameters
$(\alpha_k, \beta_k, \gamma_k)$	Normal	μ_p, Σ_p
$\sigma_{u,k}$	Inverse-Gamma	$t_{0,u}, s_{0,u}$

The hyperparameters μ_p, Σ_p are set equal to their ols counterpart. The prior on $\sigma_{u,k}$ is inverse-gamma with $\sigma_{u,k} \sim IG(t_{0,u}/2, s_{0,u}/2)$ where $t_{0,u} = 2.1$ and $s_{0,u} = t_{0,u}k_H^2$, with $k_H = 0.01$, which is a rather flat prior. In the filtering step of h_t, we use the same Normal-mixture approximation for the $\chi 2$-distribution of ϵ_t^2 as in Kim et al. (1998).

Priors for BVAR

Table 3.3: Priors for Multivariate SV Model

Parameter	Prior	Hyperparameters
A_0	Normal	μ_0, Σ_0
A_i for $i = 1, ..., p$	Normal	μ_i, Σ_i
σ_u	Gamma	$t_{0,u}, s_{0,u}$
σ_ξ	Gamma	$t_{0,\xi}, s_{0,\xi}$

We follow closely the existing literature on density forecast with VARs. The hyperparameters take the values, for $i = 1, ..., p$, $\mu_i = \mathbf{0}$ and $\Sigma_i = \mathbb{I}$, i.e. Minnesota-type priors for the coefficient matrices A_i. For the intercept the hyperparameters do not impose shrinkage as $\mu_0 = \mathbf{0}$ and $\Sigma_0 = 100\mathbb{I}$. For the priors and hyperparameters on (σ_u, σ_ξ), we follow Chan and Eisenstat (2017) and choose a Gamma prior (instead of inverse-gamma) with hyperparameters $t_{0,u} = t_{0,\xi} = \frac{1}{2}$ and $s_{0,u}, s_{0,\xi} = 0.01$.[10]

3.2.5 Data and Estimation Strategies

The real-time data on all predictors, besides the term-spread, is taken from the "Real-Time Data Research Center" of the Federal Reserve Bank of Philadelphia

[10]The estimation and predictions of the BVAR is based on the code of Chan and Eisenstat (2017).

and was originally compiled by Croushore and Stark (2001).[11]. We define the term spread to be the difference between the *10-year treasury constant maturity rate* and the *3-month treasury bill: secondary market rate*, for which the data is taken from FRED. Our sample, restricted by the length of the term spread, starts in 1962:Q1 and ends in 2017:Q4.

The timing of the vintages is such that the vintage in quarter t contains data for $t-1$. The data of the vintage in quarter t becomes available in the middle of quarter t, i.e. typically at the beginning of February, May, August and November.[12] We want to highlight that for each forecast we use only data that would have been available at that point in time in the past. For the one quarter-ahead horizon, we define that a forecast based on the quarter t vintage is a prediction for quarter t data (which could also be labeled a *nowcast*). The four quarter-ahead forecast is the year-on-year quarterly growth of GDP. Following the same logic as for the one quarter-ahead prediction, the forecast for $t+3$ is based on t, i.e. we use year j, quarter i data to predict the growth up to (and including) year $j+1$, quarter i. Table 3.4 display the predictors, which includes the major macroeconomic indicators and the term spread, and their transformations.

Table 3.4: List of Predictors

Predictor	Transformation
Real GDP	$\Delta\log$
Real Consumption	$\Delta\log$
Real Residential Investment	$\Delta\log$
Real Business Fixed Investment	$\Delta\log$
Unemployment Rate	Δ
Non-Farm Employment	$\Delta\log$
Real Government Spending	$\Delta\log$
GDP Price Deflator	$\Delta\log$
Term Spread	-

NOTE. – The transformation Δ denotes simple differences and $\Delta\log$ denotes log-differences. The Term Spread is not transformed.

Following the notation of Koenig et al. (2003), let $y(t), x^{(k)}(t)$ denote the true value of the predictand and predictor (k). Let $y_s(t), x_s^{(k)}(t)$, with $s \geq t$ denote the *estimate* of the variables by the statistical agency, i.e. the observation available at time s. We use two different strategies to forecast different vintages of the data. To forecast the first-release data, we use what Koenig et al. labeled *strategy 1*, which implies

[11] https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data

[12] See the description on the website of the Philadelphia FED.

using only first-release data for the predictand and the predictor, such that

$$y_t(t) = \alpha + \beta x_t^{(k)}(t) + e_t \tag{3.13}$$

Strategy 2 denotes what is typically used in the literature, i.e. using the latest-vintage available at the forecast origin on the left- and right-hand side.[13]

$$y_P(t) = \alpha + \beta x_P^{(k)}(t) + e_t \tag{3.14}$$

where P denotes forecast origin. We evaluate the forecasts of *strategy 1* and *strategy 2* against the first-release realizations of the predictand. and we evaluate only the forecasts of *strategy 2* against the final-release realizations.

3.3 Empirical Results

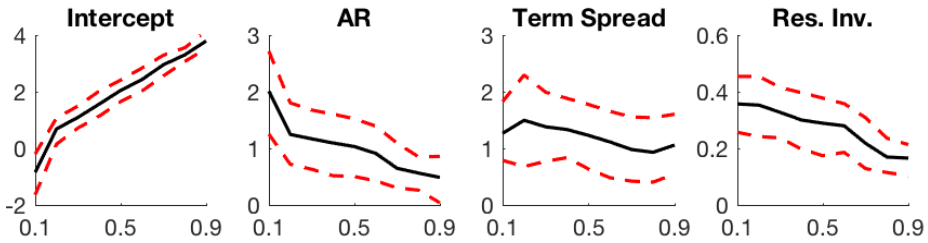
3.3.1 In-Sample Results of the Quantile Regressions

Before we move to the evaluation of the out-of-sample forecasts, we would like to draw the attention to some results of the in-sample estimation. Figure 3.1 displays in-sample parameter estimates at different quantiles at the forecast origin 2013:Q1 and the four-quarter-ahead regression. The upper panel shows results for strategy 1 and the lower panel for strategy 2. Several features of the plot are important. First, there is strong evidence for quantile effects in the regressions. The coefficients of the predictors exhibit a downward slope, which is overall stronger for lower quantiles, in particular for the autoregressive coefficient.[14] The intercept exhibits a kink in the lower quantiles, which is an indicator for skewness in the distribution. Second, the measurement error in the vintages clearly affects in-sample parameter estimation, in particular for the term spread and the autoregressive coefficient. Section 3.3.2 investigates to what extent this affects the forecasting performance. Figure 3.2 shows parameter estimates (strategy 2) at two different forecast origins. The results indicate that the quantile coefficients are not constant over time. The coefficients from the term spread shifted downwards, indicating a weaker relationship between the spread and GDP growth, and the autoregressive coefficients exhibit stronger lower tail dependence - potentially driven by the great recession period.

