ESSAYS IN TIME SERIES ECONOMETRICS

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
doctor rerum politicarum
(Dr. rer. pol.)

am Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Elias R. A. Wolf, M.Sc.

Berlin, Mai 2023

Dekanin:	Prof. Dr. Natalia Kliewer
Erstgutachter:	Prof. Dr. Dieter Nautz Freie Universität Berlin
Zweitgutachter:	Prof. Frank Schorfheide (Ph.D.) University of Pennsylvania
Tag der Disputation:	16. Juni 2023

Acknowledgements

In the course of writing this dissertation I have benefited from many people. First and foremost, I want to express my gratitude to my supervisors, Dieter Nautz and Frank Schorfheide for their excellent advice and support during my dissertation. Working as a research assistant at the Chair of Econometrics and visiting the University of Pennsylvania as a PhD student have been invaluable experiences. I am grateful to both of them for their many insightful remarks, stimulating conversations and for their assistance at the different stages of my dissertation.

Second, I want to thank Carlos Montes-Galdón for his great mentorship during my time at the ECB and beyond. I am also immensely grateful to my other co-authors: Yves Schüler, Frieder Mokinski, Till Strohsal and Joan Paredes. Collaborating with them on the different projects of this thesis has greatly helped me to develop my research and writing skills.

I am thankful to Lars Winkelmann for the numerous successful courses we co-taught over the past 4 years and for his feedback on my projects. I also like to thank Helmut Lütkepohl for being a member of my PhD committee and for his valuable advice during the PhD seminars. I appreciate Isabella's excellent organizational assistance. I want to thank Catalina, Max, Winnie and Lea, my fellow and former PhD colleagues at the FU Berlin for many inspiring discussions. Likewise, I want to thank my fellow trainees at the ECB Astrid, Alice, Antoine, Camille, Marie, Pepe and Redwan for many lunches, coffees and the occasional after work drink. I also want to thank Min, Kathleen, Leon, Marco, Aaron, Priyanka, Joao, Yoshiki and Boyuan who welcomed me into their group and made my stay at Penn a great experience. Despite the demanding and at times frustrating aspects of writing a PhD thesis, all of these connections made the last 4 years an enjoyable and exciting time.

Furthermore, I want to thank my family, especially my parents and brothers for giving me the ambition, persistence and most importantly the passion to complete this long and often stressful challenge.

However, the most important person I want to thank is my wife Lea, who has always supported me with her love, patience and suggestions, despite the toll this dissertation took on countless weeknights, weekends, and vacations.

Berlin, Mai 2023 *Elias Wolf*

Declaration of Co–Authorship

This dissertation consists of four (working) papers. One paper was written in collaboration with one co-author and two papers were written in collaboration with two co-authors. The contribution in conception, implementation and drafting for the four chapters can be summarized as follows:

1. On adjusting the one-sided HP filter
   with Yves Schüler and Frieder Mokinski
   Contribution: 33.$\bar{3}$%

2. Data Revisions to German National Accounts: Are Initial Releases Good Nowcasts?
   with Till Strohsal
   Contribution: 50%

3. Estimating Growth at Risk with Skewed Stochastic Volatility Models
   Contribution: 100%

4. Conditional Density Forecasting: A Tempered Importance Sampling Approach
   with Carlos Montes-Galdón and Joan Paredes
   Contribution: 33.$\bar{3}$%

Contents

List of Figures

List of Tables

Introduction and Overview

Life is rich with moments of uncertainty about the future. In economics, predicting the future state of the economy is a key issue for the decision making of economic agents and policy makers alike. On the one hand, firms and households require information about future economic conditions to determine their savings and investment decisions. On the other hand, accurate forecasts of key macroeconomic variables such as output, inflation or unemployment rates are important determinants for policy makers such as central banks, international organizations or fiscal authorities. By providing a "language to model stochastic dynamics" (Diebold 2019), time series econometrics is a natural workhorse to tackle the task of economic forecasting and the cornerstone of this dissertation.

Economic forecasting frequently goes beyond predicting the future values of economic variables to additionally forecast the uncertainty and risk around their expected future paths in order to obtain an even more complete picture of the future. All of these moments are effectively captured by a predictive density. Therefore, density forecasts have developed into an essential area of economic forecasting. The Bank of England's regular publishing of density forecasts for inflation is a notable example (Clements 2004). Additionally, the seminal article of Tobias Adrian, Nina Boryaschenko and Domenico Giannone (2019) has renewed attention on the significance of density forecasting for policy makers with their recent study on asymmetric risks to future US GDP growth.

Yet, while the importance of prediction has led to a plethora of methods for economic forecasting (see for example Elliott and Timmermann (2013) or Clements and Hendry (2011)) it turns out that knowing the current state of the economy poses an equally important problem to policy makers. As shown for example in Orphanides (2001) or Croushore (2011), data revisions due to measurement errors, incomplete information or time lags pose a serious challenge to policy decision making. This has motivated a large literature on so called "Nowcasting" that analyses the nature of revisions and aims to predict the revision making process to macroeconomic variables (Clements and Galvão 2019). Closely tied to the significance of nowcasting is also the real-time assessment of economic fluctuations, such

as business or financial cycles. This is crucial for assessing the status of the economy and accurately determine macroprudential and economic policies. However, as demonstrated by Orphanides and van Norden (2002) or Edge and Meisenzahl (2011) problems to accurately measure econometric fluctuations in real time go beyond data revisions and are also related to the time series methods that are commonly applied by researchers and policy makers.

With the COVID pandemic and the Russian invasion of Ukraine putting a sudden end to a long period of economic stability, econometric methods to accurately measure the state of our economy as well as to predict and quantify the future distributions of key macroeconomic variables play an important role in the toolkit of national and international economic policy institutions and statistical agencies. Therefore, this thesis aims to apply and develop new time series methods to determine the current state of the economy in real time as well as to predict the future path of the economy together with its associated risks in form of density forecasts. Following a chronological order (along several dimensions), the first two chapters focus on the analysis of the present state of the economy, while chapter three and four seek to provide new methods in the field of density forecasting. Another overarching topic of the last two chapters is the focus on modelling and predicting time variation in the higher moments of the forecasting distribution such as the volatility and skewness, a subject that has recently gained high popularity amongst academics and policy makers alike.
From a methodological point of view, Chapter 1 and 2 rely on well established time series methods using spectral analysis and linear state space methods in a frequentist framework. Chapter 3 and 4 move to more recent methods using state of the art particle filtering and Sequential Monte Carlo methods developed by Herbst and Schorfheide (2014) and Herbst and Schorfheide (2019) to capture non-linear and non-Gaussian behavior of economic time-series in a Bayesian estimation framework.

The first chapter, which is joint work with Yves Schüler and Frieder Mokinski, investigates the properties of the one-sided Hodrick-Prescott filter (HP-1s) a popular tool to measure the business cycle in real time. The two-sided Hodrick-Prescott filter (HP-2s) is a symmetric moving average and has been heavily criticized for its use of future observations to extract the cyclical component of time series. This leads to revisions and spurious correlations and makes HP-2s unsuited for policy makers who need to operate in real time (Hamilton 2018). HP-1s uses only observations up to the present. Using methods from spectral analysis, we show that one should not use the one-sided Hodrick-Prescott filter as the real-time

version of the two-sided Hodrick-Prescott filter: First, in terms of the extracted cyclical component, HP-1s fails to remove low-frequency fluctuations to the same extent as HP-2s. Second, HP-1s dampens fluctuations at all frequencies – even those it is meant to extract. As a remedy, we propose two adjustments to HP-1s, aligning its properties closely with those of HP-2s: (1) a lower value for the smoothing parameter and (2) a multiplicative rescaling of the extracted cyclical component. For example, for HP-2s with $\lambda = 1{,}600$, the value commonly used to extract business cycles (Hodrick and Prescott 1997), the adjusted one-sided HP filter uses $\lambda^* = 650$ and rescales the extracted cyclical component by a factor of 1.1513. We illustrate the relevance of these adjustments using simulated and empirical data for different countries. For instance, financial cycles may appear to be 70% more volatile than business cycles, where in fact volatilities differ only marginally.

The second chapter is joint work with Till Strohsal and aims to nowcast German macroeconomic variables using vintages for quarterly and monthly data provided by the Bundesbank. While well-behaved revisions should be unbiased, small and unpredictable (Aruoba 2008), we find that revisions to German national accounts are biased, large and predictable. Moreover, using the different filtering techniques presented in Kishor and Koenig (2012) to process data subject to revisions, we find that the real-time forecasting performance of initial releases can be increased by up to 23%. For total real GDP growth, however, the initial release is an optimal forecast. Yet, given the results we find for disaggregated variables, we conclude that the averaging-out of biases and inefficiencies at the aggregate GDP level appears to be good luck rather than good forecasting.

Moving from nowcasting to forecasting, the third chapter proposes a Skewed Stochastic Volatility (SSV) model to estimate asymmetric macroeconomic tail risks in the spirit of Adrian et al. (2019) seminal paper "Vulnerable Growth". In contrary to their semi-parametric approach, the SSV model captures the evolution of the conditional density of future US GDP growth in a parametric, non-linear, non-Gaussian state space model. This allows to statistically test the effect of exogenous variables on the different moments of the conditional distribution and provides a law of motion to predict future values of volatility and skewness. The model is a non-linear and non-Guassian state space model that can be estimated using a particle MCMC algorithm (Kim et al. 1998). To increase the estimation accuracy of the algorithm, I use the tempered particle filter introduced by Herbst and Schorfheide (2019). Given the asymmetric distribution of the measurement error, I modify the filter to take the time-varying volatility and asymmetry of the densities into account. This reduces the

number of tempering steps. Estimating the model on the same data as Adrian et al. (2019), I find that financial conditions affect the mean, variance and skewness of the conditional distribution of future US GDP growth. Adding to the debate about the importance of time-varying asymmetries in US GDP growth rates (see Hasenzagl et al. (2020)), I find that with a Bayes ratio of 1612.18, the Skewed Stochastic Volatility model is strongly favored by the data over a symmetric Stochastic Volatility model.

The fourth chapter is joint work with Carlos Montes-Galdón and Joan Paredes. We propose a new and robust methodology to obtain conditional density forecasts, based on information not contained in an initial econometric model. The methodology allows to condition on expected marginal densities for a selection of variables in the model, rather than just on future paths as it is usually done in the conditional forecasting literature (see for example in Waggoner and Zha (1999) or Bańbura et al. (2015)). The proposed algorithm, is based on tempered importance sampling introduced into the econometric literature by Herbst and Schorfheide (2014) and provides a robust alternative the the entropic tilting methodology of Robertson et al. (2005). As an example, this paper shows how to implement the algorithm by conditioning the forecasting densities of a BVAR and a DSGE model on information about the marginal densities of future oil prices. We obtain this information from implied probability functions of oil price options with different maturities (see for example (Breeden and Litzenberger 1978)). Most importantly, these options implied densities indicate that the distribution of future oil prices is positively skewed at different horizons. Our results show that increased asymmetric upside risks to oil prices result in upside risks to inflation as well as higher core-inflation over the considered forecasting horizon. Additionally, a real-time forecasting exercise yields that introducing market-based information on the oil price improves inflation and GDP forecasts during crises times such as the COVID pandemic. These results are in line with the findings of the third chapter, where results indicate that asymmetries seem to matter especially in times of economic turmoil.

Chapter 1

On Adjusting the One-Sided Hodrick-Prescott Filter*

with Yves Schüler and Frieder Mokinski

1.1 Introduction

The one-sided Hodrick Prescott filter (HP-1s) Hodrick and Prescott (1981) is commonly used as the real-time version of the regular, two-sided HP filter (HP-2s), a popular tool for detrending macroeconomic time series. As a "one-sided" or "real-time" filter, HP-1s uses only observations dated t and earlier to filter the time series observation y_t.
This makes HP-1s appealing for policy making, where decisions are made in real time and revisions of past estimates are undesirable. The Basel III regulations are a prominent example. These regulations recommend the use of HP-1s to calculate the so-called credit-to-GDP gap (see Basel Committee on Banking Supervision (2010)). The purpose of the credit-to-GDP gap is to inform the calibration of the countercyclical capital buffer, a cyclical capital requirement for banks. National authorities rely on this gap to inform the setting of their countercyclical capital buffer (CCyB) rates. HP-1s also has broader applications beyond Basel III regulations, including alternative credit-to-GDP gaps based on different credit data (e.g. Detken et al. (2014); Basset et al. (2015); Tente et al. (2015)), house price gaps for risk assessments recommended by the European Systemic Risk Board (European

*This chapter has benefited from valuable comments and suggestions by Robert De Jong, Alessandra Luati, Elmar Mertens, Christian Schumacher, Lars Winkelmann, Dieter Nautz and Helmut Lütkepohl as well as participants of the Deutsche Bundesbank research seminar, the EEA 2020 and CMES 2021. An earlier version of this chapter is published as Bundesbank Discussion Paper 11/2020.

Systemic Risk Board (2019)), and funding gaps or other types of gaps (e.g. Anundson et al. (2016); Giordani et al. (2017); Lang et al. (2019)). Additionally, HP-1s is often used as a benchmark for real-time output gap estimates (e.g. Coibion et al. (2018); EU Independent Fiscal Institutions (2020); Furlanetto et al. (2023)).

In this paper, we demonstrate that HP-1s should not be used as a real time version of HP-2s due to important differences in the way HP-1s and HP-2s extract a cyclical component. Through frequency domain analysis, we identify two important deviations of HP-1s from HP-2s that have implications for its practical use. First, HP-1s fails to eliminate low-frequency fluctuations to the same extent as HP-2s, resulting in a cyclical component that is contaminated with more low-frequency fluctuations. Consequently, when using the same smoothing parameter, HP-1s extracts longer-duration cycles than HP-2s. Our finding contradicts the widely held belief that HP-1s with $\lambda = 400,000$ (as specified in Basel III) extracts fluctuations four times the duration of business cycles, and HP-1s with $\lambda = 1,600$ extracts business cycle frequencies, just like HP-2s (e.g. Drehmann et al. (2010); Kauko and Tölö (2020); Jylhä and Lof (2022); Drehmann and Alfaro (2023)). Second, we find that HP-1s attenuates the higher frequencies, and therefore the variance of the cyclical component, with the degree of attenuation depending on the chosen value of the smoothing parameter. This differs from HP-2s, which does not distort the variance of the cyclical component for the frequencies one aims to extract. Implicitly assuming that HP-1s behaves similar to HP-2s, studies compare HP-1s gaps estimated with different smoothing parameters against similar thresholds, such as the one used for the activation of the CCyB (credit-to-GDP gap $>$ 2 p.p.; e.g. Borio and Lowe (2002); Drehmann et al. (2010); Galan (2019)). Our finding suggests that this practice in incorrect. Therefore, we argue that HP-1s should not be use as the one-sided analogue of HP-2s, contrary to common practice (see also Stock and Watson (1999); Orphanides and van Norden (2002); Christiano and Fitzgerald (2003); Edge and Meisenzahl (2011); Hamilton (2018); Kamber et al. (2018); Quast and Wolters (2022)). Ultimately, our results highlight that differences between cyclical components estimated with HP-1s and HP-2s may be due to deviations of HP-1s from HP-2s, beyond those attributable to the use of a one-sided filter.[1]

To address this issue, we propose an adjusted one-sided HP filter (HP-1s*). HP-1s* makes two simple adjustments to HP-1s that align its properties more closely with those of HP-2s: (1) a lower value for the smoothing parameter and (2) a multiplicative rescaling of the cyclical component. For instance, consider HP-2s with λ = 1,600 (6.25; 400,000). In this

[1]Unlike two-sided filters, one-sided filters shift phases, altering the timing relationships of fluctuations at different frequencies.

case, HP-1s* uses $\lambda^* = 650$ (2.45; 163,101) as a smoothing parameter and rescales the cyclical component by a factor of 1.1513 (1.7962; 1.0360).
Using simulated and empirical data, we show the relevance of these adjustments. For instance, given discussions on volatile financial cycles (e.g. Borio (2014)), we show that financial cycles may appear to be 72% more volatile than business cycles, where in fact volatilities are only marginally different.

Our paper is specifically targeted at researchers who use the Hodrick-Prescott (HP) filter and assume that HP-1s, the real-time version of HP-2s behaves like HP-2s. This is because we align the properties of HP-1s* closely to those of HP-2s by minimizing the squared distance of HP-1s' power transfer function (PTF) with the PTF of HP-2s. It is essential to emphasize these differences because many researchers believe that HP filtering is misguided due to spurious cycles, ad hoc filter settings, and end-of-sample biases. This view is echoed in the literature, with previous critiques by researchers such as Harvey and Jaeger (1993), Cogley and Nason (1995), Orphanides and van Norden (2002), and more recently Hamilton (2018). Despite these criticisms, the HP filter continues to be widely used, as also acknowledged by Hamilton (2018), and several recent studies suggest that it outperforms other filters for a wide range of data generating processes (DGPs; e.g. Hodrick (2020); Jönsson (2020); Canova (2021); Schüler (2021); Phillips and Shi (2021)). Therefore, our aim is to provide the adjusted HP-1s to researchers who actively use the HP filter in real-time settings, without attempting to convince critics of the HP filter otherwise.
The first chapter is structured as follows: Sections 1.2 and 1.3 provide a theoretical overview of HP Filter and the frequency domain methods we use for our analysis. The differences between HP-2s and HP-1s as well as our adjustment are explained in sections 1.4 and 1.5 respectively. Sections 1.5.1 and 1.5.2 provides simulation results and our application. Section 1.6 concludes.

1.2 The Hodrick-Prescott Filter

The HP filter decomposes the time series $y_{1:T} = (y_1, \ldots, y_T)'$ into a cyclical component $\psi_{1:T} = (\psi_1, \ldots, \psi_T)'$ and a trend component $\tau_{1:T} = (\tau_1, \ldots, \tau_T)'$:

$$y_t = \tau_t + \psi_t, \tag{1.1}$$

where $t = 1, \ldots, T$ and T denotes sample size.

The regular, two-sided filter: The *two-sided* HP filter (HP-2s) estimates the trend component by solving the following minimization problem:

$$\left\{\widehat{\tau}_{1|T,\lambda},\ldots,\widehat{\tau}_{T|T,\lambda}\right\} = \underset{\tau_1,\ldots,\tau_T}{\arg\min}\left(\sum_{s=1}^{T}(y_s-\tau_s)^2+\lambda\sum_{s=2}^{T-1}(\tau_{s+1}-2\tau_s+\tau_{s-1})^2\right), \tag{1.2}$$

where λ controls the smoothness of the estimated trend component $\widehat{\tau}_{t|T,\lambda}$: The higher its value, the smoother the extracted trend component will be. $\lambda = 1{,}600$ is a common choice to extract business cycle fluctuations in quarterly data.
Notice that we have denoted the estimate of the trend component for period t by $\widehat{\tau}_{t|T,\lambda}$ to illustrate that it depends on the full sample of data $(1,\ldots,T)$ and on the choice of the smoothing parameter λ, see Equation (1.2). Accordingly, $\widehat{\psi}_{t|T,\lambda}$ is the estimate of the cyclical component obtained as $\widehat{\psi}_{t|T,\lambda} = y_t - \widehat{\tau}_{t|T,\lambda}$.
As demonstrated by Hamilton (2018), for instance, the dependency on the full sample leads to spurious correlations in the resulting trend and cycle components (spurious predictability) and also to revisions in $\widehat{\tau}_{t|T,\lambda}$ and $\widehat{\psi}_{t|T,\lambda}$ once new data become available (end-point bias). Since both facts pose problems when it comes to real-time applications, the *one-sided* HP-Filter (HP-1s) is commonly used as the real-time version of the HP filter.

The one-sided filter: The idea of HP-1s is to decompose y_t into trend (τ_t) and cycle (ψ_t) components based only on observations dated t and earlier (and not beyond t, as with HP-2s). To stress this idea, we denote the corresponding estimates by $\widehat{\tau}_{t|t,\lambda}$ and $\widehat{\psi}_{t|t,\lambda}$. The trend component is extracted by solving the following expression for all values of $t > 2$:

$$\widehat{\tau}_{t|t,\lambda} = \underset{\tau_t}{\arg\min}\left(\min_{\tau_1,\ldots,\tau_{t-1}}\left(\sum_{s=1}^{t}(y_s-\tau_s)^2+\lambda\sum_{s=2}^{t-1}(\tau_{s+1}-2\tau_s+\tau_{s-1})^2\right)\right). \tag{1.3}$$

This procedure is equivalent to applying HP-2s recursively on an expanding sample and keeping, from each recursion step, only the estimate of the trend component for the latest period. Analogously to HP-2s, the cyclical component is obtained as $\widehat{\psi}_{t|t,\lambda} = y_t - \widehat{\tau}_{t|t,\lambda}$. This means that, by construction, HP-1s does not suffer from spurious predictability or end-point biases.

1.2.1 Hodrick-Prescott Filter as a Linear Moving Average

Both HP-1s and HP-2s are linear filters. This means that one can express the trend components and the cyclical components as weighted averages of the data (see, for instance,

Danthine and Girardin (1989); De Jong and Sakarya (2016); Cornea-Madeira (2017); Hamilton (2018); Phillips and Jin (2021)). Specifically, the trend components can be expressed as

$$\text{HP-1s:} \quad \widehat{\tau}_{t|t,\lambda} = \sum_{s=1}^{t} w_{t|t,s,\lambda} \cdot y_s \quad = W_{t|t,\lambda}(L) \cdot y_t, \tag{1.4}$$

$$\text{HP-2s:} \quad \widehat{\tau}_{t|T,\lambda} = \sum_{s=1}^{T} w_{t|T,s,\lambda} \cdot y_s \quad = W_{t|T,\lambda}(L) \cdot y_t, \tag{1.5}$$

where L is the lag operator with $L^k \tau_s = \tau_{s-k}$. $W_{t|t,\lambda}(L) = \sum_{s=1}^{t} w_{t|t,s,\lambda} L^{t-s}$ and $W_{t|T,\lambda}(L) = \sum_{s=1}^{T} w_{t|T,s,\lambda} L^{t-s}$ are linear filter polynomials. Analogously, the cyclical components are

$$\text{HP-1s:} \quad \widehat{\psi}_{t|t,\lambda} = y_t - \sum_{s=1}^{t} w_{t|t,s,\lambda} \cdot y_s = (1 - W_{t|t,\lambda}(L)) \cdot y_t \quad = \overline{W}_{t|t,\lambda}(L) \cdot y_t, \tag{1.6}$$

$$\text{HP-2s:} \quad \widehat{\psi}_{t|T,\lambda} = y_t - \sum_{s=1}^{T} w_{t|T,s,\lambda} \cdot y_s = (1 - W_{t|T,\lambda}(L)) \cdot y_t \quad = \overline{W}_{t|T,\lambda}(L) \cdot y_t. \tag{1.7}$$

The notation of the weights ($w_{t|t,s,\lambda}$, $w_{t|T,s,\lambda}$) illustrates that they depend on the observation t to be filtered, the sample length t (or T in the two-sided case), the position of the weighted observation in the sample s, and the value of λ.
In this paper, we use these filter polynomials to analyse the frequency domain properties of HP-1s and HP-2s. Table 1.1 summarizes them as a brief overview.

Table 1.1: Notation for the Filter Polynomials of the Hodrick-Prescott Filter

Filter	Component	Filter polynomial
HP-1s	Trend ($\widehat{\tau}_{t|t,\lambda}$)	$W_{t|t,\lambda}(L)$
HP-1s	Cycle ($\widehat{\psi}_{t|t,\lambda}$)	$\overline{W}_{t|t,\lambda}(L)$
HP-2s	Trend ($\widehat{\tau}_{t|T,\lambda}$)	$W_{t|T,\lambda}(L)$
HP-2s	Cycle ($\widehat{\psi}_{t|T,\lambda}$)	$\overline{W}_{t|T,\lambda}(L)$

Specifically, we consider the closed form expressions for the HP-2s filter polynomials for the cyclical component provided by Phillips and Jin (2021) and King and Rebelo (1993). Therefore, the filter polynomial of HP-1s for the cyclical component is given by

$$\overline{W}_{t|t,\lambda}(L) = e_t' Q_t' \left(Q_t Q_t' + \frac{1}{\lambda} I_{t-2} \right)^{-1} Q_t \eta_t, \tag{1.8}$$

for $t = 3, \ldots, T$ and where $\eta_t = (L^{t-1}, \ldots, L^1, 1)'$, $e_t = (0, \ldots, 0, 1)'$ is a t-dimensional column vector, I_{t-2} is an identity matrix with size $(t-2) \times (t-2)$, and Q_t is second order

differencing matrix with size $(t-2) \times t$. Specifically,

$$Q_t = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}. \tag{1.9}$$

A derivation of the weights is given in Appendix 1.A. In this paper, we use a sample size of $t = T = 1{,}000$ to avoid problems related to the small sample properties of the filter weights (Schüler (2020)). We ensure that increasing the sample size, does not alter the frequency domain properties of the filter polynomial for given value of the smoothing parameter.
For the cyclical component of HP-2s, we use the large sample result for the filter polynomial given in King and Rebelo (1993), i.e.,

$$\overline{W}_{t|T,\lambda}(L) := \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}. \tag{1.10}$$

This assumes that $T \to \infty$ and t being far away from the beginning and the end of the sample.

1.3 Filtering from a Frequency-Domain Perspective: The Power Transfer Function

We use the concept of the *power transfer function (PTF)* to study the cyclical properties of HP-1s and HP-2s (similar to Baxter and King (1999) or Christiano and Fitzgerald (2003)). PTFs allow us to analyze the extent to which different variants of the HP filter succeed in eliminating lower frequencies and preserving higher frequencies. For the HP filters, PTFs summarize the cyclical properties of the filter independently of any data generating process (DGP). This is because the filter weights do not depend on the data that is filtered.[2]

Assume that $\mathcal{W}(L)$ is a linear filter polynomial with $\mathcal{W}(L) = \sum_{j=-\infty}^{\infty} w_j L^j$ and absolutely summable polynomial coefficients $\sum_{j=-\infty}^{\infty} |w_j| < \infty$. The PTF of the linear filter polynomial

[2]A counterexample for this is the Hamilton (2018) filter. The PTF of this filter varies with the DGP of the series to be filtered (see Schüler (2021)).

$\mathcal{W}(L)$ is defined as

$$PTF_{\mathcal{W}}(\omega) = |\mathcal{W}(e^{-\mathrm{i}\omega})|^2, \tag{1.11}$$

with $\mathrm{i}^2 = -1$. $PTF_{\mathcal{W}}(\omega)$ is a non-negative and real-valued scalar function that measures how $\mathcal{W}(L)$ dampens ($PTF_{\mathcal{W}}(\omega) < 1$), passes ($PTF_{\mathcal{W}}(\omega) = 1$), or amplifies ($PTF_{\mathcal{W}}(\omega) > 1$) movements at specific frequencies ω of the time series the researcher aims to filter.
To illustrate the interpretation of $PTF_{\mathcal{W}}(\omega)$, assume y_t is a stationary stochastic process with autocovariances $\gamma_k = \mathrm{Cov}(y_t, y_{t-k})$ and define the autocovariance-generating function of y_t as $g_y(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$, where z denotes a complex scalar and $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$. Evaluating the autocovariance-generating function at $z = e^{-\mathrm{i}\omega}$ and dividing by 2π yields the spectral density of y_t:

$$S_y(\omega) = \frac{1}{2\pi} g_y(e^{-\mathrm{i}\omega}). \tag{1.12}$$

Integrating the spectral density over the interval $[-\pi, \pi]$ gives the variance of y_t, i.e.

$$\mathrm{Var}(y_t) = \int_{-\pi}^{\pi} S_y(\omega)\mathrm{d}\omega = 2\int_{0}^{\pi} S_y(\omega)\mathrm{d}\omega,$$

where we can interpret the value of $\omega \in [0, \pi]$ as a cycle frequency measured in radians. This suggests that we can decompose the variance of y_t into portions related to movements at different frequencies. For instance, integrating the spectral density over the interval $[0, \omega_1]$ with $\omega_1 < \pi$, i.e. $2\int_0^{\omega_1} S_y(\omega)\mathrm{d}\omega$, would give the portion of variance related to movements at frequencies less than or equal to ω_1.
From this, it is possible to show that the spectral densities of y_t and the filtered series $x_t = \mathcal{W}(L)y_t$ are related through

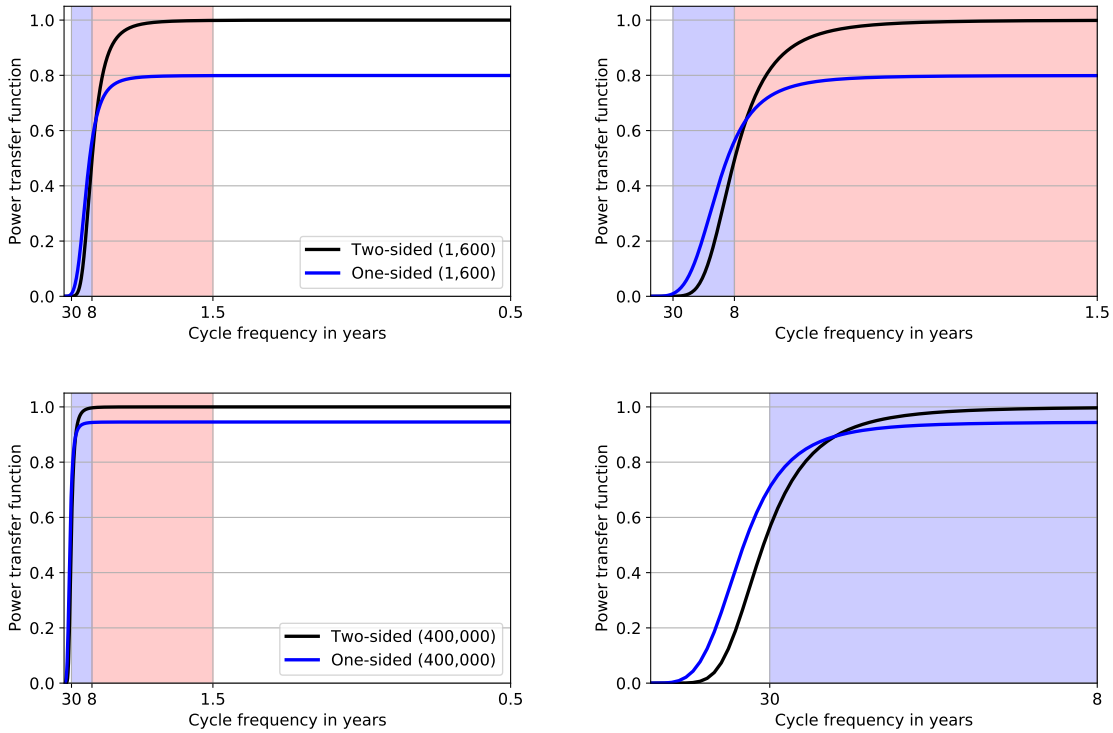
$$S_x(\omega) = PTF_{\mathcal{W}}(\omega) \cdot S_y(\omega). \tag{1.13}$$

It is in this sense that $PTF_{\mathcal{W}}(\omega)$ dampens ($PTF_{\mathcal{W}}(\omega) < 1$), passes ($PTF_{\mathcal{W}}(\omega) = 1$), or amplifies ($PTF_{\mathcal{W}}(\omega) > 1$) movements at specific frequencies ω of y_t.

1.4 Why Adjust the One-Sided HP Filter?

The top two panels of Figure 1.1 show the PTFs of HP-1s and HP-2s for $\lambda = 1{,}600$. The PTF of HP-2s is that of a high pass filter: Higher-frequency fluctuations in the range from 8 to 0.5 years pass the filter with only moderate or virtually no dampening. Lower-frequency fluctuations, by contrast, are dampened almost entirely: At a cycle length of approximately

Figure 1.1: PTFs of HP-1s and HP-2s for $\lambda = 1{,}600$ and $\lambda = 400{,}000$



Notes: Panels depict PTFs of HP-1s ("One-sided") and HP-2s ("Two-sided") for different values of the smoothing parameter λ. The top panels refer to $\lambda = 1{,}600$, the bottom panels refer to $\lambda = 400{,}000$. Right-hand panels are magnified versions of the left-hand panels. Red-shaded areas highlight business cycle frequencies (1.5 to 8 years), while purple areas highlight financial cycle frequencies (8 to 30 years).

17 years, for instance, its PTF reaches a value of 0.01, meaning that the filter dampens variations at this frequency by 99 percent. As argued by Baxter and King (1999), these are desirable features of HP-2s because the filter approximates the ideal band pass filter for this frequency range. While the filter effectively cancels fluctuations at frequencies that are lower than the targeted frequency range through a sharp cut-off, the frequency components above the cut-off pass the filter and remain unaffected.

The PTF of HP-1s differs in two respects.[3] First, to the left of the two curves' intersection, the PTF of HP-1s runs above that of HP-2s. This implies that HP-1s fails to dampen lower frequencies to the same extent as HP-2s. In fact, differences are relatively pronounced: At a cycle length of 17 years, for instance, the value of the PTF of HP-1s is 0.08, which is eight times the value of the PTF of HP-2s. As a consequence, cyclical components extracted using

[3]In the time domain, differences between HP-2s and HP-1s have been implicitly pointed out as well. For instance, decomposing the HP-2s weights into eight components, De Jong and Sakarya (2016) show that one component is only important near the end of the sample.

HP-1s feature low-frequency fluctuations to a much larger extent than cyclical components extracted using HP-2s. This also implies that the cyclical components of HP-1s tend to be more persistent.[4] Clearly, this is an undesirable property of HP-1s, as it implies that its output is more heavily contaminated by the fluctuations that one aims to remove.
Second, to the right of the two curves' intersection, the PTF of HP-1s starts to run horizontally at a level of approximately 0.8. The fact that this value is less than 1 implies that HP-1s dampens higher frequencies. As a consequence, higher-frequency fluctuations are less prevalent in cyclical components extracted using HP-1s than they are in the original data. By contrast, these fluctuations pass HP-2s without dampening, as its PTF approaches a value of one as the frequency rises. This is another drawback of HP-1s compared to HP-2s: HP-1s dampens exactly the fluctuations that one aims to extract.
Together, these two differences imply that the variability of filtered series will differ between HP-1s and HP-2s. Yet the direction is not clear and depends on the spectral density of the series we intend to filter. We elaborate on this in the empirical part of the paper (see Section 1.5.2).
Lastly, the bottom two panels of Figure 1.1 show the PTFs of HP-1s and HP-2s for $\lambda = 400{,}000$, which is the parameter value recommended in Basel III regulations to construct the credit-to-GDP gap. The respective PTFs resemble those for $\lambda = 1{,}600$, though there are some quantitative differences: Both the lack of dampening of lower frequencies and the excessive dampening of high frequencies that occur when using HP-1s appear less pronounced for $\lambda = 400{,}000$ than for $\lambda = 1{,}600$. This suggests that differences between HP-1s and HP-2s diminish as λ increases.
In this section, we have shown that the PTFs of HP-1s and HP-2s differ and that differences are more pronounced for small values of the smoothing parameter λ. So why are these differences important? The reason is that HP-1s is regularly used as the real-time version of HP-2s, wrongly assuming that it has the same desirable properties. The next section proposes two adjustments to solve this issue.

1.5 The Adjusted One-Sided HP Filter

We propose two adjustments to HP-1s that harmonize its properties with HP-2s: First, in order to eliminate the (relatively constant) dampening of higher-frequency fluctuations, we

[4]We use the word "tend" because persistence, for example, as measured by first-order autocorrelation, reflects only a proxy for cycle length. First-order autocorrelation crucially depends on the actual DGP of the series, such as the presence of a unit root.

rescale the cyclical component. Second, to harmonize PTFs in the range of lower frequencies, we choose a lower value of λ for HP-1s than for HP-2s.

