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

Manuel Schick†, Anne Opschoor‡

October 24, 2025

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 enhanced. Modeling a constant correlation matrix further 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, providing a transparent assessment of 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, and Julius Schoelkopf, 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), and the Workshop on Macro- and Financial Econometrics at the School of Business, Örebro University (November 2025), for their helpful comments and suggestions.

†Department of Economics, Heidelberg University, 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 are essentially forward-looking forecasts that typically rely on judgment-driven projections at long horizons or on a small set of reported quantiles. As a result, they offer only a partial view of the joint predictive distribution and, crucially, leave the accuracy of such forecasts untested.

Despite their widespread use, existing approaches to scenario analysis face important limitations. Institutional exercises are often grounded in structural models such as vector autoregressions (VARs), which impose strong assumptions about dynamics and shocks, or in judgment-driven projections that lack statistical transparency. Even in the academic literature, recent innovations remain constrained. For example, Chavleishvili and Manganelli (2024) develop a quantile VAR to stress-test GDP growth under adverse financial shocks, but note that stress testing in central banks often lacks a rigorous econometric foundation. 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. While this provides a clever synthesis of model-based forecasting and judgmental narratives, the resulting scenario densities are difficult to evaluate statistically.

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. These approaches provide a valuable tool to assess the exposure of GDP growth to severe but plausible global or domestic shocks. However, they typically focus on GDP growth alone and condition on exogenously stressed latent factors, rather than estimating a joint predictive density across macroeconomic variables directly. This highlights a broader limitation of existing scenario tools: few provide transparent,

multivariate predictive densities that can be formally evaluated out-of-sample.

In this paper, we extend the Growth-at-Risk (GaR) framework (Adrian et al., 2019) into a multivariate setting to generate 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. Methodologically, we incorporate Survey of Professional Forecasters (SPF) expectations into the conditional mean within a parametric AR-GARCH framework (Schick, 2024), capturing time-varying volatility and cross-variable correlations. A central strength of our approach is that the underlying parametric joint predictive densities can be used not only to transparently construct scenarios, but also directly evaluated out-of-sample, a feature largely absent in existing scenario exercises.

In the GaR literature, conditional quantile regression (QR) has been the predominant tool to model tail risk, relying on observable predictors such as financial conditions (see, 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). By contrast, the GARCH framework imposes no assumptions about the source of shocks and focuses purely on the statistical properties of predictive distributions. As Brownlees and Souza (2021) demonstrate, such volatility-based approaches can match the performance of QR models in capturing macro-financial risks. While this comes at the cost of structural interpretation, the statistical agnosticism makes GARCH particularly well suited for transparent and flexible scenario construction.

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 consensus 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: 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 embedding 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 as an anchor for 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 for scenario forecasting. Methodologically, we generate joint predictive densities from a multivariate GARCH framework anchored by SPF expectations, allowing for transparent scenario construction and out-of-sample evaluation. Empirically, we demonstrate the approach on US data and highlight its policy relevance by constructing scenarios before and during the COVID-19 pandemic. This framework provides policymakers with a transparent, statistically grounded tool to assess risks across multiple macroeconomic variables simultaneously.

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. Finally, the results are robust to alternative distributional assumptions, including specifications allowing for fat-tailed innovations.

The remainder of the paper is structured as follows. Section 2 describes the real-time data used in the analysis. Section 3 introduces the methodological framework, including the multivariate GARCH model and scenario construction. Section 4 presents the evaluation metrics used to assess the predictive densities. Section 5 reports the empirical results, including out-of-sample performance and policy-relevant scenario forecasts. Section 6 concludes with a discussion of the findings and their implications.

2 Real-time data and predictors

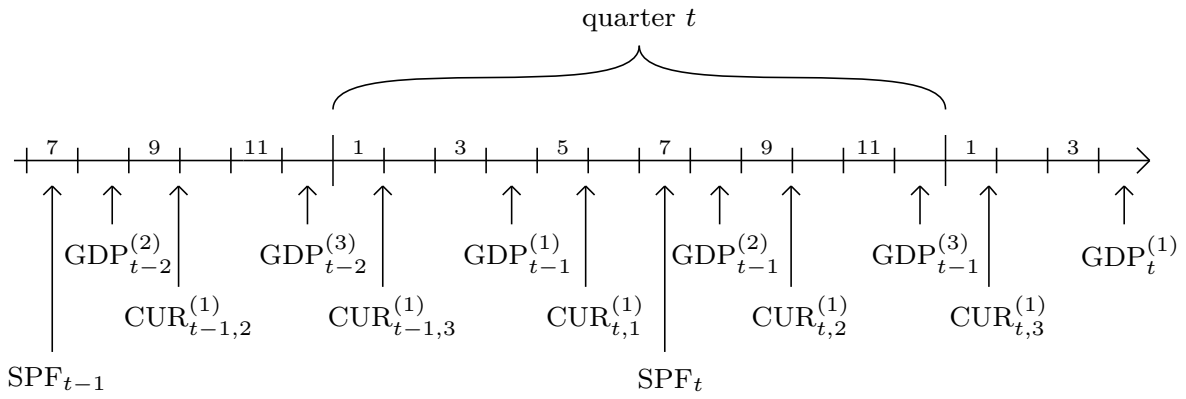
In order to avoid look-ahead bias, we conduct a real-time out-of-sample forecasting analysis. To this end, we use monthly real-time vintage data on real GDP, the GDP Price Index, and the unemployment rate, obtained from the Real-Time Data Set for Macroeconomists (RTDSM) provided by the Federal Reserve 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, $\text{GDP}_{t-1}^{(1)}$, becomes available around week 4 of quarter t. One and two months later, the BEA releases the *preliminary* and *final* estimates, denoted $\text{GDP}_{t-1}^{(2)}$ and $\text{GDP}_{t-1}^{(3)}$, 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. 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 $\text{CUR}_{t,1}^{(1)}$, becomes available in week 5. One and two months later, the

[1]Prior 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 civilian rate of unemployment unemployment (CUR) together with the release dates of the Survey of Professional Forecasters (SPF). The numbers on the time line indicate the weeks in each quarter.