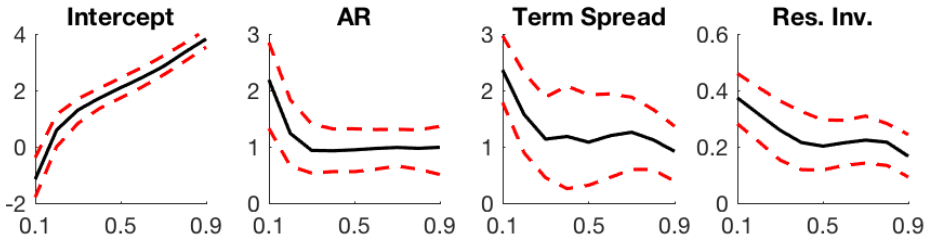
[13]The first-vintage data point for real GDP for 1995:Q4 is missing. We substituted the data point by the reported real GDP for 1995:Q4 in the 1996:Q2 vintage.

[14]Note that an autoregressive coefficient larger than one in absolute value does not imply explosive behavior in this case. The autoregressive coefficient here denotes the coefficient for the last observed quarter-on-quarter growth rate while the left-hand side variable is the year-on-year growth rate.

Figure 3.1: Coefficients



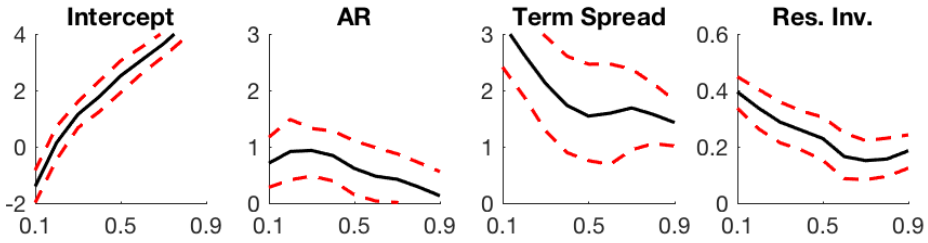
(a) Strategy 1



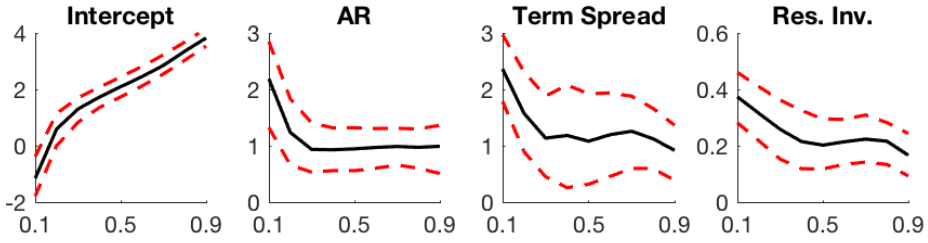
(b) Strategy 2

NOTE. – The dashed line depicts pointwise 90% confidence bands. The solid line displays the mean of the posterior of the respective coefficient for a forecast origin of 2013:Q1 and the four quarter-ahead regression. The x-axis denotes quantiles at which the parameters were estimated.

Figure 3.2: Coefficients



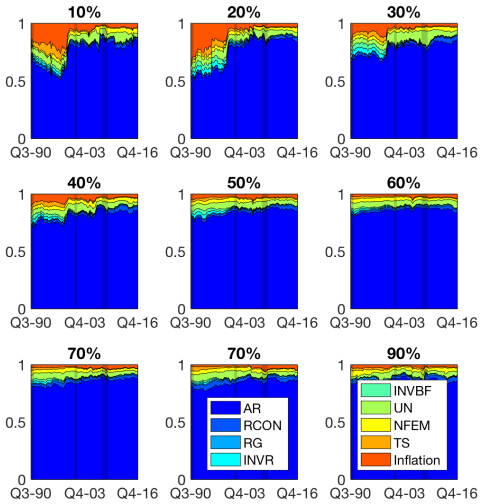
(a) 1995:Q1



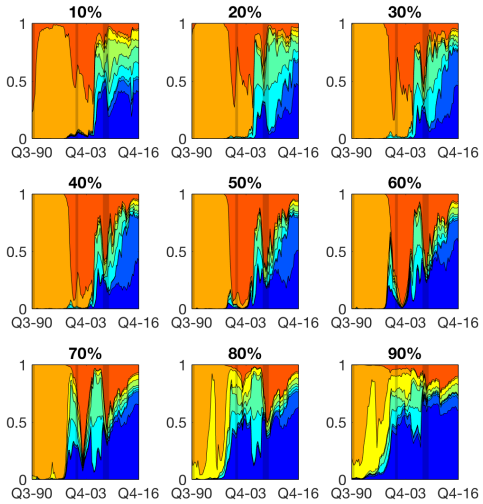
(b) 2013:Q1

NOTE. – The dashed line depicts pointwise 90% confidence bands. The solid line displays the mean of the posterior of the respective coefficient for a forecast origin of 2013:Q1 and the four quarter-ahead regression. The x-axis denotes quantiles at which the parameters were estimated. The dates denote the forecast origin.