More formally, the filter polynomial for the cyclical component of the *adjusted one-sided HP filter* (HP-1s*) is:

$$\widetilde{W}_{t|t,\lambda}(L) = \kappa \cdot \overline{W}_{t|t,\lambda^*}(L), t = 3, \ldots, T \tag{1.14}$$

where $\widetilde{W}_{t|t,\lambda}(L)$ denotes the filter polynomial of HP-1s* that is harmonized with the two-sided HP filter (HP-2s) with smoothing parameter λ. $\kappa = k(\lambda)$ is the scaling factor, and $W_{t|t,\lambda^*}(L)$ is the standard filter polynomial of the one-sided HP filter (HP-1s) with the adjusted smoothing parameter $\lambda^* = l(\lambda)$. To clarify notation, consider HP-1s* for $\lambda = 1{,}600$. As we show below, in this case we have

$$\widetilde{W}_{t|t,1600}(L) = 1.1513 \cdot \overline{W}_{t|t,650}(L),$$

i.e. HP-1s* uses the weight polynomial of the unadjusted one-sided HP filter (HP-1s) with smoothing parameter $\lambda^* = 650$ and scales the weights by the factor $\kappa = 1.1513$.

We obtain $\kappa = k(\lambda)$ and $\lambda^* = l(\lambda)$ by minimizing the squared distance between the PTFs of HP-1s* and HP-2s for $\kappa, \lambda^* \in \mathbb{R}^2_{>0}$.[5] Specifically, we solve the following minimization problem

$$\min_{\kappa,\lambda^*} \left(\int_0^\pi \left(PTF_{W_{t|T,\lambda}}(\omega) - \kappa^2 \cdot PTF_{W_{t|t,\lambda^*}}(\omega) \right)^2 d\omega \right). \tag{1.15}$$

This is similar to Baxter and King (1999) approach to finding an approximate bandpass filter for economic time series.

Table 1.2 gives an overview of the most commonly used smoothing parameters of HP-2s for business and financial cycles together with the corresponding parameters we find for HP-1s*.[6]

Figure 1.2 shows the PTFs of HP-1s* and HP-2s. Though some visible differences remain between the two PTFs, these are clearly smaller than for HP-1s in Figure 1.1. Notably, in the immediate neighbourhood of the intersection of the two PTF curves, HP-1s* continues to show a slight lack of dampening at lower frequencies (left of intersection) and a slight excess of dampening at higher frequencies (right of intersection). Despite these differences, in terms of the proximity of PTFs, HP-1s* is a more closely harmonized real-time version of HP-2s than HP-1s.

[5] Restricting the optimization problem to $\mathbb{R}^2_{>0}$ insures that κ is unique such that the optimization problem is well defined. We also provide contour plots of loss function for the solutions in Table 1.2 in Appendix 1.B.

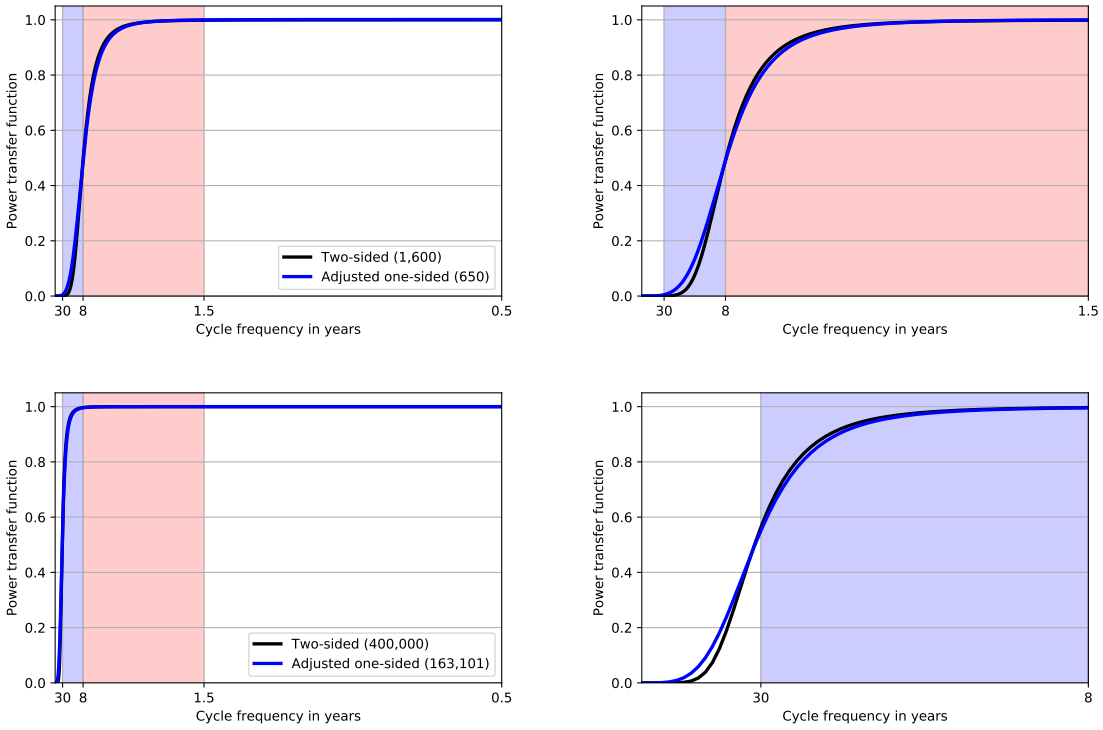
[6] Appendix 3.A offers a concise overview of adjustment parameters for a broad range of smoothing parameters of HP-2s. In addition, software implementing the adjusted HP filter in R, Python and MATLAB can be downloaded from `https://sites.google.com/site/yvesschueler/research`

Table 1.2: Typical Parameters of the Adjusted One-Sided Hodrick-Prescott Filter

	Two-sided filter (HP-2s)	Adjusted one-sided filter (HP-1s*)	
	λ	λ^*	κ
Business cycles (yearly data)	6.25	2.45	1.7962
Business cycles (quarterly data)	1,600	650	1.1513
Financial cycles (quarterly data)	400,000	163,101	1.0360

Notes: λ denotes smoothing parameter of the two-sided HP filter; λ^* and κ denote, respectively, the corresponding smoothing parameter and scaling factor of the adjusted one-sided HP filter. The scaling factor is multiplied by the extracted cyclical component of the one-sided HP filter.

Figure 1.2: PTFs of HP-1s* and HP-2s for $\lambda = 1{,}600$ and $\lambda = 400{,}000$



Notes: Panels depict PTFs of HP-1s* ("Adjusted one-sided") and HP-2s ("Two-sided") for different values of the smoothing parameter λ. The top panels refer to $\lambda = 1{,}600$, the bottom panels refer to $\lambda = 400{,}000$. Right-hand panels are magnified versions of the left-hand panels. Red-shaded areas highlight business cycle frequencies (1.5 to 8 years), while purple areas highlight financial cycle frequencies (8 to 30 years).

Interestingly, the "fourth-power rule" rule of Ravn and Uhlig (2002) also holds for HP-1s*. In the context of HP-2s, it is used to adjust the value of the smoothing parameter for different sampling frequencies or cycle lengths. For instance, the conversion factor when going from a quarterly to a yearly sampling frequency is $s = (1/4)^4$. Given our value of $\lambda^* = 650$ for quarterly data, this gives $\lambda^* = 2.45$ for yearly data. Furthermore, when targeting financial cycles that are four times as long as business cycles, the conversion factor to mirror this

increase in cycle length is $s = 4^4$. Applying this to $\lambda^* = 650$, yields a value of $\lambda^* = 166,400$. Since both converted values only differ marginally from the values we find as solutions to Equation (1.15) (for instance, see Table 1.2), we conclude that the rule also applies to HP-1s*.

Why do we adjust HP-1s in the manner of Equation (1.14)? Suppose we adjust only the smoothing parameter λ^* but apply no scaling factor, i.e. we set $\kappa = 1$ in Equation (1.14). This adjustment does not produce a good fit, because a change in the smoothing parameter alone will have opposing effects on the degree to which PTFs are harmonized in the range of higher frequencies vs. the range of lower frequencies. This can be seen from Figure 1.1: Choosing a higher value of λ reduces the undesirable dampening of higher frequencies but, at the same time, increases the undesirable lack of dampening at lower frequencies. By contrast, in the specification of Equation (1.14), the scaling factor κ eliminates the dampening of higher frequencies, where the PTF is almost horizontal, whereas the adjusted smoothing factor λ^* harmonizes PTFs in the low-frequency band.

1.5.1 Simulation Exercise

Using simulated data, we analyze how closely the cyclical components of HP-1s* and HP-1s match the cyclical component of HP-2s, considering volatility and persistence. To this end, we consider three alternative classes of DGPs:

(i) A **white noise** process with $y_t = \varepsilon_t$ and $\varepsilon_t \sim N(0,1)$, in which all frequencies contribute equally to the overall variance;

(ii) A **random walk** process with $y_t = y_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim N(0,1)$, which is a good representation of the time series behaviour of many macroeconomic variables;

(iii) Four **ARIMA** processes fitted to log US real GDP and the US credit-to-GDP ratio, both at a yearly and a quarterly sampling frequency. The sample period of the yearly (quarterly) data is 1880 to 2019 (1952Q2 to 2019Q4). The lag orders, parameter estimates and more details on the data are provided in Appendix 1.D.

Although a white noise process is stationary, we use its uniform spectral density where all frequencies contribute equally to the overall variance to investigate the effects of the three filters. As argued for example in Hamilton (2018), the second DGP is a good representation of the time series behaviour of many macroeconomic variables. Roughly speaking, in a random walk process, lower frequencies contribute relatively more to the total variance

with a pole at frequency $\omega = 0$. Eventually the fitted ARIMA processes provide a good approximation of real world data. Additionally, they can generate more complex cyclical properties than white noise or random walk processes. For instance, they can mirror the distinct feature of real GDP of having fluctuations at business cycle frequencies (1.5 to eight years) that contribute heavily to its overall variance. Furthermore, they can mirror the fact that the credit-to-GDP ratio has fluctuations at medium-term frequencies (eight to 30 years) that contribute heavily to its overall variance.[7]

For filtering, we use $\lambda = \{6.25,\ 1{,}600,\ 400{,}000\}$ as values for the smoothing parameter of HP-2s and HP-1s. The respective parameters for HP-1s* are $\lambda^* = \{2.45, 650, 163{,}101\}$ and $\kappa = \{1.7962, 1.1513, 1.0360\}$. The simulations use 1,000 draws of 100 observations each. In order to judge the performance of HP-1s* and HP-1s, we calculate the standard deviation and first-order autocorrelation relative to HP-2s. Specifically, we consider the (i) relative difference in standard deviations: $(\hat{\sigma}_{\text{HP-1s}(*)} - \hat{\sigma}_{\text{HP-2s}})/\hat{\sigma}_{\text{HP-2s}}$, where $\hat{\sigma}_{\text{HP-2s}}$ and $\hat{\sigma}_{\text{HP-1s}(*)}$ are the estimated standard deviations of the cyclical components of HP-2s and either HP-1s* or HP-1s and (ii) the difference in autocorrelations: $\hat{\rho}_{\text{HP-1s}(*)} - \hat{\rho}_{\text{HP-2s}}$, where $\hat{\rho}_{\text{HP-2s}}$ and $\hat{\rho}_{\text{HP-1s}(*)}$ are the estimated first order autocorrelations of the cyclical components of HP-2s and either HP-1s* or HP-1s. Table 1.3 shows the results of the simulation experiment. In square brackets, we report the fraction of cases where the statistic of the perspective one-sided filter is closer to the statistics of HP-2s. In terms of our measure of volatility, the relative difference in standard deviations, we find that, over all different simulations HP-1s* produces an estimate that is closer to HP-2s in about 79% of the cases. The differences are especially pronounced for yearly data and business cycle frequencies. For example for yearly data, the standard deviation of the cyclical component obtained with HP-1s is 20% lower compared to the one of HP-2s while HP-1s* matches the second moment well. In terms of our measure of persistence, the first order autocorrelation, the results favor HP-1s* even more strongly. We find that HP-1s* dominates HP-1s in 89% of the cases. The differences between HP-2s and HP-1s matter especially for business cycles filtering with quarterly data. These simulation results are also robust with regards to larger and smaller sample sizes as can be seen from the additional results in Table 1.D.3 in the Appendix. Furthermore, as shown in Table 1.D.2 in the Appendix of this chapter, differences in the persistence of the cyclical components are similar or even larger if the comparison is extended to include higher order autocorrelation coefficients. We conclude that HP-1s* outperforms HP-1s with regard to the two criteria.

[7]See, for example, Galati et al. (2016); Schüler (2018) and Schüler (2019) and Strohsal et al. (2019).

Table 1.3: Simulation Exercise

	Relative difference in SD		Difference in autocorrelations	
λ	HP-1s*	HP-1s	HP-1s*	HP-1s
		T = 100		
White noise				
6.25	**0.01** [1.00]	**-0.50** [0.00]	**0.02** [0.99]	**0.07** [0.01]
1,600	**0.01** [1.00]	**-0.10** [0.00]	**0.02** [0.87]	**0.03** [0.13]
400,000	**0.00** [0.93]	**-0.03** [0.07]	**0.02** [0.67]	**0.02** [0.33]
Random walk				
6.25	**0.04** [1.00]	**-0.33** [0.00]	**0.07** [1.00]	**0.20** [0.00]
1,600	**0.06** [0.55]	**0.06** [0.45]	**0.04** [1.00]	**0.09** [0.00]
400,000	**0.02** [0.45]	**0.03** [0.55]	**0.02** [0.70]	**0.03** [0.30]
ARIMA				
Yearly:				
6.25	**0** [0.94]	**-0.19** [0.06]	**0** [1.00]	**0** [0.00]
1,600	**0.08** [0.71]	**0.11** [0.29]	**0.09** [1.00]	**0.17** [0.00]
Quarterly:				
1,600	**0.08** [0.73]	**0.1** [0.27]	**0.09** [1.00]	**0.17** [0.00]
400,000	**0.07** [0.55]	**0.09** [0.45]	**0.1** [0.66]	**0.11** [0.34]

Notes: The table compares properties of cyclical components extracted using the adjusted one-sided HP filter (columns 'HP-1s*') and the unadjusted one-sided HP filter (columns 'HP-1s'), both relative to the two-sided HP filter. Column λ gives the value of the smoothing parameter used with the two-sided HP filter and the unadjusted one-sided HP filter. The adjusted one-sided HP filter uses the corresponding smoothing and scaling parameter (see, for instance, Table 1.C.1 in Online Appendix 3.A). 'Relative difference in standard deviations': $(\hat{\sigma}_{\text{HP-1s}(*)} - \hat{\sigma}_{\text{HP-2s}})/\hat{\sigma}_{\text{HP-2s}}$, where $\hat{\sigma}_{\text{HP-2s}}$ and $\hat{\sigma}_{\text{HP-1s}(*)}$ are the estimated standard deviations of the cyclical components of HP-2s and either HP-1s* or HP-1s. 'Difference in autocorrelations': $\hat{\rho}_{\text{HP-1s}(*)} - \hat{\rho}_{\text{HP-2s}}$, where $\hat{\rho}_{\text{HP-2s}}$ and $\hat{\rho}_{\text{HP-1s}(*)}$ are the estimated first order autocorrelations of the cyclical components of HP-2s and either HP-1s* or HP-1s. Numbers are the averages of the summary statistics across simulations. In square brackets, we report the fraction of cases where the summary statistic of a one-sided filter is in absolute value smaller than the summary statistic of the other one-sided filter, i.e. matches HP-2s more closely.

1.5.2 Application: How Different Are Financial and Business Cycles for G7 Countries?

In this section, we illustrate how the use of HP-1s and HP-1s* leads to different conclusions about the properties of financial cycles relative to business cycles. The question of how the properties of financial cycles relate to those of business cycles is well researched. Borio (2014) documents that financial cycles have a greater amplitude and are more persistent than business cycles.[8] These differences may matter, for example, in the calibration of macro-econometric models and in the debate on financial market regulation (see, for example, Hiebert et al. (2018)).

In our application, we proxy financial cycles with the HP-filtered credit-to-GDP ratio (λ = 400,000 for quarterly and λ = 1,600 for yearly data) and business cycles with HP-

[8]Further studies are, for example, Claessens et al. (2012), Aikman et al. (2015), Schüler et al. (2015,0) or Strohsal et al. (2019).

Table 1.4: Financial Cycle Statistics Relative to Business Cycle Statistics

	Relative difference in standard deviations						
	Canada	France	Germany	Italy	Japan	UK	US
Yearly data							
HP-2s	0.34	1.31	0.77	-0.11	-0.04	1.78	0.25
HP-1s*	0.31	1.14	0.67	-0.11	-0.05	1.70	0.25
HP-1s	0.78	1.64	1.18	0.28	0.29	2.55	0.72
Quarterly data							
HP-2s	2.83	2.15	2.20	3.23	6.25	2.90	2.65
HP-1s*	2.70	2.16	2.17	3.27	6.02	3.04	2.57
HP-1s	2.91	2.62	2.45	4.06	6.79	3.76	2.79
	Difference in autocorrelations						
	Canada	France	Germany	Italy	Japan	UK	US
Yearly data							
HP-2s	0.51	0.63	0.50	0.61	0.47	0.59	0.53
HP-1s*	0.45	0.53	0.43	0.58	0.44	0.52	0.49
HP-1s	0.36	0.38	0.33	0.51	0.37	0.40	0.41
Quarterly data							
HP-2s	0.25	0.45	0.29	0.31	0.28	0.40	0.22
HP-1s*	0.24	0.44	0.28	0.29	0.24	0.38	0.20
HP-1s	0.20	0.39	0.24	0.26	0.18	0.33	0.17

Notes: The table compares the properties of financial cycles (measured by filtered credit-to-GDP ratio) relative to business cycles (measured by filtered log real GDP), relating the two cyclical components of the two-sided HP filter (rows 'HP-2s'), the two cyclical components of the adjusted one-sided HP filter (rows 'HP-1s*'), and the two cyclical components of the unadjusted one-sided HP filter (row 'HP-1s'). HP-2s and HP-1s use $\lambda = \{1{,}600, 400{,}000\}$ ($\lambda = \{6.25, 1{,}600\}$) for estimating yearly and quarterly financial (business) cycles, thereby conducting a comparison in the spirit of Borio (2014). The adjusted one-sided HP filter uses the corresponding smoothing and scaling parameter (see, for instance, Table 1.C.1 in Online Appendix 3.A). 'Relative difference in standard deviations' is the estimated standard deviation of financial cycles ($\hat{\sigma}_{\text{fc}}$) minus the estimated standard deviation of business cycles ($\hat{\sigma}_{\text{bc}}$), relative to the standard deviation of business cycles, i.e. $(\hat{\sigma}_{\text{fc}} - \hat{\sigma}_{\text{bc}})/\hat{\sigma}_{\text{bc}}$. 'Difference in autocorrelations' is the difference of the estimated first order autocorrelation of financial cycles ($\hat{\rho}_{\text{fc}}$) and the estimated first order autocorrelation of business cycles ($\hat{\rho}_{\text{bc}}$), i.e. $\hat{\rho}_{\text{fc}} - \hat{\rho}_{\text{bc}}$. The input for the numbers of the table are in Appendix 1.E Table 1.E.1.

filtered log real GDP (λ = 1,600 for quarterly and λ = 6.25 for yearly data). We also include the Covid-19 pandemic in the sample period, because structural breaks do not affect the weights of the HP filter. Therefore, our yearly (quarterly) sample is 1880 until 2020 (1952Q2 until 2022Q2), where starting dates across G7 countries are not equal however.[9] Note that, at the time of writing, the yearly database only covers the time period until 2020. For further details, please see Appendix 1.E.

Table 1.4 shows our results. The first metric, the relative difference in standard deviations, measures how much more volatile the financial cycle is relative to the business cycle. It is evident from the table that HP-1s has a tendency to exaggerate – relative to HP-2s and HP-1s – the differences between financial and business cycles. For HP-1s*, the differences

[9]For the yearly dataset, France starts in 1990. The starting dates for the quarterly dataset are Canada (1991Q1), France (1969Q4), Germany and Italy (1960Q4), Japan (1964Q4), United Kingdom (1963Q1), and United States (1952Q2).

are quite similar to the differences reported by HP-2s. For example, HP-1s suggests that the US financial cycle is 72% more volatile than the US business cycle, looking at the yearly data. By contrast, HP-2s and HP-1s* indicate that the difference is less than half of that, at about 25%. The second metric, the difference in first-order autocorrelation, measures how much more persistent financial cycles are relative to business cycles. Here, again, the conclusions from using HP-2s and HP-1s* are relatively similar. In comparison to those filters, smaller differences in the relative persistence emerge when applying HP-1s.
Overall, we conclude the adjustments succeed at harmonizing the properties of HP-1s with the properties HP-2s, also when considering actual time series for G7 countries.

1.6 Conclusion

Should the cyclical component obtained from the standard HP-1s be used as the real-time version of HP-2s' cyclical component? This paper argues that it should not. The reason is that important properties of the standard HP-1s are quite different from those of HP-2s. Specifically, the standard one-sided filter (1) fails to remove low-frequency fluctuations to the same extent as the two-sided filter and (2) has the undesirable feature of dampening precisely those fluctuations that one wishes to extract.
As a remedy, this paper proposes HP-1s*, making two easy-to-implement adjustments to HP-1s: (1) a lower smoothing parameter and (2) a rescaling of the cyclical component. Together, these two adjustments address the above-mentioned problems with HP-1s.
We confirm this in applications of HP-1s* to both simulated and empirical data. Our simulation results show that the adjustments yield cyclical components that are more similar to HP-2s in terms of volatility and persistence. Furthermore, the empirical application demonstrates that HP-1s needs to be adjusted as it otherwise distorts comparisons of different economic cycles with regard to the latter properties.

Appendix

1.A Closed Form Expression for the Filter Polynomials

To derive the closed form of the filter polynomials, we use insights of Hamilton (2018) and Phillips and Jin (2021). Specifically, the minimization problem of Equation (1.2) has the matrix representation

$$\hat{\tau}_{1:T,\lambda} = \underset{\tau_{1:T}}{\arg\min} \left\{ (y_{1:T} - \tau_{1:T})'(y_{1:T} - \tau_{1:T}) + \lambda \tau_{1:T}' Q_T' Q_T \tau_{1:T} \right\}, \tag{1.16}$$

where $\hat{\tau}_{1:T,\lambda} = (\hat{\tau}_{1|T,\lambda}, \ldots, \hat{\tau}_{T|T,\lambda})'$, $\tau_{1:T} = (\tau_1, \ldots, \tau_T)'$, and $y_{1:T} = (y_1, \ldots, y_T)'$. Taking the derivative with respect to $\tau_{1:T}$ yields the FOC

$$-2(y_{1:T} - \tau_{1:T}) + 2\lambda Q_T' Q_T \tau_{1:T} = 0 \tag{1.17}$$

Rearranging gives the solution

$$\hat{\tau}_{1:T,\lambda} = (I_T + \lambda Q_T' Q_T)^{-1} y_{1:T} = \left\{ I_T - Q_T'(Q_T Q_T' + \lambda^{-1} I_{T-2})^{-1} Q_T \right\} y_{1:T}. \tag{1.18}$$

where the second equality is given by the Woodbury matrix identity. Since the estimated cyclical component is $\hat{\psi}_{1:T,\lambda} = y_{1:T} - \hat{\tau}_{1:T,\lambda}$, similarly defining $\hat{\psi}_{1:T,\lambda} = (\hat{\psi}_{1|T,\lambda}, \ldots, \hat{\psi}_{1|T,\lambda})'$, it follows that

$$\begin{aligned}
\hat{\psi}_{1:T,\lambda} &= \left\{ I_T - \left(I_T - Q_T'(Q_T Q_T' + \lambda^{-1} I_{T-2})^{-1} Q_T \right) \right\} y_{1:T} && (1.19)\\
&= Q_T'(Q_T Q_T' + \lambda^{-1} I_{T-2})^{-1} Q_T y_{1:T} && (1.20)\\
&= Q_T'(Q_T Q_T' + \lambda^{-1} I_{T-2})^{-1} Q_T \eta_T y_T, && (1.21)
\end{aligned}$$

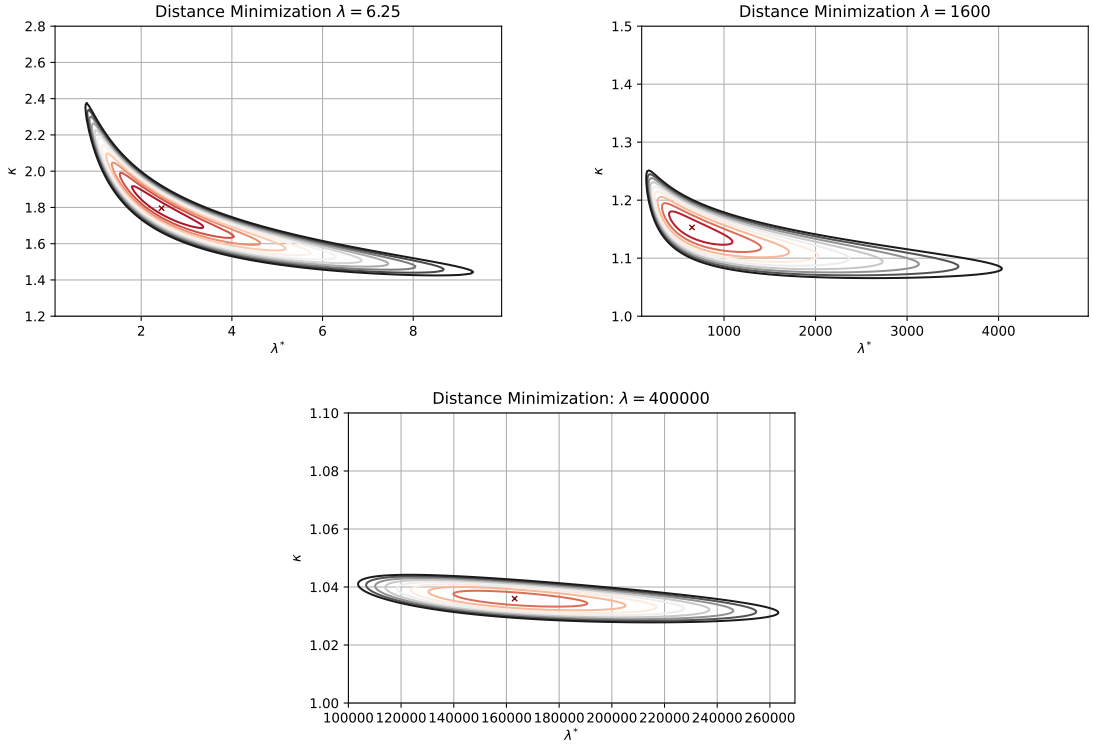
where $\eta_T = (L^{T-1}, \ldots, L^1, 1)'$. Finally, the cyclical component of HP-1s $\hat{\psi}_{t|t,\lambda}$ is obtained by setting the information set of $\hat{\psi}_{1:T,\lambda}$ to t (for $t = 3, \ldots, T$) and picking its last element, i.e.,

$$\hat{\psi}_{t|t,\lambda} = e_t' \hat{\psi}_{1:t,\lambda} = e_t' Q_t' (Q_t Q_t' + \lambda^{-1} I_{t-2})^{-1} Q_t \eta_t y_t, \tag{1.22}$$

where $e_t = (0, \ldots, 0, 1)'$ is a t-dimensional column vector.

1.B Optimization Results

Figure 1.B.1: Contour lines of the loss-function



Notes: Optimization results for most common reference values of the two-sided HP-Filter λ corresponding to business cycles with yearly and quarterly data as well as financial cycles with quarterly data. Red crosses denote the optimal values reported in table 1.2. The contour lines indicate that the optimization problem is well defined with unique minima in the domain of the parameters $\mathbb{R}^2_+$.

1.C Adjustment Parameters for the One-Sided Hodrick-Prescott Filter

Table 1.C.1: Smoothing and Rescaling Parameters λ^* and κ of HP-1s

λ	λ^*	κ	λ	λ^*	κ	λ	λ^*	κ	λ	λ^*	κ
1.00	0.35	2.7174	55.00	22.15	1.3921	1000	406	1.1718	55000	22411	1.0598
1.25	0.45	2.5372	57.50	23.16	1.3869	1250	508	1.1617	57500	23431	1.0591
1.50	0.55	2.4111	60.00	24.17	1.3820	**1600**	**650**	**1.1513**	60000	24450	1.0584
1.75	0.65	2.3165	62.50	25.19	1.3774	1750	711	1.1477	62500	25469	1.0578
2.00	0.75	2.2420	65.00	26.20	1.3730	2000	813	1.1425	65000	26488	1.0572
2.25	0.85	2.1814	67.50	27.21	1.3688	2250	915	1.1381	67500	27507	1.0567
2.50	0.95	2.1308	70.00	28.23	1.3648	2500	1017	1.1342	70000	28526	1.0562
2.75	1.04	2.0877	72.50	29.24	1.3610	2750	1118	1.1309	72500	29546	1.0557
3.00	1.14	2.0503	75.00	30.25	1.3574	3000	1220	1.1279	75000	30565	1.0552
3.25	1.24	2.0176	77.50	31.26	1.3540	3250	1322	1.1252	77500	31584	1.0547
3.50	1.34	1.9885	80.00	32.28	1.3506	3500	1424	1.1227	80000	32603	1.0543
3.75	1.44	1.9625	82.50	33.29	1.3475	3750	1525	1.1205	82500	33623	1.0538
4.00	1.55	1.9390	85.00	34.30	1.3444	4000	1627	1.1184	85000	34642	1.0534
4.25	1.65	1.9177	87.50	35.32	1.3415	4250	1729	1.1166	87500	35661	1.0530
4.50	1.75	1.8981	90.00	36.33	1.3387	4500	1831	1.1148	90000	36680	1.0527
4.75	1.85	1.8802	92.50	37.35	1.3359	4750	1933	1.1132	92500	37700	1.0523
5.00	1.95	1.8636	95.00	38.36	1.3333	5000	2034	1.1116	95000	38719	1.0519
5.25	2.05	1.8483	97.50	39.37	1.3308	5250	2136	1.1102	97500	39738	1.0516
5.50	2.15	1.8339	100	40	1.3283	5500	2238	1.1089	100000	40758	1.0512
5.75	2.25	1.8206	125	51	1.3077	5750	2340	1.1076	125000	50951	1.0484
6.00	2.35	1.8080	150	61	1.2919	6000	2442	1.1064	150000	61145	1.0462
6.25	**2.45**	**1.7962**	175	71	1.2792	6250	2543	1.1053	175000	71340	1.0444
6.50	2.55	1.7851	200	81	1.2688	6500	2645	1.1042	200000	81534	1.0429
6.75	2.65	1.7746	225	91	1.2599	6750	2747	1.1032	225000	91730	1.0416
7.00	2.75	1.7647	250	101	1.2523	7000	2849	1.1022	250000	101925	1.0405
7.25	2.85	1.7552	275	111	1.2456	7250	2951	1.1012	275000	112120	1.0396
7.50	2.95	1.7463	300	122	1.2397	7500	3053	1.1003	300000	122316	1.0387
7.75	3.05	1.7377	325	132	1.2343	7750	3154	1.0995	325000	132512	1.0379
8.00	3.15	1.7296	350	142	1.2295	8000	3256	1.0986	350000	142708	1.0372
8.25	3.25	1.7218	375	152	1.2251	8250	3358	1.0978	375000	152904	1.0366
8.50	3.35	1.7143	400	162	1.2211	8500	3460	1.0971	**400000**	**163101**	**1.0360**
8.75	3.45	1.7072	425	172	1.2174	8750	3562	1.0964	425000	173297	1.0354
9.00	3.55	1.7004	450	183	1.2140	9000	3664	1.0956	450000	183494	1.0349
9.25	3.65	1.6938	475	193	1.2108	9250	3765	1.0950	475000	193691	1.0344
9.50	3.76	1.6875	500	203	1.2078	9500	3867	1.0943	500000	203887	1.0340
9.75	3.86	1.6814	525	213	1.2051	9750	3969	1.0937	525000	214084	1.0336
10.00	3.96	1.6755	550	223	1.2024	10000	4071	1.0930	550000	224281	1.0332
12.50	4.96	1.6265	575	233	1.2000	12500	5089	1.0878	575000	234478	1.0328
15.00	5.97	1.5897	600	243	1.1976	15000	6108	1.0837	600000	244675	1.0324
17.50	6.98	1.5607	625	254	1.1954	17500	7127	1.0804	625000	254873	1.0321
20.00	7.99	1.5370	650	264	1.1933	20000	8145	1.0776	650000	265070	1.0318
22.50	9.00	1.5172	675	274	1.1913	22500	9164	1.0753	675000	275267	1.0315
25.00	10.01	1.5001	700	284	1.1894	25000	10183	1.0733	700000	285465	1.0312
27.50	11.02	1.4853	725	294	1.1876	27500	11202	1.0715	725000	295662	1.0309
30.00	12.03	1.4723	750	304	1.1859	30000	12221	1.0699	750000	305860	1.0307
32.50	13.04	1.4606	775	315	1.1842	32500	13240	1.0685	775000	316057	1.0304
35.00	14.05	1.4502	800	325	1.1826	35000	14259	1.0672	800000	326255	1.0302
37.50	15.06	1.4407	825	335	1.1811	37500	15278	1.0660	825000	336453	1.0299
40.00	16.08	1.4320	850	345	1.1796	40000	16297	1.0649	850000	346650	1.0297
42.50	17.09	1.4241	875	355	1.1782	42500	17316	1.0639	875000	356848	1.0295
45.00	18.10	1.4167	900	365	1.1768	45000	18335	1.0629	900000	367046	1.0293
47.50	19.11	1.4099	925	376	1.1755	47500	19354	1.0621	925000	377244	1.0291
50.00	20.12	1.4036	950	386	1.1743	50000	20373	1.0612	950000	387442	1.0289
52.50	21.14	1.3977	975	396	1.1730	52500	21392	1.0605	975000	397640	1.0287
									1000000	407838	1.0285

Notes: λ denotes the smoothing parameter of the two-sided HP filter. λ^* is the corresponding adjusted smoothing parameter, used as an input to the one-sided HP filter.

1.D Simulation Exercise

Table 1.D.1: Estimated ARIMA Models

	log GDP
Yearly	$\Delta y_t = 3.27 + 0.83\Delta y_{t-1} - 0.82\Delta y_{t-2} - 0.08\Delta y_{t-3} + \varepsilon_t - 0.4\varepsilon_{t-1} + 0.9\varepsilon_{t-2}$
Quarterly	$\Delta y_t = 0.74 - 0.36\Delta y_{t-1} + 0.33\Delta y_{t-2} + \varepsilon_t + 0.68\varepsilon_{t-1}$
	Credit-to-GDP ratio
Yearly	$\Delta y_t = 0.32 + 0.57\Delta y_{t-1} - 0.84\Delta y_{t-2} + 0.29\Delta y_{t-3} + \varepsilon_t + 0.07\varepsilon_{t-1} + 0.93\varepsilon_{t-2}$
Quarterly	$\Delta y_t = 0.38 + 0.92\Delta y_{t-1} + 0.99\Delta y_{t-2} - 0.93\Delta y_{t-3} + \varepsilon_t - 0.64\varepsilon_{t-1} - 0.92\varepsilon_{t-2} + 0.57\varepsilon_{t-3}$

Notes: The table gives the estimated model equations for each series in logs. We choose the lag orders of the AR and MA polynomials using the Akaike information criteria. Quarterly data was obtained from the FRED and BIS databases, yearly series where retrieved from the Jordà-Schularick-Taylor Macrohistory Database (see Jordà et al. (2017)). The sample period of the yearly (quarterly) data is 1880 to 2019 (1952Q2 to 2019Q4).

