



Multivariate Economic Tail Risk and Scenario Analysis using the Survey of Professional Forecasters

Manuel Schick

Anne Opschoor

AWI DISCUSSION PAPER SERIES NO. 771

April 2026

Multivariate Economic Tail Risk and Scenario Analysis using the Survey of Professional Forecasters*

Manuel Schick[†], Anne Opschoor[‡]

April 15, 2026

Abstract

This paper proposes forecasting joint tail risks for key macroeconomic indicators, GDP growth, inflation, and unemployment, using the US Survey of Professional Forecasters (SPF). By incorporating SPF consensus forecasts into the conditional mean of AR-GARCH-type models, the accuracy of univariate and multivariate predictive densities is significantly improved. Modeling a constant correlation matrix captures strong dependencies, particularly between GDP growth and unemployment. Using US data from 1990 to 2024, we show that the joint modeling framework enables scenario-based analysis in which predictive densities, conditioned on adverse developments in other variables, differ substantially from the baseline marginal distributions. The framework allows for a formal out-of-sample evaluation of joint predictive densities and a transparent assessment of conditional tail risks.

Keywords: Growth-at-Risk, Multivariate Predictive Densities, GARCH, Tail Risk, Macroeconomic Forecasting

JEL Classification: C22, C52, C53

*We would like to thank Sarah Arndt, Christian Conrad, Timo Dimitriadis, Fabian Krueger, Julius Schoelkopf, and Bernd Schwaab, as well as seminar and conference participants at the Macro & Econometrics Seminar at Heidelberg University (May 2024), the HKMetrics Research Network Workshop at the University of Mannheim (June 2025), the Workshop on Macro- and Financial Econometrics at the School of Business, Örebro University (November 2025), and the Finance@VU lunch seminar at Vrije Universiteit Amsterdam (March 2026), for their helpful comments and suggestions.

[†]Heidelberg University and Vrije Universiteit Amsterdam, Email: manuel.schick@awi.uni-heidelberg.de.

[‡]Vrije Universiteit Amsterdam and Tinbergen Institute, Email: a.opschoor@vu.nl.

1 Introduction

Policymakers routinely rely on scenario analysis to assess risks to the macroeconomy. Prominent examples include the *World Economic Outlook* of the International Monetary Fund (IMF), which presents baseline and adverse scenarios, and the Federal Reserve’s *Tealbook*, which provides baseline and alternative scenarios for monetary policy deliberations. Yet these scenario exercises often rely on judgment or calibration to stress indicators, presenting forward-looking projections without formally treating them as forecasts. As a result, they offer only a partial view of the joint predictive distribution and, crucially, lack a formal statistical assessment of forecast accuracy.

In this paper, we extend the Growth-at-Risk (GaR) framework ([Adrian et al., 2019](#)) into a multivariate setting to generate distribution-based macroeconomic forecast scenarios in real time. Scenarios are defined as conditional predictive distributions, for example GDP growth conditional on inflation and unemployment being in their upper deciles, in the spirit of conditional Value at Risk (CoVaR) (see, e.g., [Girardi and Ergün, 2013](#); [Adrian and Brunnermeier, 2016](#)). Methodologically, we incorporate Survey of Professional Forecasters (SPF) expectations into the conditional mean within a parametric AR-GARCH framework—henceforth SPF-GARCH—([Schick, 2024](#)), extending the univariate approach to a multivariate system that jointly models GDP growth, inflation, and unemployment and captures time-varying volatility and cross-variable correlations. A central strength of our approach is that the underlying parametric joint predictive densities can be used to transparently construct scenarios and to evaluate the models’ out-of-sample forecasting performance, a feature absent in existing macroeconomic scenario exercises. However, in line with recent results showing that CoVaR does not admit a strictly consistent scoring function ([Fissler and Hoga, 2024](#)), the conditional scenarios themselves cannot be evaluated directly, motivating our focus on the joint predictive density.

Despite their widespread use, existing approaches to scenario analysis face important limitations. Institutional exercises often rely on structural models such as vector autoregressions (VARs), which impose strong assumptions about dynamics and shocks, or on

judgment-driven projections that lack statistical transparency. In the academic literature, [Chavleishvili and Manganelli \(2024\)](#) develop a quantile VAR to stress-test GDP growth under adverse financial shocks, noting that stress testing in central banks often lacks a rigorous econometric foundation. Related work by [Chavleishvili et al. \(2026\)](#) develops a Bayesian structural quantile VAR for macro-prudential policy assessment that also allows for counterfactual scenario analysis. [Corradi and Llorens-Terrazas \(2026\)](#) extend the concept of GaR to joint tail risks for growth and inflation. Alternatively, [Adrian et al. \(2025\)](#) propose an entropic tilting approach to construct full scenario densities from institutional inputs, such as a baseline and a handful of reported quantiles, providing a clever synthesis of model-based forecasting and judgmental narratives, but still relying on externally imposed scenario paths. We address these limitations by constructing scenarios from an estimated multivariate predictive density rather than imposed paths.

A related strand of work develops the Growth-in-Stress (GiS) framework introduced by [González-Rivera et al. \(2019\)](#) and extended by [González-Rivera et al. \(2024\)](#), which constructs conditional growth densities under stressed macro-financial factor scenarios using dynamic factor models. Their approach offers a way to assess the exposure of GDP growth to severe yet plausible global or domestic shocks. However, the GiS framework typically focuses on GDP growth alone and conditions on exogenously stressed latent factors, rather than estimating a joint predictive density across macroeconomic variables directly. While it provides meaningful tail scenarios for GDP growth, it does not allow for a formal out-of-sample evaluation of the resulting predictive density.

In the GaR literature, conditional quantile regression (QR) has been the predominant tool for modeling tail risk, relying on observable predictors such as financial conditions (e.g., [Giglio et al., 2016](#); [Adrian et al., 2019](#); [Figueres and Jarociński, 2020](#); [Adrian et al., 2022](#); [Ferrara et al., 2022](#); [Castelnuovo and Mori, 2025](#)). These approaches require specifying which predictors drive downside risks, and the predictive content of such variables can vary substantially across episodes. By contrast, the GARCH framework does not rely on explicit tail-risk predictors and instead allows tail probabilities to evolve directly

from the estimated conditional mean and variance. As [Brownlees and Souza \(2021\)](#) demonstrate, volatility-based approaches can match the performance of QR models in capturing macro-financial risks even without incorporating financial-stress indicators. This predictor-agnostic flexibility makes GARCH a natural foundation for constructing jointly modeled scenario distributions.

A key ingredient of our framework is the use of consensus forecasts from the SPF. Instead of building a real-time forecasting model from macroeconomic and financial indicators, typically released with publication lags and at mixed frequencies, we incorporate the current outlook directly through SPF median forecasts, implying minimal data requirements. At the same time, SPF forecasts are empirically competitive: particularly at short horizons, the consensus often outperforms standard time-series models ([Stark, 2010](#); [Faust and Wright, 2013](#)), likely because expert aggregation quickly adapts to structural breaks and incorporates a broad information set ([Ang et al., 2007](#)).

Prior work has also highlighted the value of SPF-based approaches for density forecasting and risk assessment. For instance, [Krüger et al. \(2017\)](#) combine survey expectations with model forecasts using entropic tilting to construct density forecasts, while [Adams et al. \(2021\)](#) use financial conditions to explain fluctuations in the uncertainty around SPF projections to construct macroeconomic risk measures. Building on this literature, [Schick \(2024\)](#) shows that using SPF expectations in a GARCH framework achieves reliable real-time estimates of GDP tail risk, with density nowcast accuracy comparable to more sophisticated and data-rich methods such as the Bayesian mixed-frequency VAR with stochastic volatility studied by [Carriero et al. \(2022\)](#). These results motivate our use of SPF expectations to model the conditional mean in a multivariate density framework.

Our contribution is threefold. Conceptually, we extend the GaR framework from univariate tail-risk measures to a multivariate, distribution-based approach that delivers economically interpretable macroeconomic forecast scenarios. Methodologically, we generate joint predictive densities from a multivariate GARCH framework anchored by SPF expectations, allowing for transparent scenario construction and formal out-of-sample

evaluation. In particular, we model the distribution of GDP growth, inflation, and unemployment with time-varying volatilities and correlations, enabling conditional scenario analysis along multiple macroeconomic dimensions. Empirically, we apply the framework to US data and highlight its policy relevance by constructing real-time scenarios both before and during the COVID-19 pandemic. Overall, the framework provides policymakers with a transparent and statistically grounded tool to assess joint macroeconomic risks across multiple variables.

We find that incorporating SPF median forecasts significantly improves the accuracy of joint predictive densities for GDP growth, inflation, and unemployment. Forecast gains are particularly pronounced around the survey release dates in the middle of the quarter, when SPF-GARCH log-scores improve markedly for both nowcasts and one-step-ahead forecasts. Moreover, the SPF substantially shapes scenario distributions compared to a purely backward-looking AR-GARCH benchmark, with conditional scenario densities differing from their marginal counterparts in ways that are policy relevant. Our focus is on time-varying tail probabilities arising from movements in the conditional mean and variance, and the results are robust to alternative distributional assumptions, including specifications allowing for skewness and fat-tailed innovations.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature, followed by a description of the real-time data in Section 3. The methodological framework, which includes the multivariate GARCH model and the construction of conditional scenarios, is introduced in Section 4. Section 5 outlines the evaluation metrics used to assess the predictive densities, and Section 6 reports the empirical findings, including out-of-sample performance and scenario analyses. Section 7 concludes.

2 Related Literature

A growing strand of the literature considers scenario-based assessments of macro risk by constructing conditional distributions. The GiS framework of [González-Rivera et al. \(2019\)](#)

and [González-Rivera et al. \(2024\)](#) uses dynamic factor models to generate conditional GDP growth densities under stressed macro-financial conditions, assessing how the economy behaves under adverse but plausible scenarios. Complementary approaches use quantile VARs to trace how downside risks to GDP respond to adverse financial shocks ([Chavleishvili and Manganelli, 2024](#)). Earlier work has also developed VAR-based conditional forecast densities that impose restrictions on future values of endogenous variables to generate scenarios (see, e.g., [Waggoner and Zha, 1999](#); [Bańbura et al., 2015](#); [Antolín-Díaz et al., 2021](#)). These methods are powerful for scenario construction but do not deliver a parametric joint predictive density that can be evaluated out-of-sample. Our contribution bridges this gap by estimating SPF-anchored multivariate predictive densities, enabling both conditional scenario analysis and formal forecast evaluation.

This paper also relates to the GaR literature, which models downside risks to GDP growth using predictive quantile regressions ([Adrian et al., 2019](#)) and has been extended to Inflation- and Unemployment-at-Risk (see, e.g., [Adams et al., 2021](#); [Carriero et al., 2024](#); [López-Salido and Loria, 2024](#)). While most GaR models rely on financial variables, recent evidence shows that much of the variation in GDP tail risk can be explained by time-varying mean and variance rather than changing skewness ([Carriero et al., 2022, 2024](#)). Building on this insight, [Brownlees and Souza \(2021\)](#) and [Schick \(2024\)](#) develop GARCH-based GaR frameworks that capture evolving macroeconomic uncertainty without conditioning on financial predictors. We generalize this approach to a multivariate setting, estimating the joint predictive density of GDP growth, inflation, and unemployment.

Methodologically, this paper is also related to systemic risk analyses based on CoVaR. Applied to our setting, [Adrian and Brunnermeier \(2016\)](#) would define a CoVaR-type measure for GDP growth as the quantile of its predictive distribution conditional on unemployment being exactly at some extreme quantile level. Our scenario construction instead follows the definition of [Girardi and Ergün \(2013\)](#), where the conditioning variable exceeds a threshold rather than equaling it. Conditioning on exceedances is generally more plausible, as it captures increasingly severe realizations and avoids the counter-intuitive

behavior where the implied risk level becomes less severe when dependence strengthens.

However, conditional quantile scenarios of this type cannot be directly backtested, since CoVaR is not elicitable and therefore does not admit strictly consistent scoring functions. By contrast, our focus on multivariate tail risk relies on the full predictive density, which can be evaluated out-of-sample. Nonetheless, VaR and CoVaR can be evaluated jointly in a lexicographical sense (see, e.g., [Fissler and Hoga, 2024](#); [Dimitriadis and Hoga, 2025](#)), although such joint evaluation is beyond the scope of this paper.

The analysis further connects to the literature that integrates survey information into model-based measures of macroeconomic uncertainty. Studies such as [Reifschneider and Tulip \(2019\)](#) and [Clark et al. \(2020\)](#) show that the variance of forecast errors varies over time, suggesting that uncertainty around professional forecasts is itself state dependent. Survey expectations, particularly from the SPF, provide accurate short-horizon projections and often outperform standard time-series models ([Ang et al., 2007](#); [Giannone et al., 2008](#); [Stark, 2010](#); [Bańbura et al., 2021](#)). Combining such expectations with econometric models has proven effective for constructing density forecasts ([Krüger et al., 2017](#); [Adams et al., 2021](#); [Schick, 2024](#)). This paper builds on that evidence by using SPF projections as timely signals of professional forecasters' expectations about macroeconomic conditions while allowing for time-varying forecast uncertainty (see also [Clark et al., 2025](#)). By linking consensus forecasts with conditional density modeling, our framework provides a practical tool for real-time risk monitoring, in line with the policy-oriented applications of GaR at institutions such as the IMF ([Prasad et al., 2019](#)).

Lastly, this paper draws on the literature on real-time macroeconomic forecasting. Studies such as [Giannone et al. \(2008\)](#) and [Bańbura et al. \(2013\)](#) show that incorporating incoming data throughout the quarter improves GDP nowcasts, while high-frequency indicators enhance both point and tail risk forecasts ([Andreou et al., 2013](#); [Ferrara et al., 2022](#)). [Carriero et al. \(2022\)](#) further demonstrate that using broad real-time macroeconomic information strengthens Growth-at-Risk estimates. Our framework complements this work by producing real-time, multivariate predictive densities that evolve with professional

forecasts and macroeconomic uncertainty.

3 Real-time data and predictors

In order to avoid look-ahead bias, we conduct a real-time out-of-sample forecasting analysis. To this end, we use monthly real-time vintage data on real GDP, the GDP Price Index, and the unemployment rate, obtained from the Real-Time Data Set for Macroeconomists (RTDSM) provided by the Federal Reserve Bank of Philadelphia. Information on the release dates of these monthly snapshots is available via ALFRED.

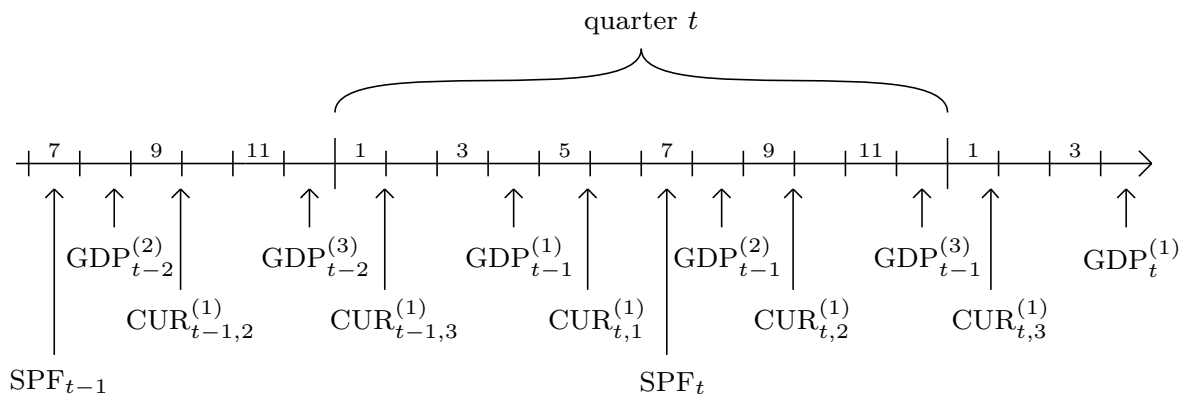
Figure 1 provides an overview of the release schedule for GDP and unemployment data in weekly terms. Typically, the Bureau of Economic Analysis (BEA) publishes the *advance* estimate of quarterly GDP, i.e., the first official release, at the end of the first month of the following quarter. Accordingly, the advance estimate, $GDP_{t-1}^{(1)}$, becomes available around week 4 of quarter t . One and two months later, the BEA releases the *preliminary* and *final* estimates, denoted $GDP_{t-1}^{(2)}$ and $GDP_{t-1}^{(3)}$, respectively. However, even the final estimate remains subject to subsequent revisions and methodological updates, which are reflected in the monthly vintage data. For inflation, we use the GDP Price Index which is highly correlated with other inflation measures. Since its release coincides with that of GDP, this simplifies the tracking of release dates.¹

Monthly unemployment data are released more timely than GDP. The Civilian Unemployment Rate (CUR) is typically published at the end of the first week of the subsequent month. As a result, the first estimate of the unemployment rate for the first month of quarter t , denoted $CUR_{t,1}^{(1)}$, becomes available in week 5. One and two months later, the corresponding first estimates for the second and third months of the quarter, $CUR_{t,2}^{(1)}$ and $CUR_{t,3}^{(1)}$, are released. This timing enables the construction of a preliminary estimate of quarterly unemployment for quarter t early in the subsequent quarter.