Figure 3.3: BMA Weights - Strategy 1



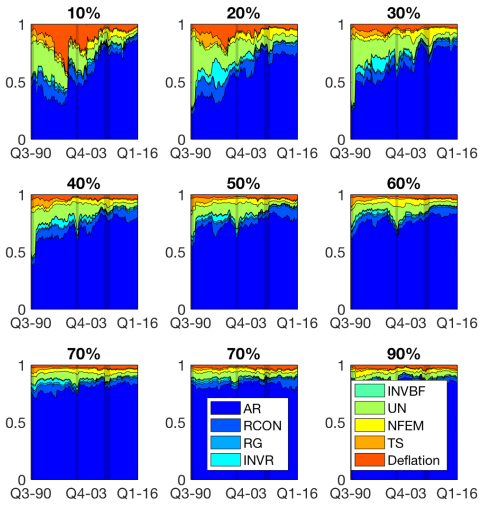
(a) h=1



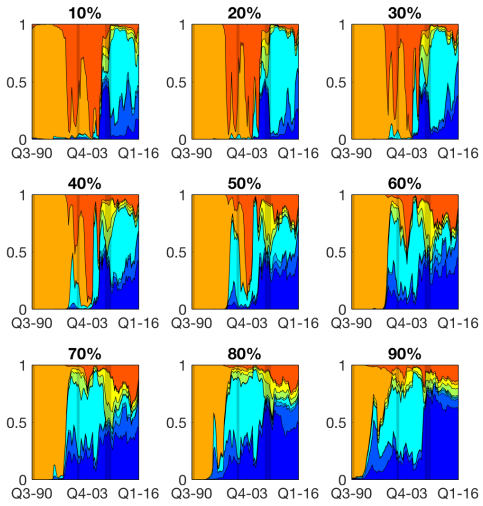
(b) h=4

NOTE. – The picture displays the evolution of the weights obtained through bayesian model averaging. Grey shaded areas depict the NBER recession periods. Figure (a) displays the results for the one quarter-ahead forecasts. Figure (b) displays the results for the four quarter-ahead horizon. The time-axis is origin based, i.e. display the weight used at the time of the origin of the forecast. The color code for (a) and (b) is identical.

Figure 3.4: BMA Weights - Strategy 2



(a) h=1



(b) h=4

NOTE. – The picture displays the evolution of the weights obtained through bayesian model averaging. Grey shaded areas depict the NBER recession periods. Figure (a) displays the results for the one quarter-ahead forecasts. Figure (b) displays the results for the four quarter-ahead horizon. The time-axis is origin based, i.e. display the weight used at the time of the origin of the forecast. The color code for (a) and (b) is identical.

Figure 3.3 and 3.4 show the model weights obtained using (3.5). For the one-quarter-ahead forecast, the simple autoregressive forecast dominates for both, strategy 1 and strategy 2. In contrast, for the four-quarter-ahead forecast, the weights displays considerable time-variation and differences between strategy 1 and strategy 2. For strategies, the term spread model evolved from being the dominant model in the 90s to being only dominant for lower quantiles until mid-2000s, to eventually receiving very small weights. In general, the autoregressive model, without additional predictors received higher weights at all quantiles and in particular for quantiles above the median. Again, the weights differ markedly across the estimation strategies, in particular with regard to the importance of residential investment.

3.3.2 Out-of-Sample Results - First-Release

To evaluate out-of-sample predictions of quantiles, we will use the tick loss function. The function is defined as

$$\mathrm{L}_\tau(\hat{y}, y) = (y - \hat{y})\big(\tau - \mathbb{1}(y < \hat{y})\big) \tag{3.15}$$

where y denotes a realization of the predictand, $\hat{y}$ denotes the forecast, $\mathbb{1}$ denotes the indicator function and τ denotes the quantile to be evaluated. Importantly, in this section, y will be first-release data. Table 3.5 and 3.6 show the results of a forecast comparison based on individual losses obtained from (3.15). For the purpose of interpretation, the numbers display the ratio of the loss of a benchmark model to the loss of the model defined in the respective column. In Panel A, the benchmark model is the bayesian quantile predictions with equal-weighting, $Q^{\text{ew}}_{t+h}(\tau)$ (BQ-EW), and estimated with first-release data only, i.e. strategy 1. In Panel B the benchmark is $Q^{\text{bma}}_{t+h}(\tau)$ (BQ-BMA). The models labeled BQ-EW-EOS and BQ-BMA-EOS denote the bayesian quantile predictions using *end-of-sample* vintages in the estimation (strategy 2). BVAR and SV-EW denote the bayesian VAR and the equal-weighting univariate SV models, all of which are estimated using strategy 1. A ratio below zero implies that the benchmark model is more accurate. A number in boldface denotes statistical significance of a (two-sided) Giacomini and White (2006) test of equal predictive ability at the 10% level.[15] The benchmark model in Table 3.5, Panel A, statistically significantly outperforms the bayesian quantile forecasts using strategy 2 at higher quantiles and obtains a smaller loss for lower quantiles. The evidence for a comparison between bma and equal-weights is

[15]The test is applied to the loss differential, which is defined to be $\Delta\mathrm{L}_\tau(\hat{y}_1, \hat{y}_2, y) = \mathrm{L}_\tau(\hat{y}_1, y) - \mathrm{L}_\tau(\hat{y}_2, y)$ where $\hat{y}_1, \hat{y}_2$ are forecasts from two competing models.

mixed and the quantile models exhibit a better predictive ability than the stochastic volatility models, especially for quantiles above the median. Overall, the results of the one quarter-ahead forecasts speak in favor of the quantile regression. Table 3.6 has the same structure as Table 3.5, but displays results for a horizon of four quarters-ahead. Evidence is more mixed for several quarters ahead. While strategy 1 quantile regressions still outperform strategy 2, the evidence is weaker. In addition, the ratio of the losses relative to the univariate models now point to the direction of a better performance of the stochastic volatility models. We find similar evidence for the bma quantile predictions.

Testing many quantiles individually is prone to a multiple testing problem. We therefore turn to a joint evaluation of the quantile predictions using the continuous ranked probability score (CRPS). Following Gneiting and Ranjan (2011), the CRPS is defined as

$$\begin{aligned} \text{CRPS}(f, y) &= \int_{-\infty}^{\infty} \big(F(z) - \mathbb{1}(y < z)\big)^2 \, \mathrm{d}z \\ &= \int_{-\infty}^{\infty} \mathrm{L}_\tau(F^{-1}(\tau), y) \, \mathrm{d}\tau \end{aligned} \tag{3.16}$$

where f denotes a density forecast, F and F^{-1} the respective cdf and inverse-cdf, y the realization and τ a quantile. The second equality shows that the CRPS is equal to evaluating the tick loss function jointly at all quantiles of the distribution. As our forecasts are restricted to a finite number of quantiles, we use the method described in section 3.2.2 to obtain a density forecast f. Table 3.7 displays the ratios of the CRPS of the benchmark model and the competitors.[16] A boldface number indicates that the results of a (two-sided) Giacomini & White test on equal predictive ability is rejected at the 10% level. The columns indicate the benchmark models BQ-EW and BQ-BMA. Table 3.7 overall confirms the results from Table 3.5 and 3.6. The *strategy 1* quantile regressions somewhat outperform the competitor stochastic volatility models and the *strategy 2* quantile predictions. Overall, the evidence is less strong for horizons of four quarters-ahead, in particular, when using bayesian model averaging as the weighting scheme.