Table 1.D.2: Simulation Exercise: Higher-Order Autocorrelation

	RMSE(4)		RMSE(8)	
λ	HP-1s*	HP-1s	HP-1s*	HP-1s
		T = 100		
White noise				
6.25	**0.01** [1.00]	**0.04** [0.00]	**0.01** [0.88]	**0.03** [0.12]
1,600	**0.01** [1.00]	**0.02** [0.00]	**0.01** [0.88]	**0.02** [0.12]
400,000	**0.01** [1.00]	**0.01** [0.00]	**0.01** [1.00]	**0.01** [0.00]
Random walk				
6.25	**0.06** [1.00]	**0.14** [0.00]	**0.05** [1.00]	**0.12** [0.00]
1,600	**0.07** [1.00]	**0.16** [0.00]	**0.08** [1.00]	**0.17** [0.00]
400,000	**0.04** [1.00]	**0.05** [0.00]	**0.06** [1.00]	**0.09** [0.00]
ARIMA				
Yearly:				
6.25	**0.08** [0.75]	**0.16** [0.25]	**0.07** [0.88]	**0.16** [0.00]
1,600	**0.08** [1.00]	**0.16** [0.00]	**0.10** [1.00]	**0.20** [0.00]
Quarterly:				
1,600	**0.08** [1.00]	**0.16** [0.00]	**0.10** [1.00]	**0.20** [0.00]
400,000	**0.04** [1.00]	**0.04** [0.00]	**0.08** [1.00]	**0.09** [0.00]
		T = 1000		
White noise				
6.25	**0.01** [1.00]	**0.04** [0.00]	**0.01** [0.88]	**0.03** [0.12]
1,600	**0.00** [1.00]	**0.01** [0.00]	**0.00** [0.88]	**0.01** [0.12]
400,000	**0.00** [1.00]	**0.01** [0.00]	**0.00** [1.00]	**0.00** [0.00]
Random walk				
6.25	**0.05** [1.00]	**0.14** [0.00]	**0.05** [1.00]	**0.12** [0.00]
1,600	**0.04** [1.00]	**0.14** [0.00]	**0.06** [1.00]	**0.16** [0.00]
400,000	**0.02** [1.00]	**0.05** [0.00]	**0.03** [1.00]	**0.08** [0.00]
ARIMA				
Yearly:				
6.25	**0.07** [0.75]	**0.15** [0.25]	**0.07** [0.88]	**0.15** [0.12]
1,600	**0.04** [1.00]	**0.13** [0.00]	**0.07** [1.00]	**0.18** [0.00]
Quarterly:				
1,600	**0.05** [1.00]	**0.13** [0.00]	**0.07** [1.00]	**0.18** [0.00]
400,000	**0.00** [1.00]	**0.01** [0.00]	**0.01** [1.00]	**0.02** [0.00]

Notes: The table compares properties of cyclical components extracted using the adjusted one-sided HP filter (columns 'HP-1s*') and the unadjusted one-sided HP filter (columns 'HP-1s'), both relative to the two-sided HP filter. We evaluate the deviation of the higher order autocorrelations of the one-sided filter based on the RMSE given as $RMSE(h) = \sqrt{1/h \sum_{i=1}^{h} (\bar{\hat{\rho}}^{(i)}_{\text{HP-1s}^{(*)}} - \bar{\hat{\rho}}^{(i)}_{\text{HP-2s}})^2}$, where $\bar{\hat{\rho}}^{(i)}_{\text{HP-2s}}$ and $\bar{\hat{\rho}}^{(i)}_{\text{HP-1s}^{(*)}}$ are the averages over the estimated autocorrelations at the i^{th} lag of the cyclical components of HP-2s and either HP-1s* or HP-1s. In square brackets, we report the fraction of lags where the summary statistic of a one-sided filter is in absolute value smaller than the summary statistic of the other one-sided filter, i.e. matches HP-2s more closely.

Table 1.D.3: Simulation Exercise (continued)

	Relative difference in SD		Difference in autocorrelations	
λ	HP-1s*	HP-1s	HP-1s*	HP-1s
		T = 50		
White noise				
6.25	**0.01** [1.00]	**-0.49** [0.00]	**0.03** [0.96]	**0.08** [0.04]
1,600	**0.01** [0.95]	**-0.10** [0.05]	**0.03** [0.80]	**0.04** [0.20]
400,000	**-0.01** [0.88]	**-0.04** [0.12]	**0.03** [0.59]	**0.03** [0.41]
Random walk				
6.25	**0.05** [0.97]	**-0.32** [0.03]	**0.08** [1.00]	**0.21** [0.00]
1,600	**0.08** [0.51]	**0.06** [0.49]	**0.08** [0.95]	**0.14** [0.05]
400,000	**0.06** [0.43]	**0.04** [0.57]	**0.05** [0.62]	**0.05** [0.38]
ARIMA				
Yearly:				
6.25	**0.08** [0.84]	**-0.18** [0.16]	**0.00** [1.00]	**0.00** [0.00]
1,600	**0.12** [0.65]	**0.13**[0.35]	**0.11** [0.96]	**0.17** [0.04]
Quarterly:				
1,600	**0.12** [0.64]	**0.13** [0.45]	**0.10** [0.95]	**0.16** [0.05]
400,000	**0.14** [0.41]	**0.13** [0.59]	**0.19** [0.50]	**0.20** [0.50]
		T = 500		
White noise				
6.25	**0.00** [1.00]	**-0.51** [0.00]	**0.01** [1.00]	**0.07** [0.00]
1,600	**0.00** [1.00]	**-0.11** [0.00]	**0.00** [1.00]	**0.02** [0.00]
400,000	**0.00** [1.00]	**-0.03** [0.00]	**0.00** [0.91]	**0.01** [0.09]
Random walk				
6.25	**0.04** [1.00]	**-0.33** [0.00]	**0.06** [1.00]	**0.20** [0.00]
1,600	**0.04** [0.49]	**0.04** [0.51]	**0.02** [1.00]	**0.07** [0.00]
400,000	**0.04** [1.00]	**0.11** [0.00]	**0.01** [1.00]	**0.02** [0.00]
ARIMA				
Yearly:				
6.25	**0.06** [1.00]	**-0.20** [0.00]	**0.00** [0.92]	**0.00** [0.08]
1,600	**0.05** [0.93]	**0.08** [0.07]	**0.07** [1.00]	**0.16** [0.00]
Quarterly:				
1,600	**0.05** [0.90]	**0.07** [0.10]	**0.07** [1.00]	**0.16** [0.00]
400,000	**0.02** [0.99]	**0.07** [0.10]	**0.02** [1.00]	**0.04** [0.00]
		T = 1000		
White noise				
6.25	**0.00** [1.00]	**-0.51** [0.00]	**0.00** [1.00]	**0.07** [0.00]
1,600	**0.00** [1.00]	**-0.11** [0.00]	**0.00** [1.00]	**0.02** [0.00]
400,000	**0.00** [1.00]	**-0.03** [0.00]	**0.00** [0.98]	**0.01** [0.02]
Random walk				
6.25	**0.04** [1.00]	**-0.33** [0.00]	**0.06** [1.00]	**0.20** [0.00]
1,600	**0.04** [0.50]	**0.04** [0.50]	**0.02** [1.00]	**0.07** [0.00]
400,000	**0.04** [1.00]	**0.11** [0.00]	**0.01** [1.00]	**0.02** [0.00]
ARIMA				
Yearly:				
6.25	**0.06** [1.00]	**-0.20** [0.00]	**0.00** [0.85]	**0.00** [0.15]
1,600	**0.05** [0.97]	**0.07** [0.03]	**0.05** [1.00]	**0.13** [0.00]
Quarterly:				
1,600	**0.05** [0.97]	**0.07** [0.03]	**0.05** [1.00]	**0.13** [0.00]
400,000	**0.01** [1.00]	**0.05** [0.00]	**0.03** [1.00]	**0.10** [0.00]

Notes: The table compares properties of cyclical components extracted using the adjusted one-sided HP filter (columns 'HP-1s*') and the unadjusted one-sided HP filter (columns 'HP-1s'), both relative to the two-sided HP filter. For a description of the summary statistics ('relative difference in standard deviations' and 'difference in autocorrelations'), please see Section 1.5.1. Numbers are the averages of the summary statistics across simulations. In square brackets, we report the fraction that the summary statistic of a one-sided filter is in absolute value smaller than the summary statistic of the other one-sided filter, i.e. matches HP-2s more closely.

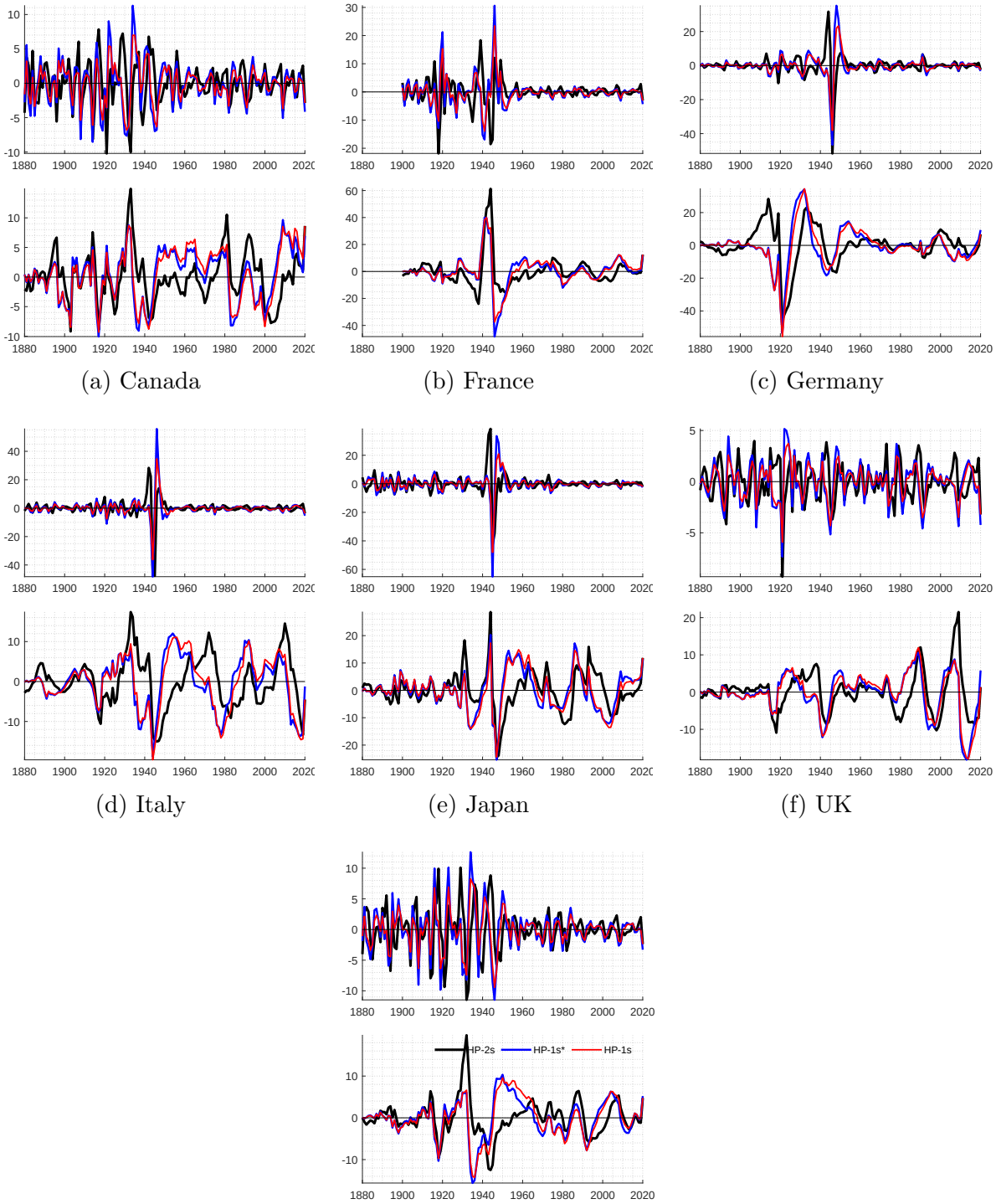
1.E Application: How Different Are Financial and Business Cycles for G7 Countries?

Table 1.E.1: Business and Financial Cycle Statistics for G7 Countries

	Standard deviation (%)						
	Canada	France	Germany	Italy	Japan	UK	US
log GDP							
Yearly data							
HP-2s	3.06	4.90	6.35	7.03	7.26	1.93	3.49
HP-1s*	3.29	5.32	6.86	7.19	7.55	2.04	3.65
HP-1s	2.49	4.10	5.33	5.20	5.58	1.56	2.74
Quarterly data							
HP-2s	1.56	1.67	1.67	1.89	1.58	2.29	1.58
HP-1s*	1.60	1.68	1.68	1.91	1.68	2.32	1.61
HP-1s	1.58	1.60	1.63	1.81	1.73	2.23	1.59
Credit-to-GDP ratio							
Yearly data							
HP-2s	4.09	11.34	11.26	6.27	6.95	5.35	4.35
HP-1s*	4.30	11.37	11.49	6.37	7.16	5.52	4.57
HP-1s	4.43	10.82	11.60	6.65	7.17	5.52	4.72
Quarterly data							
HP-2s	5.99	5.25	5.34	8.02	11.48	8.94	5.78
HP-1s*	5.93	5.32	5.33	8.14	11.81	9.38	5.75
HP-1s	6.17	5.79	5.62	9.18	13.47	10.64	6.03
	Autocorrelation						
	Canada	France	Germany	Italy	Japan	UK	US
log GDP							
Yearly data							
HP-2s	0.27	0.21	0.39	0.29	0.28	0.29	0.34
HP-1s*	0.35	0.31	0.47	0.32	0.34	0.38	0.39
HP-1s	0.49	0.48	0.58	0.41	0.45	0.51	0.50
Quarterly data							
HP-2s	0.70	0.50	0.69	0.67	0.71	0.59	0.77
HP-1s*	0.72	0.51	0.70	0.69	0.76	0.61	0.78
HP-1s	0.77	0.57	0.74	0.73	0.81	0.66	0.82
Credit-to-GDP ratio							
Yearly data							
HP-2s	0.78	0.84	0.89	0.90	0.76	0.88	0.87
HP-1s*	0.80	0.84	0.90	0.90	0.77	0.90	0.88
HP-1s	0.85	0.86	0.92	0.92	0.82	0.91	0.91
Quarterly data							
HP-2s	0.96	0.94	0.98	0.98	1.00	0.98	0.99
HP-1s*	0.97	0.95	0.98	0.98	0.99	0.98	0.99
HP-1s	0.97	0.96	0.98	0.99	1.00	0.99	0.99

Notes: The table gives the underlying input for the comparison of financial and business cycles of Table 1.4. 'HP-2s' refers to the cyclical component obtain via the two-sided HP filter, HP-1s* the adjusted one-sided HP filter, and HP-1s the unadjusted one-sided HP filter. For more details, please see Section 1.5.2. 'Autocorrelation' refers to first-order autocorrelation.

Figure 1.E.1: Yearly Hodrick-Prescott Filtered Data



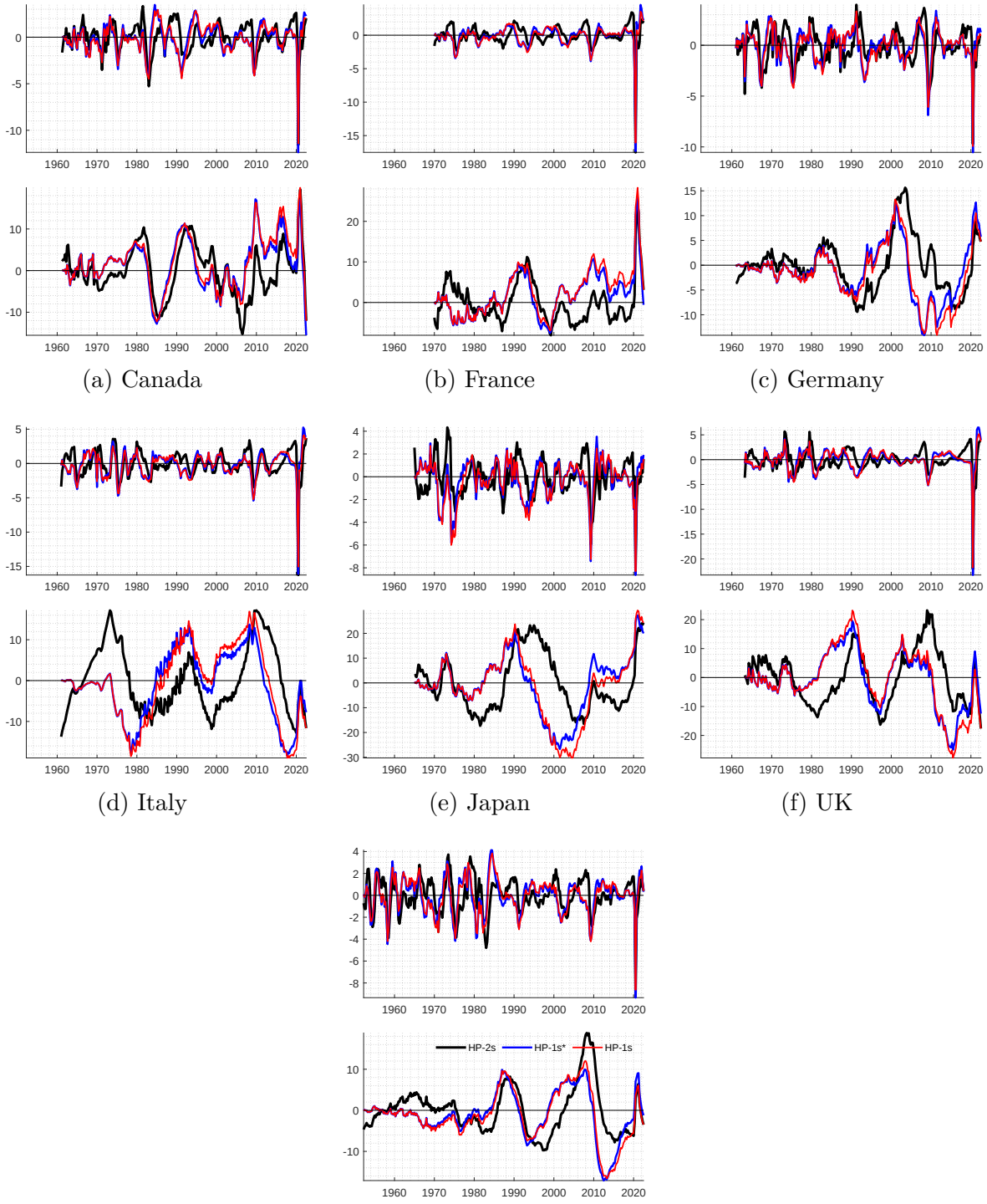
(a) Canada (b) France (c) Germany

(d) Italy (e) Japan (f) UK

(g) US

Notes: For each country, the top graph shows filtered log real GDP (×100). The bottom graph depicts the filtered credit-to-GDP ratio.

Figure 1.E.2: Quarterly Hodrick-Prescott Filtered Data



(a) Canada (b) France (c) Germany

(d) Italy (e) Japan (f) UK

(g) US

Notes: For each country, the top graph shows filtered log real GDP (×100). The bottom graph depicts the filtered credit-to-GDP ratio.

Chapter 2

Data Revisions to German National Accounts: Are Initial Releases Good Nowcasts?

with Till Strohsal

Chapter 3

Estimating Growth at Risk with Skewed Stochastic Volatility Models*

3.1 Introduction

John Maynard Keynes already wrote in his General Theory that economic recessions tend to be more volatile and more severe than economic expansions. Since then, different nonlinear or non-Gaussian features of the business cycle have been described in the economic literature (Pitt et al. (2012)). A recent study on non-linearities in US GDP growth that gained considerable attention amongst academics and policy makers alike is the seminal paper 'Vulnerable Growth' by Tobias Adrian, Nina Boryaschenko and Domenico Giannone (2019). Based on the results in the Macro-Finance literature, the authors link the non-linear and non-Gaussian features of economic growth to a country's national financial conditions. Borrowing from the Value at Risk concept in financial econometrics, they capture time-varying tail risks to GDP growth conditional on a country's national financial conditions using a semi-parametric estimation procedure. They find that a deterioration in national financial conditions increases the volatility and skewness of the conditional distribution of

*This chapter has greatly benefited from valuable comments and suggestions by Frank Schorfheide, Frank Diebold, Carlos Montes-Galdón, Joan Paredes, Dieter Nautz, Lars Winkelmann, Helmut Lütkepohl, Yves Schüler, Lea Wolf, Max Diegel, Andrea de Polis, Laura Liu, Minchul Shin, Thorsten Drautzburg, Simon Freyaldenhoven, Matteo Cicarelli, Michelle Lenza, Mareck Jarochinski, Marta Bańbura as well as other participants of the University of Pennsylvania Econometrics Lunch Seminar, the reserach seminar at the Federal Reserve Bank of Philadelphia, the EABCN and Bundesbank Conference on Challenges in Empirical Macroeconomics in May 2022, the ECB Brownbag Seminar in December 2021, the 11^{th} RCEA Money Macro Finance Conference in June 2021, and the "Topics in Time Series Econometrics"-Workshop 2020 and 2021 in Tornow. This chapter has been selected as one of the winners of the PhD paper competition of the 13^{th} ECB Forecasting Conference. An earlier version of this chapter is published as FU Discussion Paper No. 2022/2.

future GDP in the US. Most notably, their results indicate that conditional forecasting densities of US GDP growth rates are not always symmetric but left skewed in times of financial distress.

Yet, while the semi-parametric approach proposed by Adrian et al. (2019) is easy to implement, it is difficult to construct confidence intervals to statistically test the impact of national financial conditions or potentially other exogenous variables on the different moments of the conditional densities. Furthermore, since their approach does not assume a parametric law of motion, it is also not possible to conduct multi-step forecasts to predict of the evolution of time-varying risks around US-GDP growth several periods in the future. Nevertheless, as outlined in Prasad et al. (2019) the concept of Growth at Risk has provided policy makers all over world with an easy risk measure to evaluate a country's economic stability.

However, other recent studies by Hasenzagl et al. (2020) or Brownlees and Souza (2021) have put the results of Adrian et al. (2019) into question. While Hasenzagl et al. (2020) document little evidence for the effects of national financial conditions on higher moments such as variance or skewness, the results of Brownlees and Souza (2021) find that quantile regressions in the style of Adrian et al. (2019) show no predictive gains over symmetric GARCH models for macroeconomic tail risks.

Adding to the debate about time-varying asymmetries in the conditional densities of macroeconomic variables, this paper proposes a parametric modeling approach to estimate macroeconomic tail risks conditional on financial conditions. I build a Skewed Stochastic Volatility model (SSV) that can capture the variation of the full conditional density of future US GDP growth parametrically while being equally flexible. In this model, skewness arises from the assumption that errors follow a skewed normal distribution as introduced by Azzalini (2013). The model is a non-linear and non-Gaussian state space model that captures the effects of exogenous variables such as the national financial conditions index on the first and second moment of the conditional forecasting distribution of US GDP growth. Additionally, the SSV model provides a law of motion for the volatility and skewness of the conditional distributions. This makes it easy to iterate the evolution of the conditional densities forward in time to obtain multi step forecasts of tail risks. The model also nests a generic Stochastic Volatility (SV) model with symmetric densities, and does not impose skewness a priori.

Building on established Bayesian estimation methods for non-linear, non-Gaussian state space models, the SSV model can be estimated using a particle Monte Carlo Markov Chain (MCMC) algorithm that treats the model coefficients and the unobserved volatility and

skewness symmetrically as random variables.[14] This allows to easily construct credible sets for both objects to conduct statistical inference. Additionally, the Bayesian estimation approach makes it possible to compare and select different model specifications based on the respective marginal data densities. To increase the efficiency of the estimation, I use the recently introduced tempered particle filter by Herbst and Schorfheide (2019) to obtain robust estimates of the likelihood and latent states. As discussed in Pitt et al. (2012) the accuracy of the particle filtering approximation is important for the efficiency of the particle MCMC algorithm. Improving on the well-documented weakness of standard bootstrap particle filters to be sensitive to extreme values (see for example Doucet et al. (2001)), the tempered particle filter is more accurate in periods of high volatility but also more computationally expensive. Building on the work of Herbst and Schorfheide (2019), I modify the tempering schedule to take the asymmetry of the measurement density into account. This results in less tempering iterations and allows to increase the targeted accuracy of the estimated states by reducing the runtime of the filter.

Estimating the model using US data, the SSV model provides further statistical evidence that national financial conditions have an impact on the second and third moment of the conditional forecasting distribution of future US GDP growth. Furthermore, the tempered particle filter provides significant estimates of the time variation in the variance and skewness of the forecasting density of US GDP growth rates.

The results are also in line with other recent studies in the Growth at Risk literature such as delle Monache et al. (2021) or Montes-Galdon and Ortega (2022). With a Bayes ratio of 1612.18, the SSV model is strongly favoured by the data over a symmetric SV model. Comparing the conditional densities based on entropy measures developed in Adrian et al. (2019), the higher marginal likelihood of the SSV model can be attributed to its ability to better capture the strong increase in downside risks in periods of economic turmoil. These results provide further statistical evidence that asymmetries are an essential feature of the conditional densities of macroeconomic variables in times of economic crises.

The chapter is structured as follows: Section 3.2 provides an overview of the Growth at Risk concept and the methodology developed by Adrian et al. (2019). The Skewed Stochastic Volatility model and its estimation is introduced in Section 3.3 and 3.4 respectively. Section 3.5 and 3.6 discuss the results for the SSV model based on US data. Section 3.7 concludes.

[14] Flury and Shephard (2011) have documented that these estimation methods work well for SV models with symmetric densities in discrete time.

3.2 Growth at Risk

The concept of Growth at Risk was introduced in the seminal paper by Adrian et al. (2019) who analyse the variation of the one-period ahead forecast distribution of US GDP growth (gdp_{t+1}) conditional on the national financial conditions index (dubbed $nfci_t$) to analyse macro-financial risks in the US economy.[15] To obtain an estimate of the one-period ahead forecasting distribution of US-GDP, the authors develop a semi-parametric approach that consists of two steps.
In the first step, the 5, 25, 75 and 95% quantiles of the conditional distributions of gdp_{t+1} are estimated by running quantile regressions of the form

$$gdp_{t+1} = \beta_{0,\tau} + \beta_{1,\tau} nfci_t + \varepsilon_{t+1} \tag{3.1}$$

where the vector of parameters $\boldsymbol{\beta}_\tau = (\beta_{0,\tau}, \beta_{1,\tau})'$ depends on a predetermined quantile $\tau \in (0,1)$. The parameters are found by minimizing the Koenker Bassett loss (Koenker and Bassett (1978)) defined as

$$\mathcal{L}_{KB}(\boldsymbol{\beta}_\tau) = \sum \Big(\tau \cdot \mathbf{1}_{(gdp_{t+1} \geq x_t' \beta_\tau)} |gdp_{t+1} - x_t' \boldsymbol{\beta}_\tau| + (1-\tau) \cdot \mathbf{1}_{(gdp_{t+1} < x_t' \beta_\tau)} |gdp_{t+1} - x_t' \boldsymbol{\beta}_\tau| \Big)$$

where the vector $x_t = (1, nfci_t)'$ contains the explanatory variable plus intercept and $\mathbf{1}_A(...)$ denotes the indicator function. In a second step, the authors match the predicted 5, 25, 75 and 95 percent quantiles from the regressions to the theoretical moments of the skew T distribution developed in Azzalini (2013). The respective density function is defined as

$$skew\ T(y|\xi, \omega, \alpha, \nu) = \frac{2}{\omega} \cdot t(z|\nu) \cdot T(\alpha z | \nu + 1) \quad \text{with} \quad z = \frac{y - \xi}{\omega} \tag{3.2}$$

with ξ, ω, α, ν controlling the location, scale, shape and kurtosis of the distribution. The resulting densities are flexible and not constrained to be symmetric. Estimating the model based on US data from the 1970s up to 2017, Adrian et al. (2019) document the following properties of the one period ahead forecast distributions:

(1) Lower quantiles of the conditional forecast distribution vary a lot over time while the upper quantiles remain relatively stable.

(2) A deterioration of national financial conditions coincides with increases in the interquartile range and decreases the mean.

[15] The National Financial Conditions Index is given by the fist principal component of a large number of financial variables and released by the Federal Reserve Bank of Chicago.

(3) Distributions are more symmetric in normal times and become left skewed in recessionary periods.

Based on their work, the proposed two-step approach has also been applied to analyze time-varying forecast distributions of European growth rates (de Santis and van der Veken (2020)) or other macroeconomic variables such as inflation (López-Salido and Loria (2020)). However, while the semi-parametric approach by Adrian et al. (2019) provides coefficients of the quantile regressions in the first step, it does not provide a parametric law of motion that describes the evolution of the time variation in the volatility and asymmetry of the conditional forecast distributions. This prevents to directly capture the effect of the national financial conditions or any other explanatory variable a researcher might include on the second and third moment. Consequently, it is also not possible to conduct statistical inference to determine the statistical relevancy of different variables on the moments of the distributions. The same is true for the estimation uncertainty of the time-varying parameters of the skew T distribution at each point in time. Furthermore, without a parametric law of motion that governs the variation in the higher moments over time it is not possible to obtain multistep forecasts of the evolution of the conditional densities and predict future risks. [16] Last but not least, it is well known that standard quantile regressions based on the Koenker Bassett Loss do not insure monotonicity in the estimated coefficients of the different quantile regressions. Hence, the regression lines for different quantiles can intersect, a problem that is commonly denoted as quantile crossing (see for example Chernozhukov et al. (2010)). To remedy these shortcomings, this paper proposes to estimate the evolution of the full forecasting density in a fully parametric way using a Skewed Stochastic Volatility Model.

3.3 Skewed Stochastic Volatility Model

The Skewed Stochastic Volatility Model (SSV) model is a non-linear, non-Gaussian state space model with measurement equation

$$y_t = \gamma_0 + \sum_{l=1}^{L} \gamma_l x_{t,l} + \sum_{p=1}^{P} \beta_p y_{t-p} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim skew\, \mathcal{N}(0, \sigma_t, \alpha_t) \tag{3.3}$$

[16] Compared to the approach in this paper, an alternative way to obtain multistep forecasts using the semi-parametric approach of Adrian et al. (2019) could be to estimate a separate model to obtain a direct forecast for each horizon in the spirit of a local projections approach. However, this method fails to capture the uncertainty around the predictions.

and latent states

$$\ln(\sigma_t) = \delta_{1,0} + \sum_{j=1}^{J_\sigma} \delta_{1,j} x_{t,j} + \sum_{k=1}^{K_\sigma} \beta_{1,k} \ln(\sigma_{t-k}) + \nu_{1,t} \tag{3.4}$$

$$\alpha_t = \delta_{2,0} + \sum_{j=1}^{J_\alpha} \delta_{2,j} x_{t,j} + \sum_{k=1}^{K_\alpha} \beta_{2,k} \alpha_{t-k} + \nu_{2,t} \tag{3.5}$$

The innovations of the latent states $\nu_{1,t}$ and $\nu_{2,t}$ are assumed to be uncorrelated Gaussian White Noise

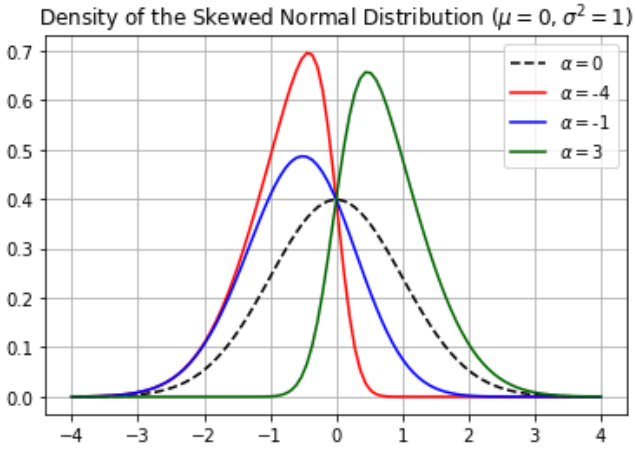
$$\begin{bmatrix} \nu_{1,t} \\ \nu_{2,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\nu_1} & 0 \\ 0 & \sigma_{\nu_2} \end{bmatrix} \right) \tag{3.6}$$

Most importantly, the errors in the measurement equation (3.3) are distributed according to the skew normal distribution of Azzalini (2013). Similar to the normal distribution it has parameters for the location (μ) and scale (σ) plus an additional shape parameter α that determines the symmetry of the density function. The probability density function of the skew normal distribution is given by

$$skew\ \mathcal{N}(y|\mu,\sigma,\alpha) = \frac{2}{\sqrt{(2\pi)}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \int_{-\infty}^{\alpha \frac{y-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\, \mathrm{d}z \quad \text{with} \quad z = \frac{y-\mu}{\sigma}. \tag{3.7}$$

Figure 3.3.1 shows how different values for α affect the skewness of the distribution function. While $\alpha < 0$ results in a left skewed distribution, $\alpha > 0$ tilts the distribution to the right. Setting $\alpha = 0$ recovers the symmetric Normal distribution. The evolution of the density

Figure 3.3.1: Skewed Normal Distribution for Different Values of the Shape Parameter α.



Notes: Negative values of α tilt the distribution to the left while positive values skew the distribution to the right. Setting $\alpha = 0$ recovers the standard normal distribution.

of the measurement errors over time is governed by state equations (3.4) and (3.5). The set of exogenous driving variables is not limited to be equal across Equations (3.3) - (3.5). Additionally, the lagged states can capture serial correlation in the evolution of the state variables. To guarantee the stability of the model I restrict the autoregressive coefficients such that

$$\sum_{p=1}^{P} \beta_p < 1 \quad \text{and} \quad \sum_{k=1}^{K_\alpha} \beta_{1,k} < 1 \quad \text{and} \quad \sum_{k=1}^{K_\sigma} \beta_{2,k} < 1.$$

Depending on the value of the static coefficients of the model, Equation (3.4) will affect the interquartile range of the conditional densities while Equation (3.5) introduces time-varying asymmetries.

The skew normal distribution can also capture excess kurtosis in the conditional densities of the SSV model. Given the derivations of Azzalini (2013) the kurtosis of the skew normal is a function of the scale and shape parameters. Hence, skew normal distribution can capture all characteristics found by Adrian et al. (2019) while maintaining a parsimonious modelling approach and without imposing skewness a priori. Furthermore, the SSV model nests a symmetric SV model with symmetric densities if $\delta_{2,0} = \delta_{2,j} = \beta_{2,k} = \sigma_{\nu_2} = 0 \quad \forall j, k$. Last but not least, given the estimation approach outlined in Section 3.4, it is also possible to allow for non-zero correlation of the error terms ν_1 and ν_1 by including their covariance in the set of static parameters that need to be estimated.

Estimating the model using Bayesian methods yields credible sets for both, the static parameters of the model as well as the value of the latent states at each time t. This allows to directly test the significance of the static and time-varying parameters. Additionally, it is easy to iterate the latent states forward in time to forecast the evolution of the conditional density of y_t several periods ahead.

The results of Adrian et al. (2019) quickly spawned a number of papers that are related to the research of this paper and that can be categorized in their parameter versus observation-driven modelling approach. As defined in Koopman et al. (2016), the observation-driven framework captures time-variation of parameters as deterministic functions of lagged dependent as well as exogenous variables. Prominent examples are the ARCH and GARCH models by Engle (1982) and Bollerslev (1986). In contrast to these models, the parameter-driven approach models time-varying parameters as independent stochastic processes with idiosyncratic errors. A notable example is the stochastic volatility model or the unobserved component models discussed in Durbin and Koopman (2012). Using a parameter driven approach, Carriero et al. (2020) work with a time-varying volatility specification in large BVAR that can capture variation in the second moment of the conditional forecast-

ing densities. Yet, skewness can only arise in the unconditional density of GDP growth. Montes-Galdon and Ortega (2022) analyze asymmetric macroeconomic risks in the Euro Area using an SVAR model with structural errors that follow a multivariate skew normal distribution with time varying shape parameters. However, the authors limit the effect of national financial conditions to the shape parameter of the structural errors and restrict the scale parameter to be constant over time. Similarly, Iseringhausen (2021) develops a panel model with time-varying skewness. Yet, national financial conditions are restricted to affect the skewness of the conditional distributions while volatility evolves as a random walk. Eventually, Hasenzagl et al. (2020) also estimate a time-varying parameter model using Hamiltonian Monte Carlo methods and a skew T distribution. The authors find that the time-varying moments cannot be estimated precisely and are generally insignificant. However, compared to the analysis in this paper, the authors focus on marginal forecasting gains of financial variables over measures of real activity.