corresponding first estimates for the second and third months of the quarter, $\text{CUR}^{(1)}_{t,2}$ and $\text{CUR}^{(1)}_{t,3}$, 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 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 real-time information flow from quarterly 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:

$$y_{s,w} = 400 \cdot (\log(\text{GDP}_{s,w}) - \log(\text{GDP}_{s-1,w})), \tag{1}$$

where $\text{GDP}_{s,w}$ 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:

$$\pi_{s,w} = 400 \cdot (\log(\text{PGDP}_{s,w}) - \log(\text{PGDP}_{s-1,w})), \tag{2}$$

where $\text{PGDP}_{s,w}$ 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 (w)	Variables
1	$\text{GDP}_{t-2,12}$, $\text{PGDP}_{t-2,12}$, $\text{CUR}_{t-1,3,12}$, $\text{SPF}_{t+h\|t-1}$, $\text{NFCI}_{t-1,12}$
3	$\text{GDP}_{t-2,12}$, $\text{PGDP}_{t-2,12}$, $\text{CUR}_{t-1,3,12}$, $\text{SPF}_{t+h\|t-1}$, $\text{NFCI}_{t,2}$
5	$\text{GDP}_{t-1,4}$, $\text{PGDP}_{t-1,4}$, $\text{CUR}_{t,1,5}$, $\text{SPF}_{t+h\|t-1}$, $\text{NFCI}_{t,5}$
7	$\text{GDP}_{t-1,4}$, $\text{PGDP}_{t-1,4}$, $\text{CUR}_{t,1,5}$, $\text{SPF}_{t+h\|t}$, $\text{NFCI}_{t,6}$
9	$\text{GDP}_{t-1,8}$, $\text{PGDP}_{t-1,8}$, $\text{CUR}_{t,2,9}$, $\text{SPF}_{t+h\|t}$, $\text{NFCI}_{t,8}$
11	$\text{GDP}_{t-1,8}$, $\text{RGDP}_{t-1,8}$, $\text{CUR}_{t,2,9}$, $\text{SPF}_{t+h\|t}$, $\text{NFCI}_{t,10}$

Notes: This table demonstrates the latest vintages of real GDP (GDP) and the Price Index (PGDP), SPF projections, and NFCI data available for odd numbered weeks throughout quarter t.

At quarter t, to construct real-time forecasts of GDP and inflation, the index $s =$

$1, \ldots, t-j$ 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,

$$u_{s,w} = \frac{1}{3} \sum_{m=1}^{3} \mathrm{CUR}_{s,m,w}, \tag{3}$$

where $\mathrm{CUR}_{s,m,w}$ 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, $u_{t-1,w}$, 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 $u_{s,w}$ on $u_{s-1,w}$ would ignore new information arriving during the current quarter. To reflect the data available in week w, we define the predictor $\tilde{u}_{s-1,w}$ as

$$\tilde{u}_{s-1,w} \begin{cases} u_{s-1,w} & \text{if } w = 1 \\ \frac{1}{M} \sum_{l=1}^{M} \mathrm{CUR}_{s,m,w} & \text{if } w > 4, \end{cases} \tag{4}$$

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, $\mathrm{CUR}_{t,1,w}$, has typically been released, so $M = 1$. Once the second month's data becomes available, $\tilde{u}_{t-1,w}$ 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 1990–2023, based

on the first vintage available in 2024:Q4. 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.

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:

$$SPF^i_{t+h,w} = \begin{cases} SPF^i_{t+h|t-1} & \text{if } w \leq w^\star \\ SPF^i_{t+h|t} & \text{if } w > w^\star, \end{cases} \tag{5}$$

where $SPF^i_{t+h|t}$ denotes the median h-step-ahead forecast from the survey released in quarter t, and $w^\star$ 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 $t-1$, $SPF^{i}_{t+1|t-1}$, 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:

$$SPF^{IQR,i}_{t+h,w} = \begin{cases} SPF^{0.75,i}_{t+h|t-1} - SPF^{0.25,i}_{t+h|t-1} & \text{if } w \leq w^{\star} \\ SPF^{0.75,i}_{t+h|t} - SPF^{0.25,i}_{t+h|t} & \text{if } w > w^{\star}, \end{cases} \tag{6}$$

where $SPF^{q,i}_{t+h|t}$ 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 (density) 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:

$$NFCI_{s,w} = \frac{1}{w} \sum_{n=1}^{w} \text{NFCI}_{s,w-n}, \tag{7}$$

where $\text{NFCI}_{s,w-n}$ 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 $NFCI_{s,2} = \frac{1}{2}(\text{NFCI}_{s,1} + \text{NFCI}_{s,0})$, where $\text{NFCI}_{s,0}$ corresponds to week 12 of quarter $s-1$.

3 Methodology

3.1 Forecasting models

We model the joint distribution of real GDP (y_t), inflation (π_t), and unemployment (u_t) 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

$$\mu_{t+h|t,w} = c_w^h + \boldsymbol{\phi}_w^h \mathbf{x}_{t,w}^{\mu} \tag{8}$$

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

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. 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 instance, if $h = 1$ and $w = 4$, $\mu_{s+1|t,4}$ are the one-step-ahead conditional means for quarters $s = 1, \ldots, t$ given the latest available information in the fourth week of quarter t.

The choice of lag j for $y_{t,w}$ aligns with direct forecasting, depending on the indicator's

release schedule. Thus, if previous quarter's GDP is not available, we use the second lag, i.e., $j = 2$, otherwise $j = 1$. The predictor set $\mathbf{x}_{t,w}^{\mu}$ for the conditional mean includes lagged values of y_t and potentially other explanatory variables. For the conditional variance, additional predictors are included via the exponential function, $\exp(\cdot)$, where $x_{t,w}^{\sigma}$ is standardized to have mean zero and variance one.

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

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

where $\mathbf{Y}_{t+h|t,w} = (y_{t+h|t,w}, \pi_{t+h|t,w}, u_{t+h|t,w})^{\top}$, 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,

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

where $\mathbf{D}_{t+h|t,w}$ is a diagonal matrix with conditional standard deviations $\sigma_{t+h|t,w}$ 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

$$Q_{t+h|t,w} = (1 - a_w^h - b_w^h)\overline{Q} + a_w^h \eta_{t-j,w}\eta_{t-j,w}^{\top} + b_w^h Q_{t+h-j|t,w} \tag{12}$$

$$R_{t+h|t,w} = \operatorname{diag}(Q_{t+h|t,w})^{-1/2}\ Q_{t+h|t,w}\ \operatorname{diag}(Q_{t+h|t,w})^{-1/2}, \tag{13}$$

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 $\mathbf{\Sigma}_{t+h|t,w}$. 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).