[16]The stochastic volatility models do not provide a closed form expression to evaluate their cdf. Krueger et al. (2017) show in a recent paper that, under very mild conditions, calculating the CRPS using the empirical cdf, based on a large number of draws from the posterior, provides a consistent approximation to the CRPS calculated under the unknown true posterior distribution.

Table 3.5: One quarter-ahead Forecasts - Tick Loss Results

Panel A										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	BQ-EW-EOS	0.89	0.93	0.98	1.00	0.99	0.99	**0.92**	**0.87**	**0.84**
	BQ-BMA	0.98	0.97	1.00	1.00	1.00	1.01	**1.02**	1.01	1.01
	BVAR	0.98	0.96	0.97	0.96	0.92	**0.88**	**0.83**	**0.80**	0.84
	SV-EW	**0.88**	0.93	0.98	1.01	0.99	0.99	**0.92**	**0.88**	**0.84**

Panel B										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	BQ-BMA-EOS	0.92	**0.92**	0.96	1.00	0.99	0.94	**0.88**	**0.84**	**0.82**
	BVAR	1.00	0.98	0.97	0.96	0.93	0.88	**0.81**	**0.80**	0.83
	SV-EW	**0.88**	0.96	0.98	1.01	1.00	0.98	**0.91**	**0.88**	**0.84**

NOTE. – Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential, i.e. the difference between the benchmark and the competing model. For ease of interpretation, the numbers show the ratio of the tick-loss of the benchmark model relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation. The out-of-sample (origin) goes from 1990:Q3 to 2016:Q4, a total of 106 matches of realized observations and forecasts. Panel A displays the results when the BQ-EW model is the benchmark. Panel B displays the results when the BQ-BMA model is the benchmark.

Table 3.6: Four quarter-ahead Forecasts - Tick Loss Results

Panel A										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4	BQ-EW-EOS	**0.93**	1.01	0.99	1.00	1.00	0.97	0.95	0.97	1.01
	BQ-BMA	0.99	0.93	**0.86**	**0.85**	**0.83**	**0.82**	**0.81**	**0.86**	0.96
	BVAR	0.91	0.86	0.82	0.80	**0.78**	**0.77**	**0.76**	**0.75**	0.80
	SV-EW	1.09	1.09	1.03	1.03	1.01	1.00	1.01	1.02	1.10

Panel B										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4	BQ-BMA-EOS	0.96	**0.92**	0.96	0.97	0.97	0.96	1.01	1.03	0.99
	BVAR	0.92	0.93	0.95	0.95	0.95	0.93	0.93	0.87	0.82
	SV-EW	1.1	**1.17**	**1.20**	**1.20**	**1.22**	**1.22**	**1.23**	1.19	1.12

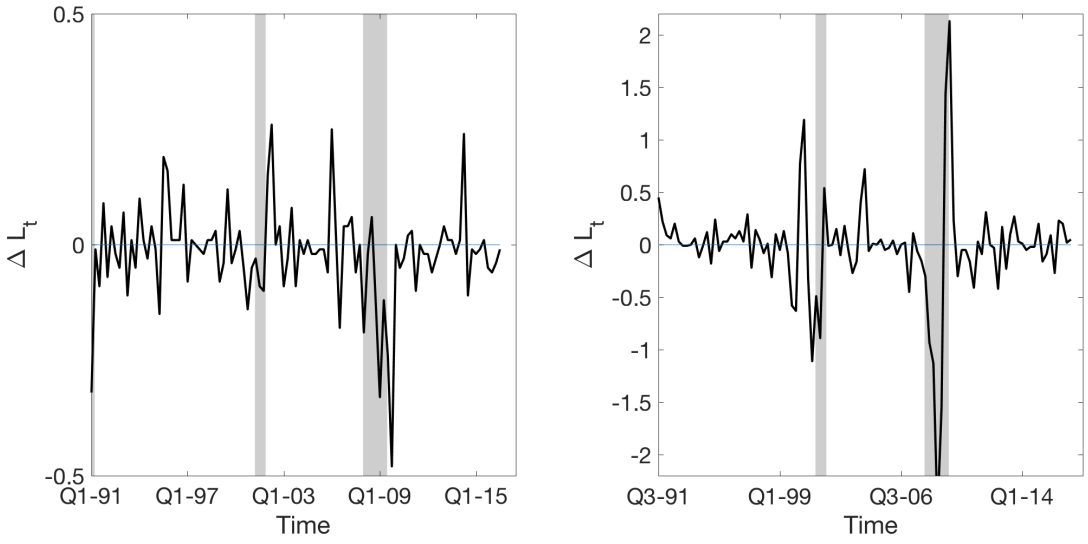
NOTE. – Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential, i.e. the difference between the benchmark and the competing model. For ease of interpretation, the numbers show the ratio of the tick-loss of the benchmark model relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation. The out-of-sample (origin) goes from 1990:Q3 to 2016Q4, a total of 106 matches of realized observations and forecasts. Panel A displays the results when the BQ-EW model is the benchmark. Panel B displays the results when the BQ-BMA model is the benchmark.

Table 3.7: Density Forecast Comparison via CRPS

	h=1		h=4	
Model	BQ-EW	BQ-BMA	BQ-EW	BQ-BMA
BQ-EW-EOS	0.97	**0.97**	0.99	**0.97**
BQ-BMA	1.00	1	**0.89**	1
BVAR	**0.90**	**0.90**	**0.81**	0.91
SV-EW	**0.95**	**0.95**	1.04	**1.16**

NOTE. – The columns indicate the benchmark models BQ-EW and BQ-BMA. Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential. The numbers show the ratio of the CRPS of the benchmark model, relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation.