The RTDSM provides quarterly vintages of the monthly unemployment series, reflecting

¹Prior to 1992, the BEA reported Gross National Product (GNP) and the GNP Price Index instead of GDP.

Figure 1: Release calendar



Notes: This graph sketches the release and revision dates of GDP and the GDP deflator, the civilian unemployment rate (CUR), and the release dates of the Survey of Professional Forecasters (SPF). GDP data refer jointly to the national accounts release of real GDP and the GDP deflator. For example, $GDP_t^{(1)}$ denotes the initial release for quarter t , while $CUR_{t,1}^{(1)}$ denotes the first release of the unemployment rate for the first month of quarter t . SPF releases are not subject to revisions. The numbers on the time line indicate the weeks within each quarter.

data as available around the middle of each quarter. Each quarterly vintage includes the most recent estimate as well as minor revisions to earlier observations, typically stemming from seasonal adjustment updates or definition changes. In practice, these revisions are small and usually negligible. Thus, constructing the monthly information flow from quarterly real-time vintages is not expected to introduce look-ahead bias, despite the underlying series being monthly.

Lastly, Figure 1 also illustrates the release schedule of the SPF. The SPF is a quarterly survey of U.S. macroeconomic forecasts, originally conducted by the American Statistical Association and the National Bureau of Economic Research. Since 1990, the Federal Reserve Bank of Philadelphia has administered the survey, which has since comprised a panel of approximately 40 professional forecasters (Clements et al., 2023). The survey began in the fourth quarter of 1968 and is released around the middle of each quarter. Respondents provide forecasts for the current quarter and up to four quarters ahead, covering key macroeconomic indicators. Specifically, the SPF reports projections for the quarter-over-quarter annualized growth rates of real GDP and the GDP Price Index, as well as quarterly averages of the monthly Civilian Unemployment Rate.

We aim to produce one-step-ahead forecasts and nowcasts with forecast origin in quarter t , updating these forecasts on a weekly basis. Based on the release schedules of the relevant macroeconomic indicators, Table 1 summarizes the latest available data vintages in each week of quarter t . Accordingly, in a given week w , we use the most recent vintage of real GDP available at that time to construct a real-time snapshot of annualized quarterly growth rates:

$$y_{s|t,w} = 400 \cdot \left(\log(\text{GDP}_{s|t,w}) - \log(\text{GDP}_{s-1|t,w}) \right), \quad (1)$$

where $\text{GDP}_{s|t,w}$ denotes quarterly GDP in quarter s , taken from the latest vintage available in week w of quarter t . Analogously, the quarterly inflation rate is defined as:

$$\pi_{s|t,w} = 400 \cdot \left(\log(\text{PGDP}_{s|t,w}) - \log(\text{PGDP}_{s-1|t,w}) \right), \quad (2)$$

where $\text{PGDP}_{s|t,w}$ is the GDP Price Index for quarter s from the latest vintage available in week w of quarter t .

Table 1: Real-time vintages available in quarter t

week	Variables			
1	$\text{GDP}_{t-2 t,12} = \text{GDP}_{t-2}^{(3)}$	$\text{CUR}_{t-1,3 t,12} = \text{CUR}_{t-1,3}^{(1)}$	$\text{SPF}_{t+h t-1}$	$\text{NFCI}_{t-1,12 t,1}$
3	$\text{GDP}_{t-2 t,12} = \text{GDP}_{t-2}^{(3)}$	$\text{CUR}_{t-1,3 t,12} = \text{CUR}_{t-1,3}^{(1)}$	$\text{SPF}_{t+h t-1}$	$\text{NFCI}_{t,2 t,3}$
5	$\text{GDP}_{t-1 t,4} = \text{GDP}_{t-1}^{(1)}$	$\text{CUR}_{t,1 t,5} = \text{CUR}_{t,1}^{(1)}$	$\text{SPF}_{t+h t-1}$	$\text{NFCI}_{t,4 t,5}$
7	$\text{GDP}_{t-1 t,4} = \text{GDP}_{t-1}^{(1)}$	$\text{CUR}_{t,1 t,5} = \text{CUR}_{t,1}^{(1)}$	$\text{SPF}_{t+h t}$	$\text{NFCI}_{t,6 t,7}$
9	$\text{GDP}_{t-1 t,8} = \text{GDP}_{t-1}^{(2)}$	$\text{CUR}_{t,2 t,9} = \text{CUR}_{t,2}^{(1)}$	$\text{SPF}_{t+h t}$	$\text{NFCI}_{t,8 t,9}$
11	$\text{GDP}_{t-1 t,8} = \text{GDP}_{t-1}^{(2)}$	$\text{CUR}_{t,1 t,9} = \text{CUR}_{t,2}^{(1)}$	$\text{SPF}_{t+h t}$	$\text{NFCI}_{t,10 t,11}$

Notes: This table demonstrates the latest vintages of real GDP (GDP), the Civilian Unemployment Rate (CUR), h -quarters-ahead SPF projections, and NFCI data available for odd numbered weeks throughout quarter t . For instance, $\text{GDP}_{t-2}^{(3)}$ denotes the third release of GDP for quarter $t-2$ from the latest vintage in the respective week of quarter t . Similarly, $\text{CUR}_{t-1,3}^{(1)}$ denotes the first release of the CUR of the third month of quarter $t-1$ from the latest real-time vintage.

At quarter t , to construct real-time forecasts of GDP and inflation, the index $s = 1, \dots, t-j$ runs up to $t-j$, where $j = 2$ prior to the release of the previous quarter's

GDP, and $j = 1$ otherwise. This ensures that forecasts are based only on information available at the time, avoiding look-ahead bias.

The quarterly unemployment rate is constructed as the average of the monthly Civilian Unemployment Rates,

$$u_{s|t,w} = \frac{1}{3} \sum_{m=1}^3 \text{CUR}_{s,m|t,w}, \quad (3)$$

where $\text{CUR}_{s,m|t,w}$ denotes the unemployment rate for month $m = 1, 2, 3$ in quarter s , taken from the most recent vintage available in week w of quarter t . Because the monthly figures are released in a timely manner, a first estimate of the previous quarter’s unemployment rate, $u_{t-1|t,w}$, typically becomes available about one week after the end of quarter $t - 1$. As a result, the estimation sample generally includes data up to $s = t - 1$.

Although revisions to the unemployment data are generally small or negligible, the real-time flow of monthly releases still matters for constructing the appropriate “lag” of unemployment. Simply regressing $u_{s|t,w}$ on $u_{s-1|t,w}$ would ignore new information arriving during the current quarter. To reflect the data available in week w , we define the predictor $\tilde{u}_{s-1|t,w}$ as

$$\tilde{u}_{s-1|t,w} = \begin{cases} u_{s-1|t,w} & \text{if } w \leq 4 \\ \frac{1}{M} \sum_{m=1}^M \text{CUR}_{s,m|t,w} & \text{if } w > 4, \end{cases} \quad (4)$$

where $M \in \{1, 2\}$, depending on the number of monthly observations available by week w . For example, in week $w = 1$, only the full quarterly unemployment figure from $s = t - 1$ is available. By week $w = 5$, however, the unemployment rate for the first month of quarter t , $\text{CUR}_{t,1|t,5}$, has typically been released, so $M = 1$. Once the second month’s data becomes available, $\tilde{u}_{s-1|t,w}$ is computed as the average of the first two monthly unemployment rates.

Having defined the construction of our real-time variables, Table 2 presents summary statistics for GDP growth, inflation, and the unemployment rate over 1990–2024, based on the first vintage available in 2025:Q2. Including the COVID-19 pandemic in panel (b)

highlights the extreme economic downturns and recoveries during this period, as well as the surge in inflation between 2021 and 2022. While average GDP growth remains close to 2.5% across both samples, its volatility nearly doubles when the pandemic period is included, reflecting historically large negative and positive quarterly growth rates. Inflation is close to 2% in the pre-pandemic sample but becomes markedly more volatile thereafter, and the unemployment rate exhibits a sharp spike during the lockdown period, reaching unprecedented levels.

Table 2: Summary statistics

	Panel (a): 1990 – 2019				
	Mean	Std. dev.	Min	Max	Median
Real GDP growth	2.491	2.326	-8.852	7.221	2.498
Inflation	1.974	0.910	-0.533	4.451	1.969
Unemployment rate	5.842	1.597	3.600	9.933	5.500

	Panel (b): 1990 – 2024				
	Mean	Std. dev.	Min	Max	Median
Real GDP growth	2.464	4.490	-32.957	30.156	2.691
Inflation	2.232	1.393	-1.336	8.930	2.062
Unemployment rate	5.715	1.742	3.533	13.000	5.383

Notes: Summary statistics for quarterly US real GDP growth, the GDP Price Index (inflation), and the Civilian Unemployment Rate. Panel (a) covers 1990–2019, while Panel (b) extends to 2024, capturing the COVID-19 pandemic and subsequent economic volatility. Statistics are based on the first vintage available in 2025:Q2. Real GDP growth and inflation are annualized quarter-over-quarter rates, and the unemployment rate is a quarterly average of monthly values.

Beyond historical vintages, the SPF offers forward-looking projections for these macroeconomic indicators, which we use to construct real-time forecasts and quantify forecast disagreement. The SPF projections are not subject to revision, so constructing the real-time information flow requires only knowledge of the survey release dates. The Federal Reserve Bank of Philadelphia provides these dates starting in 1990. We define the real-time

SPF h -step-ahead forecast for indicator $i \in \{y, \pi, u\}$, available in week w of quarter t , as:

$$SPF_{t+h,w}^i = \begin{cases} SPF_{t+h|t-1}^i & \text{if } w \leq w^* \\ SPF_{t+h|t}^i & \text{if } w > w^*, \end{cases} \quad (5)$$

where $SPF_{t+h|t}^i$ denotes the median h -step-ahead forecast from the survey released in quarter t , and w^* is the week of the survey's publication. The SPF is typically released in week 6 or 7 of each quarter. Hence, at the beginning of quarter t , the most recent available forecast is the h -step-ahead projection from the previous quarter's survey. For example, to construct a one-step-ahead forecast early in quarter t , only the two-step-ahead projection from quarter $t-1$, $SPF_{t+1|t-1}^i$, is available.

While the consensus (median) forecast captures the central tendency of the panel, the SPF also allows measurement of forecast disagreement across respondents. To quantify this disagreement, we use the interquartile range (*IQR*) of the individual projections:

$$SPF_{t+h,w}^{IQR,i} = \begin{cases} SPF_{t+h|t-1}^{0.75,i} - SPF_{t+h|t-1}^{0.25,i} & \text{if } w \leq w^* \\ SPF_{t+h|t}^{0.75,i} - SPF_{t+h|t}^{0.25,i} & \text{if } w > w^*, \end{cases} \quad (6)$$

where $SPF_{t+h|t}^{q,i}$ denotes the q -th quantile of the cross-sectional distribution of h -step-ahead forecasts for indicator i . Although the *IQR* is not a suitable proxy for forecast uncertainty (Glas, 2020), it offers a readily available measure of forecast dispersion at fixed horizons. By contrast, the SPF also collects histogram forecasts, but these refer to fixed-event targets for calendar years rather than rolling horizons, and are therefore not directly comparable.

Finally, Table 1 also reports the release schedule of the Chicago Fed's National Financial Conditions Index (NFCI). The NFCI is a weekly index of U.S. financial conditions constructed from a broad set of indicators covering money markets, debt and equity markets, and the banking system (Brave and Butters, 2012). It is standardized to have mean zero and unit variance, with positive values indicating tighter-than-average financial conditions. While the NFCI series extends back to 1973, the first official real-time vintage

archived on ALFRED begins in 2011. To avoid look-ahead bias before that date, we use the unofficial weekly real-time vintages constructed by [Amburgey and McCracken \(2023\)](#), which are available from 1988 onward.

The NFCI is released mid-week and reflects information up to the preceding Friday, implying an effective publication lag of approximately one week. Given weekly observations, we define the real-time quarterly NFCI in week w of quarter t as the average of the first w weekly values in each quarter s :

$$NFCI_{s|t,w} = \frac{1}{w} \sum_{n=1}^w NFCI_{s,w-n|t,w}, \quad (7)$$

where $NFCI_{s,w-n|t,w}$ is the weekly NFCI for week $w - n$ of quarter s , taken from the latest vintage available in week w of quarter t . For example, in the second week of a quarter, we have $NFCI_{s|t,2} = \frac{1}{2}(NFCI_{s,1|t,2} + NFCI_{s,0|t,2})$, where $NFCI_{s,0|t,2}$ corresponds to week 12 of quarter $s - 1$.

4 Methodology

4.1 Forecasting models

We model the joint distribution of real GDP (y), inflation (π), and unemployment (u) using a standard AR(1)–GARCH(1,1) framework augmented with additional predictors. For each indicator $i \in \{y, \pi, u\}$ and omitting the index i for notational convenience, we model the h -step-ahead conditional mean and variance according to

$$\mu_{t+h|t,w} = c_w^h + \phi_w^h \mathbf{x}_{t,w}^\mu \quad (8)$$

$$\sigma_{t+h|t,w}^2 = \alpha_w^h \varepsilon_{t-j,w}^2 + \beta_w^h \sigma_{t+h-j|t,w}^2 + \exp(\omega_w^h + \gamma_w^h x_{t,w}^\sigma), \quad (9)$$

where $\varepsilon_{s,w} = z_{s,w} - \mu_{s|t,w}$ denotes the residual, with $z_{s,w}$ referring to $y_{s,w}$, $\pi_{s,w}$, or $u_{s,w}$ depending on the variable i being modeled. Predictors enter the conditional mean and

variance through $\mathbf{x}_{t,w}^\mu$ and $x_{t,w}^\sigma$. The parameters satisfy $\alpha_w^h > 0$, $\beta_w^h > 0$, and $\alpha_w^h + \beta_w^h < 1$. The index w indicates that models are re-estimated weekly using real-time vintage data for GDP, inflation, and unemployment. For example, $\mu_{t+1|t,1}$ denotes the conditional mean of the indicator for quarter $t + 1$ based on information available in week 1 of quarter t .

The choice of lag j for $y_{t,w}$ aligns with direct forecasting and depends on the indicator's release schedule: if the previous quarter's macroeconomic indicator is not yet available, we set $j = 2$, otherwise $j = 1$. The predictor set $\mathbf{x}_{t,w}^\mu$ for the conditional mean includes lagged values of the indicator and may incorporate real-time explanatory variables such as the SPF consensus forecasts $SPF_{t+h,w}^i$ and the NFCI. For the conditional variance, additional predictors enter through the exponential function $\exp(\cdot)$, where $x_{t,w}^\sigma$ is standardized to have mean zero and unit variance. In our application, potential predictors in $x_{t,w}^\sigma$ include the SPF interquartile range $SPF_{t+h,w}^{IQR,i}$ and the NFCI.