In the observation driven framework, Adrian et al. (2019) already include a simple GARCH-type model that can capture time variation in the first two moments. Yet, resulting conditional forecast distributions remain constrained to be symmetric and Gaussian similar to Carriero et al. (2020). Using the Generalized Autoregressive Score (GAS) model framework developed by Creal et al. (2013), the paper by delle Monache et al. (2021) also allows for time varying skewness. Furthermore, older observation driven models that seek to model time-varying skewness are proposed in the work of Hansen (1994) or Engle and Manganelli (2004).

Compared to the other studies, an additional contribution of this paper is the Bayesian estimation approach of the model using advanced particle filtering methods recently introduced by Herbst and Schorfheide (2019). In contrary to the aforementioned studies, the SSV model and the Bayesian estimation strategy can capture all features documented by Adrian et al. (2019) while remaining flexible with regards to other distribution families and more complex model specifications.

The SSV model in this paper is a parameter driven model and estimated using a combination of particle filters and Markow Chain Monte Carlo methods. To the best of my knowledge, this is the first paper that estimates a SV model that also allows for time-varying skewness using state of the art particle filtering techniques in combination with MCMC methods. The estimation methods for the non-linear state space model are introduced in detail in the next section.

3.4 How to Estimate the Skewed Stochastic Volatility Model

Due to the skew normal distribution of the measurement errors as well as the non-linearity in the state equation of the scale parameter, the SSV model is a non-linear, non-Gaussian state space model. Therefore, one cannot estimate the states and static model parameters with methods such as Kalman filtering and the EM-algorithm. Yet, non-linear state space models can feasibly be estimated in a Bayesian framework using a combination of particle filters and MCMC-Algorithms (Schön et al. 2015). In particular, Kim et al. (1998) show how symmetric SV models can be estimated using a particle Metropolis Hastings algorithm. Furthermore, convergence results have been derived in Andrieu et al. (2010).

3.4.1 The Particle Metropolis Hastings Algorithm

In general a non-linear, non-Gaussian state space system consists of measurements y_t and latent states s_t that evolve according to the densities

$$y_t \sim p(y_t|s_t, \theta) \tag{3.8}$$

$$s_t \sim p(s_t|s_{t-1}, \theta). \tag{3.9}$$

In the SSV model, the latent states are the time-varying second and third moment of the measurement density determined by $s_t = (\ln \sigma_t, \alpha_t)$. The static model parameters are given in the vector

$$\begin{aligned}\theta = (&\gamma_0, \gamma_1, ..., \gamma_L, \beta_1, ..., \beta_P,\\ &\delta_{1,0}, \delta_{1,1}, ..., \delta_{1,J_\sigma}, \beta_{1,1}, ..., \beta_{1,K_\sigma}, \sigma_{\nu,1},\\ &\delta_{2,0}, \delta_{2,1}, ...\delta_{2,K_\alpha}, \beta_{2,1}, ..., \beta_{2,K_\alpha}, \sigma_{\nu,2})\end{aligned}$$

and determine the dynamics of the latent states and the mean equation of the observed measurements. Estimating the model using the the particle Metropolis Hastings Algorithm consists of two steps:

Step 1: Posterior distributions of the static model parameters θ are obtained with a Metropolis Hastings sampler that generates draws from the posterior distribution

$$p(\theta|y_{1:T}, s_{1:T}) = \frac{p(y_{1:T}|s_{1:T}, \theta)p(s_{1:T}|\theta)p(\theta)}{p(y_{1:T})}. \tag{3.10}$$

The MCMC step of the algorithm constructs a Markov Chain $\{\theta_n\}_{n=1}^N$ of length N with stationary distribution equal to the posterior $p(\theta|y_{1:T}, s_{1:T})$.

Step 2: The second step uses a particle filter to sequentially estimate the posterior distributions of the time-varying model parameters

$$p(s_t|y_{1:t}, \theta) = \frac{p(y_t|s_t, \theta)p(s_t|y_{1:t-1}, \theta)}{\int p(y_t|s_t, \theta)p(s_t|y_{1:t-1}, \theta)\mathrm{d}s_t} \tag{3.11}$$

$$\text{with } p(s_t|y_{1:t-1}, \theta) = p(s_t|s_{t-1}, \theta)p(s_{t-1}|y_{1:t-1}, \theta). \tag{3.12}$$

using importance sampling. Doucet et al. (2001) give a thorough treatment of importance sampling and particle filters. For each point in time t, M particles $\{s_{t,i}, W_{t,i}\}_{i=1}^M$ are drawn from a proposal density $q(s_t|y_{1:t})$ and resampled based on the importance weights

$$W_{t,i} = \frac{w_{t,i}}{\sum_{i=1}^M w_{t,i}} \quad \text{with} \quad w_{t,i} = \frac{p(s_{t,i}|y_{1:t}, \theta)}{q(s_{t,i}|y_{1:t})}. \tag{3.13}$$

In principle, the proposal density can be chosen freely, yet a convenient choice generates draws based on the mixture density

$$q(s_t|y_{1:t}, \theta) = \sum_{i=1}^M W_{t-1,i}p(s_{t,i}|s_{t-1,i}, \theta) \quad \text{with} \quad \sum_{i=1}^M W_{t-1,i} = 1 \tag{3.14}$$

Given that the process for s_t is markovian, this choice yields the standard bootstrap particle filter where particles are resampled proportional to the likelihood $w_{t,i} = p(y_t|s_{t,i}, \theta)$ of the measurement y_t given the proposed states $s_{t,i}$ as discussed in Doucet et al. (2001). At each t, particles are proposed given the stochastic process of the latent states and weighted based on how well they explain the observed measurement y_t.
Conditional on a draw of parameters θ_n, the particle filter generates an estimate of the

models likelihood function $p(y_{1:T}|s_{1:T}, \theta_n)$. This estimate is given by

$$\hat{p}(y_{1:T}|s_{1:T}, \theta_n) = \prod_{t=1}^{T} \frac{1}{M} \sum_{i=1}^{M} W_{t,i}. \tag{3.15}$$

The particle Metropolis Hastings Algorithm iterates between estimating $\hat{p}(y_{1:T}|s_{1:T}, \theta_n)$ given a draw θ_n with the particle filter and drawing a new vector of static model parameters θ_{n+1} from the posterior distribution $p(\theta|y_{1:T}, s_{1:T})$ using $\hat{p}(y_{1:T}|s_{1:T}, \theta_n)$ for the Metropolis Hastings step. Increasing the length N of the Markov chain as well as the number of particles M improves the accuracy of the estimation but results in more computational work and longer run times. As shown by Andrieu et al. (2010), the distribution of the resulting chain $\{\theta_n\}_{n=1}^{N}$ converges to the exact posterior distribution $p(\theta|y_{1:T}, s_{1:T})$ even if the likelihood function is estimated using the particle filter. Convergence results for the particle filter can be found in Doucet et al. (2001) or Chopin (2004).

Yet, it is well known that the proposal distribution of the bootstrap particle filter given by Equation (3.14) is suboptimal since it ignores information about the states s_t contained in y_t (Herbst and Schorfheide 2019).

For example, given the SSV model a large increase of US GDP growth will be more likely under particles that suggest high values in volatility as well as positive skewness in the next period. However, as particles are proposed conditional on $s_{t-1,i}$ only a few particles will imply large values for the latent states if volatility and skewness are small in $t-1$. Since most proposed particles will have a low likelihood under the model, this results in a high variance of the normalized weights $W_{t,i}$. Consequently, only a few particles are resampled which leads to a poor approximation of the filtering density $p(s_t|y_{1:t}, \theta)$ and the likelihood of the static parameters $\hat{p}(y_{1:T}|s_{1:T}, \theta_n)$. This phenomenon is commonly referred to as weight degeneracy (Pitt and Shephard 1999). Conversely, a smaller variance of $W_{t,i}$ implies a more uniform distribution of the weights $W_{t,i}$ and will lead to a better importance sampling approximation.

As discussed in Pitt et al. (2012) even though the particle MCMC algorithm is generally unbiased, the quality of the likelihood approximation of the particle filter is crucial for its efficiency. A low approximation accuracy results in slow mixing properties of the markow chains which yields high rejection ratios and a slow convergence to the ergodic distribution (see for example Flury and Shephard (2011)).

A common measure to gauge the accuracy of the particle approximation at time t is the

inefficiency ratio

$$Ineff_t = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{w_{t,i}}{\frac{1}{M} \sum_{i=1}^{M} w_{t,i}} \right)^2 \tag{3.16}$$

where $w_{t,i}$ are the unormalized weights.[17] A high inefficiency ratio indicates that the distribution of weights is uneven such that the approximation of the target distribution is bad, while an inefficiency ratio close to one indicates evenly distributed weights and a good approximation of $\hat{p}(s_t|y_{1:t}, \theta)$ and $\hat{p}(y_{1:T}|s_{1:T}, \theta_n)$.

The recently introduced tempered particle filter by Herbst and Schorfheide (2019) controls the inefficiency ratio by sequentially adjusting the proposal distribution in each period. This greatly improves the accuracy of the estimated states and leads to a better approximation of the likelihood function in the Metropolis Hastings step. Building on annealed importance sampling that was first proposed by Neal (2001), it is a more complex but also more accurate filtering algorithm. Given the aim of this paper to estimate tail risks, the ability to handle outliers and extreme values better then standard particle filters makes the tempered particle filter a suitable method to obtain precise estimates of the variation in the higher moments. This property of the tempered particle filter is further elaborated in the next section.

3.4.2 Adjusting the Tempered Particle Filter

The tempered particle filter proposed by Herbst and Schorfheide (2019) adjusts the proposal distribution to the observation y_t using an adaptive version of annealed importance sampling procedure for each t. Instead of directly reweighting the particles drawn from $p(s_t|s_{t-1,i}, \theta)$ proportional to the likelihood function $p(y_t|s_{t,i}, \theta)$, the particles are sequentially adapted to a more optimal proposal via a sequence of N_ϕ bridge distributions. These bridge distributions are defined by the "tempered" likelihood function $p_0(y_t|s_{t,i}, \theta)$. The tempered likelihood function has an inflated variance defined as $\sigma_{t,i}/\phi_n$ with $0 < \phi_n < 1$. Intuitively, the likelihood function is initially "flattened" to ensure that the weights of the proposed particles are equal. As described in Herbst and Schorfheide (2019), the variance of the measurement equation is then sequentially reduced to its actual level while targeting a user-defined inefficiency ratio r^*. Concurrently, the particles are adapted to a better proposal distribution using a combination of importance sampling and MCMC methods (for a detailed description of the algorithm see Herbst and Schorfheide (2019), Herbst and Schorfheide (2014) or Godsill and Clapp (2001) for a simpler outline of the basic idea). In the context

[17] It can be shown, if one can draw particles from the optimal proposal density $p(s_t|y_t, s_{t-1})$ the weights become $w_{t,i} = \frac{1}{M}$ $\forall i$ which gives $Ineff_t = 1$. In general this distribution is not available in closed form since it requires the distributions of the measurement and states to be conjugate to each other.

of the SSV model this means that for each t, the volatility is assumed to be large and subsequently shrunk towards a level that fits the data best during the tempering steps. For each tempering step a new value for ϕ_n is chosen as

$$\phi_n = \underset{\phi}{\operatorname{argmin}} \; Ineff(\phi) - r^* = 0 \tag{3.17}$$

until $\phi_{N_\phi} = 1$. If the set of particles proposed based on Equation (3.14) satisfies $Ineff(1) \leq r^*$ no tempering is required and the filtering step is equal to classic bootstrap particle filtering. Targeting a lower r^* will result in a better approximation of the latent states, but comes at the price of more tempering steps and a longer runtime.
Following the reasoning in Herbst and Schorfheide (2019), I modify the adaptive tempering schedule such that the asymmetry of $p_n(y_t|s_{t,i}, \theta)$ is also taken into account. Since the skew normal distribution converges to a symmetric normal distribution for $\alpha \to 0$, starting from a symmetric and flat distribution will result in a value for ϕ_0 that is closer to 1 reducing the number of required tempering iterations. At each step, the targeted inefficiency ratio $r*$ can be achieved with a higher value for ϕ_n. More formally, in the SSV model the unnormalized weights $w_{t,i}(\phi_0)$ are given by

$$w_{t,i}(\phi_0) = \frac{2\phi_0^{1/2}}{\sqrt{2\pi}\sigma_{t,i}} \exp\left(\frac{-\phi_0(y_t - \mu_t)^2}{2\sigma_{t,i}^2}\right) \int_{-\infty}^{\alpha_{t,i}\phi_0^{3/2}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right) \mathrm{d}z. \tag{3.18}$$

Compared to the unnormalized weights obtained from a normal distribution, the integral in Equation (3.18) introduces additional variation to $w_{t,i}$ which increases the inefficiency ratio. The tempering parameter ϕ_0 shrinks the upper bound of the integral towards 0 for $\lim \phi_0 \to 0$ and brings $p_0(y_t|s_{t,i}, \theta)$ closer to a normal distribution. Using Equation (3.16) and taking the limit for ϕ_0 shows that the Inefficiency Ratio is decreasing in ϕ and bounded from below by

$$\lim_{\phi_0 \to 0} \; Ineff_t(\phi_0) = \frac{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{\sigma_{i,t}}\right)^2}{\left(\frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma_{i,t}}\right)^2} > 1$$

by Jensen's inequality.[18] This is a special feature of the stochastic volatility model and differs from the lower bound derived in Herbst and Schorfheide (2019) for the DSGE model case. Without the stochastic volatility component, the lower bound is given by $r^* = 1$.

[18]Note that this also holds for a stochastic volatility model with symmetric densities.

Hence, I define the target inefficiency ratio as

$$r^* = \frac{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{\sigma_{i,t}}\right)^2}{\left(\frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma_{i,t}}\right)^2} + \Delta_r$$

where Δ_r is set by the researcher to determine the accuracy of the filter.
Given Expression (3.7) and Algorithm 2 in Herbst and Schorfheide (2019), the expression for the unnormalized weights at the n^{th} tempering step is given by

$$\tilde{w}_{t,i}(\phi_n) = \left(\frac{\phi_n}{\phi_{n-1}}\right)^{\frac{1}{2}} \exp\left(\frac{-(\phi_n - \phi_{n-1})(y_t - \mu_t)^2}{2\sigma_{t,i}^2}\right)\tilde{\Lambda}_{t,i}(\phi_n) \tag{3.19}$$

with

$$\tilde{\Lambda}_{t,i}(\phi_n) = \frac{\int_{-\infty}^{\alpha_{t,i}\phi_n^{2/3}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z}{\int_{-\infty}^{\alpha_{t,i}\phi_{n-1}^{2/3}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z} \tag{3.20}$$

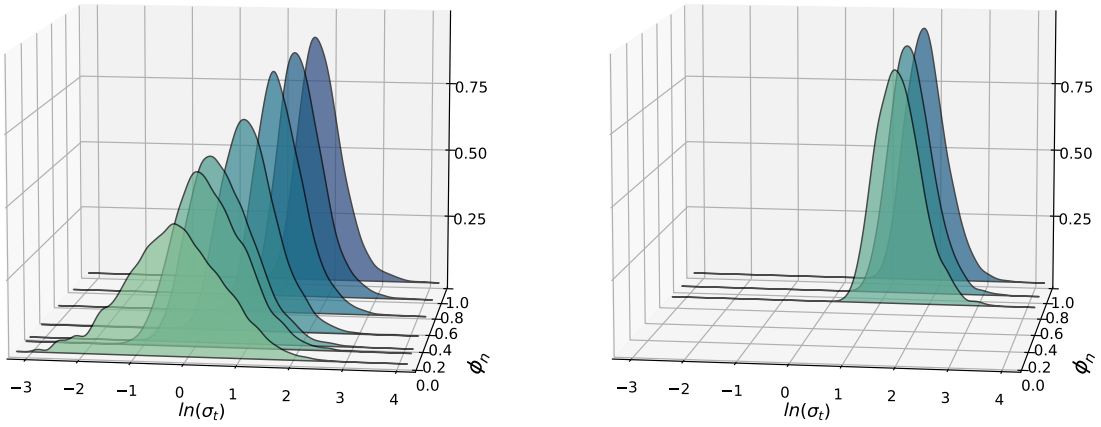
Once again, Expressions (3.19) and (3.20) show that the weights of the skew normal distribution differ from a symmetric normal density by a factor $\tilde{\Lambda}_{t,i}(\phi_n)$ that is greater or smaller than one depending on the sign of $\alpha_{t,i}(y_t - \mu_t)$ (see Appendix 3.A). This introduces additional variation into the weights and increases the inefficiency ratio for a given ϕ_n compared to the weights from a standard normal distribution. Since the limit for ϕ_n is given as

$$\lim_{\phi_n \to 0} \tilde{\Lambda}_{t,i}(\phi_n) = 1, \tag{3.21}$$

additionally tempering the skewness of the likelihood brings this factor closer to one and results in weights that are more uniform and less tempering steps.

The example in Figure 3.4.1 illustrates the idea of tempering and provides a comparison of the two different tempering schedules based on simulated data from the SSV model. The panels shows an approximation of the filtering distribution based on the proposed particles for $\ln(\sigma_{i,t}^2)$ at each tempering step. In each iteration the particles are reweighted and adjusted to the final measurement. The particles are more spread out in the beginning and slowly moved to the final filtering density at $\phi_{N_\phi} = 1$. The improvements of the skewness tempering are obvious. If the skewness of the measurement distribution is not taken into account, the mean of the filtering distribution is mutated from values of -0.2 to a final value of approximately 1.2 in 7 tempering steps. Yet, if the skewness is tempered as well it only takes 3 tempering steps and the approximation of the filtering density is more accurate at

Figure 3.4.1: Tempering of the State Distributions of $\ln(\sigma_{t,i}^2)$



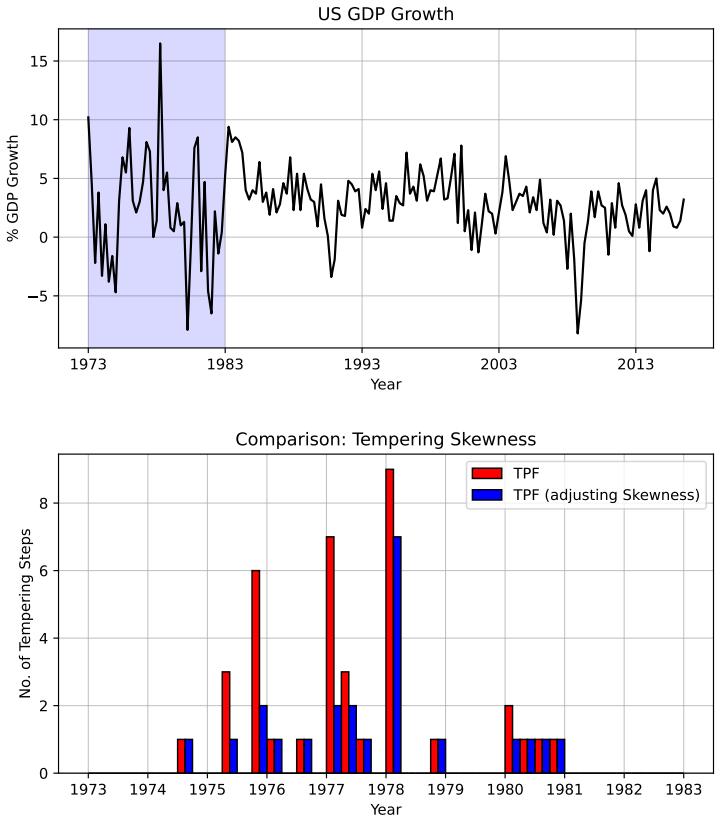
Notes: Illustration of tempering based on simulated data. The mean of the distribution moves from -0.2 (left) or 1.5 (right) to about 1.9. While tempering only the scale of the distribution (left) requires 7 iterations, additionally tempering the shape parameter $\alpha_{t,i}$ reduces the tempering steps to only 3 iterations (right). Furthermore, in case of skewness tempering the optimal ϕ_0 is much closer to one.

the beginning of the tempering. Furthermore, with an optimal value of 0.67, the initial ϕ_0 is already much closer to 1 in case of skewness tempering.

Figure 3.4.2 shows a comparison of the total number of tempering steps for US data from 1973 to 1983 for the tempered particle filter with and without skewness tempering. The filter is run using the model introduced in Section 3.3 and the estimated parameters from Section 3.5. As highlighted in the upper panel, this decade represents a particularly volatile period of the sample with large jumps in US GDP growth rates. The lower panel shows the number of tempering steps that are adaptively chosen in each time period t. Tempering increases especially if the values of subsequent observations are far apart. In periods where the standard tempering schedule requires a large number of tempering steps, additionally tempering the skewness is most effective. For example, in 1975Q2, 1975Q4 or 1977Q1 skewness tempering yields a similar reduction of tempering steps as with simulated data in Figure 3.4.1. The number of tempering steps required in 1975Q2, 1975Q4 decreases by more than 60%. While exclusively tempering the scale of the measurement density takes six or seven tempering iterations additionally tempering the symmetry of the distribution requires only two iterations.[19]

[19]Additionally, a comparison of the performance of the Bootstrap and the Tempered Particle Filter based on simulated data from the SSV model is given in 3.C. The tempered particle filter leads to a clear improvement of the MSE for both filtered states.

Figure 3.4.2: Total Number of Tempering Steps for both Tempering Variants



Notes: Tempering steps increase during times of high volatility. The plot shows that additionally tempering the shape of the measurement density requires fewer tempering steps.

3.4.3 Data and Priors

The proposed model is estimated on the same data set as used by Adrian et al. (2019). Measurements y_t are chosen to be one-period ahead realizations of GDP growth (gdp_{t+1}). To compare the model with the results of Adrian et al. (2019), I use contemporaneous realizations of the national financial conditions index $nfci_t$ as exogenous driving variable in the measurement and state equations. This introduces time variation of the variance and skewness of the conditional densities of the one-period ahead US GDP growth. The lag order of the latent states is determined using the Bayes Ratio as selection criteria. I use a mixture of uninformative and informative priors on the static parameters. Table 3.D.1 in Appendix 3.D gives a comprehensive overview of the prior specification of the static parameters as well as the data.

The tempered particle filter is tuned to use $M = 10,000$ particles with a targeted inefficiency ratio $\Delta_r = 0.01$ and 2 mutation steps in each tempering iteration. Draws for the static

model parameters are generated using a standard random walk proposal with four chains ran in parallel on the HPC-Cluster at the Freie Universität. To increase the efficiency of the Metropolis Hastings algorithm the constrained parameters such as the autoregressive coefficients $-1 < \beta < 1$ and the variances $\sigma_{\nu_i} > 0$ are mapped to the real line using the following transformations

$$\beta = \tanh(\psi) \in [-1, 1] \tag{3.22}$$

$$\sigma = \exp(\zeta) \in \mathbb{R}^+ \tag{3.23}$$

where ψ and ζ can be drawn from the set of real numbers $\mathbb{R}$. This allows to obtain samples from the transformed unconstrained target distribution of

$$\begin{aligned}\tilde{\theta} = (&\gamma_0, \gamma_1, ..., \gamma_L, \psi_1, ..., \psi_P, \\ &\delta_{1,0}, \delta_{1,1}, ..., \delta_{1,J_\sigma}, \psi_{1,1}, ..., \psi_{1,K_\sigma}, \zeta_{\nu,1}, \\ &\delta_{2,0}, \delta_{2,1}, ...\delta_{2,J_\alpha}, \psi_{2,1}, ..., \psi_{2,K_\alpha}, \zeta_{\nu,2}) \in \mathbb{R}^S\end{aligned}$$

were $S = 5 + L + P + J_\sigma + K_\sigma + J_\alpha + K_\alpha$. As described in Schön et al. (2015) this requires to correct the acceptance ratio for the Jacobians of the inverse functions

$$\frac{\mathrm{d}\tanh^{-1}(\psi)}{\mathrm{d}\psi} = \frac{1}{1-\psi^2} \quad \text{and} \quad \frac{\mathrm{d}\log(\zeta)}{\mathrm{d}\zeta} = \frac{1}{\zeta}$$

based on the change of variables rule. The posterior distributions of the constrained model parameters can then be recovered using Equations (3.22) and (3.23). To further improve the mixing properties of the chains an initial estimate of $Var(\tilde{\theta}) = \Omega$ is obtained based on a pre-run with 5000 draws. The proposal variance is scaled to target an acceptance ratio between 20% and 30% as suggested in Roberts and Rosenthal (2001).

3.5 Results

Table 3.5.1 presents the estimates for the static model parameters from the particle Metropolis Hastings algorithm described in the previous section. The estimated coefficient γ_1 gives a negative impact of $nfci_t$ of about -0.69 on the one period ahead realization of gdp_{t+1}. Furthermore, with an impact of 0.24 on the scale ($\delta_{1,1}$) and -0.29 on the shape ($\delta_{2,1}$) of the skew normal distribution, the effect of national financial conditions on the different moments of the conditional densities is in line with the stylized facts described in Adrian

Table 3.5.1: Posterior Means, Standard Deviations (SD) and 68% and 90% Credible Sets

Model Parameter	Mean	SD	q16	q84	q05	q95
γ_0	2.285	0.398	1.898	2.672	1.623	2.94
γ_1	-0.686	0.362	-1.045	-0.335	-1.311	-0.119
$\delta_{1,0}$	0.865	0.285	0.573	1.164	0.446	1.372
$\delta_{1,1}$	0.242	0.096	0.147	0.338	0.102	0.412
$\beta_{1,2}$	0.108	0.278	-0.192	0.396	-0.375	0.522
$\delta_{2,0}$	0.218	0.221	0.006	0.429	-0.143	0.595
$\delta_{2,1}$	-0.290	0.226	-0.477	-0.103	-0.603	0.042
σ_{ν_1}	0.092	0.059	0.037	0.14	0.023	0.209
σ_{ν_2}	0.020	0.020	0.006	0.032	0.004	0.058

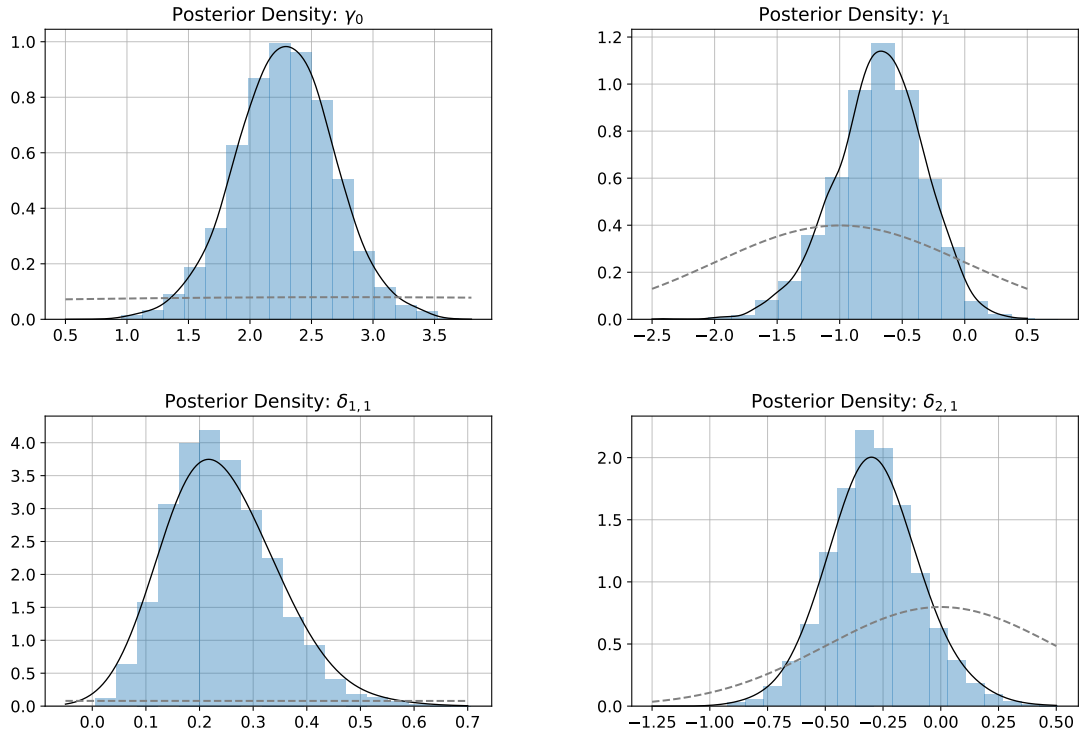
Notes: The model is estimated using $N = 20,000$ draws of the tempered particle Metropolis Hastings Algorithm. The first half of the sample is discarded as burn in. The model specification containing a lagged value of α_t in the state equation of the shape parameter is strongly rejected against a model specification without an autoregressive term based on a Bayes Ratio of about 200. The marginal data densities are estimated using the modified harmonic mean estimator of Geweke (1999))

et al. (2019). Consequently, the estimated coefficients of the SSV model imply an inverse relationship of the second and third moment of the conditional densities. As financial conditions deteriorate, the expected growth rate decreases while the interquartile range widens and downside risks of future GDP growth increase.

However, while the 90% credible sets of the coefficients for the effect of current national financial conditions on the mean and variance of gdp_{t+1} do not overlap the zero, this is not the case for the effect on the shape parameter. Significance of the impact of $nfci_t$ on the skewness of the conditional densities is only given by the 68% credible set of $\delta_{2,1}$. This raises the question about the importance of time-varying asymmetry of the conditional densities, a topic that will be further investigated in Section 3.6. Additionally, Figure 3.5.1 shows the sample approximations and prior distributions of the parameters γ_0 and γ_1 in the measurement equation as well as the parameters that capture the effect of $nfci_t$ on the shape and scale of the conditional distribution of gdp_{t+1}. All posteriors are well-behaved, uni-modal and clearly centered away from zero. The distributions also differ sufficiently from the priors indicating that the effects are well identified by the data.

Given the posterior mean estimates of the static model parameters, Figure 3.5.2 shows the filtered states and the respective 68% and 90% credible sets, an additional feature of the SSV model that is not available with the quantile approach of Adrian et al. (2019). The sharp increase in volatility and downside risks in the 1980s as well as during the financial crises in 2009 is well captured by the evolution of the two latent states. Compared to the results in Hasenzagl et al. (2020), both states indicate significant time variation in

Figure 3.5.1: Posterior Distributions Obtained Using the Particle MCMC Algorithm



Notes: Posterior distributions for the parameters of the measurement equation γ_0 and γ_1 as well as the parameters that capture the effect $nfci_t$ on the scale ($\delta_{1,1}$) and shape ($\delta_{2,1}$). All posteriors are well-behaved, unimodal and clearly centered away from zero. The coefficients γ_0, γ_1 and $\delta_{1,1}$ are significantly different from zero based on the 90 % credible set constructed from the posterior draws. For $\delta_{2,1}$ the 68% credible set does not overlap the 0. Grey dashed lines indicate the prior distributions.

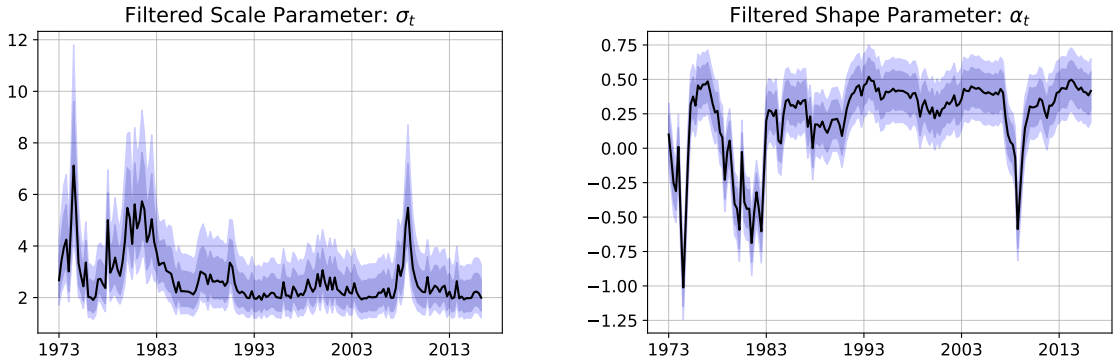
the second and third moments of the conditional densities of gdp_{t+1} based on the 90% credible sets. Yet, Hasenzagl et al. (2020) approach includes autoregressive components of GDP growth to investigate time-varying asymmetries from a forecaster's perspective who seeks to obtain additional forecasting gains by including financial variables. While the aim of this paper remains close to the work of Adrian et al. (2019), the SSV model and estimation algorithm can easily be extended to further investigate these questions and provide a further comparison with the results of Hasenzagl et al. (2020).[20] In general, the plots show that the proposed model is able to capture the stylized facts given in Section 3.2. However, even though the distribution is more symmetric in normal times, the estimated state of α_t also exhibits significant levels of positive skewness when levels of volatility are low. Based on the estimated states, significant upside risks to GDP growth occur during

[20]Including real macroeconomic variables such as for example the unemployment rate is a natural step left for further research.

times of economic moderation for example during the late 1980s and early 1990s. These results are similar to recent results of delle Monache et al. (2021) who find evidence for cyclical behavior in the shape of the one-step ahead conditional forecast densities. Based on a trend-cycle decomposition of the latent states in their model, the authors find that conditional forecasting distributions of future US GDP growth do not only exhibit negative skewness during recessions, but become positively skewed in expansionary periods.

Eventually, Figure 3.5.3 shows the resulting conditional densities of the estimated model

Figure 3.5.2: Filtered States Obtained with the Tempered Particle Filter



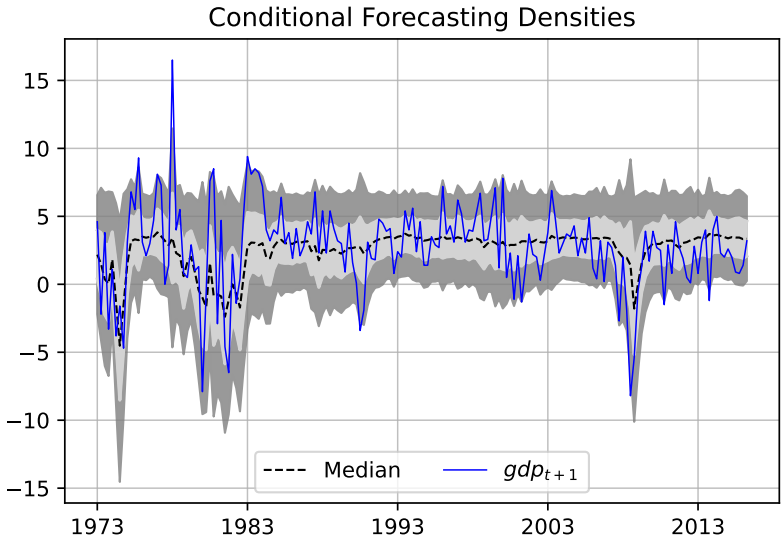
Notes: The filter is tuned to target an inefficiency ratio with $\Delta_r = 0.01$, 2 Mutation steps and $M = 10,000$ particles. The posterior means in table 3.5.1 are used for the value static model parameters. The shaded areas give the respective 68% and 90% credible sets obtained from the approximations of the filtering distributions.

with the respective lower and upper 5% and 25% quantiles. The effect of the strong increase in the scale and shape parameters during the Great Recession as well as during the oil crises in the 1970s and 80s is clearly visible in the behavior of the lower quantiles. Similar to stylized fact (1) in 3.2, the upper quantiles of the conditional forecasting distributions remain relatively stable while the lower quantiles vary significantly over time.