Figure 3.5: CRPS Differences of Different Quantile Regressions



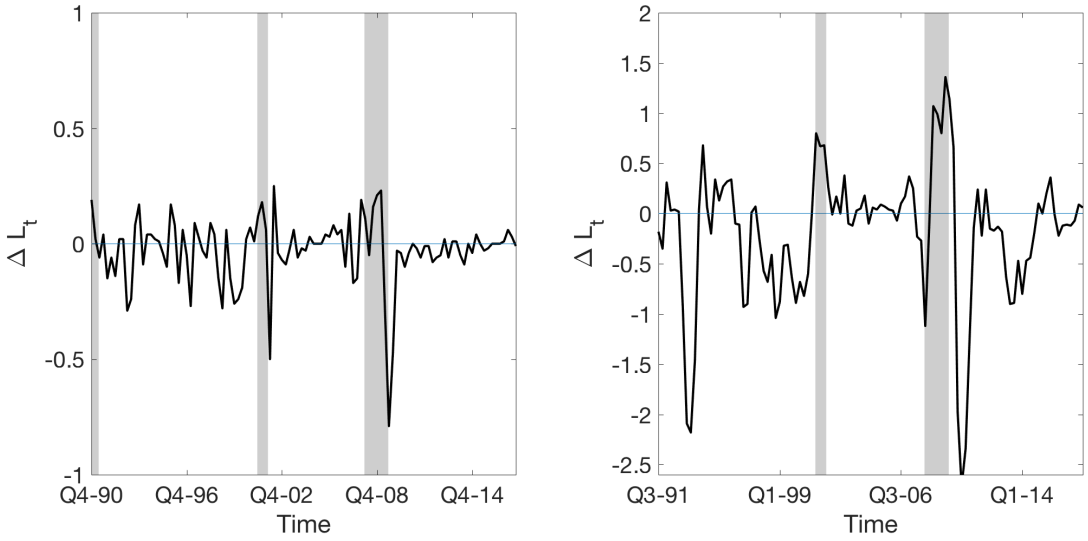
(a) h=1 (b) h=4

NOTE. – h denotes the forecast horizon. The solid line displays the difference of the CRPS of BQ-EW, estimated using strategy 1, and the BQ-EW estimated using strategy 2. The grey shaded areas display NBER recession periods.

Figure 3.5 displays the difference of the CRPS using *strategy 1* versus *strategy 2* of the BQ-EW model, i.e. a negative number indicates a more accurate prediction of *strategy 1*. For both horizons, the difference gets negative before and during the recession periods, indicating that the gain of *strategy 1* might be particularly important during economic downturns and the magnitude is quite large for a horizon of four quarters-ahead. Whereas Figure 3.5 shows the importance of using *real-time* vintages for the estimation, Figure 3.6 displays the difference of the CRPS using BQ-EW (*strategy 1*) versus BVAR (*strategy 1*). Although the full out-of-sample

test indicates superior performance of the quantile regressions, the BVAR seems to outperform the quantile regression during recessions period and more so for longer horizons.

Figure 3.6: CRPS Differences of Quantile Regression to BVAR



(a) h=1 (b) h=4

NOTE. – h denotes the forecast horizon. The solid line displays the difference of the CRPS of BQ-EW, estimated using strategy 1, and the BVAR estimated using strategy 2. The grey shaded areas display NBER recession periods.

3.3.3 Out-of-Sample Results - Final-Release

The difference from this section to section 3.3.2 is that the y to evaluate the loss functions $\mathrm{L}_\tau(\hat{y}, y)$ and the $\mathrm{CRPS}(F, y)$ will be final-release data, i.e. the most accurate estimate of the true value of y. The benchmark model is the equal-weighting quantile regression forecast, estimated using strategy 2. The competitors are the bma quantile model, the BVAR and the equal-weighting univariate stochastic volatility models, all estimated using *strategy 2*. For the one quarter-ahead horizon the BQ-EW model again emerges as the best model as it outperforms the competitors, especially for upper quantiles.

Table 3.8: One quarter-ahead Forecasts - Tick Loss Results

Panel A										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	BQ-BMA	0.98	**0.97**	**0.98**	**0.98**	**0.98**	**0.98**	0.99	0.99	1.00
	BVAR	1.01	0.94	**0.92**	**0.86**	**0.82**	**0.82**	**0.80**	**0.78**	**0.78**
	SV-EW	0.99	1.00	0.99	1.00	1.00	**0.98**	**0.95**	**0.90**	**0.86**

Panel B										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	BVAR	1.02	0.97	0.93	**0.88**	**0.85**	**0.84**	**0.81**	**0.79**	**0.79**
	SV-EW	1.01	1.02	1.02	1.01	1.02	1	0.97	**0.92**	**0.87**

NOTE. – Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential, i.e. the difference between the benchmark and the competing model. For ease of interpretation, the numbers show the ratio of the tick-loss of the benchmark model relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation. The out-of-sample (origin) goes from 1990:Q3 to 2016:Q1, a total of 103 matches of realized observations and forecasts.. Panel A displays the results when the BQ-EW model is the benchmark. Panel B displays the results when the BQ-BMA model is the benchmark.

For the four quarter-ahead horizon the stochastic volatility model tends to outperform both, the BQ-EW and BQ-BMA model. The BVAR again shows a better performance for lower tails than for upper tails.

Table 3.9: Four quarter-ahead Forecasts - Tick Loss Results

Panel A										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4	BQ-BMA	1.05	0.96	0.94	**0.9**	**0.85**	**0.82**	**0.82**	**0.83**	0.94
	BVAR	1.04	0.95	0.93	0.87	0.83	**0.8**	**0.78**	**0.75**	**0.78**
	SV-EW	1.09	1.04	**1.05**	**1.03**	1.02	0.99	0.97	0.93	1.03

Panel B										
		Quantiles								
h	Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4	BVAR	1	0.99	0.98	0.97	0.98	0.97	0.95	0.90	0.84
	SV-EW	1.04	1.08	1.12	**1.15**	**1.20**	**1.21**	**1.19**	1.13	1.1

NOTE. – Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential, i.e. the difference between the benchmark and the competing model. For ease of interpretation, the numbers show the ratio of the tick-loss of the benchmark model relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation. The out-of-sample (origin) goes from 1990:Q3 to 2016:Q1 a total of 103 observations. Panel A displays the results when the BQ-EW model is the benchmark. Panel B displays the results when the BQ-BMA model is the benchmark.