To model the joint density of the three macroeconomic indicators, we assume a multivariate normal distribution:

$$\mathbf{Y}_{t+h|t,w} \sim \mathcal{N}(\boldsymbol{\mu}_{t+h|t,w}, \boldsymbol{\Sigma}_{t+h|t,w}), \quad (10)$$

where $\mathbf{Y}_{t+h|t,w} = (y_{t+h|t,w}, \pi_{t+h|t,w}, u_{t+h|t,w})'$, and $\boldsymbol{\mu}_{t+h|t,w}$ collects the individual means of each variable according to equation (8).

The covariance matrix follows the dynamic conditional correlation (DCC)-GARCH framework of [Tse and Tsui \(2002\)](#) and [Engle \(2002\)](#), which allows for time-varying conditional correlations across variables. Specifically,

$$\boldsymbol{\Sigma}_{t+h|t,w} = \mathbf{D}_{t+h|t,w}, \mathbf{R}_{t+h|t,w}, \mathbf{D}_{t+h|t,w}, \quad (11)$$

where $\mathbf{D}_{t+h|t,w}$ is a diagonal matrix with conditional standard deviations $\sigma_{t+h|t,w}^i$ on the diagonal, and $\mathbf{R}_{t+h|t,w}$ denotes the time-varying correlation matrix. In the DCC model,

the correlation matrix evolves according to

$$\mathbf{Q}_{t+h|t,w} = (1 - a_w^h - b_w^h)\overline{\mathbf{Q}} + a_w^h\boldsymbol{\eta}_{t-j,w}\boldsymbol{\eta}'_{t-j,w} + b_w^h\mathbf{Q}_{t+h-j|t,w} \quad (12)$$

$$\mathbf{R}_{t+h|t,w} = \text{diag}(\mathbf{Q}_{t+h|t,w})^{-1/2} \mathbf{Q}_{t+h|t,w} \text{diag}(\mathbf{Q}_{t+h|t,w})^{-1/2}, \quad (13)$$

where $\boldsymbol{\eta}_{s,w}$ is a vector of standardized residuals for each indicator, $\varepsilon_{s,w}/\sigma_{s|t,w}$, and $\mathbf{R}_{t+h|t,w}$ is obtained by normalizing $\mathbf{Q}_{t+h|t,w}$. As a restricted version, when $a_w^h = b_w^h = 0$, the constant conditional correlation (CCC)-GARCH model assumes $\mathbf{R}_{t+h|t,w} = \mathbf{R}_{h,w}$, implying that correlations remain constant over time.

We estimate the model separately for each variable. The conditional mean equation (8) is estimated by ordinary least squares, and the variance equation (9) is fitted by quasi-maximum likelihood using the residuals $\varepsilon_{t+h,w}$. In the benchmark specification, we assume zero correlations across variables by setting $\mathbf{R}_{h,w} = \mathbf{I}$, which implies a diagonal covariance matrix $\boldsymbol{\Sigma}_{t+h|t,w}$. When allowing for constant conditional correlations, $\mathbf{R}_{h,w}$ is computed as the sample correlation matrix of the standardized residuals $\boldsymbol{\eta}_{s,w}$. When allowing for dynamic correlations, the parameters a_w^h and b_w^h are estimated according to equations (12) and (13).

4.2 Predictive densities and scenarios analysis

Given the joint predictive density in equation (10), we can extract risk measures such as Growth-at-Risk from the marginal distributions. Under the assumption of a multivariate normal (MVN) distribution, the marginals are also normally distributed. The h -step-ahead τ -quantile of variable i in $\mathbf{Y}_{t+h|t,w}$, available in week w , is simply

$$Q_{t+h|t,w}^i(\tau) = \mu_{t+h|t,w,i} + \sqrt{\sigma_{t+h|t,w,i}^2} F^{-1}(\tau), \quad (14)$$

where $F^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard normal, $\mathcal{N}(0, 1)$.

For scenario analysis, we are primarily concerned with conditional distributions. For instance, a bad scenario for GDP growth may be defined as a situation in which both inflation and unemployment are at elevated levels, specifically their 90% quantiles. This setup is closely related to the conditional tail-risk concepts used in the CoVaR literature (e.g., [Adrian and Brunnermeier, 2016](#)), where conditional risk measures are constructed by conditioning one variable on tail events of another. In our setting, the corresponding bad-scenario 10% quantile of GDP growth is then given by

$$Q_{t+h|t,w}^y(0.1) \mid (\pi = Q_{t+h|t,w}^\pi(0.9), u = Q_{t+h|t,w}^u(0.9)). \quad (15)$$

Under the multivariate normal assumption, this conditional distribution is analytically available via the standard formula for the conditional distribution of a multivariate normal. Letting $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, partition the mean and covariance matrix as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (16)$$

the conditional distribution of \mathbf{y}_1 given $\mathbf{y}_2 = \mathbf{c}$ is normal with conditional mean and variance

$$\mathbb{E}[\mathbf{y}_1 \mid \mathbf{y}_2 = \mathbf{c}] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{c} - \boldsymbol{\mu}_2), \quad (17)$$

$$\text{Var}[\mathbf{y}_1 \mid \mathbf{y}_2 = \mathbf{c}] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \quad (18)$$

The conditioning vector \mathbf{c} encodes the scenario. For example, if GDP growth, inflation, and unemployment are ordered in that sequence in \mathbf{Y} , then the bad scenario distribution of GDP defined above corresponds to $\mathbf{c} = (Q_{t+h|t,w}^\pi(0.9), Q_{t+h|t,w}^u(0.9))'$.

Alternatively, scenarios can be defined in a more extreme sense by conditioning on variables being more extreme than a quantile level. This follows recent developments in the CoVaR literature, which argue that conditioning on exceedances, rather than on a variable

being exactly equal to a specific threshold, provides a more natural way to represent severe stress events (Girardi and Ergün, 2013). Accordingly, bad-scenario GaR conditional on both inflation and unemployment being above their 90% quantiles is defined as

$$Q_{t+h|t,w}^y(0.1) \mid (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9)), \quad (19)$$

which requires simulation. Good-scenario GaR can be defined analogously by reversing the inequalities and considering the $(1 - \tau)$ -quantile, and one can similarly construct scenarios for inflation and unemployment.

5 Forecast evaluation

This section describes the construction of the ‘true’ realizations against which we evaluate the density forecasts, as well as the evaluation metric. An important question in forecast evaluation is which values to use as the actual outcomes. Later data vintages offer more precise estimates of the economic stance due to reduced measurement error. However, using the latest available vintage may incorporate definition changes that were unknown to the model or the SPF panelists at the time of forecasting.

To balance this trade-off, we follow Carriero et al. (2022) and define actual outcomes for quarter $t - 1$ as the values available at the end of quarter t . For real GDP and the GDP Price Index, this typically includes two revisions, up to the final estimate, i.e., growth rates are computed based on the vintage containing $\text{GDP}_{t-1}^{(3)}$ in Figure 1. In rare cases where the release of $\text{GDP}_{t-1}^{(3)}$ falls into quarter $t + 1$, the vintage includes only the first major revision, $\text{GDP}_{t-1}^{(2)}$.

Specifically, we define the actual outcomes for GDP growth and inflation under the

assumption that each quarter consists of 12 weeks as

$$y_t = 400 \cdot (\log(\text{GDP}_{t|t+1,12}) - \log(\text{GDP}_{t-1|t+1,12})), \quad (20)$$

$$\pi_t = 400 \cdot (\log(\text{PGDP}_{t|t+1,12}) - \log(\text{PGDP}_{t-1|t+1,12})), \quad (21)$$

where $\text{GDP}_{t|t+1,12}$ and $\text{PGDP}_{t|t+1,12}$ denote real GDP and the GDP Price Index for quarter t , available in the most recent vintage at the end of quarter $t + 1$.

For unemployment in quarter $t - 1$, we construct actual outcomes using the quarterly vintages from the Philadelphia Fed, which capture information as available around the middle of quarter t , approximately in week seven of Figure 1. Thus, the actual value is defined as

$$u_t = \frac{1}{3} \sum_{m=1}^3 \text{CUR}_{t,m|t+1,7}, \quad (22)$$

where $m = 1, 2, 3$ denotes the months within the quarter. Although minor revisions may occur, they are rare and of negligible magnitude. As such, the precise definition of unemployment outcomes is unlikely to impact the out-of-sample evaluation.

To evaluate the forecasting models, we rely on the log-score loss function, which is a strictly proper scoring rule (Gneiting and Raftery, 2007). The negative log-score at the realized outcomes $\mathbf{Y}_{t+h} = (y_{t+h}, \pi_{t+h}, u_{t+1})'$ is defined as

$$\text{LogS}(p_{t,w}^k, \mathbf{Y}_{t+h}) = -\log p_{t,w}^k(\mathbf{Y}_{t+h}), \quad (23)$$

where $p_{t,w}^k$ denotes the h -step-ahead real-time predictive density of model k at quarter t and week w . Since we report log-scores in their negative form, lower values indicate superior forecasting performance. While our primary focus is on evaluating full predictive densities using log-scores, directly assessing the conditional quantiles that define our scenarios is not feasible, as such conditional measures are not elicitable (Fissler and Hoga, 2024).

To test for differences in model performance, we use the Diebold and Mariano (1995)

test. The h -step-ahead loss differentials between the log-likelihood contributions of two models k and l at week w during the evaluation period are defined as

$$d_{t+h,w}^{k,l} = \log p_{t,w}^l(\mathbf{Y}_{t+h}) - \log p_{t,w}^k(\mathbf{Y}_{t+h}), \quad (24)$$

where positive values indicate superior forecasting performance of model l relative to model k . To formally evaluate the relative performance of the two models, the test statistic is

$$S_{t+h,w}^{k,l} = \frac{\bar{d}_{t+h,w}^{k,l}}{se(\bar{d}_{t+h,w}^{k,l})}, \quad (25)$$

where $\bar{d}_{t+h,w}^{k,l}$ is the average loss differential, and $se(\cdot)$ denotes its standard error. $S_{t+h,w}^{k,l}$ is asymptotically normally distributed, allowing for the application of standard critical values in pairwise comparisons.

However, because we are comparing $k = 1, \dots, K$ models, this setup presents a multiple comparison problem. To address this, we employ the Model Confidence Set (MCS) procedure proposed by Hansen et al. (2011). Unlike standard pairwise comparisons, which can inflate the probability of Type I errors, the MCS procedure allows for the joint evaluation of all models, sequentially eliminating the worst-performing models until the null hypothesis of equal predictive performance can no longer be rejected. The resulting MCS comprises the *surviving* models, i.e., models not significantly outperformed by any competitor.

Formally, let the set of models be indexed by $k = 1, \dots, K$, and denote this set as $\mathcal{M} \subset \mathcal{M}_0$. At each stage of the MCS procedure, the worst-performing model is identified through pairwise comparisons of all model combinations k and l for $k, l \in \mathcal{M}$. The test statistic is defined as

$$T_{\mathcal{M}}^w = \max_{k,l \in \mathcal{M}} |S_{t+h,w}^{k,l}|, \quad (26)$$

where $\mathcal{M} \subset \mathcal{M}_0$ denotes the set of remaining (non-eliminated) models. The elimination

process continues until $T_{\mathcal{M}}^w$ falls below the critical value corresponding to the chosen significance level α , resulting in the $1 - \alpha\%$ model confidence set \mathcal{M}^* .

Since the distribution of $T_{\mathcal{M}}^w$ is non-standard, its critical values must be determined through simulation. For all forecast horizons, we chose a block bootstrap procedure with 50,000 replications and a block length of three weeks to approximate these critical values.²

The estimation sample is determined by the availability of the SPF, which began in 1968. For models incorporating the NFCI, the sample starts in 1973. To ensure reasonably precise parameter estimates and allow for a sufficiently long out-of-sample period for the MCS to have statistical power, the evaluation period for the main analysis spans from 1990:Q1 to 2019:Q4, covering 120 quarters and excluding the COVID-19 pandemic in the main analysis. The pandemic period is examined separately in Section 6.3.

Due to the limited sample size, we estimate the models recursively using expanding windows. However, the loss differentials $d_{t+h,w}^{k,l}$ must be stationary with strictly positive variance. When models are nested, as is the case in this paper, rolling-window estimation ensures that the asymptotic distribution of $T_{\mathcal{M}}^w$ is well-behaved (Giacomini and White, 2006). In the baseline analysis, we use expanding windows to obtain more precise parameter estimates, and we assess robustness by re-estimating the models with rolling windows.

In addition, to assess the in-sample fit of various specifications, we report pairwise likelihood-ratio tests, which formally compare the relative performance of nested models. These tests evaluate whether a more complex model provides a statistically significant improvement in fit over a simpler alternative.

6 Empirical results

This section presents the empirical results. We begin by evaluating the model’s out-of-sample forecast performance over the pre-pandemic period from 1990:Q1 to 2019:Q4, which implies an initial estimation window of 85 observations and 120 out-of-sample

²We use the MFE Toolbox in MATLAB for the MCS procedure, see Sheppard (2009): <https://bashtage.github.io/kevinsheppard.com/code/matlab/mfe-toolbox/>.

forecasts. We report out-of-sample density forecast performance and complement these results with calibration evidence and in-sample diagnostics. We then extend the analysis to the post-2019 period and discuss the implications of the COVID-19 pandemic. Finally, we present a set of robustness checks.

6.1 Out-of-sample forecast evaluation

We begin by assessing the predictive content of the SPF relative to the NFCI for the joint predictive density within an AR-GARCH-type framework. For the benchmark, we assume a multivariate normal predictive density and restrict correlations across macroeconomic indicators to zero, such that $\Sigma_{t+h|t,w}$ is diagonal in equation (10). Table 3 presents the main results for the evaluation period 1990:Q1 to 2019:Q4. It reports negative predictive log-scores for one-step-ahead forecasts and nowcasts, evaluated at the end of each respective week. Within columns, grey shading indicates inclusion in the 90% Model Confidence Set for equal predictive accuracy. The first three rows report results for the benchmark AR(1)–GARCH(1,1) specification, optionally augmented with the NFCI in the conditional mean ($x_{t,w,i}^\mu$) or variance ($x_{t,w,i}^\sigma$) of indicator i . The fourth row refers to the SPF-GARCH model, which includes the SPF consensus forecast, $SPF_{t+h,w}^i$, in the conditional mean. Subsequent rows build on the SPF-GARCH model, adding one predictor at a time: the NFCI in the conditional mean or variance, or the interquartile range (IQR) of the SPF in the variance equation.

Most notably, incorporating the SPF consensus forecast into the conditional mean significantly improves the joint predictive density. All models included in the MCS contain the SPF, demonstrating that the consensus forecasts refine the entire multivariate predictive density. Comparing log-scores throughout the quarter reveals large improvements in the SPF-GARCH model following the release of the survey forecasts around the sixth week. Between weeks $w = 4$ and $w = 8$, forecasting performance considerably improves for both the one-step-ahead and nowcast distributions.

The baseline AR-GARCH model also reveals clear gains in forecast accuracy when

Table 3: Out-of-sample log-scores for alternative model specifications

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
AR-GARCH:	4.47	4.30	4.12	3.61	3.05	2.22
$+x_{t,w,i}^\mu = NFCI_{t,w}$	4.22	4.09	3.93	3.43	3.00	2.15
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.41	4.30	4.21	3.72	3.23	2.23
SPF-GARCH:	4.17	3.74	3.62	3.38	2.55	1.84
$+x_{t,w,i}^\mu = NFCI_{t,w}$	4.07	3.91	3.67	3.29	2.47	1.74
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.08	3.71	3.57	3.46	2.61	1.75
$+x_{t,w,i}^\sigma = SPF_{t,w}^{IQR,i}$	4.17	3.76	3.66	3.44	2.61	1.83

Notes: This table reports average log-scores of the multivariate normal distribution for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

incorporating releases and revisions of each macroeconomic indicator during the quarter. Predictive log-scores for both one-step-ahead forecasts and nowcasts improve as the quarter progresses. Including the NFCI in the conditional mean of the AR-GARCH model further enhances predictive performance, whereas adding it to the conditional variance does not yield improvements. Including the NFCI as a mean predictor within the SPF-GARCH model offers modest additional gains in the nowcast distribution, though these are not statistically significant. Similarly, augmenting the model with either the NFCI or the SPF interquartile range in the variance does not improve predictive accuracy (Brownlees and Souza, 2021; Schick, 2024).