To further illustrate the difference in variation of the tail risks, Figure 3.5.4 shows the expected shortfall SF_t and expected longrise LR_t for various probability levels q. The two measures give the expected GDP growth under a specific distribution and chosen probability level q. Given the inverse CDF of the skewed Normal distribution together with the estimated parameters of the SSV model denoted by $F^{-1}_{y_{t+1}|\hat{\mu}_t,\hat{\sigma}_t,\hat{\alpha}_t}(y)$ and a chosen target probability $q \in [0,1]$ values for SF_t and LR_t are calculated as

$$SF_t(q) = \frac{1}{q}\int_0^q F^{-1}_{y_{t+1}|\hat{\mu}_t,\hat{\sigma}_t,\hat{\alpha}_t}(y)\mathrm{d}y \quad \text{and} \quad LR_t(q) = \frac{1}{q}\int_{1-q}^1 F^{-1}_{y_{t+1}|\hat{\mu}_t,\hat{\sigma}_t,\hat{\alpha}_t}(y)\mathrm{d}y \tag{3.24}$$

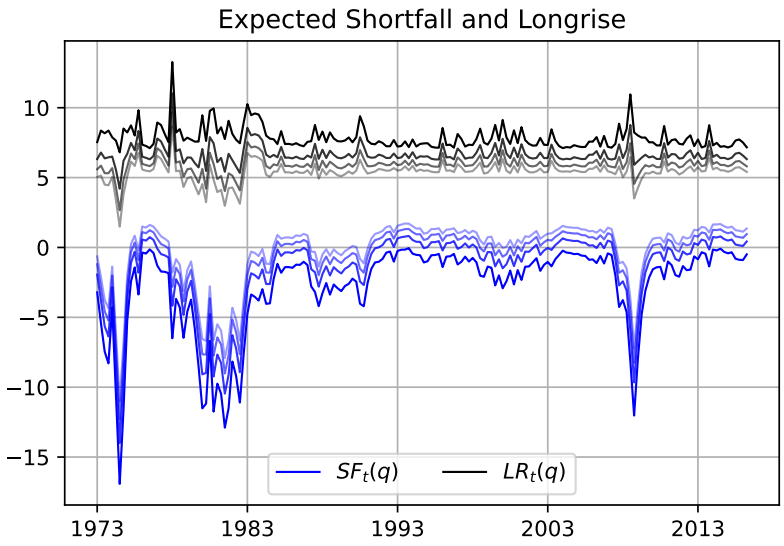
Figure 3.5.3: Conditional Distribution for One-Step Ahead GDP Growth



Notes: Lower and upper 5% and 25% percent quantiles display the same characteristics as found by Adrian et al. (2019). While the upper quantiles remain relatively stable, the lower quantiles vary strongly over time indicating increased downside risks to GDP growth in times of financial distress.

Similar to Value at Risk, expected Shortfall and Longrise measure tail risks under a given probability distribution. However, since they are an average over the outcomes up to a certain quantile rather than just the upper or lower bound, the behavior of the full tail is captured more comprehensively. Figure 3.5.4 shows that the complementary effects on the different moments of the distribution result in downside risks that are larger in size and vary more strongly over the full sample period. Especially the oil price shocks in the 1970s and early 1980s cause a large variation in downside risks compared to upside risks captured by the expected Longrise. During the Great Moderation the variation in the tails becomes more equal. The differences become more pronounced again during the financial crises in 2008 and 2009. Figure 3.5.4 shows that this feature is not only valid for the lower and upper 5% but also for the $15, 25$ and 35% levels. However, compared to the results of Adrian et al. (2019) who find values of approximately -18% for the expected 5% shortfall during the early 1980s, the predicted tail risks of the SSV model are less severe.

Figure 3.5.4: Expected Longrise and Shortfall



Notes: The plot shows expected shortfall/longrise for the $q = 5, 15, 25, 35$ percent quantiles. It gives the expected GDP growth in the worst/best q percent of the outcomes under the estimated skew normal distribution for each period. The plot shows that risks to the lower tails are larger in size and vary much more compared to upside risks to GDP growth.

3.6 Does Skewness Matter?

The estimation results of the SSV model yield significant coefficients for effect of national financial conditions on the the mean and the variance. However, in case of the skewness parameter the 90% credible sets for the static parameters overlap the zero. Furthermore, as argued in papers such as Adrian et al. (2020) or Carriero et al. (2020) the different behavior of the upper versus the lower quantiles can also be attributed to the inverse relationship of the mean and variance of the conditional distributions. For example, in the New Keynesian Volatiltiy Model of Adrian et al. (2020), symmetric conditional densities generate non-zero skewness in the unconditional distribution of GDP growth that closely aligns with the values observed in real world data. Additionally, based on the results of Brownlees and Souza (2021) symmetric GARCH models yield similar predictive gains as quantile regressions that allow for skewness when forecasting macroeconomic tail risks. This suggests that the inverse relationship of the mean and variance can capture much of the variation in downside risks. To further investigate the importance of asymmetries in the conditional densities, I estimate a symmetric SV model and compare it with the SSV model based on the Bayes ratio. Since the SV model is nested in the SSV model, this corresponds to testing a joint parameter

constraint on the latent state. The prior and posterior distributions of the estimated SV model are given in Appendix 3.E. Since both models are estimated using Bayesian methods, the two models can easily be compared using their Bayes factor. The Bayes factor determines which model is favored by the data based on the marginal data densities $p(y|\mathcal{M}_i)$. Let $\mathcal{M}_1$ be the SSV model and $\mathcal{M}_2$ denote the symmetric SV model. The Bayes factor is given by the ratio of the two marginal data densities $p(y|\mathcal{M}_i)$

$$\frac{p(y|\mathcal{M}_1)}{p(y|\mathcal{M}_2)} = \frac{\int p(y|\theta, \mathcal{M}_1)p(\theta|\mathcal{M}_1)\mathrm{d}\theta}{\int p(y|\theta, \mathcal{M}_2)p(\theta|\mathcal{M}_2)\mathrm{d}\theta} \tag{3.25}$$

The Bayes factor indicates which model describes the observed data better. In general, a value larger then one indicates that $\mathcal{M}_1$ fits the data better, while values smaller than one imply a better model fit of $\mathcal{M}_2$. The marginal data densities $p(y|\mathcal{M}_i)$ are estimated using the the modified harmonic mean estimator of Geweke (1999).[21] Based on the table of Kass and Raftery (1995) a Bayes factor lager than 10 indicates strong evidence in favor of $\mathcal{M}_1$, a Bayes factor larger than 100 implies decisive evidence. Values lower then three are considered to be inconclusive. Table 3.6.1 gives the results for the respective quantities. With a value of 1612.18 the SSV model is clearly favored by the data. Given the alterntive

Table 3.6.1: Model Selection and Evaluation

Bayes Factor	log Odds	$\log p(y\|\mathcal{M}_1)$	$\log p(y\|\mathcal{M}_2)$
1612.18	7.39	-435.78	-443.16

Notes: Bayes Factor and the log of the marginal data densities for the SSV and the SV-Model. The Bayes factor gives decisive evidence for the SSV-Model.

of a symmetric SV model, this implies that allowing for time varying skewness in the conditional densities of gdp_{t+1} increases the model fit.
Based on these results, I also compute the upside and downside entropy $\mathcal{L}_t^U$ and $\mathcal{L}_t^D$ defined in Adrian et al. (2019) to further analyze the differences of the SSV and the SV model. The upside and downside entropy compare the divergence in the probability mass of two probability distributions above and below the median. As described in Adrian et al. (2019), it is a relative measure of the divergence between two distributions in the upper and lower

[21]The Harmonic Mean Estimator is given by

$$\hat{p}(y) = \frac{1}{N}\sum_{n=1}^{N}\frac{f(\theta_n)}{p(y|\theta_n)p(\theta_n)} \tag{3.26}$$

were $\{\theta_n\}_{n=1}^N$ are the draws obtained from the Metropolis Hastings sampler.

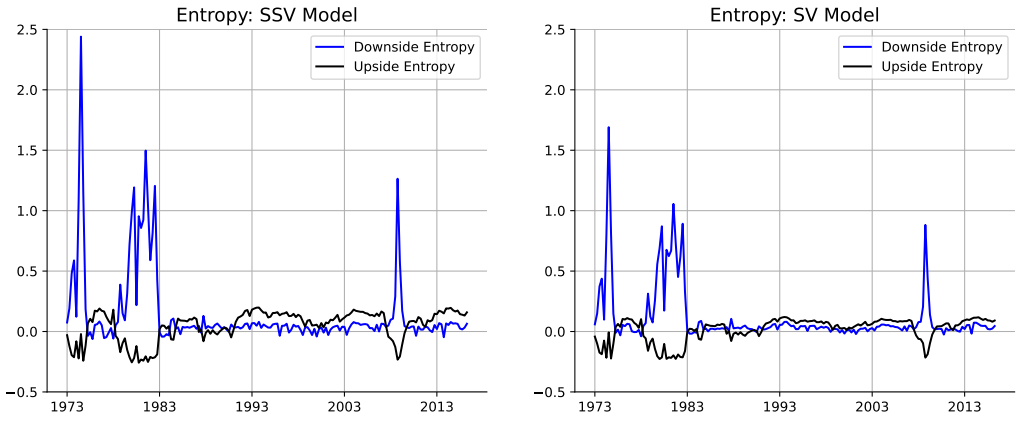
tails. Formally, $\mathcal{L}_t^U$ and $\mathcal{L}_t^D$ are given by

$$\mathcal{L}_{\mathcal{M}_i}^U = -\int_{-\infty}^{\hat{F}_{\mathcal{M}_i}^{-1}(0.5)} (\log \hat{g}(y) - \log \hat{f}_{\mathcal{M}_i}(y))\hat{f}_{\mathcal{M}_i}(y)\mathrm{d}y \tag{3.27}$$

$$\mathcal{L}_{\mathcal{M}_i}^D = -\int_{\hat{F}_{\mathcal{M}_i}^{-1}(0.5)}^{\infty} (\log \hat{g}(y) - \log \hat{f}_{\mathcal{M}_i}(y))\hat{f}_{\mathcal{M}_i}(y)\mathrm{d}y \tag{3.28}$$

where $\hat{g}(y)$ denotes the fitted unconditional density of GPD growth. The densities $\hat{f}_{\mathcal{M}_i}(y)$ are the conditional densities of the SSV model ($\mathcal{M}_1$) and the SV model ($\mathcal{M}_2$) given the estimated time-varying parameters. Upside entropy $\mathcal{L}_{\mathcal{M}_i}^U$ becomes positive in a given period if more probability mass is shifted to the upper tail compared the the unconditional distribution of GDP growth and vice versa. Similarly, high values for downside entropy $\mathcal{L}_{\mathcal{M}_i}^U$ indicate that under model $\mathcal{M}_i$, there is more probability mass in the lower tail compared to the unconditional distribution. Figure 3.6.1 shows $\mathcal{L}_{\mathcal{M}_i}^U$ and $\mathcal{L}_{\mathcal{M}_i}^U$ for both models. The

Figure 3.6.1: Upside and Downside Entropy of the SSV Model and the SV Model



Notes: Upside and Downside Entropy for the SSV model (left) and the SV model (right). Relative to the unconditional distribution, the SSV model shifts more of the probability mass to the downside compared to the SV model indicating higher tail risks.

two plots indicate that downside entropy for the SSV model is much higher in times of economic crisis compared to SV model. Yet, the upside risks of the SSV model during the Great Moderation are visible but less pronounced compared to the downside risks during economic crises. With regards to the upside entropy, both models are fairly similar. This indicates that modelling asymmetries matters especially to appropriately capture the risks to the lower tails of the distribution. This is in line with other recent finding in the literature of asymmetric forecasting distributions. For example Montes-Galdón et al. (2022)

find that introducing information on the asymmetry of forecasting densities in a BVAR strongly improves the probabilistic forecasts of GDP, inflation and core inflation in the Euro Area during times of economic crisis while there are no substantial gains in periods with stable economic conditions. In general, these results motivate future research to address the question of different risk regimes of macroeconomic variables. While models such as the SSV model can capture these characteristics, the implications of risk regimes suggest that markov switching models with exogenous driving variables that impact the transition probabilities could be another class of models to explore in subsequent studies.

3.7 Conclusion

This paper proposes a Skewed Stochastic Volatility (SSV) model as an alternative method to estimate Growth at Risk as introduced by Adrian et al. (2019). The SSV model can capture variation in the second and third moments of the conditional forecast distributions of US GDP growth and allows researchers to estimate and conduct statistical inference on the estimated parameters. The resulting state space model is non-linear with non-Gaussian errors and can be estimated with a Particle MCMC algorithm. I obtain accurate estimates of the model likelihood and the evolution of the latent states, using the tempered particle filter introduced by Herbst and Schorfheide (2019). Building on the adaptive tempering schedule proposed by the authors, I modify the tempering schedule to take the asymmetry of the distribution of the measurement error into account. This reduces the number tempering steps to save computational time while achieving the same accuracy. Estimating the model based on US data yields conditional forecast densities that closely resemble the findings by Adrian et al. (2019).

Exploiting the advantages of the proposed model, I find that national financial conditions have an effect on the moments of the forecasting distribution. The tempered particle filter provides significant estimates of the variation in the variance and skewness over time that imply a strong positive relationship between volatility and downside risks. Times with high volatility in growth rates coincide with an increase in risks to the lower tail of the conditional distributions. My results are also in line with results of other recent studies such as Montes-Galdon and Ortega (2022) or delle Monache et al. (2021). Compared to the findings of Hasenzagl et al. (2020), my results indicate that there is predictive content in national financial conditions for downside risks to US GDP growth and significant variation in the second and third moment of the conditional densities. I further analyze the importance of time-varying asymmetries by comparing the SSV model with a symmetric SV model

using Bayesian model selection criteria. With a Bayes Factor of 1612.18, the results provide decisive evidence for the SSV model. Comparing the upside and downside entropy of the two models reveals that these advantages arise mainly from the ability of the SSV model to capture the increased tail risks in times of financial and economic turmoil.
The flexibility of the proposed SSV model and the particle MCMC algorithm allows for further research to investigate asymmetric risks to macroeconomic variables. Given the different conclusions of Hasenzagl et al. (2020) and Adrian et al. (2020) in particular, extending the set of predictors to contain autoregressive components of GDP growth as well as other predictors can provide more insights on the predictability of time-varying asymmetries. Additionally, it is straight forward to extend the measurement equation of the SSV model to a VAR specification with additional variables and lift the exogeneity assumption of national financial conditions in the original model.
Given the active research field of macro at risk in macro-finance, empirical macroeconomics and econometrics, the SSV model provides a flexible toolkit for future research.

Appendix

3.A Inefficiency Ratio Under a Skew Normal Measurement Error

From Herbst and Schorfheide (2019), the weights $w_t^i(\phi_0)$ for $0 < \phi_0 \leq 1$ are given by the tempered likelihood function evaluated at the states $s_{t,i}$

Given the annealed importance sampling method described in Neal (2001) and Algorithm 2 by in Herbst and Schorfheide (2019), the expression of the unnormalized weights $w_{t,i}(\phi_n)$ are defined as the ratio of the bridge distributions

$$w_{t,i}(\phi_n) = \frac{p_n(y_t|s_t^i)}{p_{n-1}(y_t|s_t^i)} \tag{3.29}$$

Using expression (3.7) for the density of the skew normal distribution yields

$$w_{t,i}(\phi_n) = \left(\frac{\phi_n}{\phi_{n-1}}\right)^{1/2} \exp\left(\frac{-(\phi_n - \phi_{n-1})(y_t - \mu_t)^2}{2\sigma_{t,i}^2}\right) \frac{\int_{-\infty}^{\alpha_{t,i}\phi_n^{1/2}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z}{\int_{-\infty}^{\alpha_{t,i}\phi_{n-1}^{1/2}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z} \tag{3.30}$$

Expression (3.19) shows that in comparison to normally distributed measurement errors, the weights of the skew normal errors are scaled by a factor

$$\Lambda_{t,i}(\phi_n) = \frac{\int_{-\infty}^{\alpha_{t,i}\phi_n^{1/2}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z}{\int_{-\infty}^{\alpha_{t,i}\phi_{n-1}^{1/2}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z} \tag{3.31}$$

Additionally tempering the symmetry of the skew normal distribution modifies this factor

to

$$\tilde{\Lambda}_{t,i}(\phi_n) = \frac{\int_{-\infty}^{\alpha_{t,i}\phi_n^{2/3}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-t^2}{2}\right)\mathrm{d}z}{\int_{-\infty}^{\alpha_{t,i}\phi_{n-1}^{2/3}\frac{(y_t-\mu_t)}{\sigma_{t,i}}} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z} \quad \begin{cases} < 1 \text{ iff } \alpha_{t,i}(y_t-\mu_t) < 0 \\ > 1 \text{ iff } \alpha_{t,i}(y_t-\mu_t) > 0 \end{cases} \quad \forall\ 0 > \phi_n > 1. \tag{3.32}$$

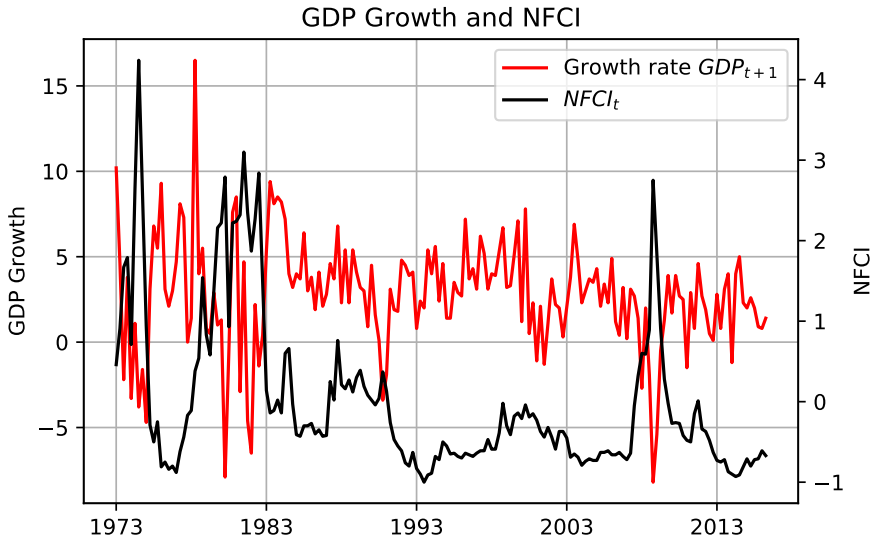
Inequality (3.32) holds since

$$\int_{-\infty}^{x} \exp\left(\frac{-z^2}{2}\right)\mathrm{d}z$$

is strictly monotonically increasing in x and $\phi_n > \phi_{n-1}$.

3.B US Data

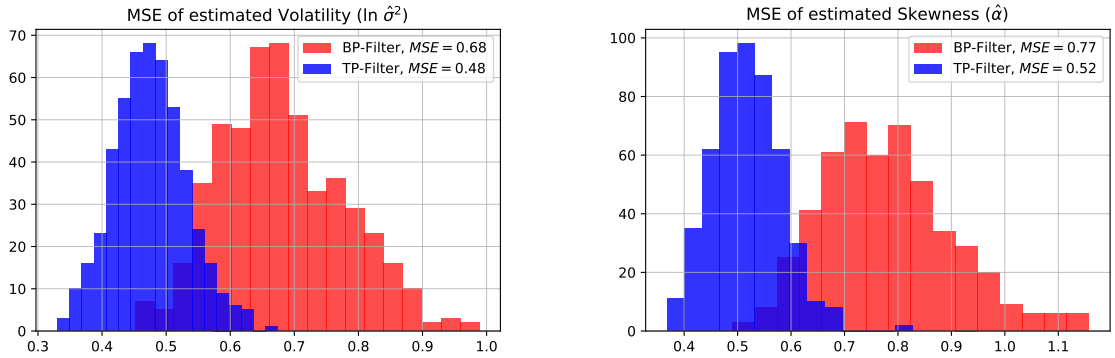
Figure 3.B.1: US GDP and National Financial Conditions Index



Notes: The sampling frequency for both series is quarterly with a sample range from 1973 Q1 to 2016 Q2.

3.C Bootstrap Particle Filter vs. Tempered Particle Filter

Figure 3.C.1: Mean Squared Errors of the Filtered States



Notes: Mean Squared Errors are calculated based on 500 simulations obtained with the Bootstrap Particle Filter and the Tempered Particle Filter. Tuning parameters of the Tempered Particle Filter were set to target an Inefficiency Ratio of $\Delta_r = 0.01$, 2 Mutation steps and $M = 10000$ particles. The superior performance of the Tempered Particle Filter is clear from the mean and standard errors of the distributions.

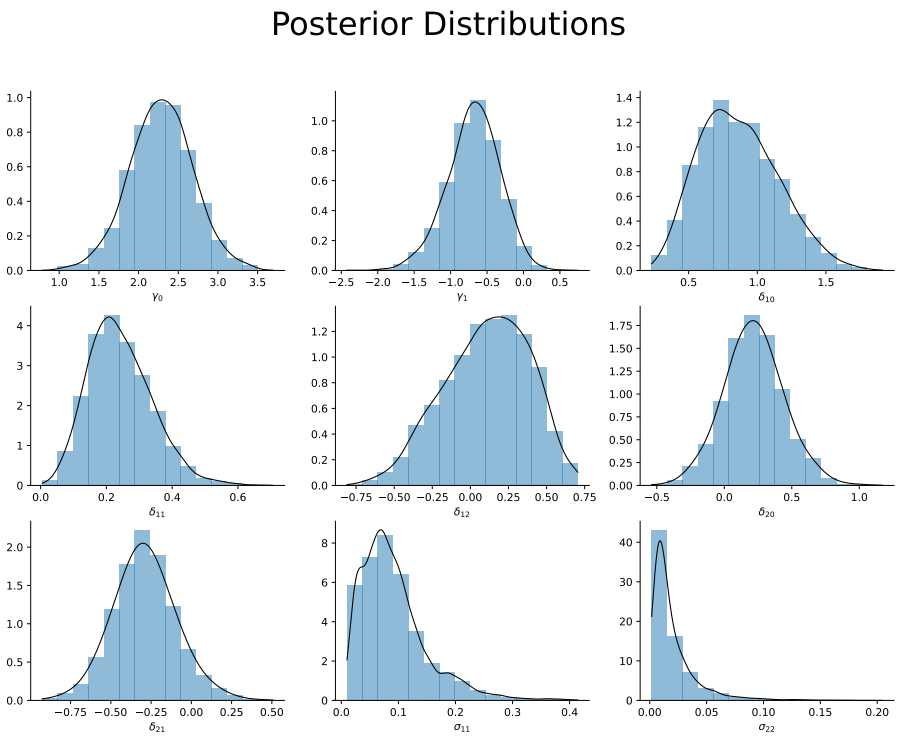
3.D SSV-Model: Prior and Posterior Distributions

Table 3.D.1: Priors for the Static Model Parameters of the SSV model.

Model Parameter	Distribution	Param 1	Param 2
γ_0	N	2.69	5
γ_1	N	-1	0.5
$\delta_{1,0}$	N	0	5
$\delta_{1,1}$	N	0	5
$\beta_{1,1}$	N	0	0.5
$\delta_{2,0}$	N	0	0.5
$\delta_{2,1}$	N	0	0.5
σ_{ν_1}	IG	1	0.25
σ_{ν_2}	IG	1	0.15

Notes: N denotes normal priors with Param 1 and Param 2 giving mean and variances. IG denotes the inverse gamma distribution with Param 1 and Param 2 for shape α and scale β.

Figure 3.D.1: Skewed Stochastic Volatility Model: Posterior Distributions



Notes: Histograms and kernel density estimates of the posteriors obtained using the particle Metropolis Hastings algorithm for the SSV model.

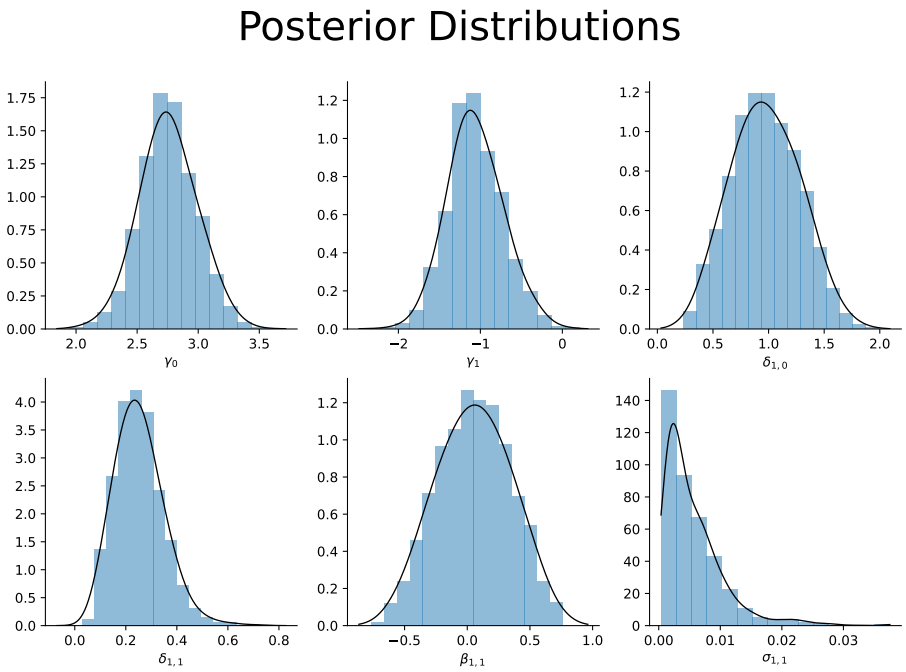
3.E SV-Model: Prior and Posterior Distributions

Table 3.E.1: Priors for the Static Model Parameters of the Symmetric SV model.

Model Parameter	Distribution	Param 1	Param 2
γ_0	N	2.69	5
γ_1	N	0	5
$\delta_{1,0}$	N	0	5
$\delta_{1,1}$	N	0	5
$\beta_{1,1}$	N	0	0.5
σ_{ν_1}	IG	1	0.25

Notes: N denotes normal priors with Param 1 and Param 2 giving the mean and variances. IG denotes the inverse Gamma distribution with Param 1 and Param 2 for the scale α and shape β.

Figure 3.E.1: Stochastic Volatility Model: Posterior Distributions



Notes: Histograms and kernel density estimates of the posteriors obtained using the particle Metropolis Hastings algorithm for the symmetric SV model.

Chapter 4

Conditional Density Forecasting: A Tempered Importance Sampling Approach*

with Carlos Montes-Galdón and Joan Paredes

4.1 Introduction

For policymakers, it is of the utmost importance to understand risks to their forecasts and how changes in the risks arising in one economic variable can translate to other variables. This has been clear for example in 2022 when it has been a challenging task for policymakers and researchers to gauge the macroeconomic impact of rising risks in energy prices due to the Russian invasion of Ukraine.

Policymakers and researchers use macroeconometric models to capture the transmission of changes in one of the model variables to the rest of variables in the model, as well as to produce forecasts and understand their balance of risks. When off-model information on the expected evolution of selected model variables is available, the literature provides several methods on how to condition model-based multivariate forecasts on the future paths of those variables. These methods usually assume that those paths correspond to a central moment of those forecast densities (see for example Waggoner and Zha (1999) or Bańbura et al.

*This paper has benefited from valuable comments of Frank Schorfheide, Matteo Ciccarelli, Marta Bańbura, Catalina Martínez Hernández and Giorgio Primiceri, Todd Clark, Michele Lenza, Ivan Petrella and participants of the 2022 Winter Meeting of the Econometric Society in Berlin. An earlier version of this chapter has been published as Working Paper No. 2754 in the ECB Working paper series. A shorter and non-technical version has also been published as SUERF Policy Brief, No 547

(2015)). This could also be understood as a scenario analysis or a conditional forecast. For instance, given an expected path of energy prices, researchers can obtain a model consistent path for inflation or real GDP.

However, limiting the information set to a central future path ignores the evolution of risks around these forecasts. So far, less research has considered conditioning forecasting densities not only on the mean but also on second or even higher moments of some of their marginal distributions (one notable approach is the entropic tilting methodology developed by Robertson et al. (2005)). For example, policymakers could have views not only about the central tendency of the evolution of a variable, but also about other moments such as the variance, which accounts for uncertainty, and the skewness, which allows to consider asymmetric upside and downside risks. Therefore, this paper proposes a robust methodology that allows researchers to condition a model based multivariate forecasting density on information about the marginal densities of some selected variables in the model, rather than just on their first moment. Specifically, researchers might want to inform the unconditional forecasting density from a macroeconometric model with market-based expectations about the marginal densities of certain economic variables. Therefore, our methodology can be understood as an extension of the conditional forecasting literature, and thus we will refer to it as conditional density forecasting. As an illustration of conditional information a researcher can include, we consider the implied probability densities obtained from option prices of futures on a specific model variable. Alternatively, information about the probability densities could be based on other econometric models (Giacomini and Ragusa (2014)), surveys or expert knowledge/judgement. Additionally, the information set could be based on assumptions made by researchers or policymakers to conduct risk analysis under certain scenarios.

Thus we first propose an algorithm that uses tempered importance sampling to adapt the model-based forecasting densities to the target marginal densities that satisfy the off-model information. While the motivation for our method is similar to the idea of entropic tilting developed in Robertson et al. (2005) and Krüger et al. (2017), our methodology is more flexible with regard to the information that can be introduced to the distributions via our algorithm. Due to our application of tempered importance sampling in our algorithm, it is also more robust with regards to the support of the densities, and avoids well-known problems that arise when using entropic tilting in practice.

Second, we demonstrate our methodology by conditioning the forecasting distribution obtained from a BVAR model and a DSGE model on asymmetric forecasting densities of the future oil price. Information about these densities at different horizons is derived from

option prices of future contracts on oil and implies asymmetric forecasting densities over the full sample period. Similar to the work of Adrian et al. (2019), we model these densities using the multivariate skew-T distribution of Azzalini and Capitanio (2003), which we fit to the option-implied moments. In a first exercise, we document the transmission of upside risks from future oil prices to inflation risks in the euro area after the onset of the Russian war in Ukraine. Since the invasion, option-implied densities exhibit high volatility and large positive skewness at all available forecasting horizons. We find that conditioning forecast densities on market-based upside risks to oil prices implies upside risks to inflation and core inflation well above the forecast implied from the BVAR model for all periods. Furthermore, the median forecast of core inflation would also remain elevated over the forecast period.
In a second exercise, we investigate if historically introducing information from oil-options would have improved the accuracy of our model-based density forecasts for GDP and inflation in a BVAR model. While the results indicate no substantial gains in moderate times, we find substantial increases in the forecasting performance during the onset of the Covid pandemic.
The chapter is structured as follows: Section 4.2 introduces the necessary information on importance sampling methods and describes our proposed algorithm. Section 4.3 provides an example of our methodology based on forecasting conditional distributions of inflation, given external information on oil prices. Section 4.4 proposes several extensions that make our method applicable for a wide range of forecasting models and their respective densities. Section 4.5 concludes and gives an outlook on further research.

4.2 Methodology

This section describes our methodology and presents the technical details on importance sampling methods that are the workhorse of our algorithm. As introduced in section 4.1 the aim of our method is to condition the forecasting distribution of a model on off-model information about the marginal densities of a subset of variables. Formally, this amounts to adapting observations generated from a model distribution Q_θ with density $q_\theta(y_i)$ and parameter vector $\theta \in \Theta$ to another parametric distribution $P_\eta(y)$ with density $p_\eta(y_i)$ and $\eta \in H$ that satisfies the assumptions made by the researcher and that come from outside the original macroeconometric model. This off-model information about $P_\eta(y)$ includes, but is not limited to, moments or parameters of $p_\eta(y_i)$. Furthermore, Q_θ and P_η are not required to belong to same family of distributions such that the parameter spaces Θ and H

can have different support and dimensions.[22] Most importantly, the external information may also be restricted to only a subset of variables included in $q_\theta(y_i)$, so that information is only available for some of the marginal densities of $p_\eta(y_i)$.

Our methodology is closely related to the entropic tilting methodology of Robertson et al. (2005). Entropic tilting uses an optimization procedure to reweight draws from distributions such that it satisfies a number of constraints on the moments. Yet, the use of entropic tilting has some drawbacks. First, while in theory entropic tilting is a powerful non-parametric tool to introduce information into a model based density forecast, its performance crucially hinges on the support of the original distribution $Q_\theta(y)$. Even if in theory the support of the distribution might be unbounded, in practice researchers work with a finite set of draws from the original distribution. If the original draws of the distribution do not have enough support for the final density $p_\theta(y_i)$, implying a large Kullback-Leibler Divergence between the two distributions, entropic tilting will yield unfavourable results and the algorithm might even fail to find a solution to the optimization problem. Therefore, instead of working directly with the moments, our algorithm imposes the constraints via a particular choice of the target density that satisfies these constraints and is chosen by the researcher. We elaborate more on this point in Section 4.2.5. Second, the methodology also struggles to find a solution as more conditions are introduced in the optimization problem. A similar point is raised in Krüger et al. (2017) who show that entropic tilting with higher dimensional distributions results in poor approximations of the final density since the weights for the reweighting step become unevenly distributed, a phenomenon commonly referred to as weight degeneracy.

Our methodology thus aims to provide a robust and flexible alternative to entropic tilting that can be applied in various circumstances. To overcome the aforementioned problems of weight degeneracy, our methodology is based on tempered importance sampling methods which we adapt to our needs. With our methodology, we are able to sequentially mutate the draws from the original distribution to the final distribution that incorporates the researcher's additional information or judgement, even if the original draws do not cover the support of the final distribution.

In the next subsections, we first describe the standard importance sampling algorithm, as well as the tempered importance sampling methodology from Herbst and Schorfheide (2014), so that we can subsequently discuss the modifications that are required for our application.

[22]For example if $Q_\theta(y)$ is a normal distribution with $\theta = (\mu, \sigma)$ and $P_\eta(y)$ is a t-distribution with η denoting the degrees of freedom, then $\Theta = \mathbb{R} \times \mathbb{R}_{>0}$ and $H = \mathbb{N}$ the set of natural numbers.