Table 3.10: Density Forecast Comparison via CRPS

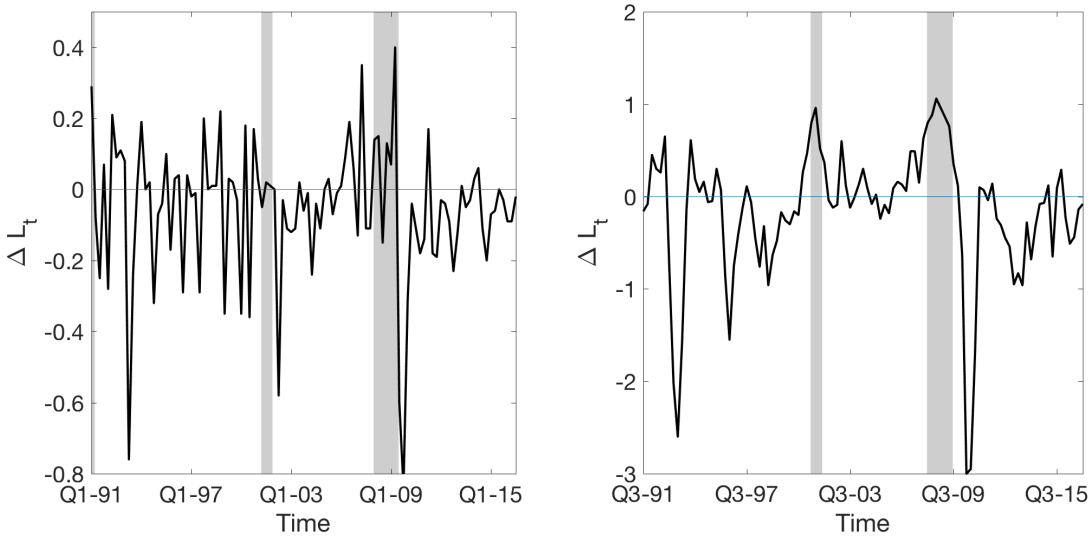
	h=1		h=4	
Model	BQ-EW	BQ-BMA	BQ-EW	BQ-BMA
BQ-BMA	**0.98**	1	**0.87**	1
BVAR	**0.85**	**0.87**	0.86	0.98
SV-EW	**0.98**	1.00	1.02	1.18

NOTE. – Numbers in boldface denote a statistic that is statistically significant different from zero, at the 10% level, when testing the loss differential. The numbers show the ratio of the CRPS of the benchmark model (EW and BMA respectively) relative to the other models. A number smaller/larger than one indicates a superior/inferior performance of the benchmark model. Standard errors are robust to heteroskedasticity and serial correlation.

We calculate again the CRPS to evaluate the quantiles jointly. The BQ-EW model significantly outperforms the competitors at the one-quarter-ahead horizon. For the four quarter-horizon, the BQ-EW model shows a superior performance albeit not statistically significant.

Figure 3.7 plots the difference between the CRPS of the BQ-EW model and the BVAR, where positive (negative) values indicate a superior (inferior) performance of the BVAR. Results are quite similar to the first-release case displayed in figure 3.6. For the four-quarter-ahead horizon, the BVAR forecasts better for recession periods but exhibits a worse performance for the period after the great recession. The evidence points in the same direction for the one-quarter-ahead forecast.

Figure 3.7: CRPS Differences of Quantile Regression to BVAR



(a) h=1 (b) h=4

NOTE. – h denotes the forecast horizon. The solid line displays the difference of the CRPS of BQ-EW, estimated using strategy 1, and the BVAR estimated using strategy 2. Positive (negative) values indicate a superior (inferior) forecast of the BVAR. The grey shaded areas display NBER recession periods.

3.4 Conclusion

We estimated *real-time* quantile and density forecast based on quantile regressions for US real GDP growth. Our findings indicate that pooled quantile regressions are competitive to stochastic volatility models, in particular for shorter horizons. We find that an equal-weighting scheme outperforms bayesian model averaging for four horizons when comparing forecasts within the class of quantile regressions. For a horizon of one quarter, the evidence is mixed. As the bma weights exhibit considerable time-variation and their performances deteriorate with the forecast horizon suggests that results could be improved by a form of dynamic model averaging. If the aim is to predict first-release data the results indicate to use first-release data in the parameter estimation of the models - the gains that we find are particularly big during the beginning of recession periods. Although in general, the quantile regressions tend to do better than the stochastic volatility models, the bayesian VAR tend to forecast better during recessions.

We leave for future work to develop a test statistic that can jointly test the performance of the models at a *finite number* of quantiles because testing different quantiles individually leads to the problem of multiple testing.

Bibliography

Aastveit, K. A., A. K. Anundsen, and E. I. Herstad, "Residential Investment and Recession Predictability," *Working Paper*, 2018.

Abel, J. et al., "The Measurement and Behavior of Uncertainty: Evidence From the ECB Survey of Professional Forecasters," *Journal of Applied Econometrics*, 2016, *31*, 533–550.

Adrian, T., N. Boyarchenko, and D. Giannone, "Vulnerable Growth," *Federal Reserve Bank of New York, Staff Reports*, 2016.

Altissimo, F. et al., "New Eurocoin: Tracking Economic Growth in Real Time," *Review of Economics and Statistics*, 2010, *92*, 1024–1034.

Andrews, D. and W. Ploberger, "Testing for Serial Correlation Against an ARMA(1,1) Process," *Journal of the American Statistical Association*, 1996, *91*, 1331–1342.

Andrews, D. W. K., "An Empirical Process Central Limit Theorem for Dependent Nonidentically Distributed Random Variables," *Journal of Multivariate Analysis*, 1991, *38*, 188–203.

Arellano-Valle, R. B., J. E. Conteras-Reyes, and M. G. Genton, "Shannon Entropy and Mutual Information for Multivariate Skew-Elliptical Distributions," *Scandinavian Journal of Statistics*, 2012, *40*, 42–62.

Arouba, S. B., F. X. Diebold, and C. Scotti, "Real Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 2009, *27*, 417–427.

Azzalini, A. and Antonella Capitano, "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-distribution," *Series B: Statistical Methodology*, 2003, *65*, 367–389.

Bernardi, M., R. Casarin, B. B. Maillet, and L. Petrella, "Dynamic Model Averaging for Bayesian Quantile Regression," *Working Paper*, 2017.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk, "Interconnections Between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model," *Journal of Applied Econometrics*, 2016, *31*, 1352–1370.