To shed light on how the SPF refines the multivariate predictive densities, Table 4 compares the out-of-sample log-likelihoods of the AR-GARCH and SPF-GARCH models. Under the assumptions of normality and zero correlations, the overall out-of-sample log-scores are simply the sum of the individual log-scores across macroeconomic indicators. Multiplying these scores by the number of observations in the evaluation period yields the corresponding out-of-sample log-likelihoods. The MCS for the overall results is based on the same loss functions as in Table 3, and thus replicates the findings reported above.

Table 4: Out-of-sample log-likelihood

	One-step-ahead			Nowcast		
	4	8	12	4	8	12
Overall						
AR-GARCH	536.26	515.88	494.02	432.92	365.95	266.17
SPF-GARCH	499.96	448.67	434.93	405.74	306.19	221.35
real GDP						
AR-GARCH	264.11	266.42	267.37	252.49	250.66	251.63
SPF-GARCH	261.84	256.61	248.97	245.70	233.11	232.10
Inflation						
AR-GARCH	190.96	194.48	193.08	178.00	180.61	180.02
SPF-GARCH	176.65	174.31	172.65	163.96	154.09	155.49
Unemployment						
AR-GARCH	81.18	54.98	33.57	2.43	-65.33	-165.47
SPF-GARCH	61.47	17.75	13.31	-3.91	-81.01	-166.24

Notes: This table reports negative log-likelihoods for individual macroeconomic indicators over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model for GDP growth, inflation, and unemployment, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations and assume a multivariate normal distribution. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Turning to individual series, the negative log-likelihoods for inflation are significantly reduced across all forecast origins. For unemployment, the one-step-ahead predictive density also improves significantly with the inclusion of the SPF. For the nowcast distribution, gains are significant except at the very shortest horizon, where the difference becomes insignificant in week 12. However, at such short horizons, unemployment becomes highly predictable due to timely monthly releases, resulting in very small forecast errors. In the final month of the quarter, forecasts are so accurate that we observe a sign switch in the out-of-sample log-likelihoods, even without conditioning on the SPF in the mean.

For GDP growth, the picture is more nuanced. The nowcast distribution from the SPF-GARCH model is significantly better in weeks 8 and 12, which coincide with the mid-quarter release of the SPF survey and the availability of its nowcast. In week 4, however, the AR-GARCH model benefits from the release of the advance GDP estimate, which

was not yet available to SPF respondents when they submitted forecasts in the previous quarter. As a result, conditional on the SPF one-step-ahead forecast, the improvement is sizable but not statistically significant.

For the one-step-ahead predictive density of GDP growth, the SPF-GARCH again performs significantly better from mid-quarter onward. At the start of the quarter, however, incorporating the SPF two-step-ahead projection does not outperform the univariate AR-GARCH model and improves the out-of-sample fit by less than three log-likelihood points. This finding aligns with [Schick \(2024\)](#), who shows that SPF forecasts improve density and quantile predictions especially at short horizons. At the same time, our results highlight limits to the predictability of GDP growth beyond a two-quarter horizon.

While the log-score evaluates overall predictive performance, the probability integral transform (PIT) assesses the calibration of the entire predictive density. Figure B.1 shows PIT values for the SPF-GARCH model for each indicator before ($w = 4$) and after ($w = 8$) the release of the SPF, along with 95% confidence bands constructed following [Rossi and Sekhposyan \(2019\)](#). For GDP growth and inflation, the predictive densities appear well calibrated across all forecast horizons. In the case of unemployment, the upper tail is well captured, but the center of the distribution is significantly underpredicted, as reflected by PIT values lying above the 45-degree line. Nonetheless, the extremely small log-scores suggest that the model effectively distinguishes between periods of high and low unemployment.

Next, we assess predictive performance under more flexible covariance structures. For comparison, the first two rows in Table 5 replicate the benchmark AR-GARCH and SPF-GARCH models, which assume a diagonal covariance matrix. Estimating a constant conditional correlation matrix (SPF-GARCH-CCC) does not improve predictive performance; however, it also does not entail a cost in terms of forecast accuracy. While this suggests limited benefits of the CCC specification from a purely forecasting perspective, it demonstrates that allowing for correlations does not worsen predictive performance, a desirable property when using these models for scenario analysis, where the assumption of

a diagonal covariance matrix is clearly restrictive.

Turning to the dynamic conditional correlations (SPF-GARCH-DCC), estimated time-varying correlations yield virtually identical predictive accuracy to the CCC approach, indicating that constant correlations are sufficient in this context. Overall, all SPF-based specifications are included in the MCS and exhibit very similar log-scores.

Table 5: Out-of-sample log-scores with and without correlation

Log-score	One-step-ahead			Nowcast		
	4	8	12	4	8	12
AR-GARCH	4.47	4.30	4.12	3.61	3.05	2.22
SPF-GARCH	4.17	3.74	3.62	3.38	2.55	1.84
SPF-GARCH-CCC	4.09	3.66	3.58	3.38	2.56	1.85
SPF-GARCH-DCC	4.08	3.66	3.59	3.39	2.56	1.85

Notes: This table reports average log-scores of the multivariate normal distribution for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). The models are AR-GARCH (baseline AR(1)–GARCH(1,1)), SPF-GARCH (adding SPF projections in the conditional mean), SPF-GARCH-CCC (assuming a constant conditional correlation matrix), and SPF-GARCH-DCC (allowing for a dynamic conditional correlation matrix). All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

At the same time, correlations matter from an in-sample perspective. Table A.1 reports in-sample log-likelihoods of the SPF-GARCH model under different covariance specifications. Stars indicate pairwise likelihood-ratio (LR) tests against the normal model with constant conditional correlations (second row). Under normality, estimating a CCC significantly improves model fit. However, allowing for DCC correlations results in similar and statistically indistinguishable in-sample log-likelihoods compared with the CCC model.

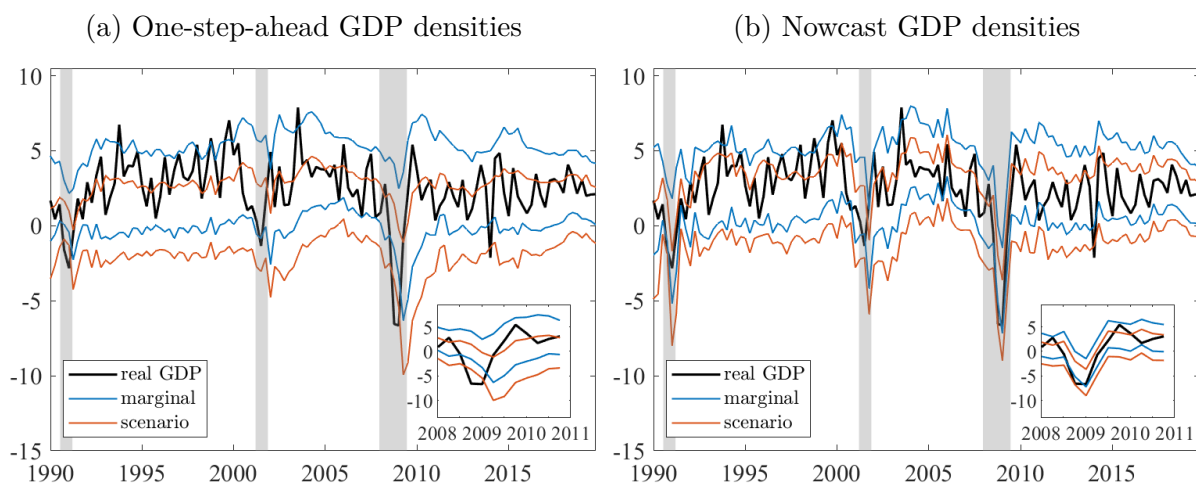
In summary, constructing scenarios based on the CCC specification appears equally justified from a forecasting perspective and sufficient in-sample, given that the data do not favor time-varying correlations. Estimating correlations adds flexibility within the sample while maintaining the same predictive performance out-of-sample.

6.2 Scenario-based risk assessment

The previous section demonstrates the strong forecasting performance of the SPF-GARCH model for relatively short forecast horizons. We now illustrate how scenarios can be constructed from the multivariate predictive densities that incorporate the SPF and a constant conditional correlation matrix.

Turning to scenario analysis, we define a bad (downside) scenario for GDP as the conditional distribution of real GDP growth given that both inflation and unemployment are unexpectedly high, i.e., above their respective 90th percent quantiles. This stagflationary setting captures the impact of simultaneous adverse surprises in prices and the labor market. Figure 2 plots the marginal distribution in blue and the scenario distribution in red, both shown by the 10%-90% quantile range estimated on an expanding window. Panels (a) and (b) use the SPF consensus one-step-ahead forecasts and nowcasts, respectively.

Figure 2: Downside GDP scenarios based on the SPF-GARCH



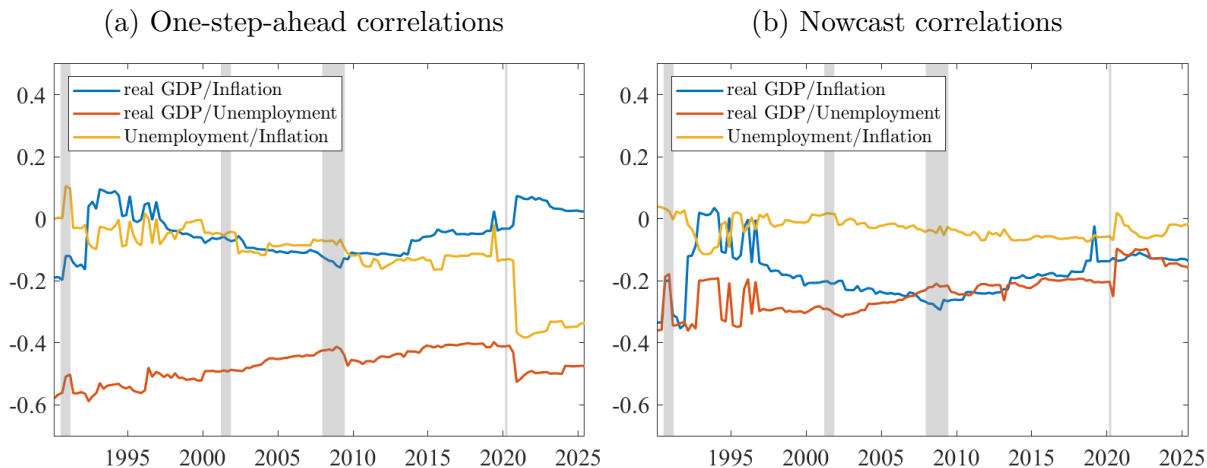
Notes: This figure plots the 10th and 90th percentiles of the marginal predictive density for real GDP growth (blue lines) and the corresponding conditional percentiles when both inflation and unemployment are in their upper deciles (red lines), i.e., $Q_{t+h|t,w}^y(\tau) \mid (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9))$ for $\tau \in \{0.1, 0.9\}$. The black line depicts realized quarterly annualized GDP growth. Panel (a) shows the one-step-ahead distribution ($h = 1$) and panel (b) the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. All forecasts are from the SPF-GARCH-CCC model assuming a multivariate normal distribution. The sample period is 1990:Q1–2019:Q4 and the model is estimated using an expanding window. Vertical gray bands indicate NBER recessions.

In both cases, the scenario distribution is clearly shifted downward. For the one-step-ahead density, the conditional 10% quantile is, on average, about 2 percentage points

lower than in the baseline distribution; for the nowcast, the shift still amounts to 1.5 percentage points. Thus, even when conditioning on SPF nowcasts, the scenario analysis points to further elevated downside risks to GDP relative to the baseline distribution. During the global financial crisis, the nowcasting scenario density in panel (b) provided an early indication of elevated downside risk. While the baseline density also reflects part of the deterioration, its 10% quantile stayed well above the realized outcome in early 2009, suggesting it placed less weight on extreme negative realizations (Schick, 2024).

Figure 3 helps explain these shifts. It shows the evolution of cross-correlations among the macroeconomic indicators, evaluated at week 8 within each quarter (after the SPF mid-quarter release). The strongest and most persistent relationship is between the standardized residuals of GDP growth and unemployment, around -0.5 for the one-step-ahead distribution and -0.3 for the nowcast. This negative correlation is the main driver of the downward shift in GDP under the stagflationary scenario.

Figure 3: Estimated correlations from the SPF-GARCH-CCC model

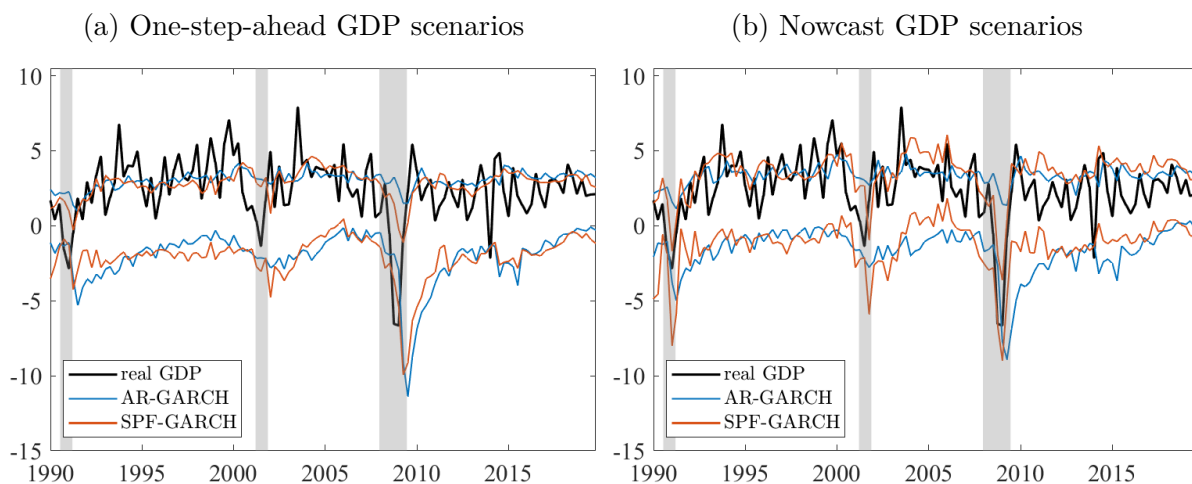


Notes: This figure plots the constant conditional correlation (CCC) coefficients for GDP growth, inflation, and unemployment estimated from the SPF-GARCH-CCC model estimated using an expanding window. Panel (a) shows results for the one-step-ahead distribution ($h = 1$) and panel (b) for the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. The sample period is 1990:Q1–2025:Q2. Vertical gray bands indicate NBER recessions.

Figure 4 shows the difference in scenario densities between the AR-GARCH and the SPF-GARCH models. Compared to the benchmark AR-GARCH specification, incorporating the current macroeconomic outlook via the SPF consensus forecasts clearly shifts

the conditional distribution, both for the one-step-ahead and the nowcast scenarios. For instance, following the global financial crisis, both models indicate substantial downside risk, but the SPF-based density adjusts more quickly and appears less pessimistic, reflecting a faster recovery in expectations. The largest differences appear for the nowcast horizon, where the SPF delivers the greatest nowcast gains and thus provides the most accurate information about the conditional mean of GDP growth, which in turn affects the quantiles of the entire predictive density.