4.2.1 Importance Sampling for Conditional Density Forecasting

The cornerstone of our methodology is importance sampling as introduced by Kloek and van Dijk (1978). Since then, importance sampling has been applied in various scientific fields and its theoretical properties are well understood. Importance sampling is helpful when a researcher wants to evaluate the properties of a distribution, but has only access to draws from other distribution that might be similar or not to the final one. In our case, we make use of importance sampling as follows: Suppose that the researcher wanted to introduce external information into a model-based density forecast. Let y_i be a vector with dimension $L \times 1$ drawn from a model based forecasting distribution $Q_\theta(y)$ with dimension L. The vector y_i^e contains a subset of elements $y_i^e \in y_i$ with dimension $L^e \times 1$ such that $L^e < L$ with information given by the distribution $P_\eta(y)$. First, given a set of i.i.d. draws $\{y_i\}_{i=1}^N$ from the model-based forecasting density $y_i \sim q_\theta(y)$, the researcher can use importance sampling to reweight these draws to adapt them to the target forecasting density, $p_\eta(y_i)$ that satisfies the information that they aim to introduce. The importance weights are calculated as the ratio

$$w_i = \frac{p_\eta(y_i^e)}{q_\theta(y_i^e)}$$

which are normalized to sum to 1 using $W_i = \frac{w_i}{\sum_{i=1}^N w_i}$. The tuples $\{y_i, W_i\}_{i=1}^N$ provide a particle approximation of $p_\eta(y_i)$ given by

$$\hat{p}_\eta(dy_i) = \sum_i^N W_i \delta_{y_i}(dy_i). \tag{4.1}$$

Resampling the draws using a multinomial distribution with support points y_i and weights W_i yields

$$\tilde{y}_i \sim \mathcal{MN}(\{y_i\}_{i=1}^N | \{W_i\}_{i=1}^N)$$

and changes the distribution of the draws y_i to $P_\eta(y)$. Formally, importance sampling constitutes a change of the measure of a random variable under a measure Q to another measure P. The weights w_i correspond to the Radon Nikodym derivative that is given by the ratio of the respective density functions (see Brereton et al. (2013)).[23]

[23] See for example Theorem 10.6 in Klebaner (2012): Two different measures Q and P are related by

$$E_P[Y] = E_Q[\Lambda Y] \quad \text{with} \quad \Lambda = \frac{dP}{dQ}.$$

The ratio Λ is called the Radon Nikodym derivative. Given the two measures are two absolutely continuous probability distributions Q and P, the Radon-Nikodym derivative is given by the ratio of the respective

In typical applications of importance sampling, $q_\theta(y_i)$ serves as a proposal density while $p_\eta(y_i)$ is the target density. Furthermore, given that samples from $q_\theta(y)$ are drawn i.i.d, the strong Law of Large Numbers ensures that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} h(\tilde{y}_i) \xrightarrow{a.s.} E_P[h(y)] = \int h(y) dP_\eta \tag{4.2}$$

This insures that moments of interest such as the mean as well as the quantiles and p-values of the target distribution can be computed from the adapted draws $\{\tilde{y}_i\}_{i=1}^N$.[24]

4.2.2 Problems

While the choice of a target density p_η avoids the optimization problem that is necessary for entropic tilting of Robertson et al. (2005), the plain vanilla importance sampling algorithm described in the previous section faces similar drawbacks as the entropic tilting methodology. It is well known that the quality of the importance sampler depends on the variance of the weights W_i. A high variance of the weights implies that only few draws are resampled, leading to a large approximation error a phenomenon commonly dubbed weight degeneracy. If both density functions are reasonably close with most of their probability mass concentrated in the same regions of space, the distribution of the weights remains even with a low variance. However, importance sampling also becomes infeasible if the Kullback-Leibler Divergence between the two distributions is large and the distribution of the weights degenerates. Consequently, the right choice of the proposal density is crucial for a good importance sampling approximation such that researchers try to choose the proposal distribution in a way that the approximation error is small.

---

density functions

$$\Lambda = \frac{p(y)}{q(y)}$$

which is equal to the unnormalized weights $w(y)$ that are calculated in the correction step of importance sampling. Rewriting the integral in Equation (4.2) as

$$E_{P_\eta}[h(y)] = \int h(y) w(y) dQ_\theta = E_{Q_\theta}[w(y)h(y)]$$

shows that importance sampling uses a finite sample approximation of Λ to change the measure of the draws y from Q and P (see also Brereton et al. (2013)).

[24] Additionally, a Central Limit Theorem can be established that enables statistical inference on the quantities computed from the draws.

Yet, in our application we assume that the proposal density is predetermined since it is a model-based (forecasting) density $q_\theta(y_i)$. Additionally, the target density $p_\eta(y_i)$ might very likely be far apart with a high Kullback-Leibler divergence. For example, external information that will improve or alter the forecast density of a model might often imply a very different mean, variance or skewness of $p_\eta(y_i)$ compared to $q_\theta(y_i)$. Additionally, this problem becomes even more severe if the dimension of the model implied density $q_\theta(y_i)$ is large such that the probability mass is concentrated in a small region of the high dimensional space (see for example the exhibition in Betancourt (2017)).
However, in contrary to entropic tilting, standard importance sampling can be extended to so called tempered importance sampling which is able to overcome the problems of weight degeneracy in many situations. We therefore build our algorithm based on this method that we present in the next section.

4.2.3 Tempered Importance Sampling

Tempering importance sampling has its roots in the annealed importance sampling methodology from Neal (2001) and was introduced to the DSGE modelling literature in Herbst and Schorfheide (2014) and Herbst and Schorfheide (2019). To remedy the aforementioned problem of uneven weights when both the proposal and target density are far away from each other, the idea of tempered importance sampling is to adapt the draws via a sequence of bridge densities that assign more equal weights to the proposed draws and eventually converges to the true target. As shown in Herbst and Schorfheide (2019), an easy way to define such a sequence of bridge distributions is to use an inflated variance that is sequentially reduced to its actual level. More formally, let $p(y_i|\mu_\eta, \Sigma_\eta)$ be a target density with first and second moments that might depend on a set of model parameters η.[25] Tempered importance sampling uses a sequence of N_ϕ bridge distributions

$$p_n(y_i|\mu_\eta, \Sigma_\eta/\phi_n) \quad \text{with} \quad 0 < \phi_1 < ... < \phi_{N_\phi} = 1 \tag{4.3}$$

that converge towards the target distribution $p(y_i|\mu_\eta, \Sigma_\eta)$ for $\phi_n \to 1$. Starting from a low value of ϕ_1, $p_1(y_i|\mu_\eta, \Sigma_\eta/\phi_1)$ assigns weights to the proposed draws $\{y_i\}_{t=1}^N$ that are more evenly distributed. It is important to note, that there is freedom in defining the tempering

[25]Note that the density can also depend on additional parameters controlling the shape or kurtosis which are dropped in the following exposition for notational convenience.

method of the function.[26] The tempered importance sampling algorithm is initialized for $\phi_0 = 0$ by proposing an initial set of N particles $\{y_i\}_{i=1}^N$ from $p_0(y_i) = q(y_i|\mu_\theta, \Sigma_\theta)$. Subsequently, a combination of importance sampling and MCMC-methods adapts the proposed particles recursivley to the target distribution via the bridge distributions by cycling through the following three steps until $\phi_{N_\phi} = 1$:

1. **Correction:** Compute new importance weights

$$W_{i,n} \propto \frac{p_n(y_{i,n-1}|\mu_\eta, \Sigma_\eta/\phi_n)}{p_{n-1}(y_{i,n-1}|\mu_\eta, \Sigma_\eta/\phi_{n-1})}$$

2. **Selection:** Resample the draws

$$\tilde{y}_{i,n} \sim \mathcal{MN}(\{y_{i,n-1}\}_{i=1}^N|\{W_{i,n}\}_{i=1}^N)$$

3. **Mutation:** Propagate the resampled particles $\{\tilde{y}_i\}_{i=1}^N$ using M steps of an MH-Algorithm with a transition Kernel

$$y_{i,n} \sim K_n(y_n|\tilde{y}_{i,n})$$

that has the stationary distribution $p_n(y_i|\mu_\eta, \Sigma_\eta/\phi_n)$

In each iteration, the correction and selection steps adapt the draws $\{y_{i,n-1}\}_{i=1}^N$ of the previous iteration to the n^{th} brigde distribution using an importance sampling step (compare Section 4.2.1). Once the particles are adapted to $p_n(y_i|\mu_\eta, \Sigma_\eta/\phi_n)$, the mutation step moves the resampled particles to a region with a higher probability density using M steps of a Metropolis Hastings sampler.

Figure 4.2.1 provides an illustration of this idea in a simple univariate example. The draws are proposed from a standard normal distribution (blue solid line) and adapted to a skew-T distribution with mean 5.16 and positive skewness (red solid line). Clearly, the probability masses of the two densities are concentrated in different regions of the real line which results

[26] In general, other tempering methods are possible. Yet, it has to be satisfied that

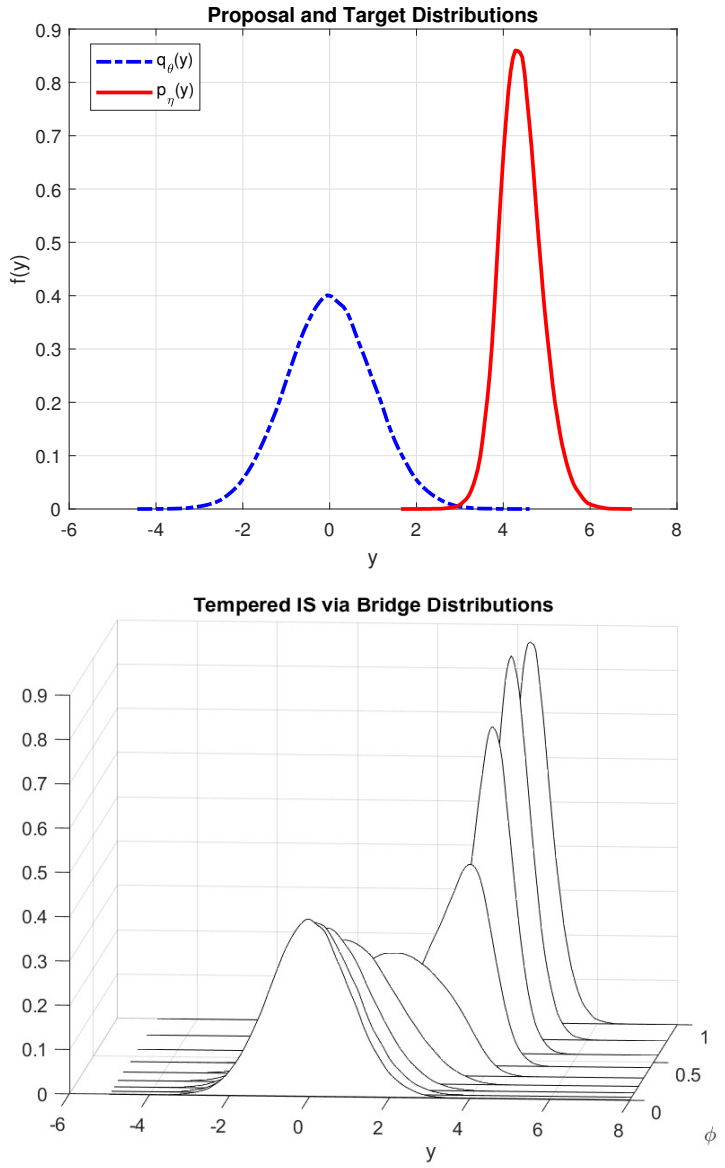
$$p_n(y_i|\mu_\eta, \Sigma_\eta) = p(y_i|\mu_\eta, \Sigma_\eta) \quad \text{for} \quad \phi_{N_\phi} = 1$$

and

$$Var[W_{i,n}] \to 0 \quad \text{for} \quad \phi_n \to 0.$$

We propose another definition in Section 4.4

Figure 4.2.1: Proposal, Target and Bridge Distributions



Notes: Illustration of tempered importance sampling based on a univariate example. The upper panel shows the proposal and target densities $q_\theta(y)$ and $p_\eta(y)$ with probability masses concentrated in different regions of the real line and a high Kullback-Leibler Divergence of 107,22. The lower panel shows how tempered importance sampling sequentially moves the draws from $q_\theta(y)$ to $p_\eta(y)$ via the bridge distributions that are close to each other by construction. The bridge distributions converge to $p_\eta(y)$ as $\Phi_n \to 1$.

in deteriorating importance weights and renders a standard importance sampler infeasible.[27] However, as can be seen from the Kernel-density estimates of the draws at different stages of the tempered importance sampling procedure in the lower plot, the intermediate bridge

[27]The Kullback-Leibler Divergence of the two densities is 107.22

distributions are close to each other by construction such that they provide a suitable proposal density for the next tempering step. This prevents the importance weights in each iteration from deteriorating. Most importantly, the Metropolis Hastings step insures that the proposed draws are concentrated in regions with high probability and sequentially mutates the draws towards the target distribution. As $p_n(y_i|\mu_\eta, \Sigma_\eta/\phi_n)$ eventually converges to the target distribution, the particles are gradually adapted to the final distribution. Convergence results and Central Limit Theory similar to simple importance sampling are based on the work of Chopin (2004) and established in Herbst and Schorfheide (2014).
It is important to note that without the mutation step, the importance sampling procedure would face the same limitations as entropic tilting or plain vanilla importance sampling. If there is no empirical support for the draws from the proposal density, the weights of the bridge distributions will deteriorate as there are no draws in the region of space where the target density has most probability mass. Therefore, the key element of the tempered importance sampling algorithm lies in the mutation step, that mutates the particles and ensures that there are sufficient draws in the region of the target densities.
In our algorithm we use the ability of tempered importance sampling to adapt draws from the model based forecasting distribution $Q_\theta(y)$ sequentially to the target distribution $P_\eta(y)$. The ability to mutate the proposed draws from the model density $q_\theta(y_i)$ towards the target density makes our methodology more robust then the entropic tilting methodology of Robertson et al. (2005).
Yet, it is important to highlight that the exposition of the tempered importance sampling steps up until now only provides an incomplete solution since the aim is to condition only on a subset of the marginals of the full forecasting distribution $Q_\theta(y)$ on external information. Compared to the algorithm of Herbst and Schorfheide (2014), we need to extend the tempered importance sampling procedure to make it suitable to our applications. First, researchers may not know a closed form solution to evaluate the model-based density $q_\theta(y_i)$. This happens for example when trying to evaluate the (marginal) forecasting density of a BVAR model accounting for parameter uncertainty. Second, external information might only be available for transformations of the model variables of interest which naturally implies a change in the respective marginal distributions of the variables that needs to be accounted for. For example, market-based options are available for the level of oil prices, but in macroeconomic models usually variables appear in log-levels or growth rates. While Herbst and Schorfheide (2014) use importance sampling as an alternative method to common MCMC methods to obtain draws from a (potentially) high dimensional posterior distribution of DSGE models, we apply tempered importance sampling as a method to

gradually change the measure of a model based forecasting density to take the off-model information about some of the model variables into account. Another recent application that uses tempered importance sampling to sequentially adapt change the measure of a set of model draws is Ho (2022) in the context of prior robustness. Furthermore, adaptive importance sampling techniques have been used in the field of credit risk modelling as outlined in Brereton et al. (2013). The next section introduces our algorithm.

4.2.4 Our Proposed Algorithm

Our algorithm proceeds in two steps: First elements of the vector y_i^e are adapted to the target density $p(y_i|\mu_\eta, \Sigma_\eta)$ that satisfies the imposed restrictions. Second, the corresponding values of the vector y_i^{-e} with length $L^{-e} = L - L^e$ that holds the remaining elements of y_i are recovered conditional on the final values of y_i^e. To overcome the problems of weight degeneracy, initial draws are proposed based on the density $q(y_i|\mu_\theta, \Sigma_\theta)$ and subsequently adapted to $p(y_i|\mu_\eta, \Sigma_\eta)$ in a sequence of tempering iterations. Since we consider a Bayesian estimation setup, we also need to take the parameter uncertainty of the model into account by sampling with replacement from the set of posterior draws for $\{\mu_\theta^{(i)}\}_{i=1}^M$ and $\{\Sigma_\theta^{(i)}\}_{i=1}^M$ for each particle. Furthermore, let the sequence $\{\phi_i\}_{i=1}^{N_{\phi_N}}$ define a tempering schedule as in (4.3). The sequence can either be determined endogenously as in Herbst and Schorfheide (2019) or predetermined as in Neal (2001). Eventually, this gives the following algorithm:

Conditional Density Forecasting Algorithm: For $i = 1, ..., N$:

1. Draw $y_{i,1} \sim q\left(y_i|\mu_\theta^{(i)}, \Sigma_\theta^{(i)}\right)$
2. Select the subset $y_{i,1}^e \in y_{i,1}$ for which there exists external information on the transformation $h(y_{i,1}^e)$.
   (a) Obtain initial importance weights $W_{i,1} \propto \frac{p_1(h(y_{i,1}^e)|\mu_\eta, \Sigma_\eta/\phi_1)}{q\left(y_i|\mu_\theta^{(i)}, \Sigma_\theta^{(i)}\right)}$
   (b) Resample $y_{i,1}^e \sim \mathcal{MN}(\{y_{i,1}^e\}_{i=1}^N|\{W_{i,1}\}_{i=1}^N)$
3. For $n = 2 : N_{\phi_N}$
   (a) **Correction**: Obtain weights $W_{i,n} \propto \frac{p_n(h(y_{i,n-1}^e)|\mu_\eta, \Sigma_\eta/\phi_n)}{p_{n-1}(h(y_{i,n-1}^e)|\mu_\eta, \Sigma_\eta/\phi_{n-1})}$
   (b) **Selection:** Resample $y_{i,n}^e \sim \mathcal{MN}\left(\{y_{i,n-1^e}\}_{i=1}^N|\{W_{i,n}\}_{i=1}^N\right)$
   (c) **Mutation:** For $j = 1 : H$

i. Draw $\hat{y}_{i,n}^e \sim q\left(y_i^e | y_{i,n}^e, \mu_\theta^{(i)}, c_n \Sigma_\theta^{(i)}\right)$

ii. Compute

$$\alpha = \frac{p_n(h(\hat{y}_{i,n}^e)|\mu_\eta, \Sigma_\eta/\phi_n)}{p_n(h(y_{i,n}^e)|\mu_\eta, \Sigma_\eta/\phi_n)} \times \left|\frac{\det(\mathcal{J}_{h^{-1}}(y_{i,n}^e))}{\det(\mathcal{J}_{h^{-1}}(\hat{y}_{i,n}^e))}\right| \tag{4.4}$$

where $\mathcal{J}_{h^{-1}}(y)$ denotes the Jacobian of the inverse transformation of $h(y)$.

iii. Draw $u \sim U(0,1)$.
Iff $u < \alpha$:

$$\text{Set} \quad y_{i,n}^e = \hat{y}_{i,n}^e$$

iv. Draw the other variables y_i^{-e} from conditional density

$$y_i^{-e} \sim q\left(y_i^{-e} | y_{i,N_\phi}^e, \mu_{\theta,-e|e}^{(i)}, \Sigma_{\theta,-e|e}^{(i)}\right)$$

Compared to previous applications of sequential importance sampling applied in the existing literature, our algorithm has two important extensions. First, since the draws are proposed based on $q(\tilde{y}_i|\mu_\theta, \Sigma_\theta)$ but evaluated given the transformation $h(y)$, we need to adjust the acceptance ratio of the Metropolis Hastings sampler to target the correct posterior density. Using a change of variables argument yields that the proposal density for the acceptance ratio α needs to be corrected by the determinant of the the Jacobian $\det(\mathcal{J}_{h^{-1}}(y)|$ of $h^{-1}(y)$ which gives

$$q(y|\mu_\theta, \Sigma_\theta) \times |\det(\mathcal{J}_{h^{-1}}(y)|. \tag{4.5}$$

An example of why we need this transformation, which we will use later in our application, is related to oil prices. Economists have access to option-based oil price forecasting densities. However, in many applications, the price of oil does not enter in our models as a level variable. Researchers usually take transformations such as a natural logarithm, a growth rate or other transformations to stationarize the level of the oil price.
Plugging in the latter expression for the proposal density leads to the acceptance ratio given in Equation 4.4. Consequently, this implies that our algorithm requires the transformation $h(y)$ to be bijective and differentiable. However, this holds for many transformations of interest such as the ones mentioned before. Second, since we only seek to include external information for a subset y_i^e of the elements of y_i the other variables in the model y_i^{-e} need to be recovered from the mutated particles. We achieve this by using the conditional distribution $q(y_i^{-e}|\mu_{\theta,-e|e}, \Sigma_{\theta,-e|e}, y_{i,N_\phi}^e)$ to mutate the other elements of y_i conditional on the final particles y_{i,N_ϕ}^e. Note that this is akin to the principle of a Gibbs sampling step. Hence, for our methodology it is necessary that one can draw from the conditional density

of $q(\tilde{y_i}|\mu_\theta, c_n\Sigma_\theta)$.
Finally, the algorithm can be used with a predetermined tempering schedule such that the sequence $\{\phi_i\}_{i=0}^{N_\phi}$ is deterministic as in Neal (2001) or Herbst and Schorfheide (2014) but also with adaptive implementations as for example in Herbst and Schorfheide (2019). The same holds for the scaling factor c_n of the covariance matrix Σ_θ of the proposal density q that can be adapted to target a specific acceptance rate of the MH-step in every new iteration.[28]

4.2.5 Relationship to Entropic Tilting

Given the aim of our new methodology as alternative to entropic tilting, this section compares our tempered importance sampling approach to the entropic tilting methodology of Robertson et al. (2005). As outlined before, the idea of entropic tilting is to reweight draws from a model-based distribution $F(y)$ to adapt them to a target distribution $F'(y)$. $F'(y)$ is obtained by solving the following optimization problem

$$D(F|F') = \int \ln\left(\frac{f'(y)}{f(y)}\right) dF'(y) \quad \text{s.t.} \quad \int g(y)dF'(y) = \bar{g} \quad \text{and} \quad \int f'(y)dy = 1 \tag{4.6}$$

Hence, the distribution $F'(y)$ is the closest density that satisfies a number of constraints $\bar{g}$ imposed by the researcher. As shown in the Appendix 4.A the solution to this problem is the Radon Nikodym derivative Λ that adjusts the probabilities under F and yields the exponentially tilted density $f'(y)$ given by

$$f'(y) \propto \Lambda f(y) \quad \text{with} \quad \Lambda \propto \exp\left(\gamma' g(y)\right) \tag{4.7}$$

where γ is the vector of Lagrange multipliers associated with the constraints $\bar{g}$ imposed under F'. Hence, the original probabilities are reweighted by Λ such that the constraints $\bar{g}$

[28] For example, following Herbst and Schorfheide (2019), c_n can be updated recursively with respect to the tempering steps using the the following formula

$$c_n = c_{n-1} f(\hat{A}_{n-1}) \quad \text{with} \quad f(x) = \alpha + \beta \frac{e^{\gamma(x-\bar{A})}}{1 + e^{\gamma(x-\bar{A})}}$$

where $\hat{A}_{n-1}$ is an estimate of the acceptance rate of the MH step in the previous iteration and $0 < \bar{A} < 1$ is the targeted acceptance ratio set be the researcher. The constants α, β and γ control the speed of the adjustment and satisfy $\alpha + 0.5\beta = 1$ and $\gamma > 0$.

under F' are satisfied. The value of γ is given by

$$\gamma = \text{argmin} \left[\int \exp\left(\gamma' g(y)\right)(g(y) - \bar{g}) dF(y) \right] \tag{4.8}$$

Since the Radon-Nikodym Derivative only exists if F' is absolutely continuous with respect to F, the optimization that is based on a finite number of draws often fails once the Kullback-Leibler divergence between F and F' is too large as there is not enough support for a distribution F' that satisfies the imposed constraints. If a solution exists, Equations (4.8) and (4.7) imply that the draws $\{y_i\}_{i=1}^N$ from F, are resampled using the normalized importance weights

$$W(y_i) = \frac{f'(y_i) \exp\left(\gamma' g(y_i)\right)}{\sum_{i=1}^N f'(y_i) \exp\left(\gamma' g(y_i)\right)} \tag{4.9}$$

While entropic tilting imposes constraints directly on the moments of the target density via the solution Λ, our methodology imposes the restrictions via the choice of a specific parametric density that satisfies the constraints. The resulting Radon Nikodym derivative is then given by the importance weights w_i.

As outlined before, while our methodology requires the additional assumption of a target density that is chosen by the researcher, it allows to move the particles in the mutation step of the tempered importance sampling procedure. This yields a more robust and versatile method to include external information into the forecast densities or perform scenario analysis based on counterfactuals.

4.3 Application: The Transmission of Oil-Price Risks to Inflation

To showcase how our algorithm functions, we use our methodology to condition the density forecast of a small euro area BVAR model and a large DSGE model on option-implied densities of oil price futures. The idea is to condition the different models on oil price forecasting densities based on market data, which might not be symmetric, and to explore how this conditional information tilts the forecasting distribution of variables such as inflation and real GDP.

Furthermore, our application contributes to the macro-at-risk literature that started with the seminal paper of Adrian et al. (2019) and has spawned an active field of research to model and evaluate asymmetric risks to macroeconomic variables such as GDP or inflation. Recent contributions from Wolf (2022) or delle Monache et al. (2021) propose different univariate

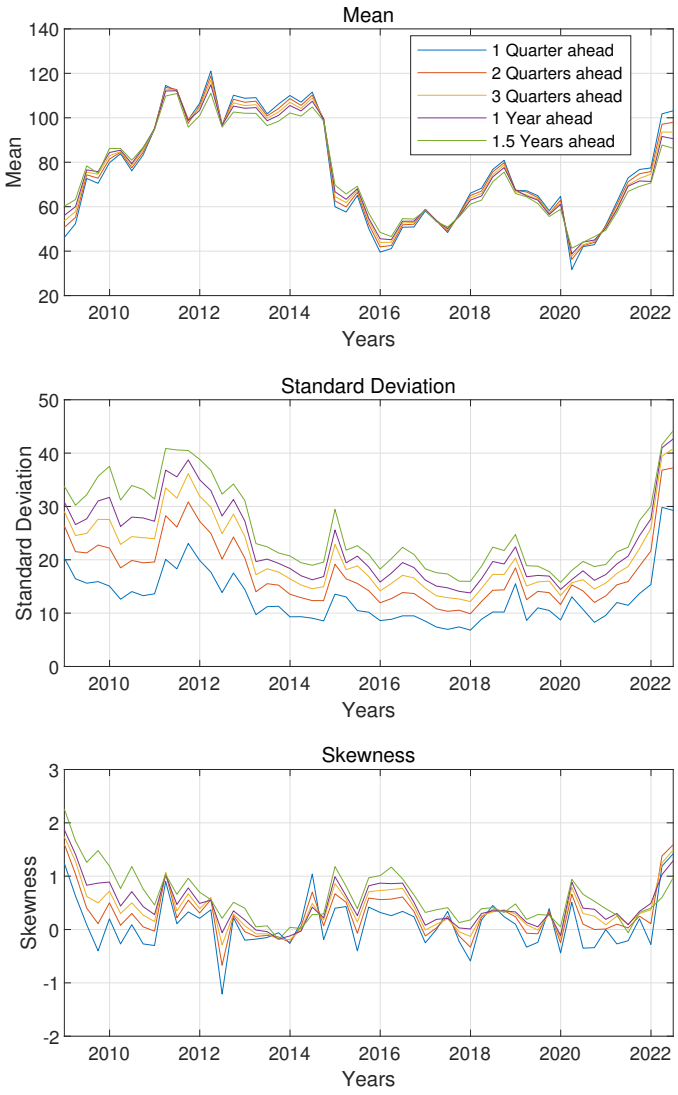
modelling approaches to model time-varying asymmetries of the forecasting densities of macroeconomic variables using a skewed distribution of the shocks whose moments are time-varying based on auto-regressive components and exogenous variables. Furthermore, the work of Montes-Galdon and Ortega (2022) extends this approach to Bayesian VAR models exploiting a specific representation of the skewed normal distribution by Azzalini (2013) for the structural shocks of the model. While the aforementioned papers seek to capture the evolution of risks based on some latent state variables, the application in this paper aims to directly introduce off-model information about the full[29] marginal distribution of some variables in the models and analyse the effect on other marginals based on correlations captured by the model. That is, we explore how possibly asymmetric risks in one variable translate to risks in other variables.

Based on the results of Breeden and Litzenberger (1978), it is possible to infer probabilities about the value of an underlying asset at the date of expiry from derivative prices observed in the market. The resulting probabilities can be used to construct the full probability density of the underlying variable at the expiration date (see for example de Vincent-Humphreys and Puigvert Gutiérrez (2010)). Option-implied probability densities are derived for different objects such as exchange rates, interest rates or oil prices, and are regularly published by the European Central Bank, the Bank of England or the Federal Reserve.

Figure 4.3.1 shows the first three moments of the option implied probability density functions of oil prices using quarterly data from 2008 to 2022 for different forecast horizons obtained from the ECB's Statistical Data Warehouse. Most notably, the probability density of the future oil price exhibits large fluctuations in the evolution of skewness for all horizons over the full sample. Remarkably, in 2022, with the beginning of the Russian invasion of Ukraine, market-based data shows that agents expected significant asymmetric upside risks to the price of oil. We first need to find a distribution that can explain those fluctuations and which will be the target distribution $P_\eta(y_i^e)$ in our methodology as defined in the previous section. Following the recent literature on skewed densities spawned by Adrian et al. (2019) we model the marginal forecasting densities of the price of oil at any time period and at different horizons as a multivariate skew-T distribution as introduced by Azzalini and Capitanio (2003). We outline further details of the density function of the multivariate skew-T distribution and our fitting procedure in the next section.

[29] In the sense of using information beyond the second moment

Figure 4.3.1: Option-Implied Moments of Future Oil-Price Densities



Notes: Mean, variance and skewness of the option implied probability density functions of oil prices using quarterly data from 2008 to 2022 for different forecast horizons. All series are obtained from the ECB's Statistical Data Warehouse. The implied densities exhibit strong time-varying asymmetries at all horizons and over the full sample.

4.3.1 Properties of the Multivariate Skew-T Distribution

A random vector $y \in \mathbb{R}^p$, follows a a multivariate skew-T

$$y \sim \mathcal{MST}(y|\xi, \Omega, \lambda, \nu)$$

where $\xi \in \mathbb{R}^p$ determines the location, Ω is a $p \times p$ covariance matrix, $\lambda \in \mathbb{R}^p$ is the shape parameter and $\nu \in \mathbb{N}$ gives the degrees of freedom. Based on the construction of skewed densities established in Azzalini (2013), the corresponding density function is given by

$$\tau(y|\xi, \Omega, \lambda, \nu) = 2t_p(y|\xi, \Omega, \nu)T_1 \left\{ \lambda' z \times \left(\frac{\nu + p}{\nu + Q(z)} \right)^{1/2} \middle| \nu + p \right\} \tag{4.10}$$

with $z = \omega(y - \xi)$, $Q(z) = z'\bar{\Omega}^{-1}z$, correlation matrix $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$ and

$$t_p(y|\xi, \Omega, \nu) = \frac{\Gamma((\nu + p)/2)}{|\Omega|^{1/2}(\nu\pi)^{p/2}\Gamma(\nu/2)} \left(1 + \frac{Q(z)}{\nu} \right)^{-(\nu+p)/2} \tag{4.11}$$

As shown in Proposition 3 in Arellano-Valle and Genton (2010) the multivariate skew-T distribution is closed under marginalization. For a partition $y = (y_1, y_2)$, with dimensions p_1 and p_2 and parameters (ξ, Ω, λ), the marginal distributions of y_i with $i = 1, 2$ are given by

$$y_i \sim \mathcal{MST}_i(\xi_i, \Omega_{ii}, \lambda_{i(j)}, \nu) \tag{4.12}$$

with

$$\lambda_{i(j)} = \frac{\lambda_i + \bar{\Omega}_{ii}^{-1}\bar{\Omega}_{ij}\lambda_j}{\sqrt{1 + \lambda_j'\tilde{\Omega}_{ii|j}\lambda_j}} \quad \text{and} \quad \tilde{\Omega}_{ii|j} = \bar{\Omega}_{jj} - \bar{\Omega}_{ji}\bar{\Omega}_{ii}^{-1}\bar{\Omega}_{ij} \tag{4.13}$$

Expression (4.13) shows that the shape parameter of the marginal distribution is a weighted sum of all elements of the vector of individual shape parameters λ, with weights depending on the correlation between y_i and y_j. Thus, $\lambda_i = 0$ does generally not imply that the marginal distribution of y_i is symmetric.[30] Figure 4.3.2 illustrates this in a simple two-dimensional example with $\xi_i = 0$ and $\omega_i = 1$ for $i = \{1, 2\}$ and a correlation coefficient of $\rho = 0.8$. The shape parameters are set to $\lambda_1 = -2$ and $\lambda_2 = 0$. Based on Equation (4.13), the values for the shape parameters of the marginal distributions are then given by

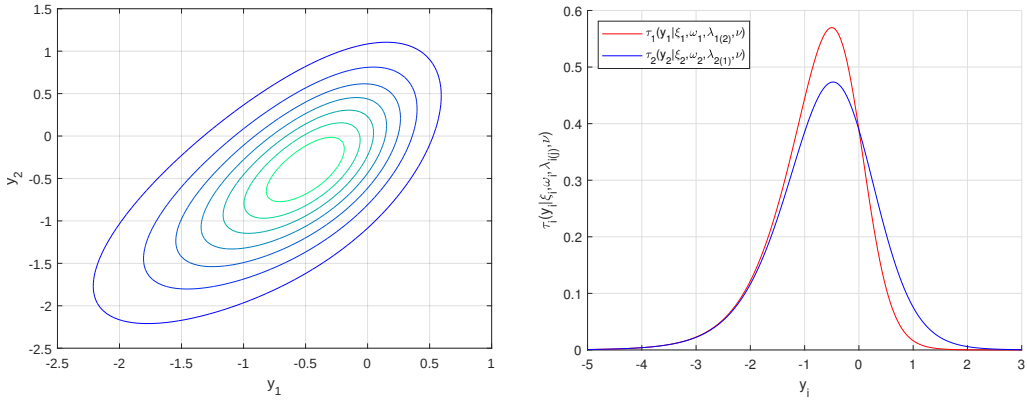
$$\lambda_{i(j)} = \frac{\lambda_i + \rho\lambda_j}{\sqrt{1 + \lambda_j^2(1 - \rho^2)}} \tag{4.14}$$

which yields $\lambda_{1(2)} = -2$ and $\lambda_{2(1)} = -1.024$. Thus, the positive correlation between y_1 and y_2 introduces negative skewness in both marginal distributions. From expression (4.14) it is also clear that a negative correlation of y_1 and y_2 has an offsetting effect on the marginal

[30] As shown in Arellano-Valle and Genton (2010) a necessary and sufficient condition for $\lambda_{i(j)} = 0$ is that $\lambda_i = -\bar{\Omega}_{ii}^{-1}\bar{\Omega}_{ji}\lambda_j$.

shape parameter such that the marginal distributions are skewed in opposite directions. We

Figure 4.3.2: Bivariate Skew-T Density with Marginal Densities



Notes: Left panel: Contour plot of the joint skew-T distribution with $\xi_i = 0$ and $\omega_i = 1$ for $i = \{1, 2\}$ and a correlation coefficient of $\rho = 0.8$. The shape parameters are $\lambda_1 = -2$ and $\lambda_2 = 0$.
Right panel: Corresponding marginal distributions of y_1 and y_2. The positive correlation between y_1 and y_2 introduces negative skewness in both marginal distributions.

use exactly this property of the multivariate skew-T distribution to analyse the effects of tail-risks in the marginal forecast densities of oil-prices on other macroeconomic variables. We use our algorithm to introduce information about the shape of the distribution of the oil price and analyze how this will affect risks to other macroeconomics variables such as GDP or measures of inflation. Depending on the correlations between the variables, conditioning on the the distribution of the oil prices based on the option-implied moments will result in asymmetric density forecasts for other variables than the oil price.