3.2 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

$$Q^i_{t+h|t,w}(\tau) = \mu_{t+h|t,w,i} + \sqrt{\sigma^2_{t+h|t,w,i}} F^{-1}(\tau), \tag{14}$$

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. The corresponding bad-scenario 10% quantile of GDP growth is then given by

$$Q^y_{t+h|t,w}(0.1) \mid (\pi = Q^\pi_{t+h|t,w}(0.9),\ u = Q^u_{t+h|t,w}(0.9)). \tag{15}$$

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

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

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

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

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

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

Alternatively, scenarios can be defined in a more extreme sense by conditioning on some variables exceeding extreme quantiles. For example, GaR conditional on both inflation and unemployment being above their 90% quantiles is defined as

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

which requires simulation.

4 Forecast evaluation

When evaluating forecasts, an important question is which values to use as the 'true' 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 panelists of the SPF 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

$$y_t = 400 \cdot (\log(\text{GDP}_{t,24}) - \log(\text{GDP}_{t-1,24})), \tag{20}$$

$$\pi_t = 400 \cdot (\log(\text{PGDP}_{t,24}) - \log(\text{PGDP}_{t-1,24})), \tag{21}$$

where $\text{GDP}_{t,24}$ and $\text{PGDP}_{t,24}$ denote real GDP and the GDP Price Index for quarter t, available in the most recent vintage at the end of quarter $t+1$ (i.e., $w = 24$).

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

$$u_t = \frac{1}{3}\sum_{l=1}^{3} \text{CUR}_{t,l,19}, \tag{22}$$

where $l = 1, 2, 3$ denotes the months within the quarter, and $w = 19$ specifies the vintage of the data used. 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

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

where $p^k_{t,w}$ 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, an ideal approach for scenario analysis would involve assessing the accuracy of conditional quantiles directly. However, due to the limited number of observations, such an approach is not feasible here.

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

$$d^{k,l}_{t+h,w} = \log\, p^l_{t,w}(\mathbf{Y}_{t+h}) - \log\, p^k_{t,w}(\mathbf{Y}_{t+h}), \tag{24}$$

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

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

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

However, because we are comparing $k = 1, \ldots, K$ 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 $k = 1, \ldots, K$, 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

$$T_{\mathcal{M}}^{w} = \max_{k,l \in \mathcal{M}} \left| S_{t+h,w}^{k,l} \right|, \tag{26}$$

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 $\mathcal{M}^{\star}$.

Since the distribution of $T_{\mathcal{M}}^{w}$ is non-standard, its critical values must be determined through simulation. 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 5.3.

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 distributional assumptions, we report pairwise likelihood-ratio tests, which formally compare the relative performance of nested

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

models. These tests evaluate whether a more complex model provides a statistically significant improvement in fit over a simpler alternative.

5 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. We then assess scenario-based risk measures and, finally, extend the analysis to include the COVID-19 pandemic.

5.1 Out-of-sample forecast evaluation

We begin by assessing the predictive content of the SPF relative to the NFCI 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 $\mathbf{\Sigma}_{t+h|t,w}$ 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 ($x^{\mu}_{t,w,i}$) or variance ($x^{\sigma}_{t,w,i}$) of indicator i. The fourth row refers to the SPF-GARCH model, which includes the SPF consensus forecast, $SPF^{i}_{t+h,w}$, 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

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

	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
AR-GARCH:	4.55	4.35	4.21	3.61	3.09	2.22
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.45	4.24	4.03	3.42	3.01	2.15
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.70	4.48	4.28	3.73	3.31	2.21
SPF-GARCH:	4.26	3.85	3.71	3.41	2.60	1.86
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.27	3.85	3.75	**3.28**	**2.49**	**1.74**
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.41	4.28	3.70	3.38	2.58	1.87
$+x^{\sigma}_{t,w,i} = SPF^{IQR,i}_{t,w}$	**4.17**	**3.76**	**3.66**	3.44	2.60	1.88

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.

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 reveals clear gains in forecast accuracy when incorporating releases and revisions of each macroeconomic indicator during the quarter. Predictive log-scores for both one-step-ahead forecasts and nowcasts improve, i.e., become smaller, 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.

The baseline AR-GARCH model also reveals clear gains in forecast accuracy when 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 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		
Overall	4	8	12	4	8	12
AR-GARCH	546.59	522.11	504.60	433.03	370.95	266.59
SPF-GARCH	**510.74**	**462.14**	**444.75**	**408.69**	**312.16**	**222.85**
real GDP	4	8	12	4	8	12
AR-GARCH	264.11	266.42	267.37	252.49	250.67	251.63
SPF-GARCH	**261.84**	**256.61**	**248.97**	**245.70**	**233.11**	**232.10**
Inflation	4	8	12	4	8	12
AR-GARCH	190.96	194.48	193.08	178.00	180.61	180.01
SPF-GARCH	**176.65**	**174.31**	**172.65**	**163.96**	**153.72**	**154.85**
Unemployment	4	8	12	4	8	12
AR-GARCH	91.52	61.21	44.16	2.54	-60.33	**-165.06**
SPF-GARCH	**72.25**	**31.22**	**23.13**	**-0.97**	**-74.67**	-164.10

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. In contrast, gains

for the nowcast distribution are mostly insignificant. 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

	One-step-ahead			Nowcast		
Log-score	4	8	12	4	8	12
AR-GARCH	4.55	4.35	4.21	3.61	3.09	2.22
SPF-GARCH	4.26	3.85	3.71	**3.41**	**2.60**	**1.86**
SPF-GARCH-CCC	4.21	3.77	3.66	3.41	2.60	1.87
SPF-GARCH-DCC	**4.20**	**3.76**	**3.66**	3.41	2.60	1.87

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.