Figure 4: Downside GDP scenarios under AR-GARCH and SPF-GARCH models



Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for real GDP growth when both inflation and unemployment are in their upper deciles, i.e., $Q_{t+h|t,w}^y(\tau) | (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (black line). Blue lines depict forecasts from the baseline AR-GARCH model; red lines depict forecasts from the SPF-GARCH model. Panel (a) shows the one-step-ahead distribution ($h = 1$) and panel (b) the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. The sample period is 1990:Q1–2019:Q4 and the models are estimated using an expanding window. Both models assume a multivariate normal distribution with constant correlations. Vertical gray bands indicate NBER recessions.

Next, Figure B.2 presents upside scenarios for inflation. As seen in Figure 3, although modest in magnitude, forecast errors of inflation are negatively correlated with those of both GDP growth and unemployment. Accordingly, we define the scenario as the conditional distribution of inflation given that GDP growth and unemployment fall below their respective 10% quantiles. The resulting conditional densities are shifted upward by, on average, 0.4 percentage points both for the one-step-ahead and nowcast distributions, respectively.

For unemployment, we construct the upside-risk scenario by conditioning on GDP growth and inflation being below their 10% quantiles. Since forecast errors for unemployment are typically small later in the quarter, we evaluate the distribution at week 4. Figure B.3 shows that unexpectedly weak growth and low inflation increase the upper tail of the unemployment distribution. On average, the conditional 90% quantile rises by 0.4 and 0.2 percentage points for the one-step-ahead and nowcast densities, respectively. In particular, the scenario densities indicate elevated upside risks to unemployment during and shortly after recessions.

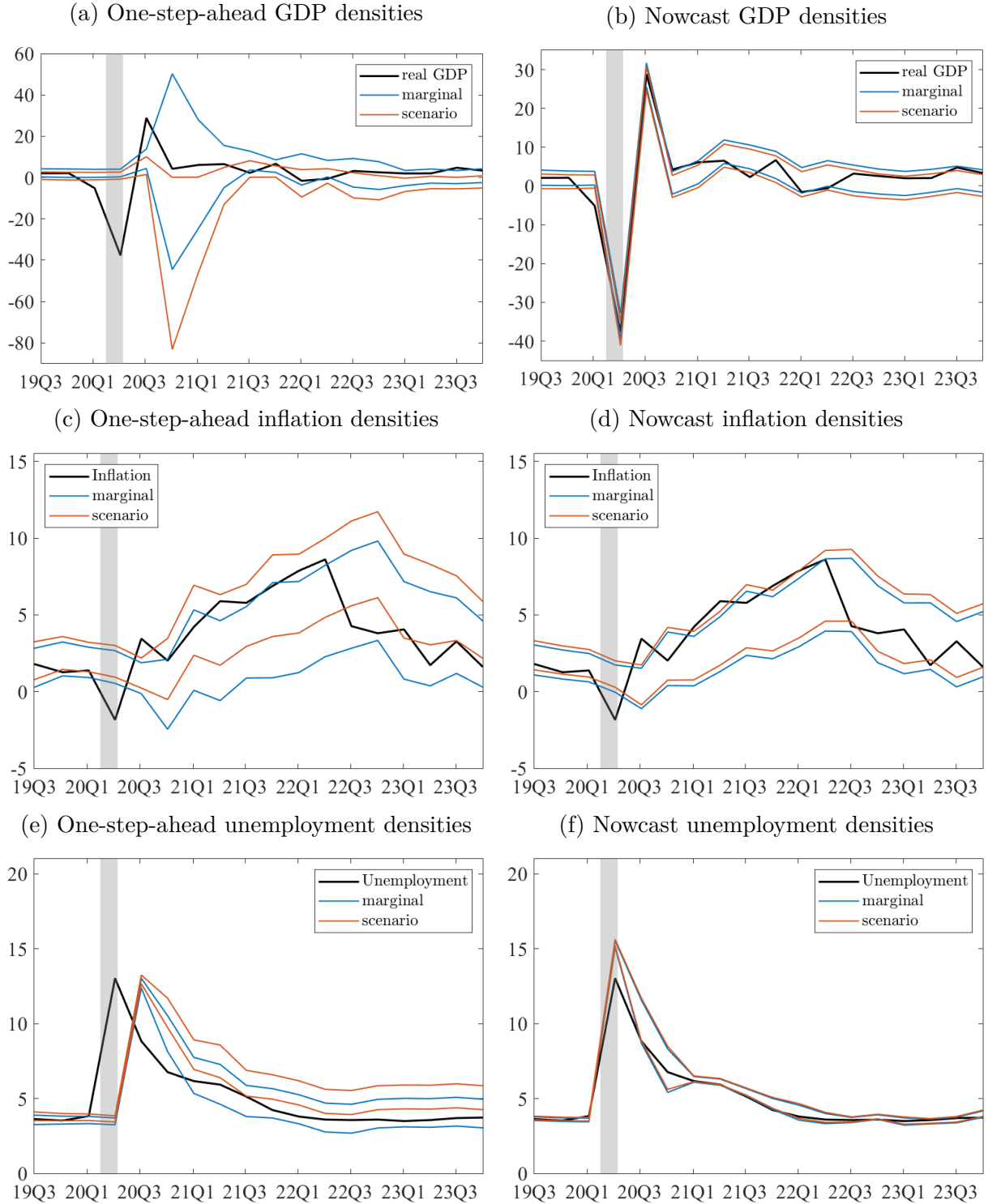
6.3 Extended-sample evaluation

We next extend the evaluation sample to 2025:Q1. Table A.2 reports the corresponding forecast evaluation results. Due to exceptionally large forecast errors during the COVID-19 pandemic, the period from 2020:Q1 to 2020:Q4 is excluded from the evaluation in this table. The results confirm that our findings are robust to the longer sample. The SPF-GARCH specification remains included in the MCS throughout. Adding the NFCI yields significant nowcast improvements only toward the end of the quarter.

Figure 5 illustrates the corresponding scenario densities during and after the COVID-19 pandemic. For one-step-ahead GDP growth in panel (a), the SPF-GARCH model generates large forecast errors due to the unforeseen pandemic. In contrast, panel (b) shows that the SPF quickly incorporated the severity of the economic downturn and the subsequent recovery. As a result, forecast errors are small, and both the baseline and scenario densities closely track GDP growth realizations in the following quarters.

Panels (c) and (d) show that the SPF-GARCH model already indicated elevated upside risks to inflation at the end of 2020. During the high-inflation period of 2021 and 2022, even the 90% quantile of the baseline one-step-ahead density closely tracked the realized inflation rates. The scenario density in panel (c) exceeds 5% inflation as early as the beginning of 2021. In panel (d), incorporating more accurate SPF projections, the nowcast density appears somewhat less pessimistic, yet remains remarkably effective in capturing

Figure 5: SPF-GARCH model-based scenarios during the COVID-19 pandemic



Notes: This figure replicates the scenarios shown in Figures 2, B.2, and B.3 for the sample period 2020:Q1–2023:Q4. Each panel plots the 10th and 90th percentiles of the conditional predictive density (red) and of the marginal distribution (blue) for one variable, alongside its realized value (black line). Panel (a) shows the one-step-ahead distribution ($h = 1$), and Panel (b) the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. All forecasts are from the SPF-GARCH-CCC model assuming a multivariate normal distribution. Vertical gray bands indicate NBER recessions.

upside risks. Both the marginal and scenario densities closely align with observed inflation outcomes.

Lastly, panels (e) and (f) of Figure 5 display the predictive densities for the unemployment rate. As seen in panel (e), the one-step-ahead density failed to anticipate the sharp increase in unemployment at the onset of the pandemic, and it was not until 2021:Q1 that the marginal distribution became centered around the realized unemployment rate. In contrast, the SPF-GARCH nowcast overpredicted the unemployment rate in 2020:Q2, resulting in 10% quantiles well above the realization. However, at this short horizon, the predictive density tracked the unemployment rate fairly closely again from the fourth quarter of 2020 onward.

6.4 Robustness analysis

6.4.1 Rolling-window estimation

As discussed in Section 5, nested models should be estimated on a rolling window to ensure the Diebold–Mariano test behaves well asymptotically. Table A.3 replicates the main results using a rolling window of 100 observations.³ Overall, the findings are robust: the SPF-GARCH model remains included in the MCS and continues to produce smaller losses than the AR-GARCH benchmark with or without the NFCI. The out-of-sample losses are broadly comparable to those reported in Table 3. However, reduced power due to higher estimation uncertainty leads to some additional AR-GARCH specifications being included in the MCS.

6.4.2 Fat-tailed innovations and copulas

The assumption of multivariate normality can be restrictive and might potentially drive the main results if the SPF only refines the predictive density through the conditional mean, while the data in fact require more flexibility such as fat tails. This would put the

³The initial estimation sample in 1990:Q1 contains 85 observations, however, a longer rolling window is used here to mitigate the low power of the MCS associated with very short estimation windows.

benchmark AR-GARCH model at an unfair disadvantage.

To demonstrate that the SPF-GARCH model outperforms the AR-GARCH benchmark not merely due to restrictive distributional assumptions, we employ a more flexible parametric specification. Specifically, we first fit the marginal densities assuming univariate Student- t distributions⁴ and then estimate a t -copula with a constant conditional correlation matrix (t -copula-CCC) to construct the joint predictive density. In total, we estimate four degrees-of-freedom (DoF) parameters, one for each marginal distribution and one for the copula, as well as three correlation coefficients. This setup allows for both tail heterogeneity and tail dependence without inflating the number of parameters. For completeness, Appendix C details the construction of the joint predictive density using Student- t marginals and a t -copula.

Table A.4 replicates the main results using the t -copula-CCC approach. Again, the MCS contains all specifications that incorporate the SPF in the conditional mean, and the out-of-sample scores are comparable to those in Table 3.

Restricting correlations to be constant, Table A.5 compares the out-of-sample log-scores of the SPF-GARCH model under normality and under the t -copula specification. The difference in forecast performance is small and, in most cases, not statistically significant. However, Table A.6 shows that the in-sample log-likelihoods are significantly higher when using the t -copula specification. Hence, the more flexible approach improves in-sample fit without imposing a cost in terms of forecast accuracy.

Interestingly, at the nowcast horizon, panel (b) of Figure B.4 shows that the downside GDP scenario density is nearly identical under both the normal and the t -copula specifications. This finding is supported by panel (a) of Figure B.5, which plots the estimated DoF parameters for the marginals and the copula. Estimating the model on an expanding window, the DoF values vary over time and are frequently large, often exceeding 100, implying that at the shortest horizon the SPF-GARCH model leaves little fatness in the tails of the joint distribution of the residuals.

⁴We also considered the Hansen Skew- t distribution to allow for skewness in the marginals. The results were qualitatively similar and are therefore not reported.

In contrast, at the one-step-ahead horizon, panel (a) of Figure B.5 indicates noticeable fat tails in the marginal distributions of GDP and unemployment. The copula’s DoF parameter typically hovers around 20, suggesting moderate deviations from Gaussian copula behavior, while inflation remains close to normality.

Comparing the one-step-ahead scenario densities in panel (a) of Figure B.4 reveals that, although the conditional 90% quantiles are nearly identical, the conditional 10% quantiles are shifted downward by about 0.7 percentage points on average. Although the effect is moderate, the downward shift suggests that accounting for fat tails and tail dependence can affect the assessment of downside risks in adverse states, which may be relevant for stress-testing and risk assessment exercises.

Overall, the t -copula-CCC approach offers additional flexibility in modeling tail dependence, while the results confirm that the main differences between scenario and marginal densities are primarily attributable to the SPF consensus forecasts.

6.4.3 Longer forecast horizons

The main results focus on the short-term predictive performance of the SPF-GARCH model for the nowcast and the one-step-ahead multivariate predictive density. We find that SPF consensus forecasts significantly improve the predictive densities for inflation and unemployment, but provide only limited refinements for the one-step-ahead marginal distribution of GDP growth when incorporating the two-step-ahead SPF projections at the beginning of the quarter. This section extends the analysis to longer horizons and evaluates the out-of-sample performance of the SPF-GARCH model for the two-quarters-ahead ($h = 2$) and three-quarters-ahead ($h = 3$) densities.

Assuming multivariate normality, Table A.7 indicates that SPF consensus projections continue to enhance the multivariate predictive density for all horizons up to $h = 3$. At the beginning of the quarter, $h = 3$ corresponds to a four-quarters-ahead forecast based on $SPF_{t|t-4}$. Augmenting the variance equation with the NFCI produces further reductions in out-of-sample log-scores; these reductions are, at times, similar in magnitude to the

improvements obtained from adding the SPF to the conditional mean and are occasionally statistically significant.

For the univariate distributions, Table A.8 shows that the SPF median forecasts for GDP growth do not improve its marginal predictive density relative to the AR-GARCH benchmark, whereas the marginal distributions of inflation and unemployment are significantly better. Thus, the improvements in the multivariate densities primarily reflect SPF information for inflation and unemployment. This confirms that predictability limits are reached for GDP growth at horizons of two or more quarters, consistent with [Stark \(2010\)](#), who document that SPF forecast performance deteriorates rapidly beyond one-quarter-ahead, especially for GDP. Regarding calibration of the SPF-GARCH model, Figure B.6 shows that GDP growth and inflation forecast distributions remain well calibrated at longer horizons, while for unemployment at least the tails are well modeled.

Table A.9 suggests that both the CCC and the t -copula approaches marginally, albeit non-significantly, refine the out-of-sample scores of the SPF-GARCH model, implying that incorporating a correlation matrix for scenario analysis does not compromise predictive accuracy. Although the out-of-sample gains are insignificant, Table A.10 suggests that the t -copula is preferred in-sample.

Turning to scenario densities for $h = 2$ and $h = 3$, Figure B.7 shows that conditional densities for all macro indicators remain substantially different from their marginals. Figure B.8 illustrates that the conditional GDP growth distributions of SPF-GARCH and AR-GARCH are fairly similar at these horizons. Nevertheless, even though SPF forecasts do not improve GDP growth predictive densities for $h = 2$ and $h = 3$, the framework still captures meaningful conditional tail risks. Notably, the lower conditional quantiles display pronounced time variation, while the upper quantile remains relatively stable. Finally, Figure B.9 shows that conditional downside risks estimated under Student- t marginals with a t -copula shift further downward relative to the multivariate normal specification.

7 Conclusion

This paper proposes to assess joint macroeconomic tail risks by modeling the multivariate predictive density of macroeconomic indicators. We incorporate SPF consensus forecasts to model the conditional mean of real GDP growth, inflation, and unemployment, and use an AR-GARCH framework to capture time-varying macroeconomic uncertainty and cross-variable correlations.

The resulting predictive densities can be evaluated out-of-sample, and we find that SPF consensus projections significantly improve density forecasts at short horizons. This highlights the value of combining survey-based expectations with econometric models to better capture short-term macroeconomic risks.

Scenarios are defined as conditional predictive densities, such as GDP growth conditional on high inflation and unemployment. We find strong dependence between GDP growth and unemployment, resulting in pronounced shifts of the scenario distributions relative to their marginals. Comparing SPF-GARCH with a purely statistical AR-GARCH shows that survey information substantially alters scenario densities, especially around the global financial crisis and subsequent recovery.

Taken together, this framework offers policymakers a statistically grounded, real-time tool to assess macroeconomic scenarios and quantify joint risks across multiple dimensions. The approach remains deliberately agnostic: although it does not deliver structural interpretation, it provides a transparent econometric basis for modeling joint macroeconomic risks and for incorporating judgment or external information.

Future extensions could broaden the set of variables included in the joint system, even in the absence of SPF projections, by integrating additional macro-financial indicators such as the term spread or credit conditions. Moreover, while our scenarios are closely related to CoVaR-type conditional functionals that are not elicitable and therefore cannot be directly evaluated, recent work shows that such conditional scenario quantiles can be assessed jointly with VaR in a lexicographical sense ([Fissler and Hoga, 2024](#)).