4.3.2 Fitting a Skew-T Distribution to Oil Price Forecasts

To fit the multivariate skew-T distribution to the option implied moments, we obtain external information on the mean μ_i^{oil}, standard deviation σ_i^{oil} and skewness γ_i^{oil} of the marginal forecast densities from derivatives on the price of Brent Crude Oil for forecast horizons from $i = 1$ up to $i = 6$ quarters.[31] From Proposition 3 in Arellano-Valle and Genton (2010), it follows that the marginal forecast densities of the oil price for each quarter are univariate skew-T distributions. Thus, we can use the results from Azzalini and Capitanio (2003) to obtain the parameters of the marginal skew T-densities. We match the option implied moments to the theoretical moments of the skew-T distribution by solving the

[31] Since options are only traded for horizons 1,...,4 and 6 quarters ahead, we interpolate the values for the 5^{th} quarter using a cubic spline.

following equations

$$\gamma_i^{oil} = \kappa_{i(j)} \left[\frac{\nu \left(3 - \delta_{i(j)}^2\right)}{\nu - 3} - \frac{3\nu}{\nu - 2} + 2\kappa_{i(j)}^2 \right] \left[\frac{\nu}{\nu - 2} - \kappa_{i(j)}^2 \right] \tag{4.15}$$

$$\kappa_{i(j)} = \frac{\sqrt{v}\Gamma\left(\frac{1}{2}(\nu - 1)\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2}\nu\right)} \delta_{i(j)} \tag{4.16}$$

$$\lambda_{i(j)} = \frac{\delta_{i(j)}}{\sqrt{1 - \delta_{i(j)}^2}} \tag{4.17}$$

$$\sigma_i^{oil} = \omega_i \sqrt{\left[\frac{\nu}{\nu - 2} - \kappa_{i(j)}^2 \right]} \tag{4.18}$$

$$\mu_i^{oil} = \xi_i + \omega_i \kappa_{i(j)} \tag{4.19}$$

with respect to the parameters $\xi_i, \omega_i, \lambda_{i(j)}$ for $i = 1, ..., 6$. To provide a maximum amount of flexibility for the PDF we set $\nu = 5$.[32] After obtaining the shape parameters of each marginal density $\lambda_{i(j)}$ we obtain the shape parameters of the joint distribution in a second step using expression (4.13). We obtain an estimate of the correlation matrix based on the draws from the VAR model we further describe in the next section. Subsequently, we jointly solve for each λ_i which is given as

$$\lambda_i = \lambda_{i(-i)} \sqrt{1 + \lambda'_{-i} \tilde{\Omega}_{ii|-i} \lambda_{-i}} - \bar{\Omega}_{ii}^{-1} \bar{\Omega}'_{-ii} \lambda_{-i} \tag{4.20}$$

Hence, while we allow for changes in the mean, standard deviation and skewness of the target distribution we assume that the model-based correlation between the different forecasting horizons does not change with regards to the target distributions. Once we have obtained the fitted parameter vectors $\hat{\xi}$, $\hat{\Omega} = \hat{\omega}^{-1} \bar{\Omega} \hat{\omega}^{-1}$ and $\hat{\lambda}$ as solutions to Equations (4.15) - (4.20), we can specify the target density of our algorithm as

$$p_\eta(y_i^{oil}) = \tau(y^{oil} | \hat{\xi}, \hat{\Omega}, \hat{\lambda}, \nu) \tag{4.21}$$

4.3.3 Introducing Market-Based Densities Information in a BVAR Model

After constructing the target density for the price of oil using the results of the multivariate skew T distribution from the previous section we need condition our model-based forecasts

[32] Based on the results in Azzalini and Capitanio (2003), this is the smallest value of ν for which the first 4 moments of the multivariate Skew-T are defined.

on these densities to infer how the forecast densities of other macroeconomic variables in our model are altered. In our application, we obtain $q_\theta(y)$ from the forecasting density of a reduced form BVAR model

$$y_t = \zeta + A_1 y_{t-1} + ... + A_s y_{t-s} + u_t \quad \text{with} \quad u_t \sim \mathcal{N}(0, \Sigma_u) \tag{4.22}$$

As the endogenous variables of the model, we include the log of the price of oil, the log of real GDP, the log of prices including and excluding energy as well as the log of the US/Dollar exchange rate, log of employment and the long and short term interest rates.

We estimate the BVAR model using Bayesian methods, under a Minnesota prior. To find the optimal hyperparameters of the prior, we use the hierarchical approach of Giannone et al. (2015) that is based on maximizing the marginal data density of the BVAR model with respect to those hyperparameters. Since we estimate the model using Bayesian methods we obtain sets with I posterior draws for both the intercept ζ_i and the slope coefficients $A_{j,i},\ j = 1, .., 5$ as well as the elements of $\Sigma_{u,i}$. Additionally, we use the novel methodology of Lenza and Primiceri (2020) to deal with the Covid period in the first quarters of 2020. We sample with replacement from the posterior draws to generate the model consistent forecasts up to h periods, $y_i = [y'_{i,T+1}, ..., y'_{i,T+h}]'$. Thus, we also incorporate parameter uncertainty from the posterior densities into our risk analysis. The density of the proposal distribution for the i^{th} draw y_i is then given by the multivariate forecasting distribution of the model and takes the form

$$q_\theta(y_i) = \varphi(y|\mu^{(i)}, \Sigma^{(i)}) \tag{4.23}$$

where $\varphi(...)$ denotes the density function of the multivariate normal distribution with mean $\mu^{(i)}$ and variance covariance matrix $\Sigma^{(i)}$ that depend on the parameters of the intercept, slope coefficients and variance covariance matrix of the reduced form errors. We provide more details on the derivation of $\mu^{(i)}$ and $\Sigma^{(i)}$ for the forecasting density in Appendix 4.B.

4.3.4 Adjusting the Forecast Densities

With the proposal and target density at hand, we use the algorithm described in Section 4.2.4 to adjust the marginal forecast densities of oil prices from the BVAR to the the option implied forecast densities. Based on $I = 25,000$ draws for $\mu^{(i)}$ and $\Sigma^{(i)}$ from the BVAR we generate $N = 50,000$ model consistent forecasts y_i that are adapted to the target density. For our algorithm, we use the the adaptive tempering schedule of Herbst and Schorfheide

(2019) to obtain the optimal values for N_ϕ and ϕ_n. In each iteration, we optimize ϕ_n such that the inefficiency ratio is equal to a target ratio $r^* > 1$

$$\phi_n = \text{argmin} \frac{1}{M} \sum_{i=1}^{M} \left[\frac{w_{i,n}(\phi)}{\frac{1}{M} \sum_{i=1}^{M} w_{i,n}(\phi)} \right]^2 - r^* \tag{4.24}$$

with $W_{i,n}$ given as in Step 3 (a) of the algorithm. Setting r^* closer to one results in a better approximation of $p_\eta(y^{oil})$ but also increases the number of tempering steps. To obtain a precise approximation of the target density we set $r^* = 1.01$. For the mutation of the particles in Step 3 (c) we set $H = 10$. Given this set-up, we first use our method to investigate the effect of the strong increase in oil prices on inflation due to the begin of the war in Ukraine in the first quarter of 2022. Subsequently, we evaluate the gains of introducing external information from options into forecasting densities in a real-time forecasting exercise.

4.3.5 Results from the BVAR

As can be seen from Figure 4.3.1, option-implied moments of the forecasting distribution of the oil price have sharply increased over the first two quarters in 2022, with option implied skewness peaking in the last quarter. We estimate the BVAR model using data up to the first quarter of 2022 and introduce the information of the option-implied densities at the $4th$ of March. The option-implied moments for all forecasting horizons h at that point are displayed in Table 4.3.1. Since there are no options with an expiry date of 15 months traded the values for the 5 period-ahead forecasting densities are interpolated using a cubic spline.

Table 4.3.1: Option-Implied Moments

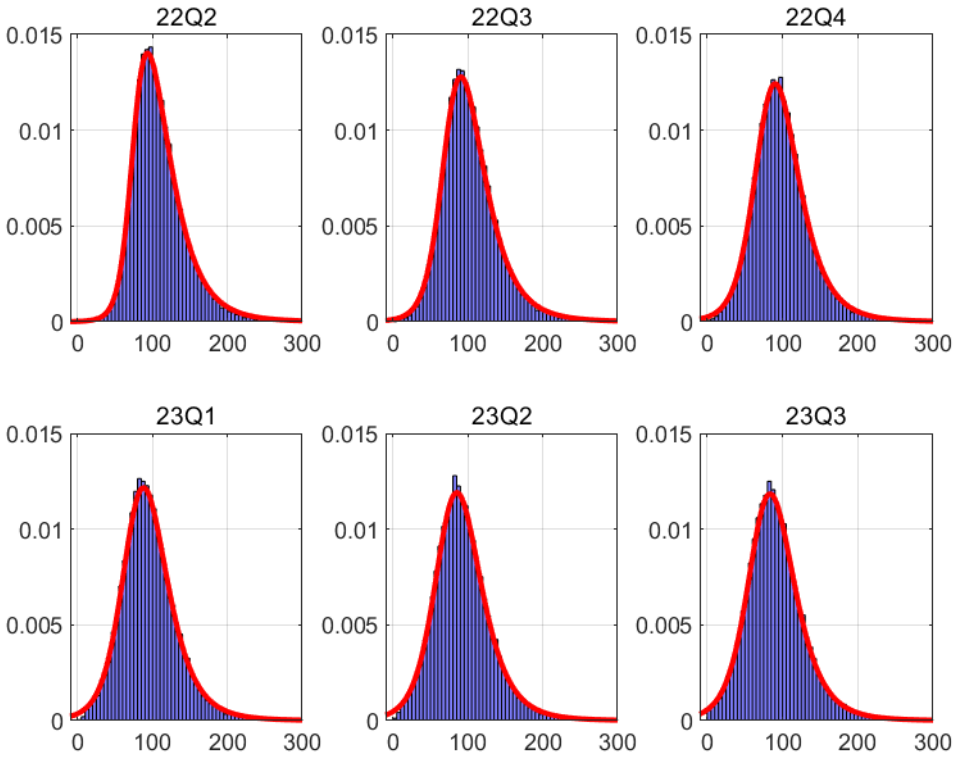
h	Mean	SD	Skewness
1	110.2	38.88	1.8
2	103.16	40.64	1.56
3	98.92	40.59	1.28
4	95.3	41.1	1.14
5	92.13	41.88	1.09
6	89.75	41.99	1.01

Notes: Option-implied moments for all forecasting horizons h at the $4th$ of March. 5 month-ahead moments are interpolated using a cubic spline. While the mean of the distribution is monotonically decreasing over the forecasting horizon, uncertainty and upside risks are increasing with h.

While the mean of the distribution is monotonically decreasing by approximately 20 euros per barrel over the forecasting horizon, uncertainty (standard deviation) is increasing

from 38.88 to 41.99 at the same time. Additionally, the distribution is significantly skewed to the right with values larger than one for all horizons indicating increased upside risks to the price of oil. Yet, the skewness is also decreasing over time, which indicates that risks become more symmetric at longer forecasting horizons. Based on the properties of the skew-T distribution described in 4.3.1, we use our algorithm to introduce the information about the forecasting densities of oil to our model to investigate if and how the forecasting distributions of inflation are affected. By introducing information about the full density, we can gauge the effects on the point forecasts as well as the effect that market-based oil upside risks have on other model variables. The skewness implied by the densities of the oil price will affect the distributions of other model variables depending on the correlations implied by the BVAR. Given the debate about the pass-through of high energy prices to inflation, we are particularly interested in the effect on both, inflation with and without energy prices. Figure 4.3.3 shows the fitted marginal skew T-densities that we obtain from the values in

Figure 4.3.3: Option-Implied Probability Density Functions

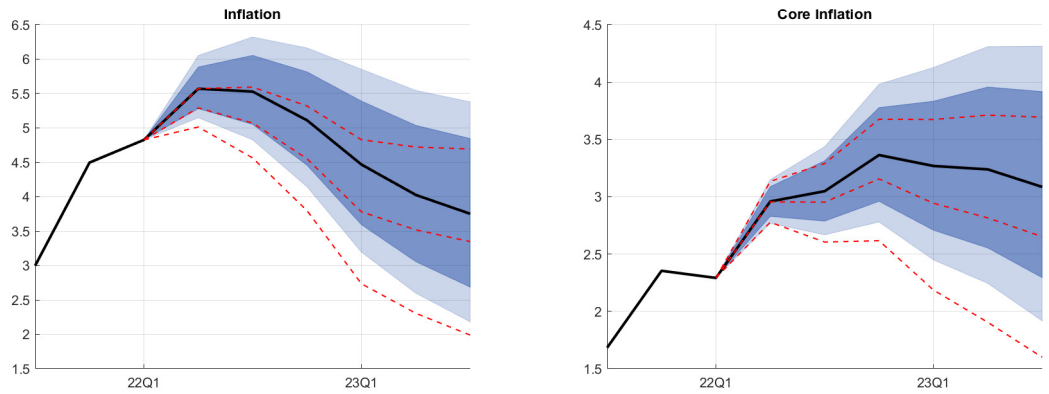


Notes: Option-implied marginal skew-T density functions for all forecasting horizons h at the $4th$ of March together with the histogram of the final particles $\{y_{i,N_\phi}^{Oil}\}_{i=1}^{N}$ with $N = 50,000$. The increased upside risks are clearly visible from the theoretical distributions as well as from the adapted draws

Table 4.3.1 together with the histogram of the final particles $\{y_{i,N_\phi}^{Oil}\}_{i=1}^{N}$ with $N = 50,000$. Based on our targeted inefficiency ratio $r^* = 1.01$, the algorithm required a number of $N_\phi = 40$ tempering steps to move the original particles form the proposal distribution to the target distribution. The draws are very well adapted to the target distribution that is implied by the options and provide a precise approximation. The positive skewness and increasing volatility is clearly visible from the theoretical distribution as well as from the adapted draws.

Figure 4.3.4 shows the resulting densities for inflation and core inflation. The shaded areas show the $16, 25, 75$ and 84 percent quantiles of resulting forecasting distribution of the annualized inflation rate together with the median given by the solid black line. Additionally, the dotted red lines show the 16 and 84 percent quantiles of the original distribution. Figure 4.D.1 in the Appendix also shows the histograms of the annualized growth rates of the oil-price, inflation and core inflation for all forecasting horizons. In

Figure 4.3.4: Option-Implied Forecasting Densities



Notes: Conditional density forecasts for inflation and core inflation. The shaded areas show the $16, 25, 75$ and 84 percent quantiles of resulting forecasting distribution of the annualized inflation rate together with the median given by the solid black line. The positive skewness in the distribution of the oil prices results in upside risks to inflation and higher mean values over the full forecasting horizon.

both cases, introducing the information of the options results in an upward shift of the full distributions. In case of inflation, the new median nearly coincides with the original 84 percent quantile of the original distribution in the first two periods. The forecasting density of core inflation is similar to the original model in the first period but subsequently deviates from the original model with significantly higher values over the rest of the forecasting horizon. Additionally, the positive skewness in the distribution of the oil prices results in upside risks to inflation. This can be seen from the plotted quantiles and the histograms of the marginal forecasting densities in Figure 4.D.1 of Appendix 4.D. Hence, while headline

inflation (i.e. including energy prices) reacts contemporaneously and the effects on the distribution are more evident over the first year of the forecast, the effects on core inflation appear later in the medium-term, reflecting second round effects arising from the upside risks in the distribution of the price of oil. Finally, the median forecast for core inflation remains persistently elevated over the forecasting horizon compared to the original forecast from the BVAR.

4.3.6 Forecasting GDP and Inflation

To evaluate the effect of conditioning on information about the market-based forecasting distribution of our BVAR model, we look at the probabilistic forecasting performance in a real time forecasting exercise to forecast GDP, inflation and core inflation. We estimate the same BVAR as in Section 4.3.3 using data vintages starting in the last quarter of 2013 up until the third quarter of 2021. With the onset of the Covid pandemic we again use the method of Lenza and Primiceri (2020). Subsequently, we use our algorithm to impose the option-implied distribution at the end of the quarter to the forecasting density of the oil-price. Since our methodology seeks to incorporate information about the full distribution, we use the continuous ranked probability score (CRPS) as the metric to evaluate the density forecasts. The CRPS generalizes the Mean Squared Error to take into account the full forecasting distribution. It can be formalised as,

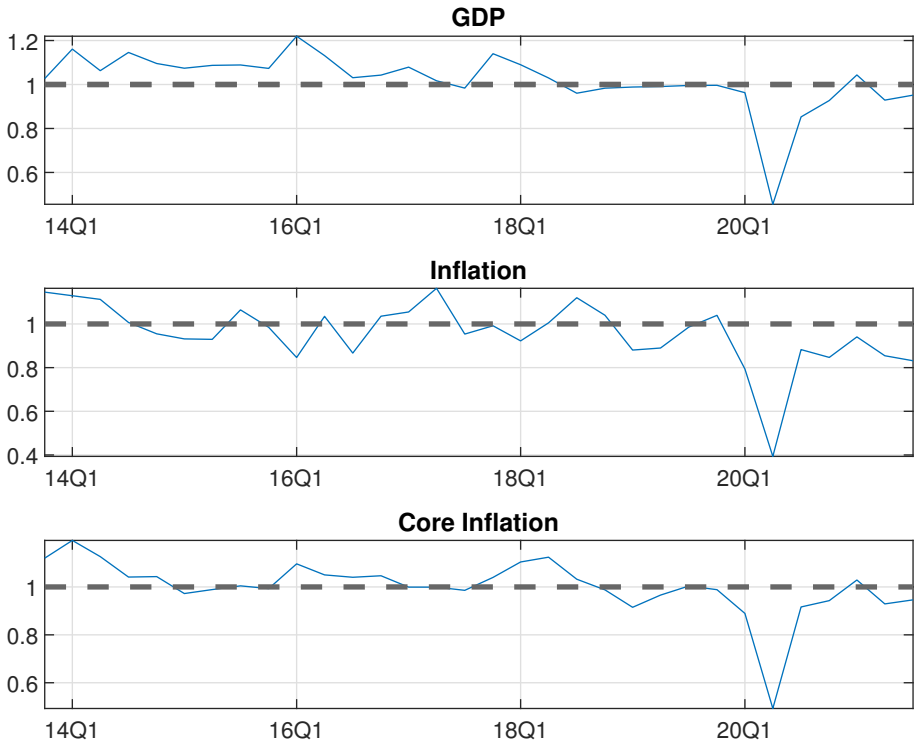
$$CRPS(F, x) = \int_{-\infty}^{\infty} \left(F(y) - \mathbf{1}(y - x)\right)^2 dy \tag{4.25}$$

where x is the realized value, $F(y)$ is the cumulative distribution function implied by the density forecast of the model and $\mathbf{1}(...)$ denotes the heavyside function. Figure 4.3.5 shows the ratios of the mean CRPS for the symmetric density forecasts under $Q_\theta(y)$ and the skew-T forecasts under $P_\eta(y)$

$$R_t = \frac{\frac{1}{P}\sum_{i=1}^{P} CRPS(Q_\theta, x_{t+i}^k)}{\frac{1}{P}\sum_{i=1}^{P} CRPS(P_\eta, x_{t+i}^k)} \tag{4.26}$$

for $k =$ GDP, inflation, core inflation. The tables with values for each period and forecasting horizon are included in Appendix 4.E. The results indicate that while including additional information on the distribution from the options does not increase predictive accuracy in moderate periods with stable economic conditions, it strongly increases the probabilistic forecast accuracy in times of economic turmoil during the onset of the Covid pandemic

Figure 4.3.5: Continuously Ranked Probability Scores



Notes: Ratios of the mean CRPS for the symmetric density forecasts under Q_θ and the skew-T forecasts under P_η for GDP, inflation, core inflation. While including additional information on the distribution from the options does not increase predictive accuracy in moderate periods, it strongly increases the probabilistic forecast accuracy in times of economic turmoil during the onset of the Covid pandemic in the first and second quarter of 2020.

in the first and second quarter of 2020. This is in line with other findings on skewed density forecasts such as Adrian et al. (2019) who show that conditional forecast densities of macroeconomic variables are symmetric in normal times but become skewed in times of crisis.

4.3.7 Tilting the Forecast Densities from a DSGE model

Finally, we show that our methodology can also be applied in large models, such as a DSGE model. Note that the reduced form solution of a DSGE model can be written as,

$$x_t = J + Qx_{t-1} + G\varepsilon_t \tag{4.27}$$

where x_t is a vector of endogenous and exogenous state variables in the model, and ε_t is a vector of i.i.d. structural shocks. Yet, note that in general the number of structural shocks in a DSGE model is smaller than the vector of state variables. Usually, for the estimation of a DSGE, there is a subset of variables that are observed, y_t, so that,

$$y_t = Hx_t \tag{4.28}$$

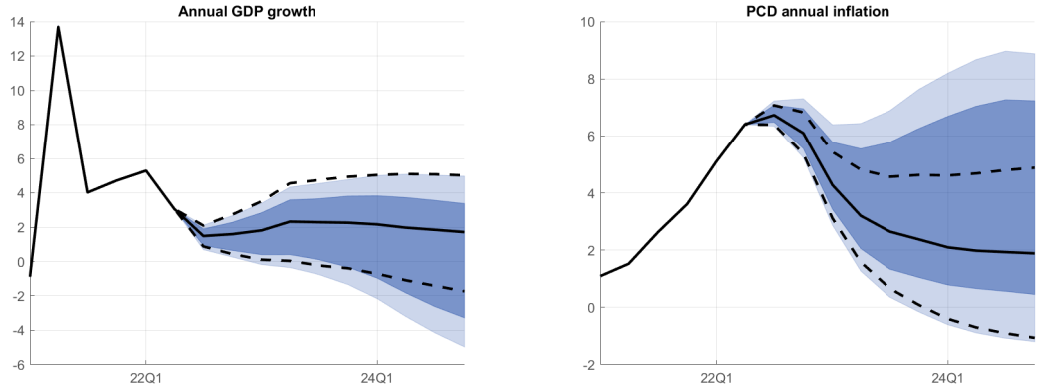
where H is a selection matrix. With these two equations in hand, we show in Appendix 4.C how to construct the proposal density $q_\theta(y)$ as in the case for the BVAR.

We then repeat the analysis in Section 4.3.5 using the ECB's New Area Wide Model II (NAWM II) from Coenen et al. (2018). The NAWM II is an estimated dynamic, stochastic, general equilibrium (DSGE) model of the euro area as a whole. The model incorporates a rich financial sector that allows for (i) a genuine role of financial frictions in the propagation of economic shocks as well as macroeconomic policies and for the presence of shocks originating in the financial sector itself, (ii) capturing the prominent role of bank lending rates and the gradual interest-rate pass-through in the transmission of monetary policy in the euro area, and (iii) providing a structural framework usable for assessing the macroeconomic impact of the ECB's large-scale asset purchases conducted in recent years. For the exercise in this section, we slightly modify the model to account for a faster pass-through of oil prices to inflation. In the original version of the model, there is one foreign intermediate-good firm that sells its goods in the domestic euro area market. The marginal costs of this firm is a weighted average of oil prices and foreign prices. Then, the firm is subject to staggered price contracts à la Calvo when setting the final domestic price, which introduces a sluggish price adjustment. For this analysis, we separate the problem into two firms. One that sets domestic prices for imported oil, and the second one that only takes care of foreign prices. The firm that sets oil prices has a smaller Calvo parameter, reflecting a faster pass-through of oil prices into final import prices, and thus, into the private consumption deflator and HICP in the model.

Figure 4.3.6 shows the results of introducing market based data on oil prices in the forecasting distribution of the NAWM II. The black lines show the model-based forecast, while the blue shaded areas represent the tilted distributions so that the distribution of the price of oil matches the market-based measures. In the model, the price of oil behaves as a supply side shock. Thus, once we incorporate information from the markets that implies that the distribution of the price of oil is skewed to the upside, the transmission channel in the model indicates significant downside risks to the real economy, which in the figure is represented

by annual GDP growth, and upside risks to inflation, represented by inflation in the private consumption deflator. The figure also shows that the final distributions are skewed, and the asymmetries are inherited from the skewness in the market-based options.

Figure 4.3.6: Option-Implied Forecasting Densities in the NAWM II



Notes: Introducing market based data on oil prices in the forecasting distribution of the ECB's New Area Wide Model II from Coenen et al. (2018). Black lines show the model-based forecast, while the blue shaded areas represent the distributions conditional on the option implied densities. Assuming that the distribution of the price of oil is skewed to the upside, the transmission channel in the model indicates significant downside risks to the real economy and upside risks to inflation.

4.4 Extensions

This section proposes several extensions of our core algorithm that make our methodology applicable in case some of the assumptions for the algorithm in section 4.2 are not satisfied. As described in Section 4.2.4, our methodology requires knowledge of the conditional distribution of the proposal distribution to recover the values of the variables y^{-e}. Second, we use a tempering method that requires a target density with a parameter to control the scale of the distribution. In this section we propose two remedies in case the application at hand does not meet these requirements.

First, if the conditional distribution is not available, but the researcher has access to draws from any arbitrary distribution, it is possible to approximate the proposal density $q(x)$ with a Gaussian mixture density of the form

$$q(x) = \sum_{k=1}^{K} \pi_k \varphi(x|\mu_k, \Sigma_k). \tag{4.29}$$

with weights $0 < \pi_k < 1$ that satisfy $\sum_{k=1}^{K} \pi_k = 1$. The Gaussian mixture density then has the conditional density

$$q(x_1|x_2) = \sum_{k=1}^{K} \left[\frac{\pi_k \varphi(x_2|\mu_{k,2}, \Sigma_{k,22})}{\sum_{l=1}^{L} \pi_l \varphi(x_2|\mu_{l,2}, \Sigma_{l,22})} \right] \varphi(x_1|x_2, \mu_{k,1|2}, \Sigma_{k,1|2}) \tag{4.30}$$

that can be used to sample from in step 3. Additionally, the recent publication of Ho (2022) motivates further research on how to construct a mutation step based on the non-parametric target density in the entropic tilting framework.
With regards to the second point, we propose another way to define the bridge distributions as given in Neal (2001)

$$p_n(y_i) = p_\eta(y_i)^{\phi_n} q_\theta(y_i)^{(1-\phi_n)} \tag{4.31}$$

In this specification, the bridge distributions are given by the geometric average of the model-implied density and the target density. Gradually increasing ϕ_n from 0 towards 1 will sequentially adapt the particles proposed by the model to the final distribution.[33] Expression 4.31 provides an attractive alternative if the scale of the distribution is not captured by a specific parameter. It is also possible to implement an adaptive tempering schedule instead of working with a predetermined sequence for $\{\phi_n\}_{n=1}^{N_\phi}$. With these remedies in hand, we consider our methodology applicable to a wide variety of problems.

4.5 Conclusion

In this paper, we develop a methodology that can be used to condition probabilistic forecasts of a model on off-model information about the marginal distributions of some of the model variables. More technically, the algorithm uses the tempered importance sampling method of Neal (2001) and Herbst and Schorfheide (2014) to adapt draws from a model-based distribution to a target distribution that satisfies the external information one intends to condition the forecast on. Our algorithm allows applications where the proposed draws are far away from the target density in a Kullback-Leibler sense as well as conditioning on information on transformations of the model variables. This makes our method superior to the entropic tilting methodology of Robertson et al. (2005) whose method is similar in spirit but less robust and less flexible.
We illustrate our algorithm by introducing off-model information about the distribution

[33] In our application, we experimented with both specifications and found that in both cases, our algorithm is able to adapt draws to the target distributions even if the Kullback-Leibler Divergence is high.

of future oil prices into the forecasting densities of a BVAR. The information is obtained from option prices and indicates significant amounts of skewness at all available forecasting horizons. Using the results of Azzalini and Capitanio (2003) we model the option-implied marginal forecasting densities as skew T and apply our methodology to investigate the transmission of upside risks to future oil prices on future inflation and core inflation in the first quarter of 2022. Due to the war in Ukraine, option-implied forecasting distributions of oil prices exhibit large positive skewness and increased volatility over the full forecasting horizon. We find that adapting the forecasting distributions of the BVAR to the option-implied densities results in upside risks to inflation and core inflation. Furthermore, median forecasts of core inflation remain persistently higher over the forecasting horizon compared to the symmetric forecast densities of the BVAR. We also investigate the forecasting accuracy of the density forecasts in real time. We focus on the probabilistic forecasts of GDP, inflation and core inflation over the period of 2013 Q4 up to 2021 Q4. Based on the CRPS, our results indicate that introducing information about the marginal distribution of oil prices improves forecasts for GDP and inflation measures during the Covid pandemic compared to symmetric forecasting distributions. This is in line with results of Adrian et al. (2019) who find skewness on conditional forecasting densities in times of economic turmoil.

Our methodology as well as our application is widely applicable and provides several extensions for further research such as introducing information from traded derivatives with other underlyings such as interest rates or exchange rates. Additionally, the model implied forecasting densities are not limited to time-series models such as BVARs but can also be applied to forecasting distributions of DSGE models or semi-structural models.

Appendix

4.A Solution to Entropic Tilting

This section shows the solution of the entropic tilting problem in Equation 4.6. Let F and F' be two distribution and let F' be absolutely continuous with respect to F. Using the representation in Ho (2022), the problem can be rewritten using in terms of expectations under the distribution F using the Radon Nikodym derivative. This yields the following Lagrangian:

$$\min_{\Lambda} \mathcal{L} = E_F[\Lambda \log(\Lambda)] - \gamma' E_F[\Lambda(g(y) - \bar{g})] - \mu E_F[\Lambda - 1] \tag{4.32}$$

where Λ is the Radon-Nikodym Derivative defined as

$$\Lambda \equiv \frac{f'(y)}{f(y)}.$$

This gives the First Order Condition:

$$\frac{\partial \mathcal{L}(\Lambda)}{\partial \Lambda} = 1 + \log(\Lambda) - \gamma'(g(y) - \bar{g}) - \mu = 0 \tag{4.33}$$

which implies

$$\Lambda \propto \exp(\gamma' g(y)). \tag{4.34}$$

From the other FOCs it follows that γ is given as

$$E_F[\exp(\gamma' g(y))(g(y) - \bar{g})] = 0 \tag{4.35}$$

and

$$\Lambda = \frac{\exp(\gamma' g(y))}{E_F[\exp(\gamma' g(y))]} \tag{4.36}$$

4.B Deriving the Proposal Density in a VAR model

Rewriting the VAR-Model as a VAR(1) gives

$$y_t = c_i + \Phi_i y_{t-1} + G_i \varepsilon_t \tag{4.37}$$

where Φ_i is the companion matrix of the i^{th} posterior draw for the slope coefficients and c_i the corresponding vector of intercepts. G_i is a lower-triangular matrix such that $G_i G_i' = \Sigma_{u,i}$ and $\varepsilon_t \sim \mathcal{N}(\varepsilon|0, I)$. Iterating the equation forward in time gives for the h step ahead forecast

$$y_{T+h} = \sum_{j=1}^{h} \Phi_i^{j-1} c_i + \Phi_i^h y_T + \sum_{j=1}^{h} \Phi_i^{h-j} G_i \varepsilon_{T+j} \tag{4.38}$$

Stacking the realizations over the full forecasting horizon in a vector y_i yields

$$\begin{bmatrix} y_{T+1} \\ y_{T+2} \\ \vdots \\ y_{T+h} \end{bmatrix} = \begin{bmatrix} \tilde{c}_{i,T+1} \\ \tilde{c}_{i,T+2} \\ \vdots \\ \tilde{c}_{i,T+h} \end{bmatrix} + \begin{bmatrix} G_i & 0 & 0 & 0 \\ \Phi_i G_i & G_i & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_i^{h-1} G_i & \Phi_i^{h-2} G_i & \cdots & G_i \end{bmatrix} \begin{bmatrix} \varepsilon_{T+1} \\ \varepsilon_{T+2} \\ \vdots \\ \varepsilon_{T+h} \end{bmatrix}$$

with $\tilde{c}_{i,T+h} = \sum_{j=1}^{h} \Phi_i^{j-1} c_i + \Phi_i^h y_T$. Redefining the terms results in the simple expression

$$y_i = \mu_i + \mathcal{G}_i \varepsilon \tag{4.39}$$

where $\mu = [\tilde{c}_{i,T+1}, ..., \tilde{c}_{i,T+h}]'$. The matrix $\mathcal{G}_i$ is a lower-triangular matrix that captures the correlations of the reduced form errors $u_t = G\varepsilon_t$ between the model variables as well as between the different time periods. These correlations depend on the values of the posterior draws for Φ_i and $\Sigma_{u,i}$. Given the distributional assumption about ε_t it follows that

$$y_i \sim \mathcal{N}\left(y|\mu^{(i)}, \Sigma^{(i)}\right) \tag{4.40}$$

with $\Sigma^{(i)} = \mathcal{G}_i \mathcal{G}_i'$. This gives the proposal density in Equation (4.23).

4.C Deriving the Proposal Density in a DSGE model

A similar reasoning as in the VAR model can be applied in the case of a DSGE model. The reduced form representation of a linearised DSGE model takes the form,

$$x_t = J + Qx_{t-1} + G\varepsilon_t \tag{4.41}$$

where x_t is a $N \times 1$ vector of endogenous and exogenous variables in the model, and ε_t is a $P \times 1$ vector of i.i.d. structural shocks. Yet, note that in general $N > P$. This implies that $\Sigma = GG'$ is a reduced rank matrix and thus not invertible. This implies that we cannot proceed exactly as in the case of the VAR model, and as the proposal density in equation (4.40) cannot be evaluated. However, as in the case of the estimation of a DSGE model, we can focus on a subset of variables in x_t which are assumed to be observed. If H is a matrix that selects some variables (or a combination of them), then we can write,

$$y_t = Hx_t = HJ + HQx_{t-1} + HG\varepsilon_t = \tilde{J} + \tilde{Q}x_{t-1} + \tilde{Q}\varepsilon_t \tag{4.42}$$

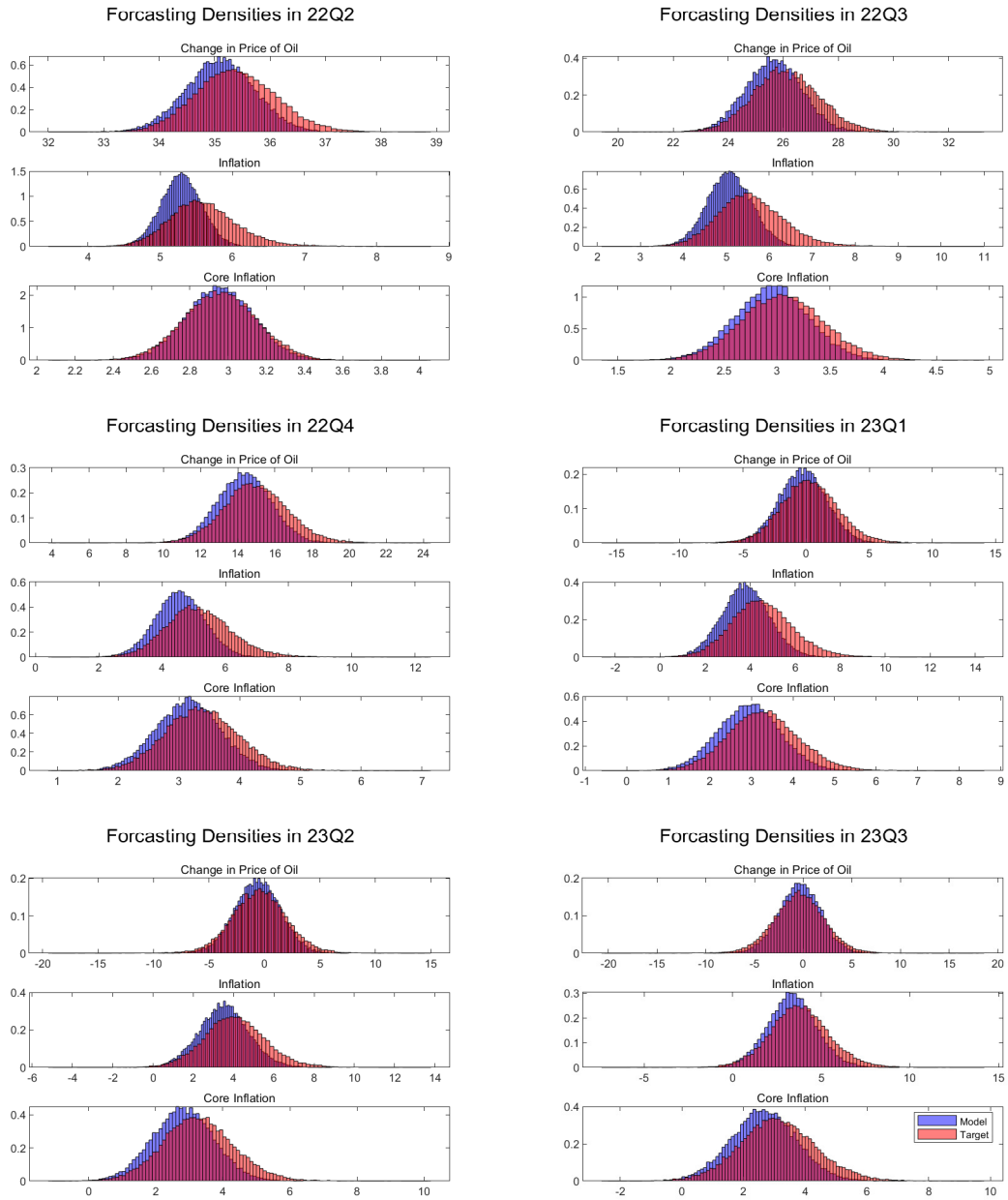
where we assume that the dimension of y_t is $P \times 1$. That is, the number of observed variables is equal to the number of fundamental shocks in the model. Then, given an initial condition for all the variables in the model, x_T, which could be recovered running a Kalman filter for example, we can proceed similarly as in the case of the VAR. Iterating forward, we can get,

$$y_{T+h} = \sum_{j=1}^{h} \tilde{Q}^{j-1}\tilde{J} + \tilde{Q}^h x_T + \sum_{j=1}^{h} \tilde{Q}^{h-j}\tilde{G}\varepsilon_{T+j} \tag{4.43}$$

And with the latter expression in hand, it is straightforward to compute an expression for the mean and the covariance matrix in Equation (4.40) using the same matrix representation as in the VAR case.