5.2 Scenario-based risk assessment

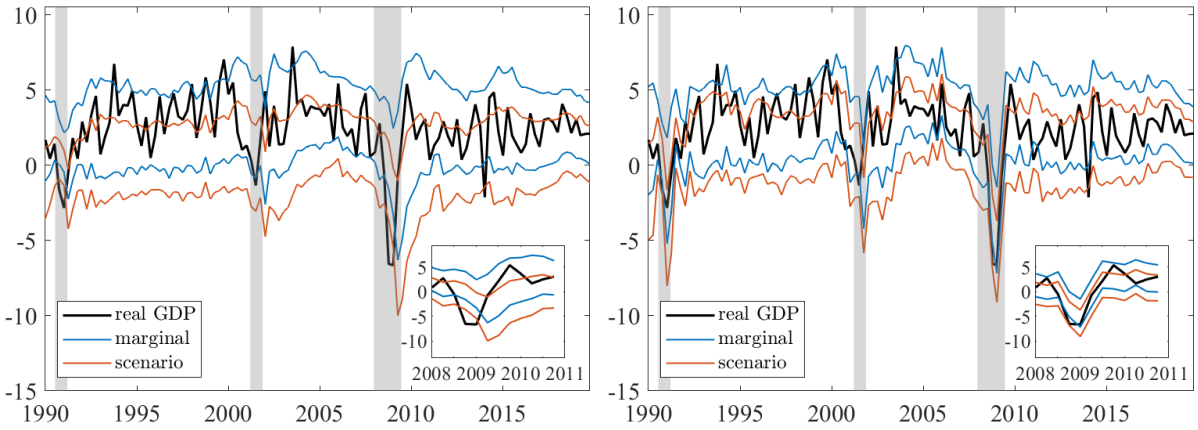
The previous section demonstrates the strong out-of-sample performance of the SPF-GARCH-CCC model for relatively short forecast horizons. We now illustrate how scenarios can be constructed from the multivariate predictive densities that incorporate the SPF.

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.

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 analysis points to elevated downside risks to GDP. 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 (cf. Schick, 2024).

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

(a) One-step-ahead GDP densities (b) Nowcast GDP densities



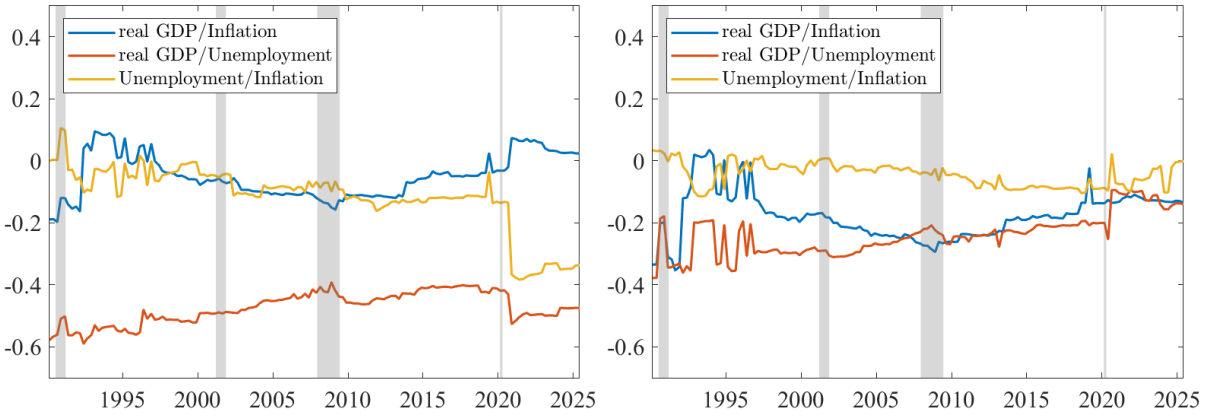
Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for real GDP growth when both inflation and unemployment are in their upper deciles, i.e., $Q^{y}_{t+h|t,w}(\tau) \mid (\pi \geq Q^{\pi}_{t+h|t,w}(0.9),\ u \geq Q^{u}_{t+h|t,w}(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (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. The sample period is 1990:Q1–2019:Q4. All forecasts are from the SPF-GARCH model assuming a multivariate normal distribution. Red lines depict the conditional ("bad-scenario") percentiles; blue lines depict the percentiles of the marginal distribution. Vertical gray bands denote NBER recessions.

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 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

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

(a) One-step-ahead correlations (b) Nowcast correlations

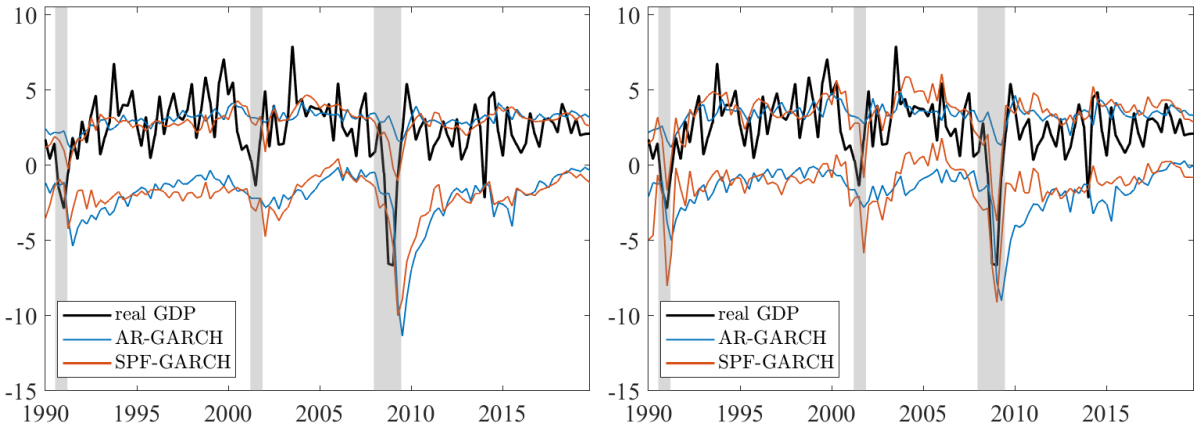


Notes: This figure plots the constant conditional correlation (CCC) coefficients for GDP growth, inflation, and unemployment estimated from the SPF-GARCH-CCC model. Panel (a) shows results for the one-step-ahead distribution ($h = 1$) and panel (b) for the nowcast distribution ($h = 0$), each evaluated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

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

(a) One-step-ahead GDP scenarios (b) Nowcast GDP scenarios



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., $Q^{y}_{t+h|t,w}(\tau) \mid (\pi \geq Q^{\pi}_{t+h|t,w}(0.9),\ u \geq Q^{u}_{t+h|t,w}(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (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. The sample period is 1990:Q1–2019:Q4. Blue lines depict forecasts from the baseline AR-GARCH model; red lines depict forecasts from the SPF-GARCH model. Both models assume a multivariate normal distribution. Vertical gray bands denote 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.