References

- Adams, P. A., T. Adrian, N. Boyarchenko, and D. Giannone (2021). Forecasting macroeconomic risks. *International Journal of Forecasting* 37(3), 1173–1191.
- Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review* 109(4), 1263–89.
- Adrian, T. and M. K. Brunnermeier (2016). CoVaR. *The American Economic Review* 106(7), 1705.
- Adrian, T., D. Giannone, M. Luciani, and M. West (2025). Scenario synthesis and macroeconomic risk. *arXiv preprint arXiv:2505.05193*.
- Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics* 14(3), 283–323.
- Amburgey, A. J. and M. W. McCracken (2023). On the real-time predictive content of financial condition indices for growth. *Journal of Applied Econometrics* 38(2), 137–163.
- Andreou, E., E. Ghysels, and A. Kourtellis (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics* 31(2), 240–251.
- Ang, A., G. Bekaert, and M. Wei (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics* 54(4), 1163–1212.
- Antolín-Díaz, J., I. Petrella, and J. F. Rubio-Ramírez (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics* 117, 798–815.
- Bañbura, 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.

- Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In *Handbook of Economic Forecasting*, Volume 2, pp. 195–237. Elsevier.
- Bańbura, M., F. Brenna, J. Paredes, and F. Ravazzolo (2021). Combining Bayesian VARs with survey density forecasts: Does it pay off?
- Brave, S. and R. A. Butters (2012). Diagnosing the financial system: Financial conditions and financial stress. *International Journal of Central Banking* 8(2), 191–239.
- Brownlees, C. and A. B. Souza (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics* 118, 312–330.
- Carriero, A., T. E. Clark, and M. Marcellino (2022). Nowcasting tail risk to economic activity at a weekly frequency. *Journal of Applied Econometrics* 37(5), 843–866.
- Carriero, A., T. E. Clark, and M. Marcellino (2024). Capturing macro-economic tail risks with bayesian vector autoregressions. *Journal of Money, Credit and Banking* 56(5), 1099–1127.
- Castelnuovo, E. and L. Mori (2025). Uncertainty, skewness, and the business cycle through the MIDAS lens. *Journal of Applied Econometrics* 40(1), 89–107.
- Chavleishvili, S., R. F. Engle, S. Fahr, M. Kremer, F. Lund-Thomsen, S. Manganelli, and B. Schwaab (2026). Macro-prudential policy under asymmetric risks: A bayesian structural quantile var approach. *Journal of Econometrics*. Forthcoming.
- Chavleishvili, S. and S. Manganelli (2024). Forecasting and stress testing with quantile vector autoregression. *Journal of Applied Econometrics* 39(1), 66–85.
- Clark, T. E., G. Ganics, and E. Mertens (2025). Constructing fan charts from the ragged edge of SPF forecasts. *The Review of Economics and Statistics*, 1–45.

- Clark, T. E., M. W. McCracken, and E. Mertens (2020). Modeling time-varying uncertainty of multiple-horizon forecast errors. *The Review of Economics and Statistics* 102(1), 17–33.
- Clements, M. P., R. W. Rich, and J. S. Tracy (2023). Surveys of professionals. In *Handbook of Economic Expectations*, pp. 71–106. Academic Press.
- Corradi, V. and J. Llorens-Terrazas (2026). Monitoring joint tail risks: An application to growth and inflation. *Journal of Econometrics* 255, 106237.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Dimitriadis, T. and Y. Hoga (2025). Dynamic CoVaR modeling and estimation. *Journal of Business & Economic Statistics* (just-accepted).
- Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics* 20(3), 339–350.
- Faust, J. and J. H. Wright (2013). Forecasting inflation. In *Handbook of Economic Forecasting*, Volume 2, pp. 2–56. Elsevier.
- Ferrara, L., M. Mogliani, and J.-G. Sahuc (2022). High-frequency monitoring of growth at risk. *International Journal of Forecasting* 38(2), 582–595.
- Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters* 191, 109126.
- Fissler, T. and Y. Hoga (2024). Backtesting systemic risk forecasts using multi-objective elicibility. *Journal of Business & Economic Statistics* 42(2), 485–498.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.

- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* 55(4), 665–676.
- Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics* 119(3), 457–471.
- Girardi, G. and A. T. Ergün (2013). Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *Journal of Banking & Finance* 37(8), 3169–3180.
- Glas, A. (2020). Five dimensions of the uncertainty–disagreement linkage. *International Journal of Forecasting* 36(2), 607–627.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), 359–378.
- González-Rivera, G., J. Maldonado, and E. Ruiz (2019). Growth in stress. *International Journal of Forecasting* 35(3), 948–966. Forecasting issues in developing economies.
- González-Rivera, G., C. V. Rodríguez-Caballero, and E. Ruiz (2024). Expecting the unexpected: Stressed scenarios for economic growth. *Journal of Applied Econometrics* 39(5), 926–942.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- 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.
- López-Salido, D. and F. Loria (2024). Inflation at risk. *Journal of Monetary Economics* 145, 103570.
- Prasad, M. A., S. Elekdag, M. P. Jeasakul, R. Lafarguette, M. A. Alter, A. X. Feng, and C. Wang (2019). *Growth at risk: Concept and application in IMF country surveillance*. International Monetary Fund.

- Reifschneider, D. and P. Tulip (2019). Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve’s approach. *International Journal of Forecasting* 35(4), 1564–1582.
- Rossi, B. and T. Sekhposyan (2019). Alternative tests for correct specification of conditional predictive densities. *Journal of Econometrics* 208(2), 638–657.
- Schick, M. (2024). Nowcasting growth-at-risk using the Survey of Professional Forecasters. *Available at SSRN 4859937*.
- Sheppard, K. (2009). MFE MATLAB function reference financial econometrics.
- Stark, T. (2010). Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters. *Research Rap Special Report* (May).
- Tse, Y. K. and A. K. C. Tsui (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics* 20(3), 351–362.
- Waggoner, D. F. and T. Zha (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics* 81(4), 639–651.

Appendices

A Tables

Table A.1: In-sample fit of SPF-GARCH

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
Normal	-844.45***	-799.27***	-795.10***	-615.00**	-540.20***	-410.09**
Normal-CCC	-827.93	-778.70	-775.80	-610.17	-533.47	-405.78
Normal-DCC	-826.28	-778.18	-775.07	-609.93	-532.95	-403.73

Notes: This table reports in-sample log-likelihoods comparing SPF-GARCH model specifications for GDP growth, inflation, and unemployment. Models are estimated on the sample 1968:Q4–2019:Q4. The baseline model (row 1) assumes a multivariate normal distribution with a diagonal covariance matrix. Row 2 shows the normal-CCC model with constant conditional correlations, and row 3 presents the normal-DCC model with dynamic conditional correlations. Stars indicate significance of likelihood-ratio (LR) tests when comparing the normal-CCC model (row 2) against the baseline diagonal covariance model (row 1) and the DCC model (row 3), imposing three and two restrictions, respectively. Significance levels are: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Out-of-sample log-scores excluding the COVID-19 pandemic

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
AR-GARCH:	4.65	4.49	4.31	3.77	3.13	2.33
+ $x_{t,w,i}^\mu = NFCI_{t,w}$	4.41	4.29	4.12	3.60	3.08	2.26
+ $x_{t,w,i}^\sigma = NFCI_{t,w}$	4.59	4.47	4.39	3.90	3.31	2.36
SPF-GARCH:	4.37	3.95	3.85	3.63	2.71	1.99
+ $x_{t,w,i}^\mu = NFCI_{t,w}$	4.28	4.10	3.90	3.54	2.64	1.89
+ $x_{t,w,i}^\sigma = NFCI_{t,w}$	4.31	3.98	3.83	3.71	2.80	1.94
+ $x_{t,w,i}^\sigma = SPF_{t,w}^{IQR,i}$	4.38	3.98	3.91	3.67	2.76	1.98

Notes: This table reports average log-scores of the multivariate normal distribution for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2025:Q1, excluding the 2020 calendar year (137 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.3: Out-of-sample log-scores based on rolling-window estimation

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
AR-GARCH:	4.60	4.41	4.23	3.60	3.00	2.16
$+x_{t,w,i}^\mu = NFCI_{t,w}$	4.42	4.25	4.08	3.57	3.06	2.17
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.56	4.48	4.38	3.78	3.25	2.27
SPF-GARCH:	4.33	3.79	3.69	3.45	2.54	1.91
$+x_{t,w,i}^\mu = NFCI_{t,w}$	4.39	3.99	3.95	3.44	2.65	1.83
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.45	4.15	3.94	3.60	2.68	1.85
$+x_{t,w,i}^\sigma = SPF_{t,w}^{IQR,i}$	4.33	3.78	3.72	3.46	2.65	1.90

Notes: This table reports average log-scores of the multivariate normal distribution for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using a rolling window of 100 observations, with the initial estimation sample consisting of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.4: Out-of-sample log-scores using the t -copula specification

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
AR-GARCH:	4.43	4.22	4.05	3.58	2.98	2.14
$+x_{t,w,i}^\mu = NFCI_{t,w}$	4.14	3.97	3.86	3.34	2.93	2.09
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.44	4.17	4.14	3.66	3.02	2.14
SPF-GARCH:	4.04	3.65	3.51	3.33	2.53	1.84
$+x_{t,w,i}^\mu = NFCI_{t,w}$	3.93	3.66	3.61	3.21	2.48	1.73
$+x_{t,w,i}^\sigma = NFCI_{t,w}$	4.01	3.63	3.52	3.31	2.59	1.74
$+x_{t,w,i}^\sigma = SPF_{t,w}^{IQR,i}$	4.05	3.69	3.61	3.44	2.58	1.86

Notes: This table reports average log-scores of multivariate predictive densities constructed from a t -copula with Student- t marginals and a constant conditional correlation matrix for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.5: Out-of-sample log-scores of SPF-GARCH model specifications using the t -copula

Log-score	One-step-ahead			Nowcast		
	4	8	12	4	8	12
Normal-CCC	4.09	3.66	3.58	3.38	2.56	1.85
t -copula-CCC	4.04	3.65	3.51	3.33	2.53	1.84

Notes: This table reports average log-scores of SPF-GARCH model specifications for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). The models are Normal-CCC (assuming a multivariate normal distribution with a constant conditional correlation matrix) and t -copula-CCC (constructed from univariate GARCH models with Student- t marginals combined via a t -copula with constant conditional correlations). All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.6: In-sample fit of SPF-GARCH using the t -copula

week	One-step-ahead			Nowcast		
	4	8	12	4	8	12
Normal-CCC	-827.93***	-778.70***	-775.80***	-610.17**	-533.47*	-405.78
t -copula-CCC	-821.10	-770.48	-766.75	-603.83	-529.45	-403.79

Notes: This table reports in-sample log-likelihoods comparing SPF-GARCH model specifications for GDP growth, inflation, and unemployment. Models are estimated on the sample 1968:Q4–2019:Q4. The baseline model (row 1) assumes a multivariate normal distribution with a constant conditional correlation (CCC) matrix. Row 2 shows the t -copula-CCC model constructed from univariate GARCH models with Student- t marginals combined via a t -copula with constant conditional correlations. Stars indicate significance of likelihood-ratio (LR) tests comparing the t -copula-CCC model (row 2) against the normal-CCC model (row 1), imposing four restrictions. Significance levels are: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Out-of-sample log-scores for longer forecast horizons

week	Three-step-ahead			Two-step-ahead		
	4	8	12	4	8	12
AR-GARCH:	5.88	5.83	5.82	5.26	4.99	4.87
+ $x_{t,w,i}^\mu = NFCI_{t,w}$	5.07	5.11	5.09	4.98	4.71	4.55
+ $x_{t,w,i}^\sigma = NFCI_{t,w}$	5.52	5.43	5.37	5.08	4.74	4.72
SPF-GARCH:	5.57	5.05	5.12	4.97	4.38	4.34
+ $x_{t,w,i}^\mu = NFCI_{t,w}$	5.04	4.96	4.95	4.83	4.34	4.31
+ $x_{t,w,i}^\sigma = NFCI_{t,w}$	5.14	4.65	4.63	4.66	4.12	4.08
+ $x_{t,w,i}^\sigma = SPF_{t,w}^{IQR,i}$	5.62	4.98	5.05	5.01	4.34	4.33

Notes: This table reports average log-scores of the multivariate normal distribution for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.8: Out-of-sample log-likelihood for longer forecasts horizon

Overall	Three-step-ahead			Two-step-ahead		
	4	8	12	4	8	12
AR-GARCH	706.00	699.95	698.30	631.04	599.22	584.91
SPF-GARCH	668.33	605.44	614.45	596.04	525.72	521.34
real GDP	4	8	12	4	8	12
AR-GARCH	279.36	279.39	280.07	275.16	275.16	275.41
SPF-GARCH	276.80	279.05	277.35	275.56	269.09	266.44
Inflation	4	8	12	4	8	12
AR-GARCH	189.06	194.72	194.44	195.90	194.47	194.99
SPF-GARCH	184.98	184.59	183.15	186.27	177.01	178.57
Unemployment	4	8	12	4	8	12
AR-GARCH	237.59	225.84	223.79	159.97	129.59	114.51
SPF-GARCH	206.55	141.80	153.95	134.21	79.63	76.32

Notes: This table reports negative log-likelihoods for individual macroeconomic indicators over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). AR-GARCH denotes the baseline AR(1)–GARCH(1,1) model for GDP growth, inflation, and unemployment, while SPF-GARCH augments the conditional mean with SPF projections. All models are estimated using an expanding window with an initial length of 85 observations and assume a multivariate normal distribution. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

Table A.9: Out-of-sample log-scores of SPF-GARCH model specifications for longer forecast horizons

Log-score	Three-step-ahead			Two-step-ahead		
	4	8	12	4	8	12
AR-GARCH	5.88	5.83	5.82	5.26	4.99	4.87
SPF-GARCH	5.57	5.05	5.12	4.97	4.38	4.34
SPF-GARCH-CCC	5.41	4.85	4.93	4.80	4.25	4.22
SPF-GARCH- <i>t</i> -copula-CCC	5.36	4.62	4.65	4.67	4.18	4.15

Notes: This table reports average log-scores of the multivariate predictive densities for GDP growth, inflation, and unemployment over the out-of-sample period 1990:Q1–2019:Q4 (120 observations). Using a multivariate normal distribution, the models are AR-GARCH (baseline AR(1)–GARCH(1,1)), SPF-GARCH (adding SPF projections in the conditional mean), and SPF-GARCH-CCC (assuming a constant conditional correlation matrix). The SPF-GARCH-*t*-copula-CCC model is constructed from univariate GARCH models with Student-*t* marginals combined via a *t*-copula with constant correlations. All models are estimated using an expanding window with an initial length of 85 observations. Columns indicate the week of the quarter in which forecasts are formed. Gray shading denotes inclusion in the 90% model confidence set; bold entries indicate the lowest average loss within each column.