Bloom, N., "The Impact of Uncertainty Shocks," *Econometrica*, 2009, *77*, 623–685.

Bullard, J., "The St. Louis Fed's New Characterization of the Outlook for the U.S. Economy," 2016.

Camacho, M., G. Perez Quiros, and P. Poncela, "Green Shots and Double Dips in the Euro Area: A Real Time Measure," *International Journal of Forecasting*, 2014, *30*, 520–535.

Carrasco, M., L. Hu, and W. Ploberger, "Optimal Test for Markov Switching in Parameters," *Econometrica*, 2014, *82*, 765–784.

Carter, A. V. and D. G. Steigerwald, "Markov Regime-Switching Tests: Asymptotic Critical Values," *Journal of Econometric Methods*, 2013, *2*, 25–34.

Chan, J. and A. Grant, "Modelling Energy Price Dynamics: GARCH versus Stochastic Volatility," *Energy Economics*, 2016, *54*, 182–189.

Chan, J. C. C. and E. Eisenstat, "Efficient Estimation of Bayesian VARMAs with time-varying coefficients," *Journal of Applied Econometrics*, 2017, *32*, 1277–1297.

Chauvet, M. and J. Piger, "A Comparison of the Real-Time Performance of Business Cycle Dating Methods," *Journal of Economics & Business Statistics*, 2008, *24*, 42–49.

_ **and S.M. Potter**, "Predicting Recessions Using the Yield Curve," *Journal of Forecasting*, 2005, *24*, 77–103.

Chesher, A., "Testing for Neglected Heterogeneity," *Econometrica*, 1984, *52*, 865–872.

Cho, J. S. and H. White, "Testing for Regime Switching," *Econometrica*, 2007, *75*, 1671–1720.

Clark, T. and K. West, "Using Out-of-sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics*, 2006, *135*, 155–186.

_ **and** _ , "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, 2007, *138*, 291–311.

_ **and M. McCracken**, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 2001, *105*, 85–110.

Clark, T. E., "Real-time density forecasts from bayesian vector autoregressions with stochastic volatility," *Journal of Business & Economic Statistics*, 2011, *29*, 327–341.

_ **and F. Ravazzolo**, "Macroeconomic Forecasting Under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics*, 2015, *30*, 551–575.

Clements, M. P. and A. B. Galvao, "Data Revisions and Real-Time Probabilistic Forecasting of Macroeconomic Variables," *Working Paper*, 2017.

Cover, T. M. and J. A. Thomas, "Elements of Information Theory," *Wiley Series in Telecommunications*, 1991.

Croushore, D., "Forecasting with Real-Time Macroeconomic Data," *Handbook of Economic Forecasting*, 2006, *1*, 961–982.

_ **and T. Stark**, "A real-time data set for macroeconomist," *Journal of Econometrics*, 2001, *105*, 111–130.

_ **and** _ , "A real-time data set for macroeconomist: Does the data vintage matter," *The Review of Economics and Statistics*, 2003, *85*, 605–617.

D'Agostino, A., L. Gambetti, and D. Giannone, "Macroeconomic Forecasting and Structural Change," *Journal of Applied Econometrics*, 2013, *28*, 82–101.

Dahlhaus, T. and T. Sekhposyan, "Monetary Policy Uncertainty: A tale of Two Tails," *Working Paper*, 2018.

D'Amico, S. and A. Orphanides, "Inflation Uncertainty and Disagreement in Bond Risk Premia," *FRB of Chicago Working Paper*, 2014, *24*.

Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 1977, *64*, 247–254.

_ , "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 1987, *74*, 33–43.

Diebold, F. X. and J. A. Lopez, "Forecast Evaluation and combination," in G. S.

Maddala and C. R. Rao, eds., *Handbook of Statistics*, Vol. 14, North-Holland: Amsterdam, 1996.

_ **and R. S. Mariano**, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 1995, *13*, 253–263.

Dovern, J., U. Fritsche, and J. Slacalek, "Disagreement Among Forecasters in G7 Countries," *Review of Economic and Statistics*, 2012, *94*, 1081–1096.

_ **,** _ **, and** _ , "Disagreement among Forecasters in G7 countries," *The Review of Economics and Statistics*, 2012, *94*, 1081–1096.

Faust, J. and J. H. Wright, "Forecasting Inflation," in G. Elliot, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2, North Holland, 2013, chapter 1, pp. 2–56.

Gaglianone, W. P. and L. R. Lima, "Constructing Density Forecasts from Quantile Regressions," *Journal of Money, Credit and Banking*, 2012, *44*, 1589–1607.

García, R., "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 1998, *39*, 763–788.

Genest, C. and Nešlehová, "A Primer on Copulas for Count Data," *Astin Bull.*, 2007, pp. 475–515.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann, "Combining Expert Forecasts: Can Anything Beat the Simple Average," *International Journal of Forecasting*, 2013, *29*, 108–120.

Geweke, N., "Measures of Conditional Linear Dependence and Feedback between Time Series," *Journal of the American Economic Association*, 1984, *79*, 907–915.

Giacomini, R. and B. Rossi, "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics*, 2010, *25*, 595–620.

_ **and H. White**, "Test of Conditional Predictive Ability," *Econometrica*, 2006, *74*, 1545–1578.

Gneiting, T. and R. Ranjan, "Comparing Density Forecasts using Threshold- and Quantile-Weighted Scoring Rules," *Journal of Business and Economic Statistics*, 2011, *29*, 411–422.

Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 1969, *37*, 424–459.

Granger, C.W.J. and P. Newbold, *Forecasting Economic Time Series*, Vol. 2, Academic Press: New York, 1986.

Greenspan, A., "Risk and Uncertainty in Monetary Policy," *American Economic Review*, 2004, *94*, 33–40.

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

_ , "Calling Recessions in Real Time," *International Journal of Forecasting*, 2011, *27*, 1006–1026.

Hansen, B. E., "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometric*, 1992, *7*, 61–82.

_ , "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 1996, *64*, 413–430.

IMF, "World Economic Outlook," *IMF Flagship Report*, 2014.

_ , "World Economic Outlook," *IMF Flagship Report*, 2016.