5.3 The COVID-19 pandemic

Next, we extend the sample to 2023:Q4. However, due to exceptionally large forecast errors during the COVID-19 pandemic, the period from 2020:Q1 to 2020:Q4 is excluded from the forecast evaluation. Table A.2 confirms that the results are robust to extending the evaluation period. The SPF-GARCH specification remains included in the MCS throughout. Adding the NFCI only yields significant nowcast improvements 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.

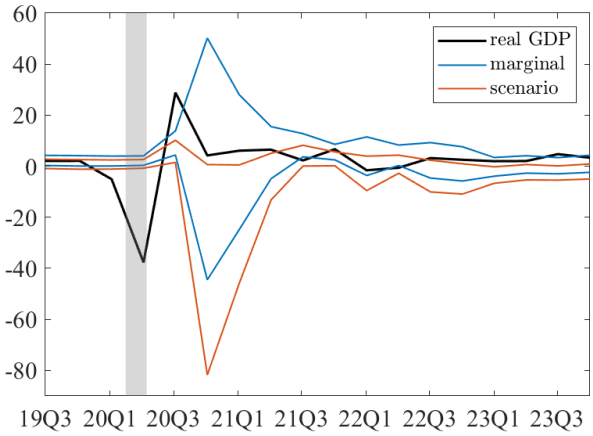
Panels (c) and (d) show that the SPF-GARCH model already indicated elevated upside risks to inflation at the end of 2020. During the high-inflation period of 2021 and 2022, even the 90% quantile of the baseline one-step-ahead density closely tracked the realized inflation rates. The scenario density 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 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, evaluated at week $w = 4$ within the quarter. This corresponds to conditioning on the two-step-ahead and one-step-ahead SPF projections, respectively. Evaluating the nowcast later in the quarter, e.g., at week 8, would yield extremely tight prediction intervals closely surrounding the realizations. However, the nowcasting scenario densities would then become nearly indistinguishable from the marginal ones, as forecast errors are extremely small at that point.

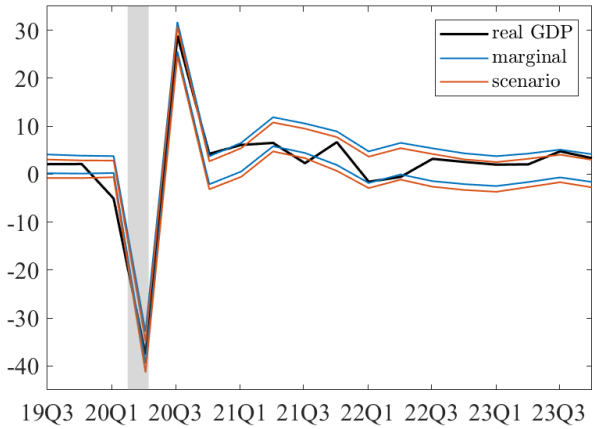
As seen in panel (e), the one-step-ahead density fails to anticipate the sharp increase in unemployment at the onset of the pandemic. The most notable result appears in panel (f), where the SPF-GARCH model produces a wide spread in both the marginal and scenario densities. This reflects the large estimated nowcast variance during the early pandemic period. Compared to the one-step-ahead specification, the GARCH equation for the nowcast places greater weight on past nowcast errors. It takes roughly four quarters until the nowcast variance falls below the forecast variance again. Overall, the scenario density for unemployment is clearly shifted upward for both forecast horizons.

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

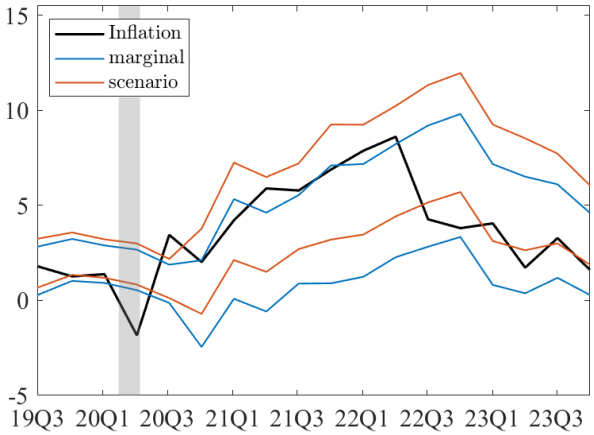
(a) One-step-ahead GDP densities



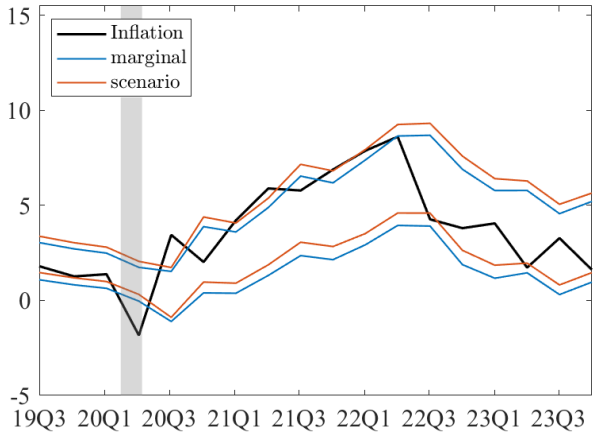
(b) Nowcast GDP densities



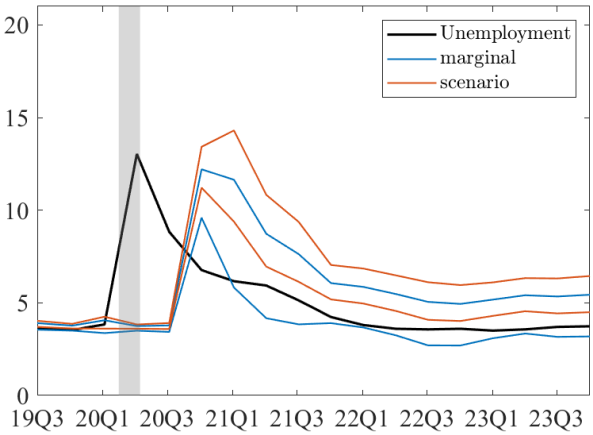
(c) One-step-ahead inflation densities



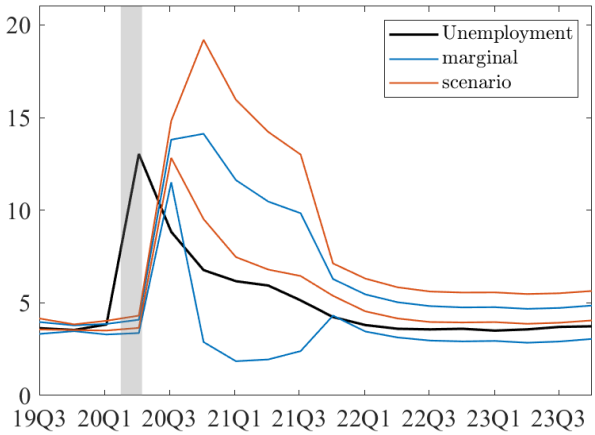
(d) Nowcast inflation densities



(e) One-step-ahead unemployment densities



(f) Nowcast unemployment densities



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 model assuming a multivariate normal distribution. Vertical gray bands denote NBER recessions.