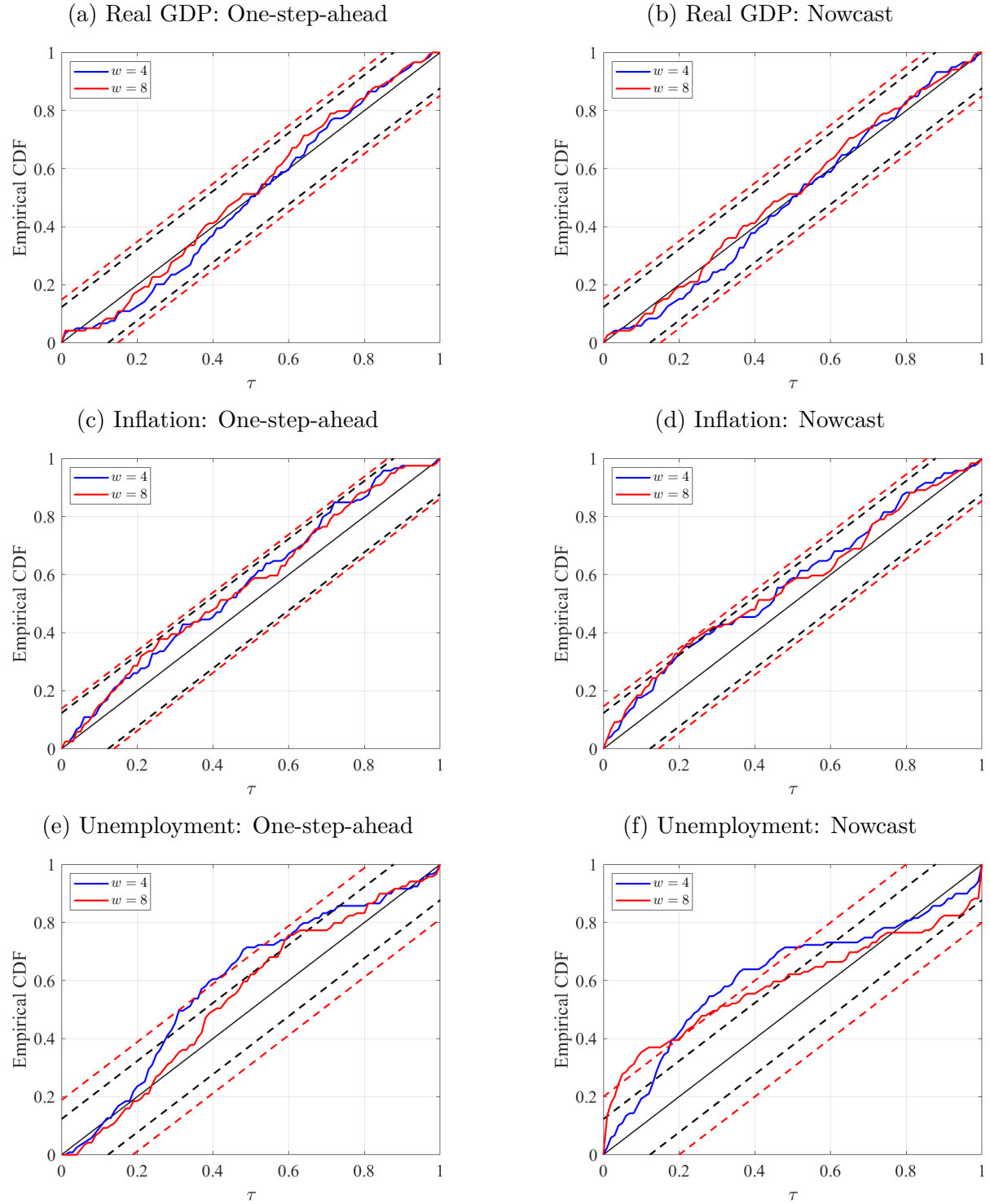
Table A.10: In-sample fit of SPF-GARCH using the *t*-copula for longer forecast horizons

week	Three-step-ahead			Two-step-ahead		
	4	8	12	4	8	12
Normal-CCC	-947.06**	-943.84***	-941.82***	-893.88***	-857.72*	-856.07*
<i>t</i> -copula-CCC	-941.25	-929.35	-927.93	-886.91	-853.30	-851.97

Notes: This table reports in-sample log-likelihoods comparing SPF-GARCH model specifications for GDP growth, inflation, and unemployment. Models are estimated on the sample 1968:Q4–2019:Q4. The baseline model (row 1) assumes a multivariate normal distribution with a constant conditional correlation (CCC) matrix. Row 2 shows the *t*-copula-CCC model constructed from univariate GARCH models with Student-*t* marginals combined via a *t*-copula with constant conditional correlations. Stars indicate significance of likelihood-ratio (LR) tests comparing the *t*-copula-CCC model (row 2) against the normal-CCC model (row 1), imposing four restrictions. Significance levels are: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

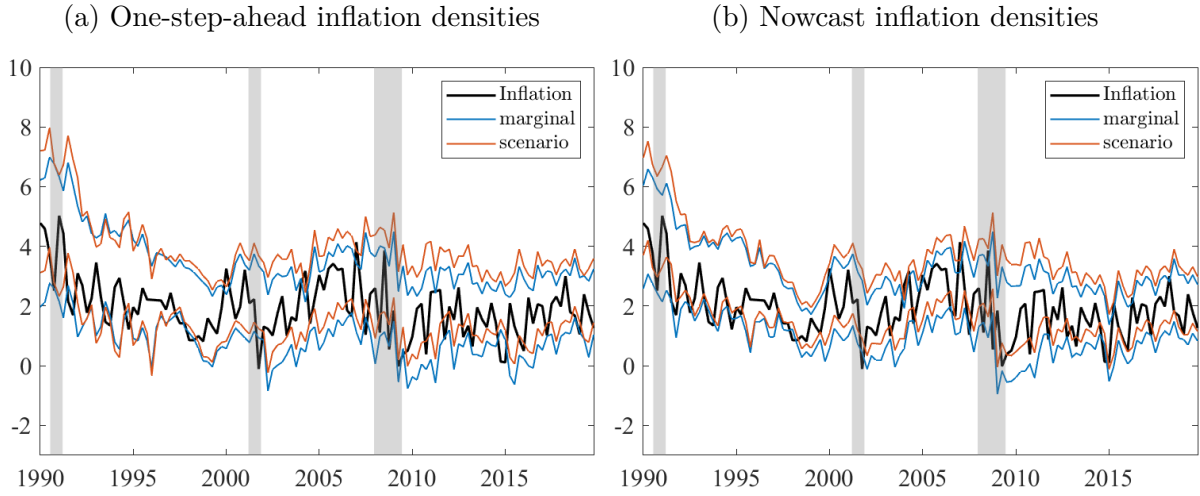
B Figures

Figure B.1: Probability integral transform of the SPF-GARCH model



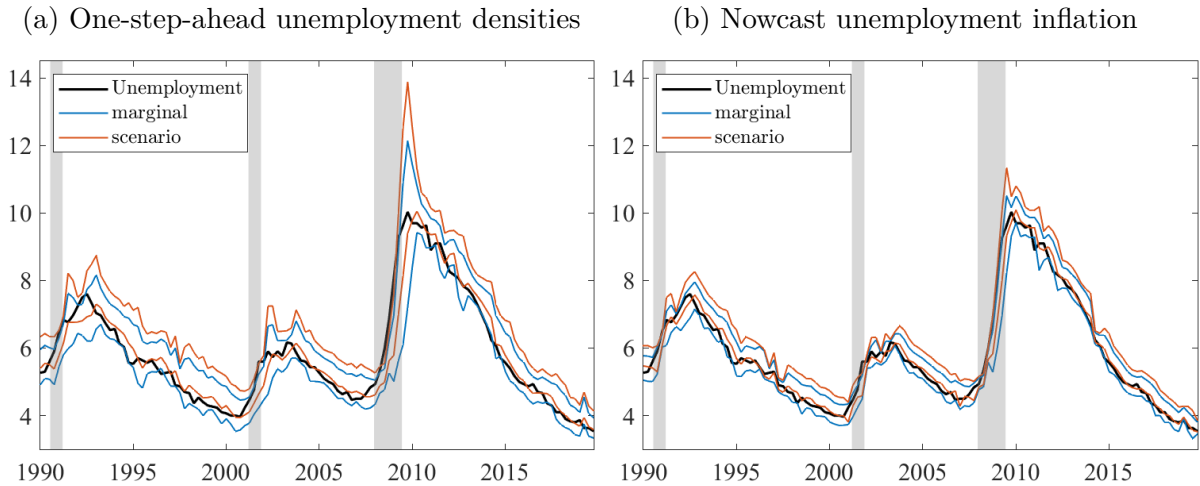
Notes: This figure displays the probability integral transform (PIT) over 1990:Q1–2019:Q4 for the SPF-GARCH model estimated on an expanding window under a multivariate normal distribution. Black dashed lines denote 95% confidence bands under the null of uniformity and independence; red dashed lines show 95% bootstrap bands assuming uniformity only, following [Rossi and Sekhposyan \(2019\)](#).

Figure B.2: Upside inflation scenarios based on the SPF-GARCH model



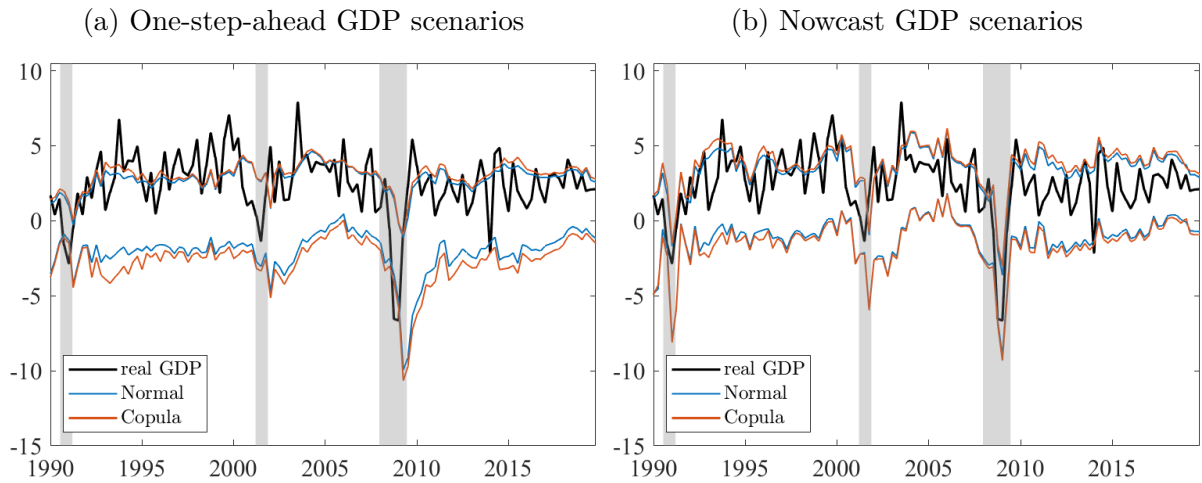
Notes: This figure plots the 10th and 90th percentiles of the marginal predictive density for inflation (blue lines) and the corresponding conditional percentiles when both GDP growth and unemployment are in their lower deciles (red lines), i.e., $Q_{t+h|t,w}^\pi(\tau) | (y \leq Q_{t+h|t,w}^y(0.1), u \leq Q_{t+h|t,w}^u(0.1))$ for $\tau \in \{0.1, 0.9\}$. The black line depicts realized quarterly annualized inflation. Panel (a) shows the one-step-ahead distribution ($h = 1$), and panel (b) the nowcast distribution ($h = 0$), each evaluated at weeks $w = 8$ within a quarter. All forecasts are from the SPF-GARCH-CCC model assuming a multivariate normal distribution. The sample period is 1990:Q1–2019:Q4 and the model is estimated using an expanding window. Vertical gray bands indicate NBER recessions.

Figure B.3: Upside unemployment scenarios based on the SPF-GARCH model



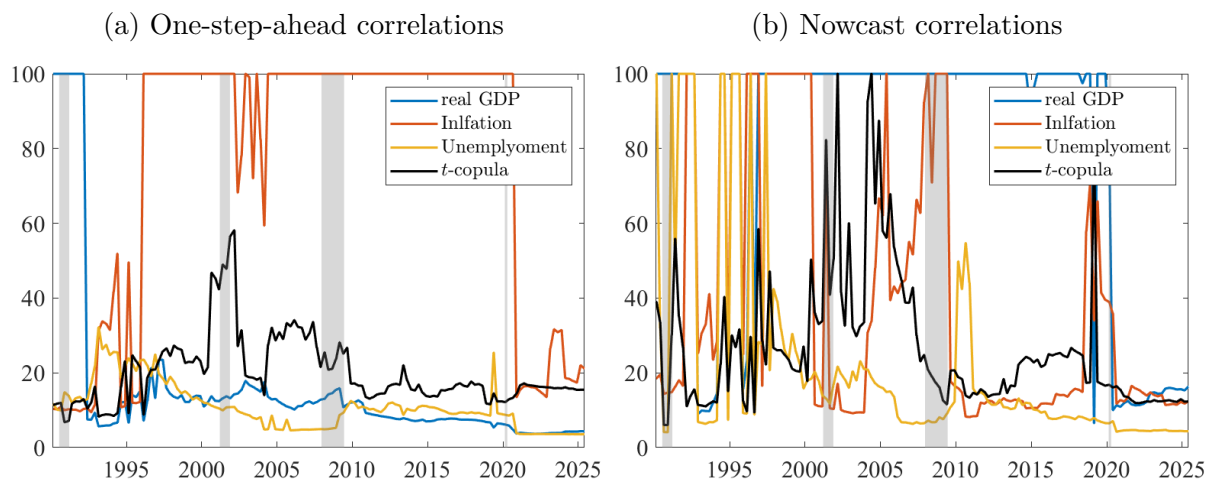
Notes: This figure plots the 10th and 90th percentiles of the marginal predictive density for unemployment (blue lines) and the corresponding conditional percentiles when both real GDP growth and inflation are in their lower deciles (red lines), i.e., $Q_{t+h|t,w}^u(\tau) | (y \leq Q_{t+h|t,w}^y(0.1), \pi \leq Q_{t+h|t,w}^\pi(0.1))$ for $\tau \in \{0.1, 0.9\}$. The black line depicts realized quarterly unemployment. Panel (a) shows the one-step-ahead distribution ($h = 1$) and panel (b) the nowcast distribution ($h = 0$), each evaluated at week $w = 4$ within the quarter. All forecasts are from the SPF-GARCH-CCC model assuming a multivariate normal distribution. The sample period is 1990:Q1–2019:Q4 and the model is estimated using an expanding window. Vertical gray bands indicate NBER recessions.

Figure B.4: Downside GDP scenarios based on the SPF-GARCH model under normality and the t -copula



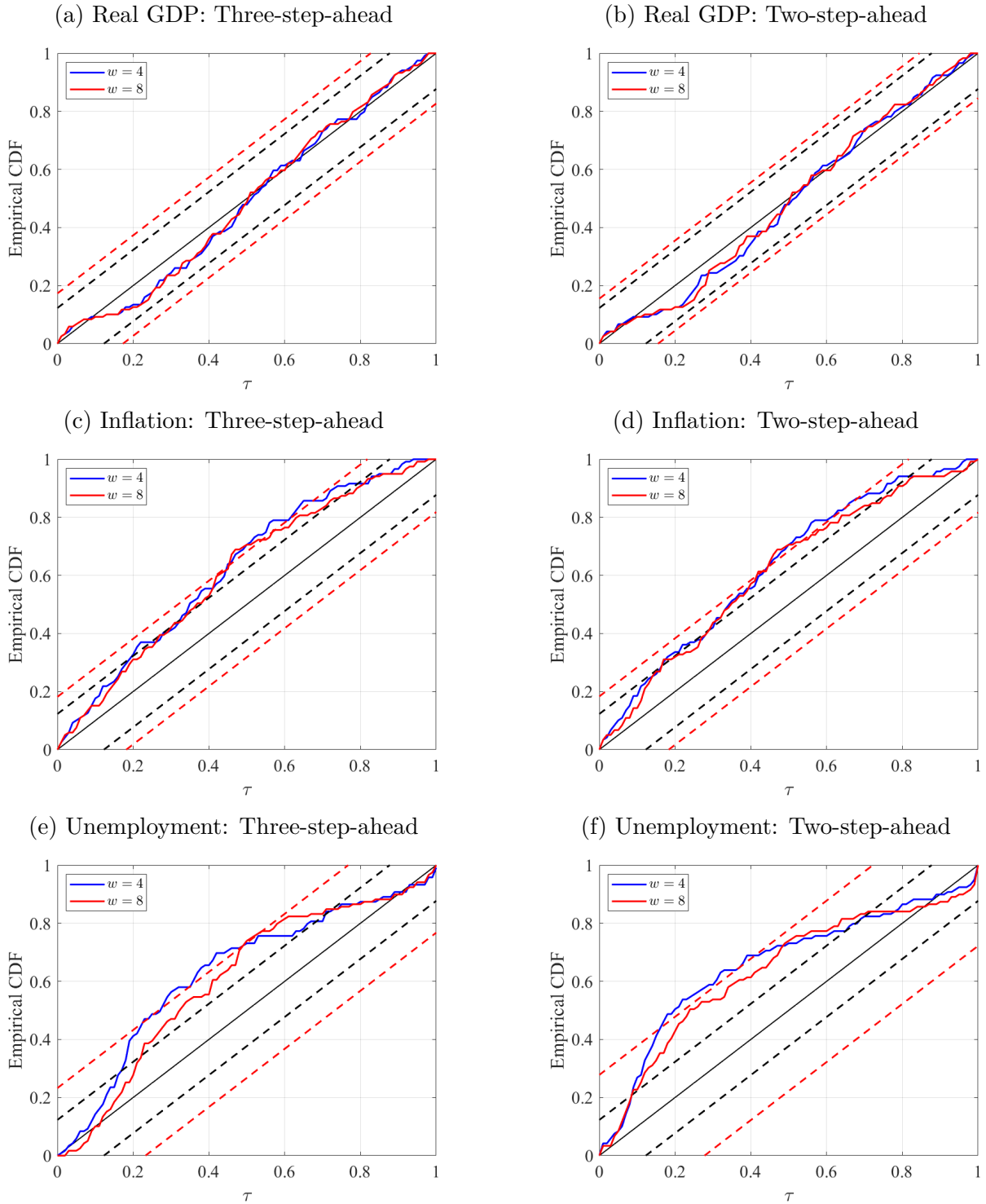
Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for real GDP growth when both inflation and unemployment are in their upper deciles, i.e., $Q_{t+h|t,w}^y(\tau) | (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (black line). Blue lines depict forecasts from the SPF-GARCH model assuming a multivariate normal distribution; red lines depict forecasts from the SPF-GARCH model with Student- t marginals combined via a t -copula with constant conditional correlations. Panel (a) shows the one-step-ahead distribution ($h = 1$) and panel (b) the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. The sample period is 1990:Q1–2019:Q4 and the models are estimated using an expanding window. Vertical gray bands indicate NBER recessions.