Jo, S. and R. Sekkel, "Macroeconomic Uncertainty Through the Lens of Professional Forecasters," *Journal of Business & Economics Statistics*, forthcoming.

Joe, H., "Multivariate Models and Dependence Concepts," *Chapman & Hall*, 1997.

Jondeau, E. and M. Rockinger, "The Copula-GARCH model of conditional dependencies: An international stock market application," *Journal of International Money and Finance*, 2006, *25*, 827–853.

Jurado, K., S. C. Ludvigson, and S. Ng, "Measuring Uncertainty," *American Economic Review*, 2015, *105*, 1177–1216.

Kenny, G., T. Kostka, and F. Masera, "Can Macroeconomists Forecast Risk: Event-Based Evidence From the Euro Area SPF," *International Journal Of Central Banking*, 2015, *11*, 1–55.

_ , _ , and _ , "Density Forecast Features and Density Forecast Performance: A panel analysis," *Empirical Economics*, 2015, *48*, 1203–1231.

Kim, S., N. Shepard, and S. Chib, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 1998, *65*, 361–393.

Koenig, E. F., S. Dolmas, and J. Piger, "The Use and Abuse of Real-Time Data In Economic Forecasting," *The Review of Economics and Statistics*, 2003, *85*, 618–628.

Koenker, R. and G. Basset, "Regression Quantiles," *Econometrica*, 1978, *46*, 33–50.

Koop, G. M. and S.M. Potter, "Estimation and Forecasting in Models With Multiple Breaks," *Review of Economic Studies*, 2007, *74*, 763–789.

Korobilis, D., "Quantile Regression Forecasts of Inflation under Model Uncertainty," *International Journal of Forecasting*, 2017, *33*, 11–20.

Kozumi, H. and G. Kobayashi, "Gibbs Sampling Methods for Bayesian Quantile Regression," *Journal of Statistical Computation and Simulation*, 2011, *81*, 1565–1578.

Krueger, F., T. E. Clark, and F. Ravazzolo, "Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts," *Journal of Business & Economic Statistics*, 2017, *35*, 470–485.

Lahiri, K. and X. Sheng, "Measuring Forecast Uncertainty By Disagreement: The Missing Link," *Journal of Applied Econometrics*, 2010, *25*, 514–538.

Manner, H. and O. Reznikova, "A survey on time-varying copulas: specification, simulations and estimation," *Empirical Economics*, 2012, *31*, 654–687.

Manski, C. F., "Measuring Expectations," *Econometrica*, 2004, *72*, 1329–1376.

_ , "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise," *NBER Macroeconomics Annual*, 2017, *12.*

Manzan, S., "Forecasting the Distribution of Economic Variables in a Data-rich Environment," *Journal of Business and Economics Statistics*, 2015, *33*, 144–164.

Mincer, J. and V. Zarnowitz, "The evaluation of economic forecasts," in Mincer JA, ed., *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, National Bureau of Economic Research: New York, 1969.

Negro, M. Del and G. Primiceri, "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum," *The Review of Economic Studies*, 2015, *82*, 1342–1345.

Nelsen, R. B., "An Introduction to Copulas," *Springer Series in Statistics*, 2006.

Oh, D. H. and A. J. Patton, "Modeling Dependence in High Dimensions With Factor Copulas," *Journal of Economics and Business Statistics*, 2017, *35*, 139–154.

Patton, A. J., "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 2006, *47*, 527–556.

Pesaran, H. and M. Weale, "Survey Expectations," in G. Elliot, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1, North Holland, 2006, chapter 14, pp. 715–776.

Pollard, D., "Empirical Processes: Theory and Applications," in "NSF-CBMS Regional Conference Series in Probability and Statistics," Vol. 2, Hayward, CA: Institute for Mathematical Statistics, 1990.

Primiceri, G., "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 2005, *72*, 821–852.

Qu, Z. and F. Zhuo, "Likelihood Ratio Based Tests for Markov Regime Switching," *Working Paper*, 2017.

Rossi, B. and T. Sekhposyan, "Forecast Rationality Tests in the Presence of Instabilities, With Applications to Federal Reserve and Survey Forecasts," *Journal of Applied Econometrics*, 2016, *31*, 507–532.

_ , _ , and M. Soupre, "Understanding the Sources of Macroeconomic Uncertainty," *Working Paper*, 2017.

Sklar, A., "Fonctions de repartition a n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris (in French)*, 1959, *8*, 229–231.

Smith, M. and S. P. Vahey, "Asymmetric Forecast Densities for U.S. Macroeconomic Variables from a Gaussian Copula Model of Cross-Sectional and Serial Dependence," *Journal of Business and Economic Statistics*, 2016, *34*, 416–433.

Stock, J. H. and M. W. Watson, "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 1989, *4*, 351–394.

_ and _ , "A Probability Model of the Coincident Economic Indicators," in K. Lahiri and G. H. Moore, eds., *Leading Economic Indicators: New Approaches and Forecasting Records*, Vol. 1, Cambridge University Press, 1991, chapter 14.

_ and _ , "Estimating Turning Points Using Large Data Sets," *Journal Of Econometrics*, 2014, *178*, 368–381.

Taamouti, A., T. Taoufik, and A. El Ghouch, "Nonparametric Estimation and

Inference for Conditional Density Based Granger Causality Measures," *Journal of Econometrics*, 2014, *180*, 251–264.

Teräsvirta, T., "Forecasting economic variables with non-linear models," in G. Elliot, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1, North Holland, 2006, chapter 8, pp. 413–457.

Timmermann, A., "Forecast Combinations," *Handbook of Economic Forecasting*, 2006, *1*, 135–196.

West, K. D., "Asymptotic Inference about Predictive Ability," *Econometrica*, 1996, *64*, 1067–1084.

_ **and M. W. McCracken**, "Regression-based tests of Predictive Ability," *International Economic Review*, 1998, *39*, 817–840.

White, H., "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 1982, *50*, 1–25.

Wiener, N., in Beckenback. E. F., ed., *The Theory of Prediction*, Vol. 1, McGraw-Hill, 1956, chapter 8.

Wright, J. H., "Forecasting US Inflation by Bayesian Model Averaging," *Journal of Forecasting*, 2009, *28*, 131–144.

Yu, K. and R. A. Moyeed, "Bayesian Quantile Regression," *Statistics and Probability Letters*, 2001, *54*, 437–447.