5.4 Robustness analysis

5.4.1 Rolling-window estimation

As discussed in Section 4, 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 85 observations, corresponding to the sample length at the start of the forecasting exercise in 1990:Q1. 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 more AR-GARCH specifications being included in the MCS.

5.4.2 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 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 distributions 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.

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 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.

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 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 Conclusion

This paper proposes assessing joint macroeconomic risks by modeling the multivariate predictive density of macroeconomic indicators. We incorporate SPF consensus forecasts to model the conditional mean 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.

Our approach is deliberately agnostic: while it lacks structural interpretation, it provides a transparent econometric framework to quantify joint risks and tail dependence across macroeconomic variables. Extensions could expand the set of variables included in the joint system even in the absence of SPF projections, for example by incorporating financial indicators such as the term spread.

Taken together, this framework offers policymakers a statistically grounded and real-time tool to assess macroeconomic scenarios and quantify risks across multiple dimensions simultaneously.

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, April). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

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, July). 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.

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.

Brave, S. and R. A. Butters (2012, June). 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.

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. and S. Manganelli (2024). Forecasting and stress testing with quantile vector autoregression. *Journal of Applied Econometrics 39*(1), 66–85.

Clements, M. P., R. W. Rich, and J. S. Tracy (2023). Surveys of professionals. In *Handbook of Economic Expectations*, pp. 71–106. Academic Press.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

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.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics 119*(3), 457–471.

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.

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.

Appendices

A Tables

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

	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
Normal	-872.59***	-817.16***	-812.86***	-614.72**	-551.49***	-416.82**
Normal-CCC	-853.90	-795.72	-793.07	-609.70	-544.22	-412.61
Normal-DCC	-852.60	-795.61	-792.84	-609.46	-544.13	-410.69

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

	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
AR-GARCH:	4.72	4.53	4.38	3.78	3.19	2.36
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.62	4.42	4.21	3.60	3.13	2.27
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.84	4.62	4.45	3.91	3.38	2.36
SPF-GARCH:	4.45	4.05	3.93	3.66	2.77	2.01
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.46	4.05	3.97	**3.54**	**2.70**	**1.90**
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.62	4.49	3.95	3.66	2.77	2.06
$+x^{\sigma}_{t,w,i} = SPF^{IQR,i}_{t,w}$	**4.38**	**3.99**	**3.91**	3.68	2.76	2.03

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

	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
AR-GARCH:	4.71	4.52	4.34	3.54	3.00	2.12
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.49	4.25	4.01	3.47	2.99	2.14
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.65	4.52	4.38	3.77	3.20	2.27
SPF-GARCH:	**4.37**	3.87	3.83	**3.40**	2.55	1.88
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.44	4.03	3.94	3.40	**2.55**	**1.83**
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.63	4.24	4.19	3.60	2.69	1.90
$+x^{\sigma}_{t,w,i} = SPF^{IQR,i}_{t,w}$	4.38	**3.81**	**3.81**	3.47	2.63	1.87

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 length 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

	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
AR-GARCH:	4.52	4.24	4.11	3.62	3.00	2.15
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.28	4.11	3.92	3.42	2.97	2.11
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.52	4.43	4.15	3.72	3.08	2.20
SPF-GARCH:	4.12	**3.67**	**3.53**	3.36	2.57	1.84
$+x^{\mu}_{t,w,i} = NFCI_{t,w}$	4.17	3.73	3.66	**3.29**	**2.52**	**1.75**
$+x^{\sigma}_{t,w,i} = NFCI_{t,w}$	4.17	3.82	3.68	3.41	2.59	1.87
$+x^{\sigma}_{t,w,i} = SPF^{IQR,i}_{t,w}$	**4.05**	3.69	3.61	3.42	2.60	1.83

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

	One-step-ahead			Nowcast		
Log-score	4	8	12	4	8	12
Normal-CCC	4.21	3.77	3.66	3.41	2.60	1.87
t-copula-CCC	**4.12**	**3.67**	**3.53**	**3.36**	**2.57**	**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

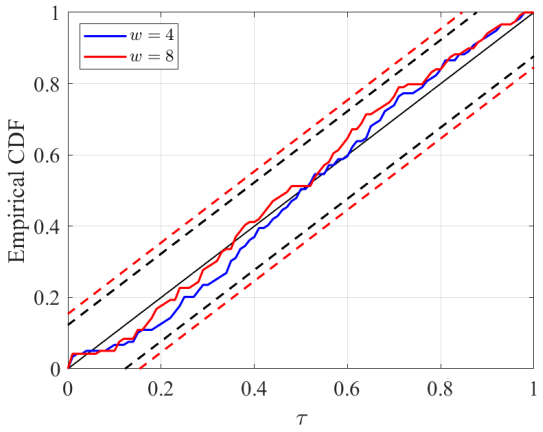
	One-step-ahead			Nowcast		
week	4	8	12	4	8	12
Normal-CCC	-853.90***	-795.72***	-793.07***	-609.70**	-544.22***	-412.61**
t-copula-CCC	-841.36	-776.71	-775.59	-603.86	-532.69	-407.29

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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