4.D Additional Plots

Figure 4.D.1: Marginal Forecasting Densities of Inflation



Notes: Histograms of the marginal forecasting distributions for each forecasting horizon. The draws in blue show the symmetric forecasting densities of the BVAR without conditioning on the option implied moments. The red draws show the forecasting distributions conditional on the option implied moments that exhibit a significant amount of skewness to the right. Therefore, the positive skewness in the distribution of the oil prices also results in upside risks to inflation and core inflation.

4.E Continuously Ranked Probability Scores

Table 4.E.1: Ratios of CRPS for GDP

T	h=1	h=2	h=3	h=4	h=5	h=6
14Q1	1.233	1.032	**0.979**	**0.851**	1.253	**0.917**
14Q2	1.258	1.065	**0.993**	1.601	1.045	1.017
14Q3	1.019	1.066	1.129	1.102	1.044	1.017
14Q4	1.023	1.089	**0.985**	1.252	1.265	1.219
15Q1	1.227	1.032	**0.948**	1.121	1.179	1.048
15Q2	1.005	1.039	**0.991**	1.258	1.013	1.144
15Q3	**0.952**	**0.936**	1.234	1.086	1.206	1.118
15Q4	1.053	1.066	1.015	1.124	1.148	1.101
16Q1	1.159	**0.979**	1.093	1.039	1.069	1.079
16Q2	1.308	1.159	1.089	1.287	1.274	1.185
16Q3	1.125	1.032	1.053	1.190	1.231	1.102
16Q4	**0.794**	**0.874**	**0.954**	1.291	1.173	1.046
17Q1	**0.859**	**0.939**	1.033	1.231	**0.946**	1.125
17Q2	1.064	1.124	1.125	**0.962**	1.184	**0.982**
17Q3	1.019	1.068	**0.901**	1.159	**0.888**	1.171
17Q4	**0.880**	1.004	1.108	**0.676**	1.358	1.210
18Q1	1.133	1.067	**0.885**	1.504	1.387	**0.962**
18Q2	1.134	0.982	1.207	1.573	**0.863**	**0.942**
18Q3	1.091	1.034	1.153	1.041	**0.984**	**0.893**
18Q4	**0.851**	**0.970**	1.034	1.033	**0.838**	**0.968**
19Q1	0.924	1.101	1.030	**0.848**	**0.963**	**0.991**
19Q2	1.135	1.042	**0.922**	**0.946**	**0.987**	1.004
19Q3	1.028	**0.919**	**0.968**	**0.987**	1.007	**0.865**
19Q4	1.041	**0.985**	**0.994**	1.005	**0.904**	**0.927**
20Q1	**0.997**	**0.996**	1.004	**0.790**	**0.847**	1.014
20Q2	1.001	**0.974**	**0.637**	**0.861**	**0.846**	**0.878**
20Q3	**0.835**	**0.439**	**0.271**	**0.296**	**0.214**	**0.146**
20Q4	**0.729**	**0.927**	**0.748**	1.122	**0.967**	1.057
21Q1	**0.696**	1.014	1.012	1.089	1.091	1.006
21Q2	**0.742**	1.086	1.181	1.112	1.177	1.084
21Q3	**0.661**	1.199	1.067	1.117	1.242	1.080
21Q4	**0.566**	1.099	1.156	1.085	1.298	1.218

Table 4.E.2: Ratios of CRPS for Inflation

T	h=1	h=2	h=3	h=4	h=5	h=6
14Q1	1.652	1.269	1.048	1.071	**0.888**	1.322
14Q2	2.008	1.310	**0.908**	**0.999**	1.220	**0.776**
14Q3	2.923	1.201	**0.974**	**0.966**	**0.874**	**0.8**
14Q4	1.463	1.077	1.163	**0.937**	**0.829**	**0.79**
15Q1	**0.935**	1.298	**0.866**	**0.893**	**0.761**	1.395
15Q2	**0.692**	1.014	**0.919**	**0.928**	1.125	1.01
15Q3	**0.982**	**0.913**	**0.871**	**0.970**	**0.899**	1.152
15Q4	1.454	**0.919**	1.171	1.201	1.174	1.075
16Q1	**0.794**	1.282	1.211	1.250	1.178	**0.936**
16Q2	**0.470**	**0.997**	**0.856**	**0.788**	1.178	1.175
16Q3	1.551	**0.813**	**0.939**	1.035	1.163	1.151
16Q4	**0.466**	**0.736**	1.142	1.332	**0.980**	1.032
17Q1	**0.589**	1.150	1.063	1.169	1.151	1.166
17Q2	**0.967**	**0.938**	1.121	1.163	1.215	1.247
17Q3	1.354	1.186	1.123	1.105	1.217	1.065
17Q4	**0.887**	**0.941**	1.079	1.121	1.056	**0.831**
18Q1	**0.946**	**0.862**	1.148	1.083	**0.850**	1.309
18Q2	**0.578**	1.015	**0.983**	**0.950**	1.234	**0.855**
18Q3	1.550	1.191	**0.922**	1.163	**0.926**	**0.875**
18Q4	2.096	1.087	1.113	1.116	1.007	**0.881**
19Q1	1.238	**0.960**	1.168	1.260	1.017	**0.875**
19Q2	**0.529**	1.250	1.163	1.030	**0.883**	**0.862**
19Q3	1.168	1.132	**0.851**	**0.877**	**0.838**	**0.783**
19Q4	1.264	**0.910**	**0.910**	1.004	**0.885**	1.046
20Q1	1.313	**0.978**	**0.912**	**0.960**	1.016	1.253
20Q2	**0.726**	1.319	**0.904**	**0.836**	**0.665**	**0.777**
20Q3	**0.188**	**0.237**	**0.518**	**0.321**	**0.395**	**0.433**
20Q4	1.72	**0.897**	**0.688**	**0.829**	**0.861**	**0.883**
21Q1	**0.666**	**0.667**	**0.899**	**0.887**	**0.915**	**0.844**
21Q2	1.01	**0.815**	1	**0.93**	**0.988**	1.135
21Q3	**0.635**	**0.872**	**0.917**	**0.827**	**0.941**	1.168
21Q4	**0.591**	**0.848**	1.019	1.213	1.041	**0.897**

Table 4.E.3: Ratios of CRPS for Core Inflation

T	h=1	h=2	h=3	h=4	h=5	h=6
14Q1	**0.928**	**0.93**	1.061	1.217	1.26	1.333
14Q2	**0.996**	1.003	1.063	1.215	1.345	1.314
14Q3	**0.926**	**0.984**	1.009	1.277	1.306	1.426
14Q4	1.033	**0.941**	**0.994**	**0.987**	1.301	1.16
15Q1	1.022	1.096	**0.988**	1.145	1.077	1.029
15Q2	1.017	1.067	1.034	**0.938**	**0.871**	1.041
15Q3	1.001	1.013	**0.94**	**0.869**	1.055	1.096
15Q4	1.025	1.044	**0.947**	**0.941**	1.104	1.065
16Q1	1.003	**0.885**	**0.986**	1.054	1.055	1.033
16Q2	1.068	1.207	1.052	**0.969**	1.146	1.162
16Q3	1.033	1.073	1.062	1.104	1.159	**0.923**
16Q4	**0.925**	1.004	**0.96**	1.125	0.933	1.175
17Q1	1.003	**0.91**	1.003	**0.949**	1.128	1.133
17Q2	**0.952**	**0.951**	**0.914**	1.105	1.107	1.081
17Q3	**0.991**	**0.987**	1.051	1.079	1.058	**0.928**
17Q4	1.028	1.062	1.052	1.105	**0.817**	**0.898**
18Q1	1	**0.983**	1.049	**0.834**	**0.922**	1.175
18Q2	1.079	1.072	**0.964**	**0.862**	1.205	1.258
18Q3	1.011	1.055	**0.962**	1.148	1.239	1.192
18Q4	1.063	1.148	**0.972**	1.167	1.165	**0.845**
19Q1	1.054	**0.944**	1.025	1.054	**0.876**	1.166
19Q2	**0.937**	**0.997**	**0.991**	**0.873**	1.037	**0.847**
19Q3	1.02	1.051	**0.926**	1.167	**0.895**	**0.903**
19Q4	**0.984**	1.011	1.077	**0.96**	**0.925**	1.021
20Q1	**0.986**	1.013	**0.933**	**0.89**	1.041	1.008
20Q2	**0.992**	**0.924**	**0.71**	**0.868**	**0.705**	**0.897**
20Q3	**0.829**	0**.633**	**0.563**	**0.384**	**0.357**	**0.248**
20Q4	**0.632**	**0.983**	**0.797**	**0.989**	**0.92**	**0.965**
21Q1	**0.837**	1.006	**0.927**	**0.986**	1.033	1.062
21Q2	1.249	**0.922**	1.057	1.013	1.123	1.055
21Q3	**0.72**	**0.893**	**0.943**	**0.98**	1.086	1.157
21Q4	**0.593**	**0.871**	1.094	**0.966**	1.203	1.253

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Adrian, T., F. Duarte, N. Liang, and P. Zabczyk (2020, May). Nkv: A new keynesian model with vulnerability. *AEA Papers and Proceedings 110*, 470–76.

Aikman, D., A. Haldane, and B. Nelson (2015). Curbing the credit cycle. *The Economic Journal 125*, 1072–1109.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Anundson, A., K. Gerdrup, F. Hansen, and K. Kragh-Sørensen (2016). Bubbles and crises: The role of house prices and credit. *Journal of Applied Econometrics 31*, 1291–1311.

Arellano-Valle, R. and M. Genton (2010, 12). Multivariate extended skew t distributions and related families. *Metron - International Journal of Statistics LXVIII*, 201–234.

Aruoba, S. B. (2008). Data revisions are not well behaved. *Journal of Money, Credit and Banking 40*(2-3), 319–340.

Azzalini, A. (2013). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society Series B 65*(2), 367–389.

Bańbura, M., D. Giannone, and M. Lenza (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting 31*(3), 739–756.

Basset, W., A. Daigle, R. Edge, and G. Kara (2015). Credit-to-GDP trends and gaps by lender- and credit-type. *FEDS Notes. Board of Governors of the Federal Reserve System*.

Baxter, M. and R. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics 81*, 575–593.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. Technical report, ArXiv e-prints.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Borio, C. (2014). The financial cycle and macroeconomics: What have we learnt? *Journal of Banking and Finance 45*, 182–198.

Borio, C. and P. Lowe (2002). Asessing the risk of banking crises. *BIS Quarterly Review*.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business 51*(4), 621–51.

Brereton, T. J., D. P. Kroese, and J. C. Chan (2013). *Monte Carlo Methods for Portfolio Credit Risk*, Chapter 7, pp. 127–152. John Wiley & Sons, Ltd.

Brownlees, C. and A. B. Souza (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics 118*, 312–330.

Canova, F. (2021). FAQ: How do I extract the output gap? A model-based perspective. *mimeo*.

Carriero, A., T. E. Clark, and M. Marcellino (2020, January). Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions. Working Papers 202002R, Federal Reserve Bank of Cleveland.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Chopin, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *The Annals of Statistics 32*(6), 2385–2411.

Christiano, L. and T. Fitzgerald (2003). The band pass filter. *International Economic Review 44*, 435–465.

Claessens, S., M. Kose, and M. Terrones (2012). How do business and financial cycles interact? *Journal of International Economics 87*, 178–190.

Clements, M. P. (2004). Evaluating the bank of england density forecasts of inflation. *The Economic Journal 114*(498), 844–866.

Clements, M. P. and A. B. Galvão (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics 28*(3), 458–477.

Clements, M. P. and A. B. Galvão (2019, 04). Data revisions and real-time forecasting. *The Oxford Research Encyclopedia of Economics and Finance*.

Clements, M. P. and D. F. Hendry (2011). *The Oxford handbook of economic forecasting*.

OUP USA.

Coenen, G., P. Karadi, S. Schmidt, and A. Warne (2018). The new area-wide model ii: an extended version of the ecb's micro-founded model for forecasting and policy analysis with a financial sector. *ECB Working Paper No 2200*.

Cogley, T. and J. Nason (1995). Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research. *Journal of Economic Dynamics and Control 19*, 253–278.

Coibion, O., Y. Gorodnichenko, and M. Ulate (2018). The cyclical sensitivity in estimates of potential output. *Brookings Paper in Economic Activity*.

Cornea-Madeira, A. (2017). The explicit formula for the Hodrick-Prescott filter in a finite sample. *The Review of Economics and Statistics 99*, 314–318.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score Models With Applications. *Journal of Applied Econometrics 28*(5), 777–795.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature 49*(1), 72–100.

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics 105*(1), 111–130.

Cunningham, A., J. Eklund, C. Jeffery, G. Kapetanios, and V. Labhard (2012). A state space approach to extracting the signal from uncertain data. *Journal of Business & Economic Statistics 30*(2), 173–180.

Danthine, J.-P. and M. Girardin (1989). Business cycles in Switzerland: A comparative study. *European Economic Review 33*, 31–50.

De Jong, R. and N. Sakarya (2016). The econometrics of the Hodrick-Prescott filter. *Review of Economics and Statistics 98*, 310–317.

de Santis, R. A. and W. van der Veken (2020). Forecasting macroeconomic risk in real time: Great and Covid-19 Recessions. Working Papers 2436, European Central Bank.

de Vincent-Humphreys, R. and J. M. Puigvert Gutiérrez (2010). A quantitative mirror on the euribor market using implied probability density functions. ECB Working Paper 1281.

delle Monache, D., A. de Polis, and I. Petrella (2021). Modeling and forecasting macroeconomic downside risk. Working Papers 1324, Banca d'Italia.

Detken, C., O. Weeken, L. Alessi, D. Bonfim, M. Boucinha, C. Castro, S. Frontczak, G. Giordana, J. Giese, N. Jahn, J. Kakes, B. Klaus, J. Lange, N. Puzanova, and P. Welz (2014). Operationalizing the countercyclical capital buffer: Indicator selection, threshold identification and calibration options. *ESRB Occasional Paper Series 5*.

Diebold, F. X. (2019). Time-series econometrics. *University of Pennsylvania, Department of Economics*.

Diebold, F. X. and G. D. Rudebusch (1991). Forecasting output with the composite leading index: A real-time analysis. *Journal of the American Statistical Association 86*(415), 603–610.

Doucet, A., N. de Freitas, and N. J. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer.

Drehmann, M. and R. Alfaro (2023). The Holt-Winters filter and the one-sided HP filter: A close correspondence. *Economics Letters 222*, 110925.

Drehmann, M., C. Borio, L. Gambacorta, G. Jiménez, and C. Trucharte (2010). Countercyclical capital buffers: exploring options. *BIS Working Paper 317*.

Durbin, J. and S. J. Koopman (2012, 05). *Time Series Analysis by State Space Methods*. Oxford University Press.

Edge, R. and R. Meisenzahl (2011). The unreliability of credit-to-GDP ratio gaps in real time: Implications for countercyclical capital buffers. *International Journal of Central Banking 7*, 261–298.

Elliott, G. and A. Timmermann (2013). *Handbook of economic forecasting*. Elsevier.

Engle, R. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics 22*, 367–381.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

EU Independent Fiscal Institutions (2020). A practitioner's guide to potential output and the output gap. *Working Paper*.

European Systemic Risk Board (2019). Methodologies for the assessment of real estate vulnerabilities and macroprudential policies: Residential real estate. *ESRB Working Group on Real Estate Methodologies*.

Flury, T. and N. Shephard (2011). Bayesian Inference based only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models. *Econometric Theory 27*(5), 933–956.

Furlanetto, F., K. Hagelund, F. Hansen, and O. Robstad (2023). Norges Bank output gap estimates: Forecasting properties, reliability, cyclical sensitivity and hysteresis. *Oxford Bulleting and Statistics 85*, 238–267.

Galan, J. (2019). Measuring credit-to-GDP gaps. The Hodrick-Prescott filter revisited. *Banco de España Documentos Ocasionales 1906*.

Galati, G., I. Hindrayanto, S. Koopman, and M. Vlekke (2016). Measuring financial cycles

with a model-based filter: Empirical evidence for the United States and the euro area. *Economics Letters 145*, 83–87.

Geweke, J. (1999). Using simulation methods for bayesian econometric models: inference, development,and communication. *Econometric Reviews 18*(1), 1–73.

Ghysels, E., C. Horan, and E. Moench (2018). Forecasting through the rearview mirror: Data revisions and bond return predictability. *Review of Financial Studies 31*(2), 678–714.

Giacomini, R. and G. Ragusa (2014). Theory-coherent forecasting. *Journal of Econometrics 182*(1), 145–155.

Giannone, D., M. Lenza, and G. Primiceri (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Giordani, P., E. Spector, and X. Zhang (2017). A new early warning indicator of financial fragility in Sweden. *Sveriges Riksbank Economic Commentaries 1/2017*.

Godsill, S. and T. Clapp (2001). *Improvement Strategies for Monte Carlo Particle Filters*, pp. 139–158. New York, NY: Springer New York.

Hamilton, J. (2018). Why you should never use the Hodrick-Prescott filter. *The Review of Economics and Statistics 100*, 831–843.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review 35*(3), 705–30.

Harvey, A. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics 8*, 231–247.

Hasenzagl, T., M. Plagborg-Møller, R. Lucrezia, and G. Ricco (2020). When is growth at risk? Working papers, Brookings Papers on Economic Activity.

Herbst, E. and F. Schorfheide (2014). Sequential monte carlo sampling for dsge models. *Journal of Applied Econometrics 29*(7), 1073–1098.

Herbst, E. and F. Schorfheide (2019). Tempered particle filtering. *Journal of Econometrics 210*(1), 26–44.

Hiebert, P., I. Jaccard, and Y. Schüler (2018). Contrasting financial and business cycles: Stylized facts and candidate explanations. *Journal of Financial Stability 38*, 72–18.

Ho, P. (2022). Global robust bayesian analysis in large models. *Journal of Econometrics*.

Hodrick, R. (2020). An exploration of trend-cycle deomposition methodologies in simulated data. *NBER Working Paper 26750*.

Hodrick, R. and E. Prescott (1981). Postwar U.S. business cycles: An empirical investigation. *Discussion Paper 451, Northwestern University*.

Hodrick, R. and E. Prescott (1997). Postwar U.S. business cycles: An empirical investigation.

Journal of Money, Credit and Banking 29, 1–16.

Howrey, E. P. (1978). The use of preliminary data in econometric forecasting. *The Review of Economics and Statistics 60*(2), 193–200.

Iseringhausen, M. (2021). A Time-varying Skewness Model for Growth-at-Risk. Working Papers 49, European Stability Mechanism.

Jacobs, J. P. A. M. and S. Van Norden (2011). Modeling data revisions: Measurement error and dynamics of "true" values. *Journal of Econometrics 161*(2), 101–109.

Jönsson, K. (2020). Cyclical dynamics and trend/cycle definitions: Comparing the HP and Hamilton filters. *Journal of Business Cycle Research 16*, 151–162.

Jordà, O., M. Schularick, and A. Taylor (2017). Macrofinancial history and the new business cycle facts. In M. Eichenbaum and J. A. Parker (Eds.), *NBER Macroeconomics Annual 2016*. Chicago: University of Chicago Press.

Jylhä, P. and M. Lof (2022). Mind the basel gap. *Journal of International Financial Markets, Institutions and Money 79*, 101605.

Kamber, G., J. Morley, and B. Wong (2018). Intuitive and reliable eatimates of the output gap from a Beveridge-Nelson filter. *The Review of Economics and Statistics 100*, 550–566.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kauko, K. and E. Tölö (2020). On the long-run calibration of the credit-to-GDP gap as a banking crisis predictor. *Finnish Economic Papers 2*, 1–12.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies 65*(3), 361–393.

King, R. and S. Rebelo (1993). Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control 17*, 207–231.

Kishor, N. K. and E. F. Koenig (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business & Economic Statistics 30*(2), 181–190.

Klebaner, F. C. (2012). *Introduction to stochastic calculus with applications* (3rd ed. ed.). Imperial College Press London.

Kloek, T. and H. K. van Dijk (1978). Bayesian estimates of equation system parameters: An application of integration by monte carlo. *Econometrica 46*(1), 1–19.

Knetsch, T. A. and H.-E. Reimers (2009). Dealing with benchmark revisions in real-time data: The case of german production and orders statistics. *Oxford Bulletin of Economics and Statistics 71*(2), 209–235.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koopman, S. J., A. Lucas, and M. Scharth (2016, 03). Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. *The Review of Economics and Statistics 98*(1), 97–110.

Krüger, F., T. E. Clark, and F. Ravazzolo (2017). Using entropic tilting to combine bvar forecasts with external nowcasts. *Journal of Business & Economic Statistics 35*(3), 470–485.

Lang, J., C. Izzo, S. Fahr, and J. Ruzicka (2019). Anticipating the bust: A new cyclical systemic risk indicator to assess the likelihood and severity of financial crises. *ECB Occasional Paper 219*.

Lenza, M. and G. Primiceri (2020, September). How to estimate a var after march 2020. Working Paper 27771, National Bureau of Economic Research.

López-Salido, D. and F. Loria (2020, February). Inflation at Risk. Finance and Economics Discussion Series 2020-013, Board of Governors of the Federal Reserve System (U.S.).

Mankiw, N. G. and M. D. Shapiro (1986). News or noise: An analysis of GNP revisions. *Survey of Current Business 66*, 20–25.

Montes-Galdon, C. and E. Ortega (2022). Skewed SVARS: Tracking the structural sources of macroeconomic tail risks. *Advances in Econometrics (forthcoming)*.

Montes-Galdón, C., J. Paredes, and E. Wolf (2022). Conditional density forecasting: a tempered importance sampling approach. Working Paper Series 2754, European Central Bank.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing 11*(2), 125–139.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review 91*(4), 964–985.

Orphanides, A. and S. van Norden (2002). The unreliability of output gap estimates in real time. *Review of Economics and Statistics 84*, 569–583.

Phillips, P. and S. Jin (2021). Business cycles, trend elimination, and the HP filter. *International Economic Review 62*, 469–520.

Phillips, P. and Z. Shi (2021). Boosting: Why you can use the HP filter. *International Economic Review 62*, 521–570.

Pitt, M. K., R. dos Santos Silva, P. Giordani, and R. Kohn (2012). On some properties of markov chain monte carlo simulation methods based on the particle filter. *Journal of Econometrics 171*(2), 134–151.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Prasad, A., S. Elekdag, P. Jeasakul, R. Lafarguette, A. Alter, A. X. Feng, C. Wang, and C. A. Gust (2019). Growth at Risk: Concept and Application in IMF Country Surveillance. Working Paper 036, International Monetary Fund.

Quast, J. and M. Wolters (2022). Reliable real-time output gap estimates based on a modified Hamilton filter. *Journal of Business and Economics Statistics*.

Ravn, M. and H. Uhlig (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics 84*, 371–380.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical Science 16*(4), 351–367.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking 37*(3), 383–401.

Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy 97*(2), 251–287.

Schön, T. B., F. Lindsten, J. Dahlin, J. Wagberg, C. A. Naesseth, A. Svensson, and L. Dai (2015). Sequential monte carlo methods for system identification. *IFAC-PapersOnLine 48*(28), 775–786. 17th IFAC Symposium on System Identification SYSID 2015.

Schüler, Y. (2018). Detrending and financial cycle facts across G7 countries: Mind a spurious medium term! *ECB Working Paper 2138*.

Schüler, Y. (2019). How should we filter economic time series? *SSRN*.

Schüler, Y. (2020). On the credit-to-GDP gap and spurious medium-term cycles. *Economics Letters 192*, 109245.

Schüler, Y. (2021). On the cyclical properties of Hamilton's regression filter. *SSRN Working Paper*.

Schüler, Y., P. Hiebert, and T. Peltonen (2015). Characterising the financial cycle: A multivariate and time-varying approach. *ECB Working Paper 1846*.

Schüler, Y., P. Hiebert, and T. Peltonen (2020). Financial cycles: Characterisation and real-time measurement. *Journal of International Money and Finance 100*, 102082.

Stock, J. and M. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*, 293–335.

Strohsal, T., C. Proaño, and J. Wolters (2019). Characterizing the financial cycle: Evidence from a frequency domain analysis. *Journal of Banking and Finance 106*, 568–591.

Tente, N., I. Stein, L. Silbermann, and T. Deckers (2015). The countercyclical capital buffer in Germany. *Deutsche Bundesbank*.

Waggoner, D. and T. Zha (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics 81*(4), 639–651.

Wolf, E. (2022). Estimating Growth at Risk with Skewed Stochastic Volatility Models. Working Paper Series 2022/2, Freie Universität Berlin.

List of Publications

Publications in peer-reviewed Journals

Strohsal, T., and Wolf, E. (2021). Data revisions to German national accounts: Are initial releases good nowcasts?.*International Journalof Forecasting 36(4)*,1252-1259.
Publication of Chapter 2.

Zusammenfassung

In den Wirtschaftswissenschaften ist die Vorhersage der künftigen Wirtschaftslage eine Schüsselfrage für die Entscheidungsfindung von Wirtschaftsakteuren und politischen Entscheidungsträgern gleichermaßen. Daher spielen ökonometrische Methoden zur genauen Messung des Zustands unserer Wirtschaft sowie zur Vorhersage und Quantifizierung der künftigen Verteilungen wichtiger makroökonomischer Variablen eine wichtige Rolle im Instrumentarium nationaler und internationaler Institutionen. Ziel dieser Dissertation ist es, Zeitreihenmethoden anzuwenden und weiterzuentwickeln, um den aktuellen Zustand der Wirtschaft in Echtzeit zu bestimmen und den zukünftigen Pfad der Wirtschaft zusammen mit den damit verbundenen Risiken in Form von Dichteprognosen vorherzusagen.

Das erste Kapitel, eine Gemeinschaftsarbeit mit Yves Schüler und Frieder Mokinski, zeigt, dass man den einseitigen Hodrick-Prescott-Filter (HP-1s) nicht als Echtzeitversion des zweiseitigen Hodrick-Prescott-Filters (HP-2s) verwenden sollte: Erstens entfernt HP-1s hinsichtlich der extrahierten zyklischen Komponente tieffrequente Schwankungen nicht im gleichen Maße wie HP-2s. Zweitens dämpft HP-1s Fluktuationen bei allen Frequenzen - sogar bei denen, die es extrahieren soll. Als Abhilfe schlagen wir zwei Anpassungen für HP-1s vor, die seine Eigenschaften an die von HP-2s angleichen sollen: (1) ein niedrigerer Wert für den Glättungsparameter und (2) eine multiplikative Neuskalierung der extrahierten zyklischen Komponente. Für HP-2s mit $\lambda = 1600$ (Wert des Glättungsparameters) verwendet der angepasste einseitige HP-Filter beispielsweise $\lambda^* = 650$ und skaliert die extrahierte zyklische Komponente um einen Faktor von $1,1513$. Anhand von simulierten und empirischen Daten veranschaulichen wir die Relevanz dieser Anpassungen. Beispielsweise können Finanzzyklen um 70% volatiler erscheinen als Konjunkturzyklen, obwohl sich die Volatilitäten in Wirklichkeit nur geringfügig unterscheiden.

Das zweite Kapitel ist eine gemeinsame Arbeit mit Till Strohsal. Wir zeigen, dass die Revisionen der deutschen Volkswirtschaftlichen Gesamtrechnungen verzerrt, großund vorhersehbar

sind. Darüber hinaus kann durch den Einsatz von Filtertechniken, die für die Verarbeitung von revisionsbehafteten Daten entwickelt wurden, die Echtzeit-Prognoseleistung von Erstveröffentlichungen um bis zu 23% gesteigert werden. Für das reale BIP-Wachstum insgesamt ist die Erstveröffentlichung jedoch eine optimale Prognose. In Anbetracht der Ergebnisse für disaggregierte Variablen scheint die Mittelwertbildung von Verzerrungen und Ineffizienzen auf der Ebene des Gesamt-BIP jedoch eher Glück als eine gute Prognose zu sein.

Im dritten Teil wird ein Modell der schiefen stochastischen Volatilität (SSV) vorgeschlagen, um asymmetrische makroökonomische Tail-Risiken im Sinne der bahnbrechenden Arbeit "Vulnerable Growth" von Adrian, Boryaschenko und Giannone (2019) zu schätzen. Im Gegensatz zu ihrem semiparametrischen Ansatz erfasst das SSV-Modell die Entwicklung der bedingten Dichte des künftigen US BIP-Wachstums in einem parametrischen, nicht-linearen, nicht-gaußschen Zustandsraummodell. Dies ermöglicht es, die Auswirkungen exogener Variablen auf die verschiedenen Momente der bedingten Verteilung statistisch zu testen und liefert ein Bewegungsgesetz zur Vorhersage künftiger Werte von Volatilität und Schiefe. Das Modell wird mit Hilfe eines Partikel-MCMC-Algorithmus geschätzt. Um die Genauigkeit der Schätzung zu erhöhen, verwende ich einen temperierten Partikelfilter, der die zeitlich variierende Volatilität und die Asymmetrie der Dichtfunktionen berücksichtigt. Ich stelle fest, dass die finanziellen Bedingungen den Mittelwert, die Varianz und die Schiefe der bedingten Verteilung des zukünftigen US BIP-Wachstums beeinflussen. Mit einem Bayes-Faktor von 1612,18 wird das SSV-Modell von den Daten stark gegenüber einem symmetrischen stochastischen Volatilitätsmodell bevorzugt.

Das vierte Kapitel ist eine gemeinsame Arbeit mit Carlos Montes-Galdón und Joan Paredes und schlägt eine neue und robuste Methode vor, um bedingte Dichteprognosen zu erhalten, die auf Informationen basieren, die nicht in einem ursprünglichen ökonometrischen Modell enthalten sind. Die Methodik ermöglicht es, auf die erwarteten Marginalverteilungen einer Auswahl von Variablen in einem Modell zu konditionieren, anstatt nur auf die zukünftigen Pfade, wie es in der Literatur für bedingte Prognosen üblich ist. Der vorgeschlagene Algorithmus, der auf dem temperierten Importance Sampling basiert, passt die modellbasierten Dichteprognosen an die Zielverteilungen an, die dem Forscher zur Verfügung stehen. Als Beispiel wird in diesem Papier gezeigt, wie der Algorithmus für die Konditionierung der Vorhersagedichten eines BVAR- und eines DSGE-Modells auf Informationen über die Marginalverteilungen zukünftiger Ölpreise implementiert werden kann. Die Ergebnisse zeigen, dass erhöhte asymmetrische Aufwärtsrisiken für die Ölpreise zu Aufwärtsrisiken

für die Inflation sowie zu einer höheren Kerninflation über den betrachteten Prognosehorizont führen. Schließlich zeigt eine Echtzeitprognose, dass die Einführung marktbasierter Informationen über den Ölpreis die Inflations- und BIP-Prognosen in Krisenzeiten wie der COVID-Pandemie verbessert.

Summary

In economics, predicting the future state of the economy is a key issue for the decision making of economic agents and policy makers alike. Therefore, econometric methods to accurately measure the state of our economy as well as to predict and quantify the future distributions of key macroeconomic variables play an important role in the toolkit of national and international policy institutions and statistical agencies. This dissertation aims to apply and develop new time series methods to determine the current state of the economy in real time as well as to predict the future path of the economy together with the associated risks in form of density forecasts.

The first chapter, which is joint work with Yves Schüler and Frieder Mokinski, shows that one should not use the one-sided Hodrick-Prescott filter (HP-1s) as the real-time version of the two-sided Hodrick-Prescott filter (HP-2s): First, in terms of the extracted cyclical component, HP-1s fails to remove low-frequency fluctuations to the same extent as HP-2s. Second, HP-1s dampens fluctuations at all frequencies – even those it is meant to extract. As a remedy, we propose two small adjustments to HP-1s, aligning its properties closely with those of HP-2s: (1) a lower value for the smoothing parameter and (2) a multiplicative rescaling of the extracted cyclical component. For example, for HP-2s with $\lambda = 1{,}600$ (value of smoothing parameter), the adjusted one-sided HP filter uses $\lambda^* = 650$ and rescales the extracted cyclical component by a factor of 1.1513. Using simulated and empirical data, we illustrate the relevance of these adjustments. For instance, financial cycles may appear to be 70% more volatile than business cycles, where in fact volatilities differ only marginally.

The second chapter is joint work with Till Strohsal. We show that revisions to German national accounts are biased, large and predictable. Moreover, using filtering techniques designed to process data subject to revisions, the real-time forecasting performance of initial releases can be increased by up to 23%. For total real GDP growth, however, the initial release is an optimal forecast. Yet, given the results for disaggregated variables, the

averaging-out of biases and inefficiencies at the aggregate GDP level appears to be good luck rather than good forecasting.

The third proposes a Skewed Stochastic Volatility (SSV) model to estimate asymmetric macroeconomic tail risks in the spirit of Adrian et. al's seminal paper "Vulnerable Growth". In contrary to their semi-parametric approach, the SSV model captures the evolution of the conditional density of future US GDP growth in a parametric, non-linear, non-Gaussian state space model. This allows to statistically test the effect of exogenous variables on the different moments of the conditional distribution and provides a law of motion to predict future values of volatility and skewness. The model is estimated using a particle MCMC algorithm. To increase estimation accuracy, I use a tempered particle filter that takes the time-varying volatility and asymmetry of the densities into account. I find that financial conditions affect the mean, variance and skewness of the conditional distribution of future US GDP growth. With a Bayes ratio of 1612.18, the SSV model is strongly favored by the data over a symmetric Stochastic Volatility (SV) model.

The fourth paper is joint work with Carlos Montes-Galdón and proposes a new and robust methodology to obtain conditional density forecasts, based on information not contained in an initial econometric model. The methodology allows to condition on expected marginal densities for a selection of variables in the model, rather than just on future paths as it is usually done in the conditional forecasting literature. The proposed algorithm, which is based on tempered importance sampling, adapts the model-based density forecasts to target distributions the researcher has access to. As an example, this paper shows how to implement the algorithm by conditioning the forecasting densities of a BVAR and a DSGE model on information about the marginal densities of future oil prices. The results show that increased asymmetric upside risks to oil prices result in upside risks to inflation as well as higher core-inflation over the considered forecasting horizon. Finally, a real-time forecasting exercise yields that introducing market-based information on the oil price improves inflation and GDP forecasts during crises times such as the COVID pandemic.

Erklärungen

Erklärung gem. §4 PO

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur ÃberprÃ¼fung vorgelegen hat.

Berlin, den 15. Mai 2023

. . . . . .

Elias Wolf

Erklärung gem. § 10 Abs. 3

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe:

- MATLAB
- Python
- R
- Curta - High-Performance Computing system of FU Berlin
- Excel
- LaTeX
- Siehe References, für die verwendete Literatur

Auf dieser Grundlage habe ich die Arbeit selbstständig verfasst.

Berlin, den 15. Mai 2023

. . . . . .

Elias Wolf