Figure B.5: Estimated degrees of freedom from the SPF-GARCH model using the t -copula



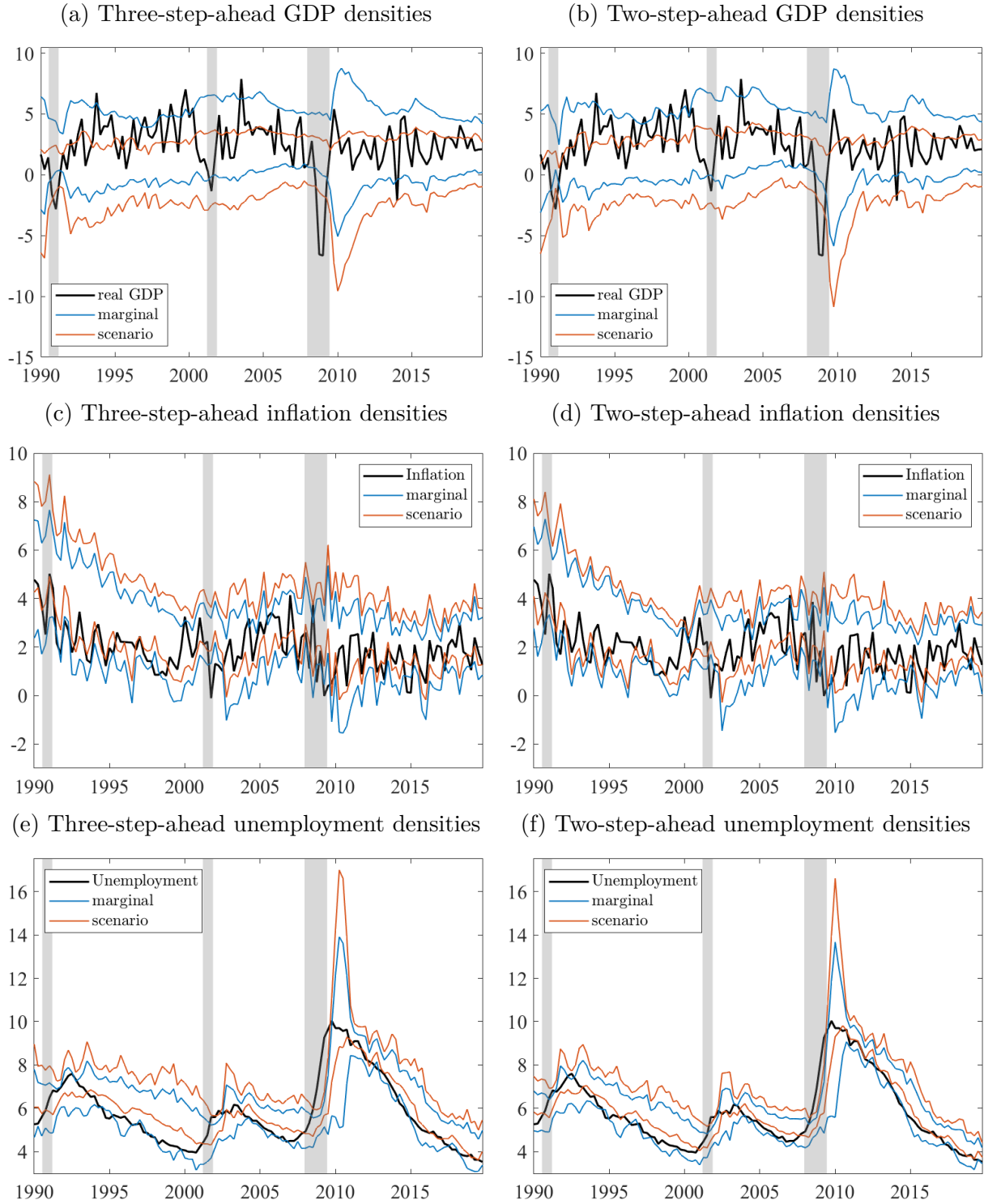
Notes: This figure plots the estimated degrees-of-freedom (DoF) parameters over time from the SPF-GARCH model using the t -copula specification with Student- t marginal distributions. Panel (a) shows results for the one-step-ahead distribution ($h = 1$) and panel (b) for the nowcast distribution ($h = 0$), each evaluated at week $w = 8$ within the quarter. Blue, red, and yellow lines correspond to the DoF parameters of real GDP growth, inflation, and unemployment, respectively, while the black line depicts the DoF parameter of the t -copula. The sample period is 1990:Q1–2025:Q3 and the model is estimated using an expanding window. Vertical gray bands indicate NBER recessions.

Figure B.6: Probability integral transform of the SPF-GARCH model



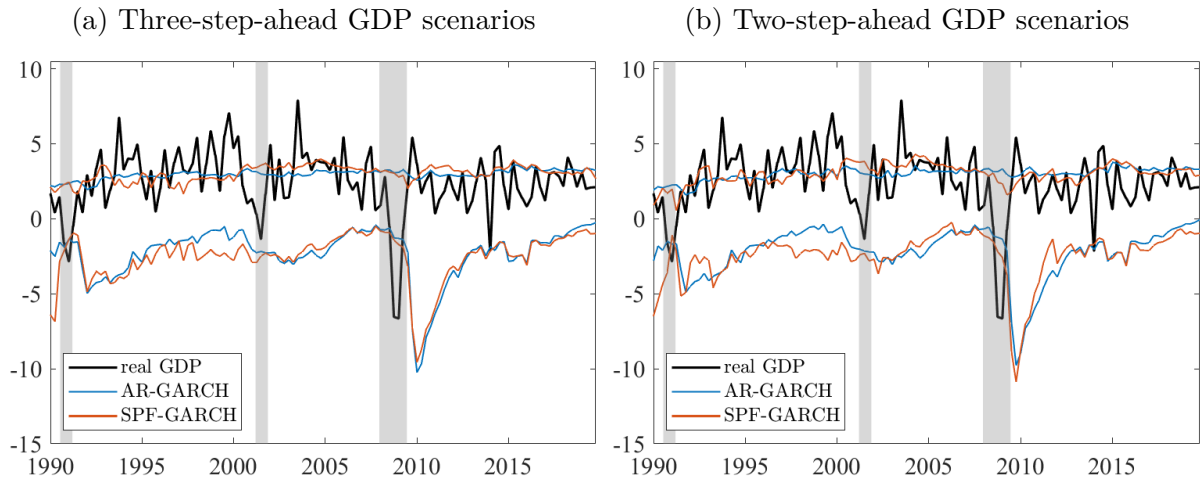
Notes: This figure displays the probability integral transform (PIT) over 1990:Q1–2019:Q4 for the SPF-GARCH model estimated on an expanding window under a multivariate normal distribution. Black dashed lines denote 95% confidence bands under the null of uniformity and independence; red dashed lines show 95% bootstrap bands assuming uniformity only, following Rossi and Sekhposyan (2019).

Figure B.7: SPF-GARCH model-based scenarios for longer forecast horizons



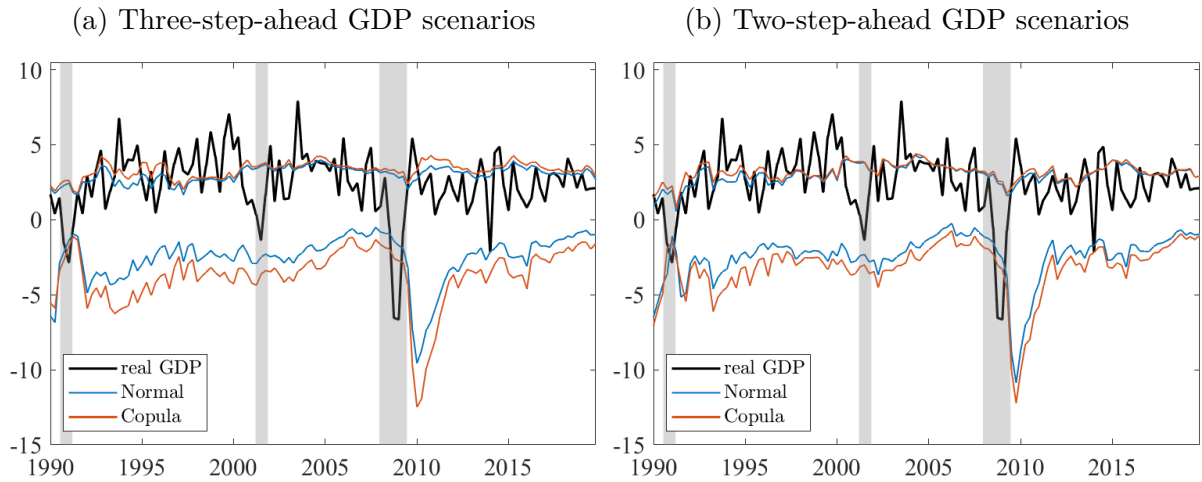
Notes: This figure replicates the scenarios shown in Figures 2, B.2, and B.3 for the sample period 1990:Q1–2019:Q4. Each panel plots the 10th and 90th percentiles of the conditional predictive density (red) and of the marginal distribution (blue), alongside its realized value (black line). Panel (a) and (b) show the three-step-ahead ($h = 3$) and the two-step-ahead distribution ($h = 2$), each evaluated at week $w = 8$ within the quarter. All forecasts are from the SPF-GARCH-CCC model estimated on an expanding window, assuming a multivariate normal distribution. Vertical gray bands indicate NBER recessions.

Figure B.8: Downside GDP scenarios under AR-GARCH and SPF-GARCH models for longer forecast horizons



Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for real GDP growth when both inflation and unemployment are in their upper deciles, i.e., $Q_{t+h|t,w}^y(\tau) | (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (black line). Blue lines depict forecasts from the baseline AR-GARCH model; red lines depict forecasts from the SPF-GARCH model. Panels (a) and (b) show the three-step-ahead ($h = 3$) and the two-step-ahead distribution ($h = 2$), each evaluated at week $w = 8$ within the quarter. The sample period is 1990:Q1–2019:Q4 and the models are estimated using an expanding window. Both models assume a multivariate normal distribution and constant correlations. Vertical gray bands indicate NBER recessions.

Figure B.9: Downside GDP scenarios based on the SPF-GARCH model under normality and the t -copula for longer forecast horizons



Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for real GDP growth when both inflation and unemployment are in their upper deciles, i.e., $Q_{t+h|t,w}^y(\tau) | (\pi \geq Q_{t+h|t,w}^\pi(0.9), u \geq Q_{t+h|t,w}^u(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (black line). Blue lines depict forecasts from the SPF-GARCH model with a multivariate normal specification; red lines depict the version with Student- t marginals combined via a t -copula with constant conditional correlations. Panels (a) and (b) show the three-step-ahead ($h = 3$) and the two-step-ahead distribution ($h = 2$), each evaluated at week $w = 8$ within the quarter. The sample period is 1990:Q1–2019:Q4, and the models are estimated using an expanding window. Vertical gray bands indicate NBER recessions.

C Copula-based robustness specification

For each indicator $i \in \{y, \pi, u\}$, the univariate h -step-ahead distributions are modeled as Student- t :

$$z_{t+h|t,w}^i \sim t\left(\mu_{t+h|t,w}^i, \sigma_{t+h|t,w}^i, \nu_{t,w}^i\right), \quad (27)$$

where the conditional mean and variance are given in equations (8) and (9), and $\nu_{t,w}^i > 2$ denotes a degrees-of-freedom parameter that is estimated using data available in week w of quarter t and held fixed within that estimation window.

Let T_{ν^i} denote the cumulative distribution function of the standard Student- t_{ν^i} distribution. The standardized residuals are given by

$$\eta_{t+h|t,w}^i = \frac{z_{t+h|t,w}^i - \mu_{t+h|t,w}^i}{\sigma_{t+h|t,w}^i}, \quad (28)$$

which are transformed to the unit interval as

$$\tilde{\eta}_{t+h|t,w}^i = T_{\nu^i}\left(\eta_{t+h|t,w}^i\right). \quad (29)$$

To construct the joint predictive distribution, we couple the marginal distributions using a Student- t copula. The Student- t copula with $\nu_{t,w}^c$ degrees of freedom and constant conditional correlation matrix $\mathbf{R}_{t,w}$ is defined as

$$C_{\nu^c, \mathbf{R}}\left(\tilde{\boldsymbol{\eta}}_{t+h|t,w}\right) = T_{\nu^c, \mathbf{R}}\left(T_{\nu^c}^{-1}\left(\tilde{\eta}_{t+h|t,w}^y\right), T_{\nu^c}^{-1}\left(\tilde{\eta}_{t+h|t,w}^\pi\right), T_{\nu^c}^{-1}\left(\tilde{\eta}_{t+h|t,w}^u\right)\right), \quad (30)$$

where $T_{\nu^c, \mathbf{R}}$ denotes the multivariate Student- t cumulative distribution function, and $\tilde{\boldsymbol{\eta}}_{t+h|t,w}$ collects the probability integral transforms defined in equation (29).

In this robustness specification, the marginal distributions are estimated separately for each indicator, yielding one degrees-of-freedom parameter $\nu_{t,w}^i$ for each marginal distribution. Dependence across variables is captured by a Student- t copula with a

constant conditional correlation (CCC) matrix $\mathbf{R}_{t,w}$ and an additional copula degrees-of-freedom parameter $\nu_{t,w}^c$. The CCC assumption implies that correlations are constant within each estimation window, which keeps the parameterization parsimonious while allowing for tail dependence through the copula. Estimation proceeds in two steps: first, the univariate Student- t GARCH models are fitted for each indicator; second, the copula parameters $(\mathbf{R}_{t,w}, \nu_{t,w}^c)$ are estimated conditional on the marginal fits using the probability integral transforms defined in equations (28) and (29). This two-step approach ensures flexibility in the marginal distributions while maintaining tractability in the joint specification.