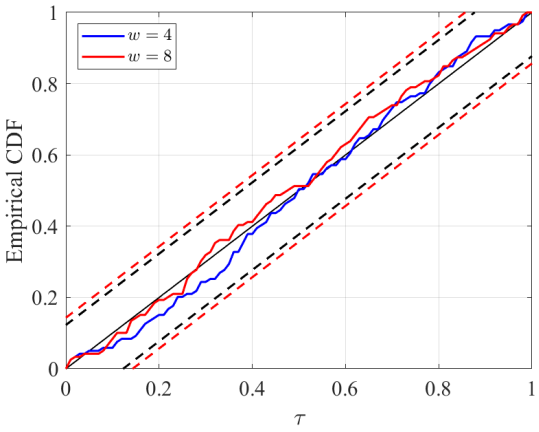
B Figures

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

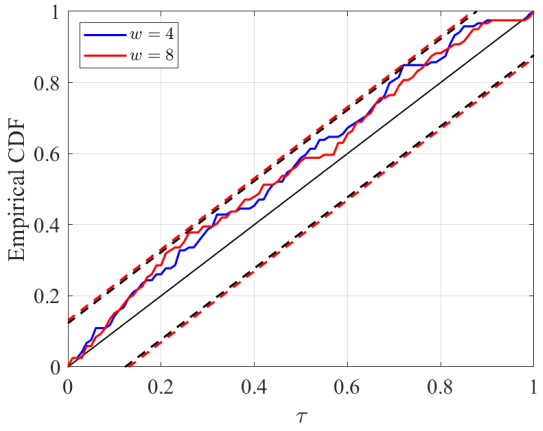
(a) Real GDP: One-step-ahead



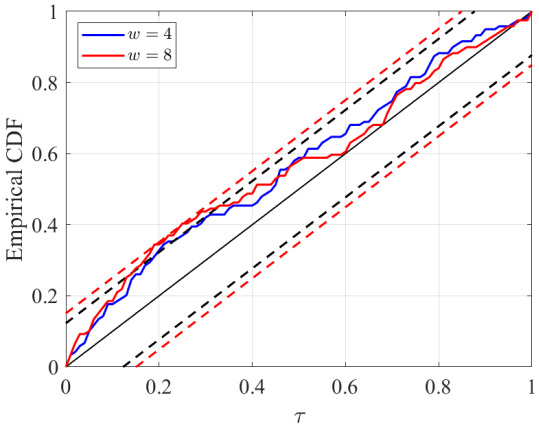
(b) Real GDP: Nowcast



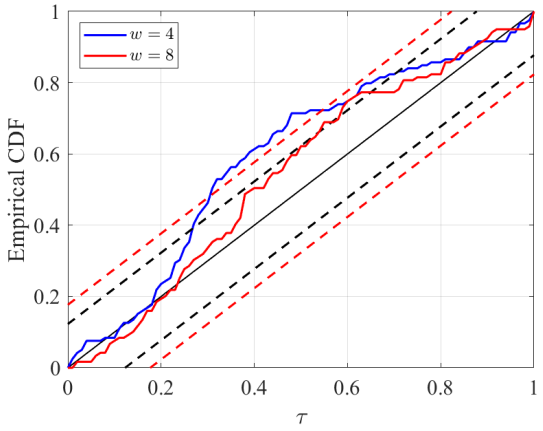
(c) Inflation: One-step-ahead



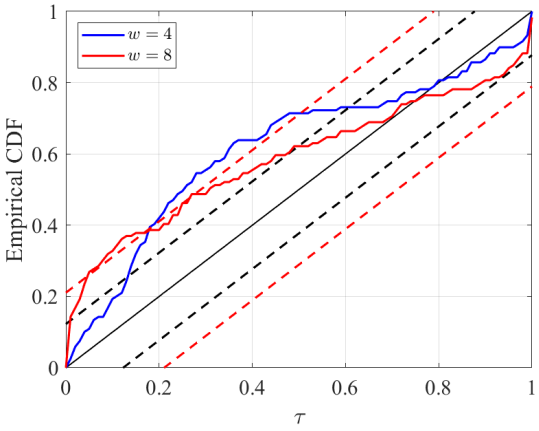
(d) Inflation: Nowcast



(e) Unemployment: One-step-ahead



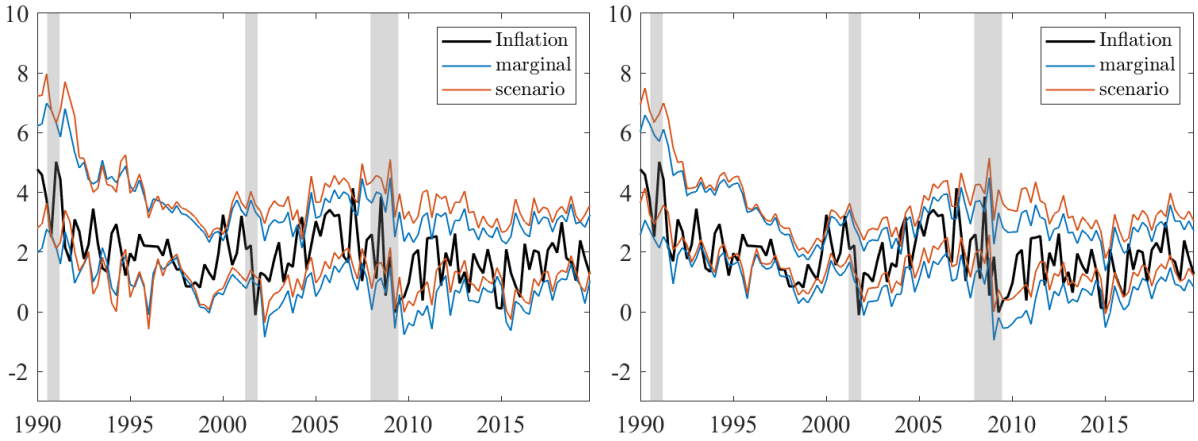
(f) Unemployment: Nowcast



Notes: This figure displays the probability integral transform (PIT) over the evaluation period from 1990:Q1 to 2019:Q4 for the SPF-GARCH model assuming 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

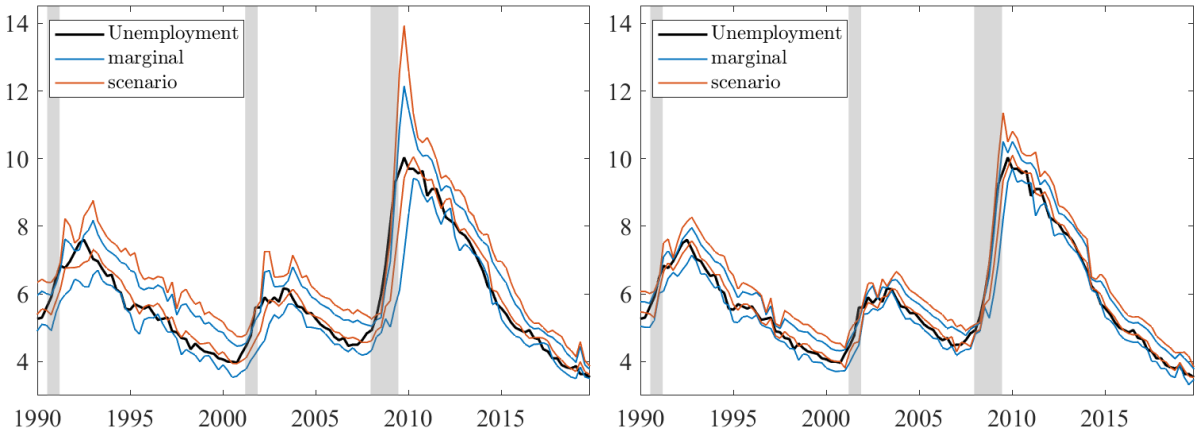
(a) One-step-ahead inflation densities (b) Nowcast inflation densities



Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for inflation when both GDP growth and unemployment are in their lower deciles, i.e., $Q^{\pi}_{t+h|t,w}(\tau) \mid (y \leq Q^{y}_{t+h|t,w}(0.1),\ u \leq Q^{u}_{t+h|t,w}(0.1))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized inflation (black line). 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. The sample period is 1990:Q1–2019:Q4. All forecasts are from the SPF-GARCH model assuming a multivariate normal distribution. Red lines depict the conditional ("bad-scenario") percentiles; blue lines depict the percentiles of the marginal distribution. Vertical gray bands denote NBER recessions.

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

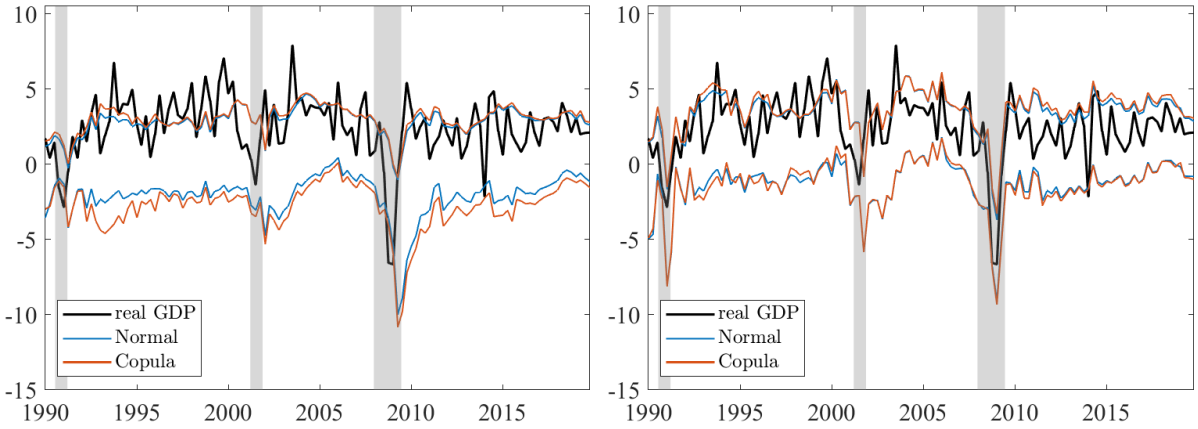
(a) One-step-ahead unemployment densities (b) Nowcast unemployment inflation



Notes: This figure plots the 10th and 90th percentiles of the conditional predictive density for unemployment when both real GDP growth and inflation are in their lower deciles, i.e., $Q^u_{t+h|t,w}(\tau) \mid (y \leq Q^y_{t+h|t,w}(0.1),\ \pi \leq Q^\pi_{t+h|t,w}(0.1))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly unemployment (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. The sample period is 1990:Q1–2019:Q4. All forecasts are from the SPF-GARCH model assuming a multivariate normal distribution. Red lines depict the conditional ("bad-scenario") percentiles; blue lines depict the percentiles of the marginal distribution. Vertical gray bands denote NBER recessions.

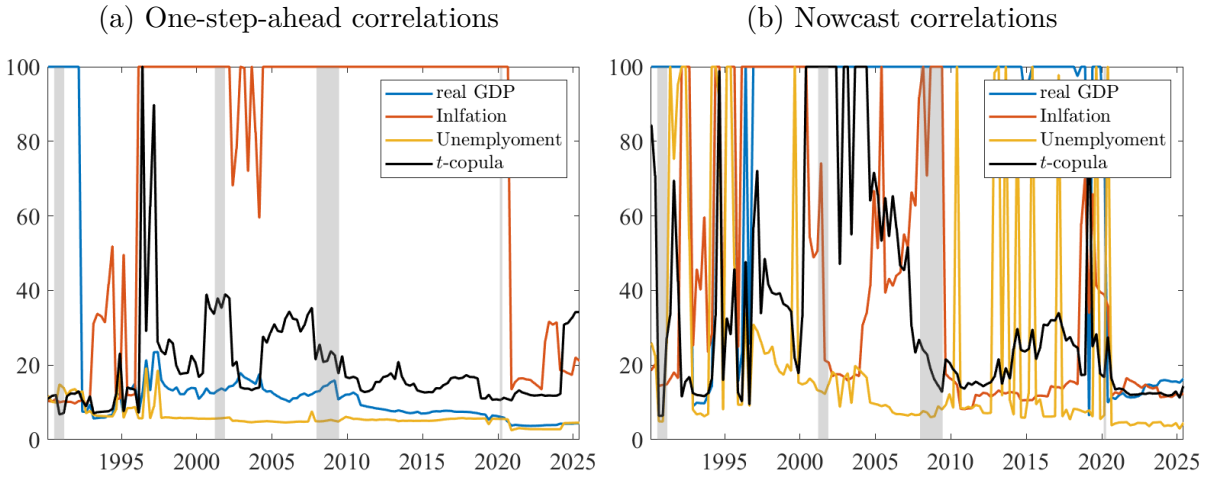
Figure B.4: Downside GDP scenarios based on the SPF-GARCH model under normality and the t-copula

(a) One-step-ahead GDP scenarios (b) Nowcast GDP scenarios



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., $Q^y_{t+h|t,w}(\tau) \mid (\pi \geq Q^\pi_{t+h|t,w}(0.9),\ u \geq Q^u_{t+h|t,w}(0.9))$ for $\tau \in \{0.1, 0.9\}$, alongside realized quarterly annualized GDP growth (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. The sample period is 1990:Q1–2019:Q4. 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. Vertical gray bands denote 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. Vertical gray bands denote NBER recessions.